

## STUDIA METODOLOGICZNE

**Beata BIESZK-STOLORZ**

### Zastosowanie metod analizy trwania w badaniu form wychodzenia z bezrobocia

**Streszczenie.** *Celem artykułu jest przedstawienie wybranych metod analizy trwania do oceny prawdopodobieństwa wyjścia z bezrobocia dla różnych rodzajów ryzyka konkurującego. W badaniu wykorzystano funkcję skumulowanej częstości i skumulowane prawdopodobieństwo warunkowe oraz dopełnienie do jedności estymatora Kaplana-Meiera. Za pomocą trzech estymatorów porównano prawdopodobieństwo wyrejestrowania z powodu podjęcia pracy, wykreślenia i pozostałych przyczyn. Analizę przeprowadzono na podstawie danych indywidualnych osób zarejestrowanych w Powiatowym Urzędzie Pracy w Szczecinie.*

**Słowa kluczowe:** estymator Kaplana-Meiera, funkcja skumulowanej częstości, skumulowane prawdopodobieństwo warunkowe, ryzyko konkurujące, bezrobocie.

**JEL:** C41, J64

---

W trakcie analizy czasu trwania zjawisk pojawia się problem występowania zdarzeń konkurujących. Na przykład w zastosowaniach inżynierskich ryzyko konkurujące związane jest z wpływem elementów składowych na cały system. Awaria któregokolwiek ze składników powoduje awarię całości. Obserwuje się, który składnik spowodował awarię i czas, w którym system zawodzi.

Gooley, Leisenring, Crowley i Storer (1999) definiują ryzyko konkurujące jako zdarzenie, którego wystąpienie wyklucza zajście innego zdarzenia lub fundamentalnie zmienia prawdopodobieństwo tego innego zdarzenia. Należy założyć, że wystąpienie zdarzenia określonego typu nie ma żadnego wpływu na prawdopodobieństwo jakichkolwiek innych zdarzeń (Crowder, 1994, 1996,

1997). Badana jednostka jest narażona na różne rodzaje ryzyka w tym samym czasie, ale przypuszcza się, że ewentualne zdarzenie wynika tylko z jednego z tych czynników, który nazywa się przyczyną niepowodzenia (Aly, Kochar i McKeague, 1994).

Analiza czasu trwania zjawisk związana jest również z występowaniem obserwacji cenzurowanych. Powstają one, gdy czas trwania dla badanej jednostki jest znany jedynie częściowo. W przypadku analizy danych kohortowych najczęściej występuje cenzurowanie prawostronne, które zachodzi wówczas, gdy zdarzenie końcowe nie wystąpiło w okresie obserwacji. Za obserwacje uznawane są również takie sytuacje, w których badana jednostka znika z pola widzenia lub występuje zdarzenie kończące obserwację, wykluczające, by mogło dojść do zdarzenia właściwego (Pepe, 1991), czyli zdarzenie konkurujące.

Kalbfleisch i Prentice (1980, s. 164) zidentyfikowali trzy różne problemy, które pojawiają się w analizie danych z ryzykiem konkurującym:

- 1) szacowanie zależności pomiędzy zmiennymi oraz współczynnika wystąpienia danego rodzaju zdarzenia;
- 2) badanie współzależności pomiędzy typami zdarzeń w ramach określonego zbioru warunków;
- 3) szacowanie ryzyka awarii z określonego powodu, biorąc pod uwagę usunięcie niektórych lub wszystkich pozostałych przyczyn awarii; problem ten jest uważany za klasyczny w analizie ryzyka konkurującego.

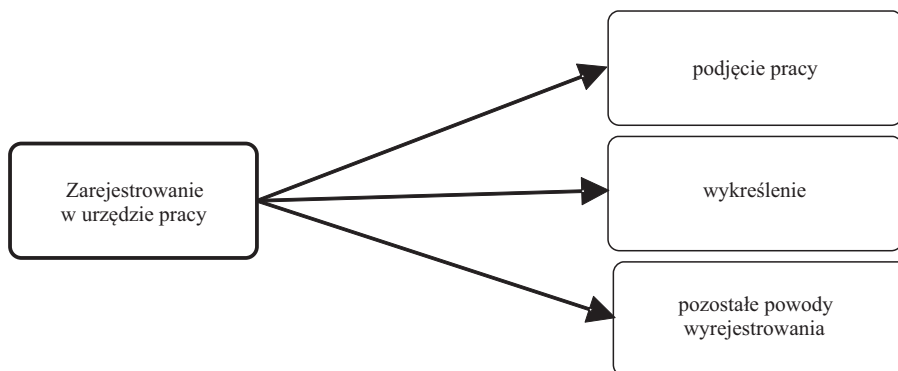
W literaturze obok zdarzeń konkurujących bezwarunkowych rozważa się również warunkowe modele ryzyka konkurującego. W przypadku pierwszego zdarzenia zakłada się jego niezależność. W przypadku tych drugich prawdopodobieństwo przejścia między stanami zależy nie tylko od zmiennych objaśniających, ale również od czasu i typu pobytu we wcześniejszym stanie (Landmesser, 2008b).

