

WIOLETTA GRZENDA

BADANIE DETERMINANT POZOSTAWANIA BEZ PRACY OSÓB MŁODYCH Z WYKORZYSTANIEM SEMIPARAMETRYCZNEGO MODELU COXA¹

1. WSTĘP

W Polsce w ostatnich latach wskaźniki zatrudnienia dla ludzi młodych, w porównaniu z ogółem społeczeństwa, są bardzo niskie. Podobną sytuację obserwuje się w całej Europie. Według danych BAEL w czwartym kwartale 2011 roku stopa bezrobocia w Polsce wynosiła 9,7%, natomiast wśród osób młodych w wieku 15-34 lata 36,6%. W roku 2009 w tym samym kwartale stopa bezrobocia wynosiła odpowiednio 8,5% i 31%.

Trudności w znalezieniu pracy w znacznym stopniu ograniczają usamodzielnienie ekonomiczne ludzi młodych, a z tym często wiąże się zależność ekonomiczna od rodziców lub opiekunów oraz odkładanie decyzji o zakładaniu własnej rodziny. W celu zwiększenia swoich szans na rynku pracy młode osoby często kontynuują edukację, jednak otrzymanie dyplomu nie zawsze poprawia ich sytuację. Ponadto młode osoby kończąc edukację często nie posiadają doświadczenia zawodowego. Aby to zmienić, w trakcie procesu kształcenia, powinny one podejmować różne formy praktyk zawodowych. Jednocześnie w dynamicznej gospodarce można obserwować niedopasowania strukturalne popytu na pracę i podaży pracy pod względem zawodów, wykształcenia, kwalifikacji oraz miejsca zamieszkania i pracy (Kwiatkowski, 2007). Istotne znaczenie mają badania dotyczące wpływu indywidualnych cech osób bezrobotnych na długość czasu poszukiwania pracy, bowiem otrzymywane oferty pracy zależą głównie od tych cech oraz sytuacji na rynku pracy. Wśród badanych cech najczęściej wyróżnia się: płeć, stan cywilny, wykształcenie, wiek oraz rodzaj wcześniej wykonywanej pracy (Collier, 2003). Celem niniejszego opracowania jest identyfikacja czynników demograficznych oraz społeczno-ekonomicznych wpływających na długość czasu pozostawania bez pracy osób młodych w wieku od 15 do 35 lat.

Analizę determinant bezrobocia najczęściej prowadzi się na podstawie standaryzowanych stóp bezrobocia (Socha, Sztanderska, 2000). W analizie przyczyn zróżnicowania długości okresu pozostawania bez pracy poszczególnych osób wykorzystuje się również metody ekonometryczne (Lancaster, 1979). Modelowe ujęcie bezrobocia oparte na modelu ekonometrycznym, z uwzględnieniem między innymi zmiennych opi-

¹ Badanie zostało zrealizowane w ramach badania statutowego „Badanie rynku pracy z wykorzystaniem bayesowskich modeli analizy historii zdarzeń” nr 03/S/0044/12, Szkoła Główna Handlowa.

sujących zatrudnienie w małych i średnich przedsiębiorstwach, nakłady na działalność innowacyjną oraz transport, zawarte jest w pracy Balcerowicz-Szkutnik i inni (2010). Przy badaniu cech demograficznych oraz społeczno-ekonomicznych, takich jak płeć, poziom wykształcenia, stan cywilny, wiek, czy miejsce zamieszkania wykorzystywane są głównie modele logitowe (Grzenda, 2011) oraz modele analizy historii zdarzeń, w tym model wykładniczy przedziałami stały (Drobnic, Frątczak, 2001), jak również wielopoziomowe modele przeżycia (Biggeri, Bini, Grilli, 2001). W ostatniej pracy analizowane są nie tylko wybrane cechy różnicujące absolwentów uczelni wyższych, ale również efektywność kształcenia na uczelniach w kontekście szans znalezienia pracy.

W niniejszej pracy do badania determinant długości czasu pozostawania bezrobotnym wykorzystano bayesowski semiparametryczny model Coxa dla danych indywidualnych. Model ten został zaproponowany przez Coxa (Cox, 1972; Cox, 1975) i najczęściej jest wykorzystywany w medycynie do badania przeżycia. W pracy (Merrick, Soyer, Mazzuchi, 2002) zaproponowano również wykorzystanie tego modelu w badaniu niezawodności, tj. analizy długości czasu właściwego działania maszyny. Obecnie często spotyka się wykorzystanie tego modelu w ekonomii i naukach społecznych do badania czasu trwania określonego zjawiska. Stosowanie tego modelu jest zalecane w badaniu kierunku i siły wpływu zmiennych na czas trwania danego zjawiska (Blossfeld, Rohwer, 2002). Ponadto model ten daje możliwość uwzględnienia w badaniu zmiennych, których stopień wpływu na czas trwania danego zjawiska jest różny w badanym okresie.

2. ZAKRES BADAŃ

Podejście bayesowskie, za pomocą rozkładów *a priori*, daje możliwość włączenia do wnioskowania wstępnych informacji spoza próby. W niniejszym opracowaniu zaproponowano wykorzystanie charakterystyk rozkładów *a posteriori* otrzymanych w roku poprzednim, jako informacji *a priori* dla modelowania danych pochodzących z roku następnego.

Dwa zbiory danych wykorzystane w badaniu pochodzą z badań Głównego Urzędu Statystycznego „Budżety Gospodarstw Domowych 2008” oraz „Budżety Gospodarstw Domowych 2009”. W pierwszym badaniu zbadano 37358 gospodarstw domowych obejmujących łączną sumę 109819 respondentów, w drugim odpowiednio 37302 gospodarstw domowych i 108038 respondentów. Ze względu na cel badania w niniejszej analizie uwzględniono tylko osoby bezrobotne w wieku od 15 do 35 lat, które poszukiwały pracy i były gotowe podjąć pracę (por. zalecenia EUROSTAT). W ten sposób wybrano do analizy 1258 osób z badania w roku 2008 oraz 1747 osób z badania w roku 2009. Wśród osób wybranych do badania, 128 osób z badania w roku 2008 oraz 130 osób z badania w roku 2009 znalazło już pracę i czekało na jej rozpoczęcie. Takie osoby to jednostki, dla których wystąpiło zdarzenie. Pozostałe zaś osoby, to jednostki ocenzone. Uwzględniając to, że różne czynniki mogą determinować bezrobocie, w zależności od czasu jego trwania, badano tylko te osoby, które pozostawały bez pracy maksymalnie 12 miesięcy.