Celem badania jest przedstawienie i zastosowanie trzech metod analizy trwania do oceny prawdopodobieństwa wyjścia z bezrobocia dla różnych rodzajów ryzyka konkurującego bezwarunkowego.

### *DANE WYKORZYSTANE W BADANIU*

W badaniu wykorzystano dane indywidualne osób zarejestrowanych w Powiatowym Urzędzie Pracy w Szczecinie (PUP).

W badaniu posłużono się anonimowymi danymi indywidualnymi 22078 osób bezrobotnych zarejestrowanych w 2013 r. w PUP i obserwowanych do końca 2014 r. Zdarzeniem finalizującym obserwację jednostki był moment wyrejestrowania z urzędu. Analizowano czas od zarejestrowania do wyrejestrowania z określonego powodu. W rejestrach urzędów pracy podaje się kilkadziesiąt powodów wyrejestrowania osoby bezrobotnej i tylko kilka z nich związanych jest z podjęciem pracy. Powody te pogrupowano i rozpatrywano trzy grupy zdarzeń konkurujących.

**Wykr. 1. GRUPY ZDARZEŃ KONKURUJĄCYCH**

Źródło: opracowanie własne na podstawie form wyjścia z bezrobocia, które przyjęto w badaniu.

Część obserwacji nie zakończyła się zdarzeniem, czyli wyrejestrowaniem w analizowanym okresie. Stanowią one obserwacje cenzurowane prawostronnie (powód cenzurowania 1). W zestawieniu podano wielkość grup przyczyn wyrejestrowania.

**ZESTAWIENIE POWODÓW WYREJESTROWANIA  
I ODPOWIADAJĄCYCH IM OBSERWACJI CENZUROWANYCH**

Zdarzenia kończące obserwację		Obserwacje cenzurowane			
powód wyrejestrowania	liczba osób	powód cenzurowania 1	liczba osób	powód cenzurowania 2	liczba osób
Praca w sensie ogólnym	9633	zdarzenie nie nastąpiło do końca 2014 r.	1856	zdarzenie nie nastąpiło do końca 2014 r. lub nastąpiło z powodu wykreślenia albo pozostałych powodów	12445
Wykreślenie	8965			zdarzenie nie nastąpiło do końca 2014 r. lub nastąpiło z powodu podjęcia pracy albo pozostałych powodów	13113
Pozostałe	1624			zdarzenie nie nastąpiło do końca 2014 r. lub nastąpiło z powodu podjęcia pracy albo wykreślenia	20454

Źródło: opracowanie własne.

Na każdy z trzech głównych powodów wyrejestrowania miały wpływ powody cząstkowe.

Pracę w sensie ogólnym można podzielić na trzy główne podgrupy: podjęcie pracy lub innego zatrudnienia, rozpoczęcie pracy subsydiowanej przez urząd (m.in. w ramach robót publicznych czy interwencyjnych, utworzonego przez

pracodawcę dodatkowego miejsca pracy z tytułu udzielonej pożyczki czy dofinansowania wynagrodzenia za zatrudnienie bezrobotnego, który ukończył 50 lat) oraz działalność gospodarczą (podjęcie pozarolniczej działalności gospodarczej, przyznanie jednorazowo środków na podjęcie działalności gospodarczej, rozpoczęcie działalności gospodarczej ze środków PFRON).

Wykreślenie jest grupą powodów, w których zarejestrowana osoba bezrobotna nie wykazała się chęcią współpracy z urzędem i została z własnej winy lub na własną prośbę wykreślona z rejestru. Chodzi m.in. o: odmowę bez uzasadnionej przyczyny przyjęcia propozycji odpowiedniego zatrudnienia lub innej pracy zarobkowej, wykonywania prac interwencyjnych lub robót publicznych albo udziału w szkoleniu, stażu, przygotowaniu zawodowym w miejscu pracy; niestawienie się w PUP w wyznaczonym terminie i niepowiadomienie w terminie do 5 dni o uzasadnionej przyczynie tego niestawiennictwa; nieprzedstawienie zaświadczenia o niezdolności do pracy wskutek choroby; odmowę bez uzasadnionej przyczyny udziału w działaniach w ramach Programu Aktywizacja i Integracja (PAI) lub przerwanie z własnej winy udziału w działaniach w ramach PAI; brak gotowości do podjęcia pracy przez okres co najmniej 10 dni; przerwanie z własnej winy szkolenia, stażu, realizacji indywidualnego planu działań oraz wniosek bezrobotnego o wykreślenie z ewidencji.