W badaniu zmienną zależną jest *czas* wyrażony liczbą miesięcy pozostawania bez pracy, natomiast jako zmienne objaśniające wybrano: płeć, stan cywilny, poziom wykształcenia, informację o tym, czy respondent nadal się doksztalca, grupę społeczno-ekonomiczną, region Polski (według podziału przyjętego przez EUROSTAT), który zamieszkuje respondent, klasę miejscowości zamieszkania oraz wiek. Poniżej podajemy charakterystykę tych zmiennych oraz problemy badawcze z nimi związane.

Wiele opracowań dotyczy równości szans na rynku pracy kobiet i mężczyzn, zatem jako pierwszą potencjalną zmienną wybrano *płeć*: 1 – mężczyzna (55,7%), 2 – kobieta (44,3%). Przypuszcza się, że mężczyźni mają potencjalnie większe szanse na znalezienie pracy niż kobiety, z uwagi na rolę, jaką pełnią kobiety w rodzinie.

Osoby młode, szczególnie kobiety, są mniej skłonne do zakładania rodziny, ponieważ obawiają się trudności związanych ze znalezieniem pracy. Przypuszcza się jednocześnie, że pracodawcy chętniej zatrudniają osoby stanu wolnego. Zatem warto zweryfikować to przypuszczenie. Zmienna *stan cywilny*: 1 – kawaler, panna, w separacji, rozwiedziony(a), wdowiec, wdowa (77,96%), 2 – żonaty, zamężna (22,04%).

Można przypuszczać, że poziom wykształcenia jest jednym z najsilniejszych czynników wpływających na szanse znalezienia pracy. Inne opracowania wskazują, że największe szanse na znalezienie pracy mają osoby z wykształceniem wyższym oraz średnim zawodowym. Zmienna *wykształcenie*: 1 – wyższe (14,54%), 2 – policealne i średnie zawodowe (27,19%), 3 – średnie ogólnokształcące (17,46%), 4 – zasadnicze zawodowe (26,96%), 5 – gimnazjalne lub podstawowe (13,85%).

Badamy osoby młode, które często doksztalczą się, zatem zweryfikujemy, czy znaczenie ma fakt, że respondent nadal doksztalca się. Informację o tym zawiera zmienna *kontynuacja nauki*: 1 – tak (14,6%), 2 – nie (85,4%).

Przypuszcza się również, że większe szanse na znalezienie pracy mają osoby zamieszkujące duże miasta niż mieszkańcy wsi. Zmienna *klasa miejsca zamieszkania*: 1 – miasto powyżej 100 tys. mieszkańców (21,47%), 2 – miasto poniżej 100 tys. mieszkańców (32,05%), 3 – wieś (46,48%).

Poszczególne obszary Polski cechuje różny rozwój gospodarczy; dlatego zbadamy, czy i w jakim stopniu podział terytorialny wpływa na szanse znalezienia pracy. Zmienna *region*: 1 – centralny (województwa: łódzkie, mazowieckie) (17,52%), 2 – południowo-zachodni (województwa: dolnośląskie, opolskie) (17,57%), 3 – południowy (województwa: małopolskie, śląskie) (20,89%), 4 – północno-zachodni (województwa: wielkopolskie, zachodniopomorskie, lubuskie) (15,34%), 5 – północny (województwa: kujawsko-pomorskie, warmińsko-mazurskie, pomorskie) (10,65%), 6 – wschodni (województwa: lubelskie, podkarpackie, świętokrzyskie, podlaskie) (18,03%).

Kolejną determinanta, którą uwzględniono w badaniu, to *grupa społeczno-ekonomiczna*: 1 – pracownicy najemni (60,78%), 2 – rolnicy (5,55%), 3 – pracujący na własny rachunek (5,84%), 4 – emeryci, renciści, utrzymujący się z niezarobkowych źródeł (27,82%). Ponadto w przypadku tej cechy warto zweryfikować sytuację osób samo zatrudniających się.

W badaniu uwzględniamy tylko osoby młode w wieku 15-35 lat. Uwzględniając potencjalną możliwość uzyskania odpowiedniego poziomu wykształcenia, podzielono respondentów na grupy według wieku. Zmienna *grupa wieku*: 1 – do 19 lat (11,56%), 2 – 20-24 lata (51,57%), 3 – powyżej 25 lat (36, 86%).

3. METODA BADANIA I ESTYMACJA MODELU

3.1. MODEL COXA I JEGO ZAŁOŻENIA

W niniejszej pracy do analizy determinant wpływających na długość czasu pozostawania bez pracy wybrano model proporcjonalnego hazardu Coxa (Cox, 1972). Model ten nie wymaga założeń dotyczących kształtu rozkładu przeżyć. W wersji klasycznej ma on następującą postać:

$$h(t; \mathbf{x}) = h_0(t) \exp(\beta \mathbf{x}),$$

lub równoważnie:

$$h(t; x_1, x_2, \dots, x_p) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p),$$

gdzie

$\mathbf{x} = (x_1, x_2, \dots, x_p)'$ – wektor zmiennych objaśniających,

$\beta = (\beta_1, \beta_2, \dots, \beta_p)$ – wektor współczynników regresji,

$h(t; x_1, x_2, \dots, x_p)$ – oznacza ryzyko (hazard) w czasie t dla p zmiennych objaśniających,

$h_0(t)$, $h_0(t) > 0$ – niewyspecyfikowany parametrycznie hazard odniesienia (*baseline hazard*).

Po obustronnym zlogarytmowaniu otrzymujemy model liniowy, którego parametry są szacowane:

$$\ln(h(t)) = \ln(h_0(t)) + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p.$$

Funkcją zależną od t jest tylko hazard odniesienia, zatem w takim modelu zakłada się, że pozostałe zmienne w badanym okresie nie zmieniają się w czasie.

Przed przystąpieniem do estymacji bayesowskiej sprawdzono metodami niebayesowskimi założenia klasycznego modelu Coxa. Pierwsze mówi, że istnieje log-liniowa zależność pomiędzy zmiennymi niezależnymi a funkcją hazardu. Drugie, zwane założeniem proporcjonalności, mówi, że iloraz hazardu dla dwóch wektorów zmiennych niezależnych jest niezależny od czasu.