Pozostałe powody wyrejestrowania, jak wykazały wcześniejsze badania, miały marginalny wpływ na prawdopodobieństwo wyrejestrowania z urzędu, dlatego postanowiono rozpatrywać je razem. Składają się nań m.in.: wyjazd za granicę na okres co najmniej 30 dni, zmiana miejsca zameldowania lub pobytu zainteresowanego poza obszar działania PUP, niezdolność do pracy wskutek choroby lub przebywania w zamkniętym ośrodku odwykowym przez nieprzerwany okres 90 dni, rozpoczęcie szkolenia organizowanego przez podmiot inny niż PUP, zgon, podjęcie nauki w szkole w systemie dziennym, nabycie prawa do emerytury, przyznanie prawa do renty, nabycie prawa do świadczenia rehabilitacyjnego, pobieranie zasiłku stałego, podleganie ubezpieczeniu emerytalno-rentowemu z tytułu stałej pracy jako domownik w gospodarstwie rolnym, pobieranie świadczenia lub dodatku do zasiłku rodzinnego z tytułu samotnego wychowywania dziecka, pobieranie zasiłku dla opiekuna oraz przyznanie prawa do pobierania świadczenia/zasiłku przedemerytalnego.

### *METODYKA BADANIA*

Analiza trwania, powszechnie stosowana w demografii i naukach medycznych do badania czasu trwania życia ludzkiego, może być wykorzystywana do oceny czasu trwania zjawisk ekonomicznych. Można w ten sposób badać: żywotność firm (Markowicz, 2012), aktywność ekonomiczną ludności (Landmesser, 2013), dynamikę ubóstwa w miejskich i wiejskich gospodarstwach domowych (Sączewska-Piotrowska, 2016), ryzyko kredytowe (Matuszyk, 2015; Wyćinka, 2015) i czas trwania bezrobocia (Bieszk-Stolorz, 2013; Bieszk-Stolorz i Markowicz, 2012a). Analizie podlega czas trwania jednostki w danym stanie

(okresy: działania firmy, bezrobocia i spłaty kredytu), który jest zmienną losową  $T$ . Podstawą tego typu analizy jest funkcja trwania, nazywana również funkcją przeżycia, zdefiniowana następująco:

$$S(t) = P(t > T) = 1 - F(T) \quad (1)$$

gdzie:

$T$  — czas trwania zjawiska,

$F(T)$  — dystrybuanta rozkładu zmiennej losowej  $T$ .

Najczęściej stosowany jest tutaj estymator Kaplana-Meiera (Kaplan i Meier, 1958):

$$\hat{S}(t) = \prod_{j: t_j \leq t} \left( 1 - \frac{d_j}{n_j} \right) \quad (2)$$

gdzie:

$d_j$  — liczba zdarzeń w momencie  $t_j$ ,

$n_j$  — liczba jednostek zagrożonych do momentu  $t_j$ .

Funkcja trwania informuje nas o tym, jakie jest prawdopodobieństwo niezajścia zdarzenia co najmniej do czasu  $t$ . W demografii i naukach medycznych, gdzie zdarzeniem jest najczęściej zgon, estymacji podlega prawdopodobieństwo przeżycia jednostki. Czasami wygodniej jest analizować dystrybuantę  $F(T)$ , która wyraża prawdopodobieństwo, że zdarzenie zajdzie najpóźniej do czasu  $t$ . Jeżeli zdarzeniem jest podjęcie pracy przez zarejestrowaną w urzędzie osobę bezrobotną, to estymator funkcji trwania informuje o tym, jakie jest prawdopodobieństwo pozostania w rejestrze, a estymator dystrybuanty pozwala określić, jakie jest prawdopodobieństwo podjęcia pracy.

W badaniu wykorzystano trzy sposoby oceny ryzyka konkurującego bezwzajemnego wywodzące się z analizy trwania.

Niech  $T$  i  $C$  będą zmiennymi losowymi ciągłymi, opisującymi odpowiednio czas do zajścia pewnego zdarzenia i czas do cenzurowania. W przypadku  $K$  konkurujących rodzajów ryzyka obserwacji podlegają pary  $(X, \delta)$ , gdzie  $X = \min(T, C)$  i  $\delta = 0, 1, \dots, K$ . Jeżeli dana obserwacja jest cenzurowana, to  $\delta = 0$  oraz  $\delta = 1, \dots, k$  dla obserwacji kończących się zdarzeniem (jednym z  $k$  konkurujących). W tym kontekście jedno ze zdarzeń  $k$  może być uznane za wydarzenie o podstawowym znaczeniu, a wszystkie inne będą rozpatrywane jako zdarzenia konkurujące.

W przypadku zdarzeń konkurujących powszechnie stosowany jest zwykły estymator Kaplana-Meiera (*KME*), jako dopełnienie do 1, określony następująco:

$$1 - KME = 1 - \hat{S}(t) \quad (3)$$

gdzie  $\hat{S}(t)$  jest estymatorem Kaplana-Meiera.

W tym przypadku jedno ze zdarzeń  $k$  jest uznane za wydarzenie o podstawowym znaczeniu, a wszystkie inne będą rozpatrywane jako obserwacje cenzurowane. Jest to prawdopodobieństwo wystąpienia zdarzenia  $k$  przy założeniu, że występuje tylko jeden rodzaj ryzyka ( $\delta = 1$ ). Zastosowanie tego estymatora prowadzi do przeszacowania wartości prawdopodobieństwa (Klein i Moeschberger, 1984; Sherif Bintu, 2004).

Drugi estymator, funkcja skumulowanej częstości występowania (*Cumulative Incidence Function — CIF*) został po raz pierwszy zaproponowany przez Kalbfleischa i Prentice'a (2002) i jest to skumulowane prawdopodobieństwo wystąpienia zdarzenia z powodu  $k$  przed czasem  $t$  przy założeniu, że jednostka jest narażona na wystąpienie któregośkolwiek ryzyka konkurującego  $k$  (Bryant i Dignam, 2004). Jest on zdefiniowany jako (Klein i Moeschberger, 2003, s. 52):

$$CIF_k(t) = P(t \leq T, \delta = k) = \int_0^t S(u)h_k(u)du = \int_0^t S(u)dH_k \quad \text{dla } k = 1, 2, 3, \dots, K \quad (4)$$

gdzie:

$H_k(t)$  — wyspecyfikowana (dla określonego  $k$ ) funkcja hazardu skumulowanego,  
 $S(t)$  — funkcja przeżycia.

Niech  $t_1 < t_2 < \dots < t_i < \dots < t_n$  będą momentami zachodzenia zdarzeń. Analogicznie do standardowej skumulowanej funkcji w analizie przeżycia, skumulowana funkcja hazardu  $H_k(t)$  dla przyczyny  $k$  może być określona przez estymator Nelsona-Aalena:

$$\hat{H}_k(t) = \sum_{j: t_j \leq t} \frac{d_{kj}}{n_j} \quad (5)$$

gdzie:

$d_{kj}$  — liczba zdarzeń z powodu wystąpienia przyczyny  $k$ ,

$n_j$  — liczba osób narażonych na ryzyko w czasie  $t_j$ .

Ogólna funkcja przeżycia  $S(t)$  może być określona na podstawie estymatora Kaplana-Meiera (2). Po połączeniu estymatorów (4) i (5) funkcję skumulowanej

częstości występowania z powodu zaistnienia przyczyny  $k$  (Marubini i Valsecchi, 1995) można oszacować jako:

$$\widehat{CIF}_k(t) = \sum_{j: t_{j-1} \leq t} \hat{S}(t_{j-1}) \frac{d_{kj}}{n_j} \quad (6)$$

Funkcja skumulowanej częstości pozwala określić wzorce zajścia zdarzenia z powodu  $k$  oraz ocenić, w jakim stopniu każdy powód przyczynia się do całkowitego wyniku.

Jeśli  $\sum_{k=1}^K d_{kj} = d_j$ , to prawdziwa jest zależność:

$$\sum_{k=1}^K \widehat{CIF}_k(t) = 1 - \hat{S}(t) \quad (7)$$

Jeżeli nie ma zdarzeń konkurujących, to zachodzi równość:

$$\widehat{CIF}(t) = 1 - \hat{S}(t) \quad (8)$$

Trzeci estymator — skumulowane warunkowe prawdopodobieństwo wystąpienia zdarzenia (*Cumulative Probability Conditional — CPC*), to zdarzenie zachodzące z powodu  $k$  przed czasem  $t$ , pod warunkiem że przed czasem  $t$  nie zaszło żadne pozostałe zdarzenie konkurujące określone wzorem (Kleinbaum i Klein, 2005, s. 420):

$$CPC_k(t) = P(t \leq T_k, T_P > t) \quad \text{dla } k = 1, 2, 3, \dots, K \quad (9)$$

gdzie  $T_p$  jest czasem do wystąpienia jakiegokolwiek zdarzenia konkurującego.

Estymator  $CPC_k$  jest określony następująco (Pepe, 1991; Pepe i Mori, 1993):

$$\widehat{CPC}_k(t) = \frac{\widehat{CIF}_k(t)}{1 - \widehat{CIF}_k(t)} \quad (10)$$

gdzie  $\widehat{CIF}_k(t)$  jest skumulowaną funkcją częstości dla wszystkich rodzajów ryzyka konkurującego różnych od  $k$ .