W modelu uwzględniamy tylko zmienne dyskretne, zatem weryfikujemy założenie proporcjonalności hazardów. W tym celu oszacowano model Coxa uwzględniający interakcję poszczególnych zmiennych ze zmienną czasową. Dodatkowe zmienne zostały określone jako iloczyny zmiennych objaśniających i zmiennej oznaczającej czas. W wyniku estymacji otrzymano, że współczynniki przy tych iloczynach są statystycznie istotne, zatem założenie proporcjonalności hazardów nie jest spełnione. Inną metodą

weryfikacji tego założenia jest metoda graficzna. Wyznaczamy wykresy funkcji hazardu dla grup wyznaczonych przez poszczególne poziomy niezależnych zmiennych jakościowych i weryfikujemy, czy są równoległe względem czasu. Zamiennie zamiast funkcji hazardu można rozważać funkcje dożycia:

$$\frac{\ln(S_1(t))}{\ln(S_2(t))} = \text{const.}$$

Otrzymane wykresy potwierdziły nasze wcześniejsze przypuszczenia, że dla rozważanych danych należy rozważać model Coxa ze zmiennymi objaśniającymi zależnymi od czasu. Wynika to z faktu, iż zmienne objaśniające w różnym stopniu wpływają na ryzyko zależnie od czasu trwania badanego zjawiska. Wówczas model ten można zapisać w następującej postaci:

$$h(t; \mathbf{x}) = h_0(t) \exp(\beta \mathbf{x}(t)).$$

Na podstawie otrzymanych wykresów funkcji przeżycia zostały do modelu wprowadzone zmienne wskaźnikowe zależne od czasu. W tym celu oszacowano dla każdej zmiennej objaśniającej kilka modeli i wybrano ten najlepszy opierając się o kryteria AIC (*Akaike's Information Criterion*) i SBC (*Shwarz's Bayesian Criterion*).

W niniejszej pracy przeprowadzono analizę determinant bezrobocia z wykorzystaniem modelu Coxa w ujęciu bayesowskim. Rozważania dotyczące modeli przeżycia w ujęciu bayesowskim można znaleźć w wielu pracach (Sinha, Dey, 1997; Ibrahim, Chen, Sinha, 2001; Sinha, Ibrahim, Chen, 2003; Congdon, 2007).

W modelach przeżycia obserwowane dane można zapisać w następującej postaci: $D = (n, \mathbf{t}, \mathbf{X}, \mathbf{v})$. Wektor $\mathbf{t} = (t_1, t_2, \dots, t_n)'$ oznacza czasy przeżycia, $\mathbf{v} = (v_1, v_2, \dots, v_n)'$ wektor zmiennych cenzurujących, przy czym $v_i = 0$, jeśli t_i jest prawostronnie oceniany oraz $v_i = 1$, jeśli dla t_i wystąpiło zdarzenie, $i = 1, \dots, n$, a \mathbf{X} oznacza macierz zmiennych objaśniających o wymiarze $n \times p$, której i -ty wiersz oznaczamy \mathbf{x}'_i .

Podstawową metodą estymacji parametrów modelu Coxa jest metoda częściowej wiarygodności (*partial likelihood method*). Jeśli nie mamy wyspecyfikowanej parametrycznie funkcji dla hazardu bazowego, to nie można stosować tradycyjnej metody największej wiarygodności. Wówczas wektor β współczynników regresji jest szacowany poprzez maksymalizację tzw. funkcji częściowej wiarygodności. Oszacowania współczynników modelu uzyskane w wyniku maksymalizacji funkcji częściowej wiarygodności mają właściwości podobne do tych uzyskanych metodą maksymalizacji standardowej funkcji wiarygodności (Cox, 1972; Cox, 1975; Blossfeld, Rohwer, 2002).

W rozważanym modelu parametrami są współczynniki regresji $\beta = (\beta_1, \beta_2, \dots, \beta_p)$. Podejście bayesowskie wymaga określenia rozkładów *a priori* dla wszystkich parametrów modelu; łączny rozkład *a priori* oznaczamy $p(\beta)$. Najczęściej dla wektora parametrów β wybierany jest p -wymiarowy rozkład normalny *a priori* $N_p(\mu_0, \Sigma_0)$, gdzie μ_0 oznacza wektor średnich, a Σ_0 macierz wariancji-kowariancji. Wnioskowanie w podejściu bayesowskim opiera się na rozkładach *a posteriori* $p(\beta|D)$ wyznaczonych

z wykorzystaniem twierdzenia Bayesa. W przypadku wykorzystania w estymacji funkcji częściowej wiarygodności otrzymuje się pewne przybliżenie rozkładu *a posteriori* dla wektora parametrów β .

W podejściu bayesowskim wnioskowanie o dowolnej współrzędnej β_i wektora parametrów $\beta = (\beta_1, \dots, \beta_p)$ odbywa się na podstawie brzegowego rozkładu *a posteriori*, otrzymanego poprzez całkowanie po pozostałych współrzędnych łącznego rozkładu *a posteriori* (Osiewalski, 2001):

$$p(\beta_i|D) = \int \dots \int p(\beta_1, \dots, \beta_p|D) d\beta_1 \dots d\beta_{i-1} d\beta_{i+1} \dots d\beta_p$$

W praktyce często wykorzystujemy metody symulacyjne, które umożliwiają generowanie prób losowych² z dowolnej gęstości *a posteriori*, dla których można wyznaczać różne charakterystyki. Obecnie dużą popularnością cieszą się metody Monte Carlo oparte na łańcuchach Markowa (*Markov Chain Monte Carlo Method* – MCMC).

Podstawą metod MCMC jest generowanie ergodycznego łańcucha Markowa, który po upływie odpowiednio długiego czasu osiąga rozkład stacjonarny, którym jest rozkład *a posteriori*. Najbardziej znanym algorytmem tych metod obliczeniowych jest algorytm Metropolisa, w niniejszej pracy zostało wykorzystane jego uogólnienie – algorytm ARMS (*Adaptive Rejection Metropolis Sampling Algorithm*). Opis użytego algorytmu zawarty jest w pracy Gilks, Best i Tan (1995).

3.2. SPECYFIKACJA I ESTYMACJA MODELU

Z uwagi na to, że zmienne objaśniające w różnym stopniu wpływają na ryzyko zależnie od czasu, do modelu oprócz zmiennych objaśniających przedstawionych w części 2 niniejszej pracy zostaną włączone zmienne wskaźnikowe zależne od czasu (Tabela 1.). Zmienne te zostały wyznaczone na podstawie wykresów funkcji dożycia.

Przed przystąpieniem do estymacji głównego modelu dla danych z roku 2009 oszacowano analogiczny model w podejściu bayesowskim dla danych z roku 2008. Chcąc osiągnąć wyniki obiektywnie poprawne, w pierwszym modelu użyto nieinformacyjnych rozkładów³ *a priori*.

² W literaturze spotyka się to pojęcie, należy jednak pamiętać, że ciągi liczb uzyskane w wyniku działania deterministycznych algorytmów są pseudolosowe, ponieważ żaden komputer nie jest w stanie wygenerować prawdziwie losowego ciągu liczb.

³ W podejściu bayesowskim szczególną rolę odgrywają tzw. nieinformacyjne rozkłady *a priori* „pozwalające danym mówić za siebie”. Rozkłady te mogą służyć do wyrażenia zupełnego braku wiedzy o szacowanych parametrach albo mogą służyć jako punkt odniesienia przy porównywaniu wyników uzyskiwanych z wykorzystaniem informacyjnych rozkładów *a priori*, wówczas zwane są one rozkładami referencyjnymi lub wzorcowymi (Osiewalski, 2001, Szreder, 1994). Jeśli otrzymane rezultaty w głównej mierze mają odzwierciedlać wpływ zaobserwowanych danych, to jako rozkłady *a priori* wybiera się rozkłady rozproszone (Pipień, 2006). Wówczas w praktyce zamiast rozkładów nieinformacyjnych, stosujemy tzw. rozkłady wystarczająco mało informacyjne. W przypadku rozkładów normalnych najczęściej dla średniej przyjmujemy wartość zero, a dla wariancji pewną dużą liczbę.

Tabela 1.

Zmienne wskaźnikowe

Zmienna	Poziom	Przedział czasu
pleć 1	1	<= 9 miesięcy
pleć 2	1	>9 miesięcy
stan cywilny 1	1	<= 3 miesięcy
stan cywilny 2	1	>3 miesięcy
kontynuacja nauki 1	1	<= 6 miesięcy
kontynuacja nauki 2	1	>6 miesięcy
klasa miejsca zamieszkania 1	1	<= 9 miesięcy
klasa miejsca zamieszkania 1	2	<= 9 miesięcy
klasa miejsca zamieszkania 2	1	>9 miesięcy
klasa miejsca zamieszkania 2	2	>9 miesięcy
grupa wieku 1	1	<= 3 miesięcy
grupa wieku 1	2	<= 3 miesięcy
grupa wieku 2	1	>3 miesięcy
grupa wieku 2	2	>3 miesięcy
grupa społeczno-ekonomiczna 1	2	<= 3 miesięcy
grupa społeczno-ekonomiczna 1	3	<= 3 miesięcy
grupa społeczno-ekonomiczna 1	4	<= 3 miesięcy
grupa społeczno-ekonomiczna 2	2	>3 miesięcy
grupa społeczno-ekonomiczna 2	3	>3 miesięcy
grupa społeczno-ekonomiczna 2	4	>3 miesięcy
wykształcenie 1	1	<= 3 miesięcy
wykształcenie 1	2	<= 3 miesięcy
wykształcenie 1	3	<= 3 miesięcy
wykształcenie 1	4	<= 3 miesięcy
wykształcenie 2	1	>3 miesięcy
wykształcenie 2	2	>3 miesięcy
wykształcenie 2	3	>3 miesięcy
wykształcenie 2	4	>3 miesięcy
region 1	1	<= 3 miesięcy
region 1	2	<= 3 miesięcy
region 1	3	<= 3 miesięcy
region 1	4	<= 3 miesięcy
region 1	5	<= 3 miesięcy
region 2	1	>3 miesięcy
region 2	2	>3 miesięcy
region 2	3	>3 miesięcy
region 2	4	>3 miesięcy
region 2	5	>3 miesięcy

Źródło: obliczenia własne⁴ na podstawie danych GUS 2009.

⁴ Wszystkie obliczenia zostały wykonane w systemie SAS 9.2.

W estymacji przyjęto nieinformacyjne rozkłady normalne *a priori* ze średnią równą 0 i wariancją 10^6 dla wszystkich parametrów regresji: $\beta \sim N(\mathbf{0}, 10^6\mathbf{I})$. W obu modelach przyjęto, że liczba cykli spalonych ma wynosić 2000, a liczba właściwych realizacji przyjętych jako próba z rozkładu *a posteriori* 10000.

Podejście bayesowskie daje możliwość uwzględnienia w badaniu, za pomocą rozkładów *a priori*, dodatkowej informacji spoza próby. Szczególne znaczenie w estymacji mają rozkłady *a priori*, które zostały skonstruowane na podstawie poprzednich empirycznych oszacowań parametrów dokonanych na bazie wcześniej posiadanych informacji. Uzyskane na podstawie wcześniej pobranej próby rozkłady *a posteriori* stają się wiedzą *a priori*, dla nowych danych. W ten sposób można dokonywać ciągłej weryfikacji ocen na podstawie nowych informacji (Szreder, 1994). W niniejszym opracowaniu charakterystyki rozkładu *a posteriori* dla roku 2008 zostały wykorzystane jako informacja *a priori* w modelu dla danych z roku 2009.

Przed przystąpieniem do wnioskowań z wykorzystaniem prób z rozkładu *a posteriori* zweryfikowano zbieżność wygenerowanych łańcuchów. Analogiczne postępowanie zostało przeprowadzone dla danych z roku 2008 i 2009. Natomiast wyniki przedstawiamy tylko dla roku 2009. Często wykorzystywanym testem do sprawdzenia, czy wygenerowany łańcuch osiągnął rozkład stacjonarny jest test Geweke'a (Geweke, 1992). Na podstawie wyników przedstawionych w tabeli 2, dla danych z roku 2009, stwierdzono brak podstaw do odrzucenia hipotezy, że wygenerowane łańcuchy dla poszczególnych parametrów modelu są zbieżne, przy poziomie istotności 0,01.

Tabela 2.

Wyniki dla testu Geweke'a

Parametr		Wartość statystyki z	p - value
pleć 1	1	-0,5852	0,5584
pleć 2	1	1,1985	0,2307
stan cywilny 1	1	-0,6410	0,5215
stan cywilny 2	1	0,4992	0,6177
kontynuacja nauki 1	1	-0,5718	0,5674
kontynuacja nauki 2	1	-0,9181	0,3586
klasa miejsca zamieszkania 1	1	-2,3914	0,0168
klasa miejsca zamieszkania 1	2	-2,0012	0,0454
klasa miejsca zamieszkania 2	1	0,0278	0,9778
klasa miejsca zamieszkania 2	2	0,5047	0,6138
grupa wieku 1	1	0,1116	0,9111
grupa wieku 1	2	0,1408	0,8880
grupa wieku 2	1	-1,3907	0,1643
grupa wieku 2	2	-1,0742	0,2827
grupa społeczno-ekonomiczna 1	2	-1,0430	0,2969
grupa społeczno-ekonomiczna 1	3	-1,6176	0,1058
grupa społeczno-ekonomiczna 1	4	-0,7212	0,4708

cd. Tabela 2.