Trzy przedstawione wyżej estymatory są związane nierównościami (Klein i Bajorunaite, 2004):

$$\widehat{CIF}_k(t) \leq 1 - KME \leq \widehat{CPC}_k(t) \quad (11)$$

przy czym równości zachodzą wtedy, gdy jest tylko jedno zdarzenie kończące obserwację (brak zdarzeń konkurujących).

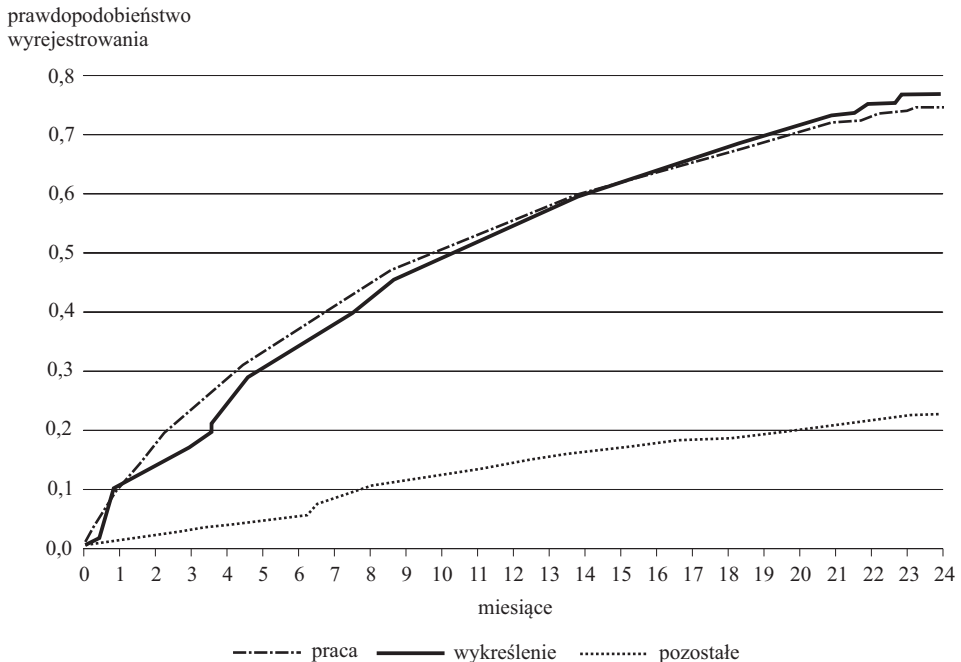
Innym sposobem oceny ryzyka konkurującego jest zastosowanie wielomianowego modelu logitowego. W analizie trwania traktuje się je jako modele ha-

zardu z czasem dyskretnym. Parametry tych modeli pozwalają wyznaczyć szansę względną zajścia zdarzeń konkurujących (Landmesser, 2008a; Bieszk-Stolorz i Markowicz, 2012b).

### *ANALIZA PRAWDOPODOBIEŃSTWA WYREJESTROWANIA Z URZĘDU PRACY*

W pierwszym etapie badania wyznaczono wartości dla estymatora Kaplana-Meiera, określonego wzorem (3). Wyznacza się go przy założeniu, że nie występuje ryzyko konkurujące. W związku z tym obserwacje cenzurowane są tutaj dwojakiego rodzaju — takie, które nie zakończyły się zdarzeniem do końca 2014 r. lub zakończyły się z innego powodu. Każda grupa powodów wyrejestrowania jest analizowana oddzielnie. Wielkość analizowanych grup i odpowiadających im obserwacji cenzurowanych jest zawarta w zestawieniu (powód cenzurowania 2). Estymatory przedstawiono na wyk. 2. Od drugiego do szesnastego miesiąca trwania w bezrobociu prawdopodobieństwo podjęcia pracy było większe niż prawdopodobieństwo wykreślenia. Następnie sytuacja uległa zmianie i od szesnastego miesiąca do końca obserwacji prawdopodobieństwo wykreślenia z rejestru urzędu pracy nieznacznie przewyższyło prawdopodobieństwo podjęcia pracy. Prawdopodobieństwo wyrejestrowania z pozostałych powodów było niskie i na koniec okresu obserwacji wyniosło maksymalnie 0,0228.