grupa społeczno-ekonomiczna 2	2	0,5396	0,5894
grupa społeczno-ekonomiczna 2	3	-0,0881	0,9298
grupa społeczno-ekonomiczna 2	4	-0,5912	0,5544
wykształcenie 1	1	-0,3599	0,7189
wykształcenie 1	2	-0,5383	0,5904
wykształcenie 1	3	-0,4091	0,6825
wykształcenie 1	4	-0,6179	0,5366
wykształcenie 2	1	-0,0952	0,9242
wykształcenie 2	2	-0,5559	0,5783
wykształcenie 2	3	-0,4902	0,6240
wykształcenie 2	4	-0,2891	0,7725
region 1	1	-0,5509	0,5817
region 1	2	0,0109	0,9913
region 1	3	-1,3753	0,1690
region 1	4	-0,5463	0,5849
region 1	5	-0,8963	0,3701
region 2	1	-0,7386	0,4602
region 2	2	-0,5522	0,5808
region 2	3	-1,4720	0,1410
region 2	4	-0,9901	0,3221
region 2	5	-1,6833	0,0923

Źródło: obliczenia własne na podstawie danych GUS 2009

W badaniu stacjonarności równie powszechna jest metoda graficzna, która potwierdziła otrzymane wnioski z testu Geweke'a. Zatem po odrzuceniu 2000 cykli wstępnych, 10000 stanów łańcucha zostało przyjętych jako próba z rozkładu *a posteriori*.

Wartości oczekiwane *a posteriori* parametrów modelu dla roku 2008 zostały przedstawione w tabeli 3, a dla roku 2009 w tabeli 4.

Tabela 3.

Charakterystyki *a posteriori* dla danych z 2008 roku

Parametr		Wartości oczekiwane <i>a posteriori</i>	Odchylenia standardowe <i>a posteriori</i>	Obszary największej wartości funkcji gęstości <i>a posteriori</i> ⁵	
pleć 1	1	0,1988	0,00543	0,1881	0,2091
pleć 2	1	0,4452	0,0177	0,4103	0,4795
stan cywilny 1	1	0,4516	0,00925	0,4334	0,4695
stan cywilny 2	1	-0,0440	0,0103	-0,0634	-0,0232
kontynuacja nauki 1	1	-0,2418	0,00888	-0,2593	-0,2246

⁵ Obszary największej wartości funkcji gęstości *a posteriori* (HPD) zostały wyznaczone przy $\alpha = 0,05$ dla wszystkich parametrów modelu.

cd. Tabela 3.

kontynuacja nauki 2	1	0,4371	0,0180	0,4022	0,4721
klasa miejsca zamieszkania 1	1	0,00384	0,00667	-0,00944	0,0166
klasa miejsca zamieszkania 1	2	-0,0774	0,00615	-0,0892	-0,0652
klasa miejsca zamieszkania 2	1	0,9785	0,0268	0,9256	1,0298
klasa miejsca zamieszkania 2	2	1,3041	0,0248	1,2551	1,3531
grupa wieku 1	1	-0,2338	0,0117	-0,2570	-0,2115
grupa wieku 1	2	-0,5760	0,00738	-0,5903	-0,5616
grupa wieku 2	1	-0,0174	0,0157	-0,0484	0,0129
grupa wieku 2	2	0,0603	0,00896	0,0427	0,0776
grupa społeczno-ekonomiczna 1	2	0,4488	0,0152	0,4193	0,4788
grupa społeczno-ekonomiczna 1	3	-0,7212	0,0165	-0,7538	-0,6892
grupa społeczno-ekonomiczna 1	4	-0,1038	0,00723	-0,1184	-0,0900
grupa społeczno-ekonomiczna 2	2	-0,3534	0,0219	-0,3967	-0,3112
grupa społeczno-ekonomiczna 2	3	-0,1373	0,0146	-0,1662	-0,1094
grupa społeczno-ekonomiczna 2	4	-0,3249	0,00929	-0,3426	-0,3066
wykształcenie 1	1	0,9720	0,0140	0,9445	0,9991
wykształcenie 1	2	0,9126	0,0132	0,8865	0,9382
wykształcenie 1	3	0,7549	0,0144	0,7261	0,7821
wykształcenie 1	4	0,8834	0,0130	0,8578	0,9087
wykształcenie 2	1	0,4980	0,0124	0,4736	0,5222
wykształcenie 2	2	-0,0686	0,0115	-0,0913	-0,0462
wykształcenie 2	3	-0,0560	0,0129	-0,0823	-0,0316
wykształcenie 2	4	-0,5485	0,0124	-0,5714	-0,5229
region 1	1	0,6351	0,00997	0,6161	0,6544
region 1	2	0,3016	0,0108	0,2809	0,3231
region 1	3	-0,2968	0,0115	-0,3192	-0,2736
region 1	4	-0,8575	0,0152	-0,8872	-0,8276
region 1	5	0,0409	0,0129	0,0142	0,0653
region 2	1	0,1681	0,0139	0,1417	0,1956
region 2	2	0,2160	0,0134	0,1902	0,2435
region 2	3	-0,0654	0,0132	-0,0912	-0,0392
region 2	4	0,7914	0,0127	0,7661	0,8156
region 2	5	-0,8038	0,0220	-0,8454	-0,7595

Źródło: obliczenia własne na podstawie danych GUS 2008

Tabela 4.