**Wykr. 2. DOPEŁNIENIA DO JEDNOŚCI ESTYMATORÓW KAPLANA-MEIERA (1-KME)**

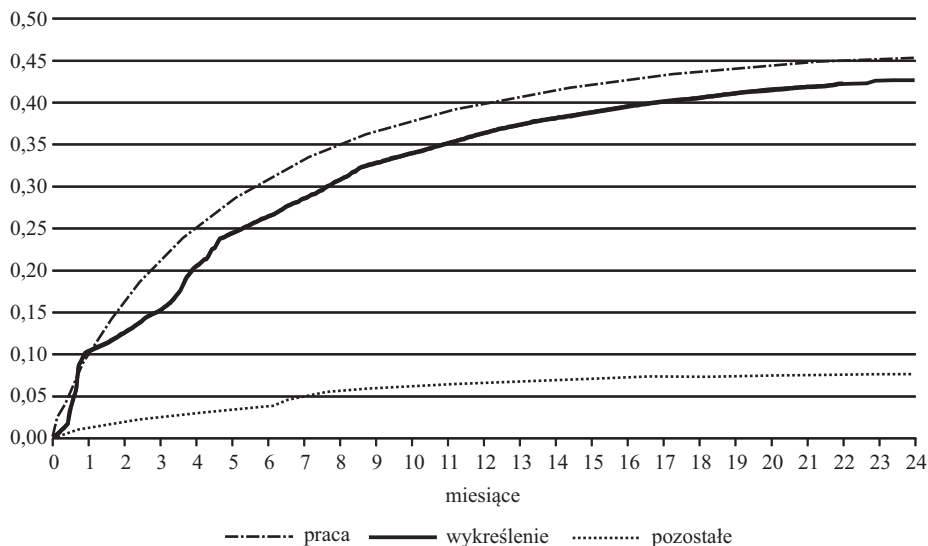




W drugim etapie analizy wyznaczono estymatory skumulowanej częstości występowania ( $CIF_k$ ) dla trzech grup powodów wyrejestrowania z urzędu pracy ( $k = 1, 2, 3$ ), określone wzorem (6). W tym przypadku przyjęto trzy rodzaje zdarzeń kończących obserwację: podjęcie pracy, wykreślenie i pozostałe, będące zdarzeniami konkurującymi. Danymi cenzurowanymi są obserwacje, które nie zakończyły się przed końcem 2014 r. (powód cenzurowania 1). Prawdopodobieństwo wyrejestrowań z powodu podjęcia pracy było wyższe od prawdopodobieństwa wykreślenia od drugiego miesiąca od momentu zarejestrowania (wykr. 3). Wpływ pozostałych powodów był marginalny, a prawdopodobieństwo ich wystąpienia nie przekraczało 0,0765.

### Wykr. 3. ESTYMATORY SKUMULOWANEJ FUNKCJI CZĘSTOŚCI ( $CIF_k$ )

prawdopodobieństwo  
wyrejestrowania



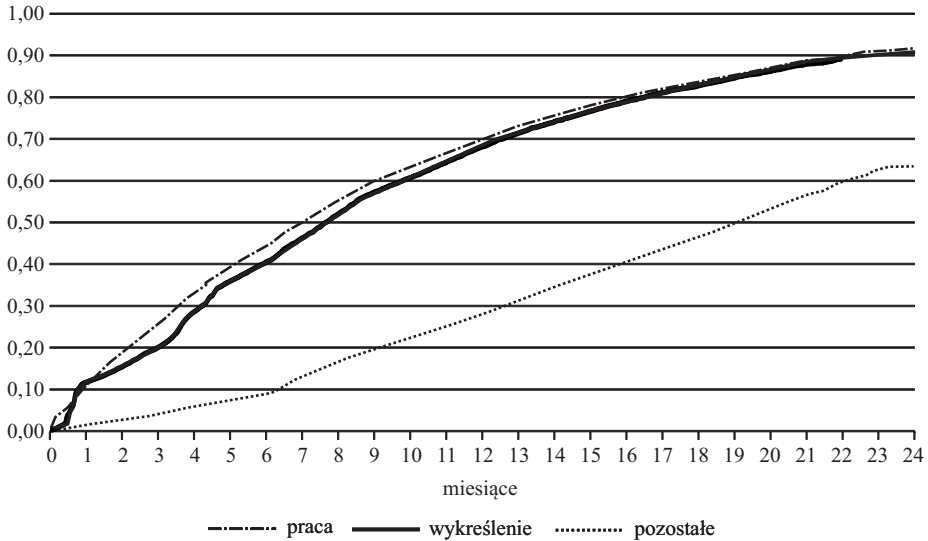
Źródło: jak przy wykr. 2.

Trzeci etap analizy polegał na wyznaczeniu estymatorów skumulowanego warunkowego prawdopodobieństwa wystąpienia zdarzenia ( $CPC_k$ ). W tym przypadku, podobnie jak w drugim etapie analizy, uwzględniono trzy grupy powodów wyrejestrowania z urzędu pracy ( $k = 1, 2, 3$ ), które stanowią zdarzenia konkurujące. Skorzystano ze wzoru (10). Obserwacje niezakończone przed końcem 2014 r. przyjęto za cenzurowane (powód cenzurowania 1). Wprawdzie od początku drugiego miesiąca od zarejestrowania prawdopodobieństwo warunkowe podjęcia pracy było wyższe od prawdopodobieństwa warunkowego wykreślenia, jednak od trzynastego miesiąca wartości obu estymatorów zbliżyły się do siebie (wykr. 4). Prawdopodobieństwo warunkowe wyrejestrowania z powodu

pozostałych przyczyn było mniejsze niż w przypadku podjęcia pracy i wykreślenia, jednak stosunkowo duże i na koniec obserwacji przyjęło wartość 0,634.

#### Wykr. 4. ESTYMATORY SKUMULOWANEGO WARUNKOWEGO PRAWDOPODOBIEŃSTWA ( $CPC_k$ )

prawdopodobieństwo  
wyrejestrowania



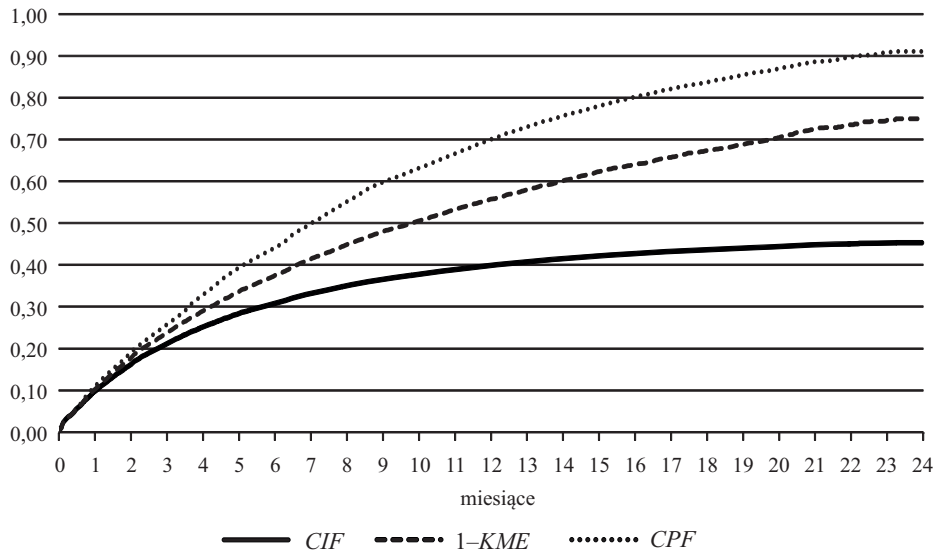
Źródło: jak przy wyk. 2.

Na wyk. 5 zestawiono trzy oszacowane estymatory dla zdarzenia podjęcie pracy. Jest to wizualizacja nierówności (11) opisującej zależność między nimi. Na uwagę zasługuje fakt regularnej krzywizny wszystkich trzech krzywych. Takiej własności nie miały estymatory prawdopodobieństwa wykreślenia i pozostałych powodów, dla których zauważalne są krótkotrwałe skoki wartości.

Szczegółowa analiza powodów wyrejestrowania pozwoliła na ocenę przyczyny zwiększonego wzrostu tych estymatorów. Dla grupy „pozostałe” skok wartości nastąpił w siódmym miesiącu od zarejestrowania. Był spowodowany zwiększoną liczbą wyrejestrowań z powodu przyznania zarejestrowanej osobie bezrobotnej prawa do pobierania świadczenia/zasiłku przedemerytalnego. Dla grupy „wykreślenie” skok w miesiącach pierwszym i czwartym był natomiast związany ze zwiększoną liczbą wyrejestrowań z powodu niestawienia się osoby bezrobotnej w PUP w wyznaczonym terminie. Jak już wcześniej wspomniano, przyjęcie uproszczenia, jakim jest zastosowanie w badaniu estymatora Kaplana-Meiera z dopełnieniem do 1, prowadzi do przeszacowania wartości ryzyka konkurującego  $CIF_k$ . W przypadku przedstawionej analizy jest ono bardzo duże i w 24. miesiącu, kończącym obserwację, jest większe o 84%.

### Wykr. 5. PORÓWNANIE ESTYMATORÓW RYZYKA KONKURUJĄCEGO DLA ZDARZENIA „PODJĘCIE PRACY”

prawdopodobieństwo  
wyrejestrowania



Źródło: opracowanie własne.