Charakterystyki *a posteriori* dla danych z 2009 roku

Parametr		Wartości oczekiwane <i>a posteriori</i>	Odchylenia standardowe <i>a posteriori</i>	Obszary największej wartości funkcji gęstości <i>a posteriori</i>	
pleć 1	1	0,0457	0,00514	0,0356	0,0558
pleć 2	1	-0,2666	0,0168	-0,2990	-0,2335
stan cywilny 1	1	-0,3218	0,00817	-0,3370	-0,3050
stan cywilny 2	1	-0,0891	0,0101	-0,1080	-0,0687
kontynuacja nauki 1	1	0,3332	0,00665	0,3204	0,3465
kontynuacja nauki 2	1	-0,6685	0,0269	-0,7214	-0,6173
klasa miejsca zamieszkania 1	1	0,0162	0,00671	0,00339	0,0298
klasa miejsca zamieszkania 1	2	0,0899	0,00579	0,0788	0,1015
klasa miejsca zamieszkania 2	1	-0,6098	0,0301	-0,6654	-0,5473
klasa miejsca zamieszkania 2	2	0,8915	0,0196	0,8543	0,9307
grupa wieku 1	1	0,7466	0,0114	0,7235	0,7678
grupa wieku 1	2	0,3490	0,00758	0,3343	0,3638
grupa wieku 2	1	-0,0611	0,0160	-0,0924	-0,0301
grupa wieku 2	2	0,0905	0,00874	0,0738	0,1079
grupa społeczno-ekonomiczna 1	2	-0,5546	0,0211	-0,5987	-0,5159
grupa społeczno-ekonomiczna 1	3	0,1872	0,0120	0,1642	0,2114
grupa społeczno-ekonomiczna 1	4	0,3389	0,00678	0,3259	0,3523
grupa społeczno-ekonomiczna 2	2	-1,1247	0,0304	-1,1825	-1,0629
grupa społeczno-ekonomiczna 2	3	1,1158	0,00940	1,0968	1,1337
grupa społeczno-ekonomiczna 2	4	-0,2941	0,00963	-0,3127	-0,2753
wykształcenie 1	1	1,1645	0,0123	1,1400	1,1879
wykształcenie 1	2	0,3591	0,0122	0,3349	0,3829
wykształcenie 1	3	0,7404	0,0119	0,7180	0,7646
wykształcenie 1	4	0,1673	0,0122	0,1428	0,1906
wykształcenie 2	1	1,8094	0,0133	1,7840	1,8335
wykształcenie 2	2	0,9653	0,0127	0,9434	0,9885
wykształcenie 2	3	0,8577	0,0129	0,8359	0,8812
wykształcenie 2	4	1,0065	0,0132	0,9797	1,0308
region 1	1	-0,1589	0,00933	-0,1782	-0,1415
region 1	2	-0,8124	0,0114	-0,8348	-0,7903
region 1	3	-0,5792	0,00991	-0,5983	-0,5595
region 1	4	-0,2035	0,00972	-0,2224	-0,1847
region 1	5	0,1402	0,0102	0,1196	0,1597
region 2	1	-0,3793	0,0131	-0,4045	-0,3535
region 2	2	0,1241	0,0108	0,1034	0,1454
region 2	3	-0,8076	0,0129	-0,8316	-0,7811
region 2	4	0,5861	0,0105	0,5660	0,6069
region 2	5	-0,4445	0,0153	-0,4746	-0,4149

Źródło: obliczenia własne na podstawie danych GUS 2009

Na podstawie otrzymanych obszarów największej wartości funkcji gęstości *a posteriori* można stwierdzić (Bolstad, 2007), że wszystkie zmienne przedstawione w tabeli 4 są statystycznie istotne.

Interpretacji poddajemy współczynniki ryzyka (*hazard ratio*), który definiujemy jako stosunek wartości ryzyka dla jednego respondenta do wartości ryzyka dla innego respondenta (tabela 5.)

Tabela 5.

Oszacowania współczynników ryzyka

Parametr		Współczynnik ryzyka
pleć 1	1	1,0468
pleć 2	1	0,7660
stan cywilny 1	1	0,7248
stan cywilny 2	1	0,9148
kontynuacja nauki 1	1	1,3954
kontynuacja nauki 2	1	0,5125
klasa miejsca zamieszkania 1	1	1,0163
klasa miejsca zamieszkania 1	2	1,0941
klasa miejsca zamieszkania 2	1	0,5435
klasa miejsca zamieszkania 2	2	2,4388
grupa wieku 1	1	2,1098
grupa wieku 1	2	1,4176
grupa wieku 2	1	0,9407
grupa wieku 2	2	1,0947
grupa społeczno-ekonomiczna 1	2	0,5743
grupa społeczno-ekonomiczna 1	3	1,2059
grupa społeczno-ekonomiczna 1	4	1,4034
grupa społeczno-ekonomiczna 2	2	0,3247
grupa społeczno-ekonomiczna 2	3	3,0520
grupa społeczno-ekonomiczna 2	4	0,7452
wykształcenie 1	1	3,2043
wykształcenie 1	2	1,4320
wykształcenie 1	3	2,0968
wykształcenie 1	4	1,1821
wykształcenie 2	1	6,1068
wykształcenie 2	2	2,6256
wykształcenie 2	3	2,3577
wykształcenie 2	4	2,7360
region 1	1	0,8531
region 1	2	0,4438
region 1	3	0,5603

cd. Tabela 5.

region 1	4	0,8159
region 1	5	1,1505
region 2	1	0,6843
region 2	2	1,1321
region 2	3	0,4459
region 2	4	1,7970
region 2	5	0,6411

Źródło: obliczenia własne

4. WYNIKI I WNIOSKI

Otrzymane wyniki nie potwierdziły jednoznacznie wcześniejszych przypuszczeń, że kobiety są w gorszej sytuacji na rynku pracy. W ciągu pierwszych 9 miesięcy pozostawania bezrobotnym młodzi mężczyźni mieli tylko o 4,68% wyższe ryzyko przejścia ze stanu bezrobotny do pracujący niż młode kobiety. Jednak po tym okresie ryzyko to dla mężczyzn było mniejsze niż dla kobiet o 23,4%. Zatem można wnioskować, że kobiety coraz lepiej radzą sobie na rynku pracy. Również według badania GUS (2009) w grupie wieku 18-19 lat stopa bezrobocia jest wyższa dla mężczyzn (29,0%) niż dla kobiet (23,4%). Wyniki zawarte w pracach dotyczących wcześniejszego okresu oraz ogółu społeczeństwa (Socha, Sztanderska, 2000) często wskazują na gorszą sytuację kobiet na rynku pracy, niż mężczyzn, mimo że kobiety są często lepiej wykształcone i bardziej aktywnie poszukują pracy.

W przypadku cechy stan cywilny, również wcześniejsze przypuszczenia nie potwierdziły się. Osoby stanu wolnego miały mniejsze szanse na znalezienie pracy niż osoby pozostające w związku. W ciągu pierwszych 3 miesięcy ryzyko przejścia ze stanu bezrobotny do pracujący było dla osób stanu wolnego niższe o 27,52%, natomiast w okresie od 4 do 12 miesięcy niższe o 8,52%. Można przypuszczać, że osoby, które posiadały rodzinę miały większą motywację do bardziej intensywnego poszukiwania pracy.

Potwierdza się, że największe szanse na znalezienie pracy miały osoby z wykształceniem wyższym. W ciągu pierwszych 3 miesięcy było przeszło trzykrotnie bardziej prawdopodobne, że znajdą one pracę niż osoby z wykształceniem gimnazjalnym lub podstawowym. Dla osób z wykształceniem policealnym i średnim zawodowym ryzyko to było wyższe o 43,2%, dla osób z wykształceniem średnim ogólnokształcącym było wyższe o 109,68%, a dla osób z wykształceniem zasadniczym zawodowym było wyższe o 18,21%. W kolejnych miesiącach było przeszło sześciokrotnie bardziej prawdopodobne, że osoby z wykształceniem wyższym znajdą pracę niż osoby z wykształceniem gimnazjalnym lub podstawowym. Dla osób z wykształceniem policealnym i średnim zawodowym, średnim ogólnokształcącym oraz zasadniczym zawodowym było przeszło dwukrotnie bardziej prawdopodobne, że znajdą pracę niż osoby z grupy referencyjnej.

Otrzymane wyniki wskazują, że w ciągu całego badanego okresu szanse na znalezienie pracy przez osoby z wykształceniem średnim ogólnokształcącym, kształtują się podobnie, natomiast dla osób o innym poziomie wykształcenia szanse na znalezienie pracy w ciągu kolejnych 9 miesięcy wzrosły około dwukrotnie w porównaniu do osób z wykształceniem gimnazjalnym lub podstawowym. Potwierdzają to również inne badania dotyczące osób młodych (GUS, 2010) oraz całego społeczeństwa (Grzenda, 2011), wskazujące, że wykształcenie wyższe oraz średnie zawodowe warunkują lepszą sytuację na rynku pracy.

W przypadku osób młodych istotne jest dokończanie się. Według GUS w badanym okresie około co piąta młoda osoba jednocześnie pracowała i uczyła się. W ciągu pierwszych 6 miesięcy, osoby, które nadal kształciły się miały o 39,54% wyższe ryzyko przejścia ze stanu bezrobotny do pracujący. Po upływie tego okresu, osoby, które nadal kształciły się, miały o 48,75% mniejszą wartość współczynnika ryzyka przejścia ze stanu bezrobotny do pracujący niż osoby, które nie dokończyły się.

W ciągu pierwszych 9 miesięcy osoby, które mieszkały w miastach miały nieznacznie większe szanse na znalezienie pracy niż osoby mieszkające na wsi, w przypadku osób, które mieszkały w miastach powyżej 100 tys. mieszkańców ryzyko przejścia ze stanu bezrobotny do pracujący było wyższe o 1,63%. W przypadku mniejszych miast ryzyko było wyższe o 9,41%. Natomiast po upływie tego okresu, osoby, które mieszkały w dużych miastach miały mniejsze szanse na znalezienie pracy, natomiast w przypadku mniejszych miast było przeszło dwukrotnie bardziej prawdopodobne, że ich mieszkańcy znajdą pracę w porównaniu do osób zamieszkujących na wsi. Zatem przypuszczenie, że większe szanse na znalezienie pracy mają osoby zamieszkujące duże miasta niż mieszkańcy wsi potwierdziło się tylko w początkowym okresie poszukiwania pracy. Otrzymany wynik potwierdzają rezultaty innych badań (Socha, Sztanderska, 2000). Wynika z nich, że różnice w poziomie bezrobocia między dużymi miastami, a wsią są niewielkie zarówno dla osób młodych, jak i w starszym wieku. Natomiast w małych miastach sytuacja osób młodych na rynku pracy jest znacznie lepsza niż na wsi.

Dla cechy region, otrzymano, że w ciągu pierwszych trzech miesięcy osoby zamieszkujące inne regiony niż północny miały mniejsze szanse na znalezienie pracy niż osoby zamieszkujące region wschodni. Natomiast w ciągu kolejnych miesięcy osoby zamieszkujące region południowo-zachodni i północno-zachodni miały większe szanse na znalezienie pracy niż osoby zamieszkujące region wschodni. Otrzymane wyniki nie wskazują jednoznacznie, w których regionach Polski osobom młodym łatwiej znaleźć pracę. Może to wynikać z tego, że są regiony, np. północny, w których w jednym województwie stopa bezrobocia jest niska – pomorskie, a w drugim wysoka – kujawsko-pomorskie (BAEL). Ponadto analizujemy osoby młode, dla których miejsce zamieszkania często nie jest tożsame z miejscem pobytu. Można również przypuszczać, że w regionach, gdzie wskaźnik bezrobocia jest niski, osoby młode mają wyższe oczekiwania dotyczące pierwszej pracy dotyczące np. wysokości wynagrodzenia, czy warunków pracy.

Uwzględniając cechę grupa społeczno-ekonomiczna otrzymano, że w ciągu pierwszych 3 miesięcy ryzyko przejścia ze stanu bezrobotny do pracujący było dla rolników

o 42,57% mniejsze niż dla pracowników najemnych, a dla pracujących na własny rachunek o 20,59% wyższe niż dla pracowników najemnych. W ciągu kolejnych 9 miesięcy sytuacja rolników wyglądała podobnie, natomiast dla osób pracujących na własny rachunek było przeszło trzykrotnie bardziej prawdopodobne, że znajdą pracę niż pracownicy najemni. Otrzymany wynik wskazuje, że samozatrudnienie jest jednym ze sposobów na rozwiązanie problemu bezrobocia, szczególnie wśród osób młodych. Ponadto według badania GUS mężczyźni częściej niż kobiety zakładali własną firmę czy działalność gospodarczą (6,0% wobec 3,9%).

Osoby w wieku do 19 lat w ciągu pierwszych 3 miesięcy miały o 110,98% wyższe ryzyko przejścia ze stanu bezrobotny do pracujący niż osoby w wieku powyżej 25 lat, natomiast po upływie tego okresu ryzyko to było o 5,93% niższe niż dla osób w wieku powyżej 25 lat. Osoby w wieku 20-24 lat w ciągu pierwszych 3 miesięcy miały o 41,76% wyższe ryzyko przejścia ze stanu bezrobotny do pracujący niż osoby w wieku powyżej 25 lat, natomiast po upływie tego okresu ryzyko to było o 9,47% wyższe w stosunku do osób w wieku powyżej 25 lat. Otrzymany wynik może wskazywać na to, że osoby najmłodsze, które ukończyły szkoły średnie lub zasadnicze zawodowe, jeśli znajdą pracę, kończą swoją edukację. Najtrudniej znaleźć pracę osobom w wieku powyżej 25 lat, otrzymany wynik może niepokoić, ponieważ w tej grupie są osoby, które ukończyły studia wyższe. Jednak rezultaty dla cechy wykształcenie pokazały, że osoby z wykształceniem wyższym mają największe szanse na znalezienie pracy. Otrzymane rezultaty są zbieżne z badaniami dla województwa śląskiego (Kryńska, 2004), zgodnie z którymi w grupie osób w wieku 18-24 w latach 2000-2003 zaobserwowano sukcesywny spadek wskaźnika bezrobocia, natomiast wśród osób w wieku 25-34 wzrost.

Wykorzystany w pracy model umożliwił zbadanie występowania zależności pomiędzy wybranymi czynnikami demograficznymi oraz społeczno-ekonomicznymi a długością czasu pozostawania bez pracy osób młodych. Zaletą modeli przeżycia w badaniu aktywności zawodowej jest uwzględnienie całej historii badanej jednostki. Ponadto zaproponowany model umożliwił analizę zmiennych, których stopień wpływu na czas trwania badanego zjawiska jest różny w badanym okresie. Podejście bayesowskie dało możliwość wykorzystania w badaniu informacji pochodzącej z wcześniejszych badań.

Szczegółowa charakterystyka sytuacji osób młodych na rynku pracy może pozwolić na odpowiednie kształtowanie polityki społeczno-gospodarczej zmierzającej do ograniczenia bezrobocia wśród tych osób.

Szkoła Główna Handlowa w Warszawie

LITERATURA

- [1] Balcerowicz-Szkutnik M., Dyduch M., Szkutnik W., (2010), *Wybrane modele i analizy rynku pracy: uwarunkowania rynku pracy i wzrostu gospodarczego*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice.
- [2] Biggeri L., Bini M., Grilli L., (2001), *The Transition from University to Work: a Multilevel Approach to the Analysis of the Time to Obtain the First Job*, Journal of the Royal Statistical Society, 164, 293-305.
- [3] Blossfeld H.P., Rohwer G., (2002), *Techniques of Event History Modeling*, New Approaches to Causal Analysis, Lawrence Erlbaum Associates, New Jersey.
- [4] Bolstad W.M., (2007), *Introduction to Bayesian Statistics*, Wiley & Sons, New Jersey.
- [5] Collier W., (2003), *The Impact of Demographic and Individual Heterogeneity on Unemployment Duration: A Regional Study*, Studies in Economics, 0302.
- [6] Congdon P., (2007), *Bayesian Statistical Modelling*, Wiley & Sons, New York.
- [7] Cox D.R., (1972), *Regression Models and Life Tables (with discussion)*, Journal of the Royal Statistical Society, 34, 187-220.
- [8] Cox D.R., (1975), *Partial Likelihood*, Biometrika, 62, 269-276.
- [9] Drobnic S., Frączak E., (2001), *Employment Patterns of Married Women in Poland*, w: Blossfeld HP., Drobnic S. (red.), *Careers of Couples in Contemporary Societies*, Oxford University Press, 281-306.
- [10] Geweke J., (1992), *Evaluating the Accuracy of Sampling-based Approaches to Calculating Posterior Moments*, w: Bernardo J., Berger J., Dawid A., Smith, A., *Bayesian Statistics*, 4, 169-193.
- [11] Gilks W., Best N., Tan K., (1995), *Adaptive Rejection Metropolis Sampling with Gibbs Sampling*, Applied Statistics, 44, 455-472.
- [12] Grzenda W., (2011), *Wykorzystanie modeli drzew decyzyjnych oraz regresji logistycznej do analizy czynników demograficznych oraz społeczno-ekonomicznych wpływających na szanse znalezienia pracy*, Studia Ekonomiczne, 95, 271-277.
- [13] GUS, (2010), *Monitoring Rynku Pracy*, Wejście ludzi młodych na rynek pracy.
- [14] Ibrahim J.G., Chen M.-H., Sinha D., (2001), *Bayesian Survival Analysis*, Springer-Verlag, New York.
- [15] Kryńska E. (red.), (2004), *Polski rynek pracy – niedopasowania strukturalne*, IPiSS, Warszawa.
- [16] Kwiatkowski E., (2007), *Bezrobocie*, Podstawy teoretyczne, WN, Warszawa.
- [17] Lancaster T., (1979), *Econometric Methods for the Duration of Unemployment*, Econometrica, 47, 939-956.
- [18] Merrick J.R., Soyer R., Mazzuchi A., (2002), *A Bayesian Semi-parametric Analysis of the Reliability and Maintenance of Machine Tools*, Technometrics 48, 58-69.
- [19] Osiewalski J., (2001), *Ekonometria bayesowska w zastosowaniach*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- [20] Pipień M., (2006), *Wnioskowanie bayesowskie w ekonometrii finansowej*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- [21] Sinha D., Dey D.K., (1997), *Semiparametric Bayesian Analysis of Survival Data*, Journal of the American Statistical Association, 92, 1195-1212.
- [22] Sinha D., Ibrahim J.G., Chen, M.-H., (2003), *A Bayesian Justification of Cox's Partial Likelihood*, Biometrika, 90(3), 629-641.
- [23] Socha M., Sztanderska U., (2000), *Strukturalne podstawy bezrobocia w Polsce*, PWN, Warszawa.
- [24] Szreder M., (1994), *Informacje a priori w klasycznej i bayesowskiej estymacji modeli regresji*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk.

**BADANIE DETERMINANT POZOSTAWANIA BEZ PRACY OSÓB MŁODYCH
Z WYKORZYSTANIEM SEMIPARAMETRYCZNEGO MODELU COXA****S t r e s z c z e n i e**

Obecnie wśród osób rozpoczynających karierę zawodową obserwuje się szczególnie dużą wartość wskaźnika bezrobocia. Celem niniejszego opracowania jest identyfikacja czynników demograficznych oraz społeczno-ekonomicznych wpływających na długość czasu pozostawania bez pracy tych osób. W badaniu wykorzystano m.in. bayesowski semiparametryczny model Coxa dla danych indywidualnych. Wykorzystanie modelu przeżycia daje możliwość analizy jednoczesnego wpływu wybranych zmiennych objaśniających na czas pozostawania bez pracy. Natomiast podejście bayesowskie umożliwia uwzględnienie w badaniu, za pomocą rozkładów a priori, dodatkowej informacji spoza próby. Estymację modeli przeprowadzono z wykorzystaniem metod Monte Carlo opartych na łańcuchach Markowa, a dokładniej algorytmu ARMS.

Słowa kluczowe: bezrobocie, semiparametryczny model Coxa, wnioskowanie bayesowskie, metody MCMC

**AN ANALYSIS OF UNEMPLOYMENT DURATION DETERMINANTS AMONG YOUNG PEOPLE
USING SEMIPARAMETRIC COX MODEL****A b s t r a c t**

High unemployment rates are observed among people beginning job careers nowadays. The aim of the work is to identify demographic and socio-economic factors influencing the unemployment duration in this age group. In this research, Bayesian semiparametric Cox model for individual data has been used. The advantage of survival model is the possibility of the analysis of the impact of selected independent variables on unemployment duration. The Bayesian approach with a priori distribution makes the use of out of the sample knowledge possible. The model has been estimated using Markov chain Monte Carlo method with ARMS algorithm.

Key words: unemployment, semiparametric Cox model, Bayesian inference, Markov chain Monte Carlo method