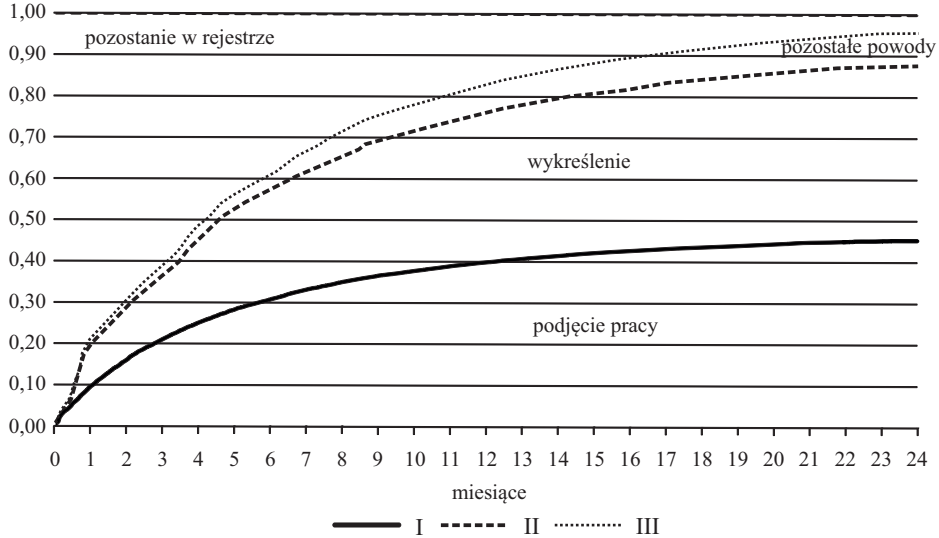
### Podsumowanie

Wstępna analiza danych indywidualnych uzyskanych z PUP w Szczecinie wskazuje na to, że wykreślenie było najczęstszą przyczyną wyrejestrowania z urzędu. Dokładniejszych wyników dostarcza badanie przeprowadzone z wykorzystaniem trzech opisanych estymatorów. Estymator skumulowanej funkcji częstości ( $CIF_k$ ) potwierdził, że praca rzeczywiście była najbardziej prawdopodobnym powodem wyrejestrowania. Pozostałe dwa wskazały, że prawdopodobieństwo wykreślenia było zbliżone do prawdopodobieństwa podjęcia pracy. Otrzymane wyniki sugerują, że chęć zatrudnienia nie była główną przyczyną zarejestrowania się w urzędzie. Pozostałe powody miały znaczenie marginalne, chociaż prawdopodobieństwo warunkowe ich zajścia było dość duże.

Aby dokonać oceny wyników, zsumowano estymatory skumulowanej funkcji częstości ( $CIF_k$ ) w sposób przedstawiony na wyk. 6. Gdyby nie było obserwacji cenzurowanych, tj. w przypadku opisanego badania wszystkie osoby z analizowanej kohorty zostałyby wyrejestrowane do końca 2014 r., to suma tych estymatorów dla wszystkich rodzajów ryzyka w 24. miesiącu byłaby równa 1, ale ponieważ jest równa 0,956, to prawdopodobieństwo pozostania w rejestrze w 24. miesiącu, czyli w stanie bezrobocia, osób z analizowanej kohorty jest równe 0,044.

**Wykr. 6. ZSUMOWANE ESTYMATORY SKUMULOWANEJ FUNKCJI CZĘSTOŚCI ( $CIF_k$ )**

skumulowane prawdopodobieństwo  
wyrejestrowania



U w a g a. I — praca, II — praca+wykreślenie, III — praca+wykreślenie+pozostałe.  
Ź r ó d ł o: opracowanie własne.

Odległości<sup>1</sup> między krzywymi wyrażają prawdopodobieństwo wyrejestrowania z określonego powodu. Odległość między krzywą III a prostą o wartości 1 jest równa prawdopodobieństwu pozostania w rejestrze. Jak widać, zmniejszało się ono w okresie badania. Po 12 miesiącach od zarejestrowania było równe 0,17. Jest to jednocześnie prawdopodobieństwo wejścia w stan długotrwałego bezrobocia. Prawdopodobieństwo pozostania w rejestrze urzędu po 24 miesiącach (koniec okresu obserwacji) było równe 0,04. Wyrejestrowanie z powodu podjęcia pracy po 12 miesiącach od zarejestrowania było równe 0,40, a po 24 miesiącach — 0,45. Analogicznie prawdopodobieństwo wykreślenia było równe 0,36 i 0,43.

Z przedstawionego badania wynika również ważna uwaga metodyczna. Jeżeli występują różne rodzaje zdarzeń kończących obserwację, to warto zastosować modele pozwalające oszacować ryzyko konkurujące. W przypadku czasu trwania bezrobocia rejestrowanego pozwalają one wyznaczyć prawdopodobieństwo podjęcia określonego rodzaju pracy oraz porównać je z innymi przyczynami wyrejestrowania. Potraktowanie innych zdarzeń kończących obserwację jako cenzurowanych może prowadzić do znacznego przeszacowania funkcji skumulowanej częstości.

**dr hab. Beata Bieszk-Stolorz** — Uniwersytet Szczeciński

<sup>1</sup> Odległość między krzywymi jest mierzona według metryki  $d((x_1, y_1), (x_2, y_2)) = |y_2 - y_1|$ .

## LITERATURA

- Aly, E.A.A., Kochar, S., McKeague, E. (1994). Some Tests for Comparing Cumulative Incidence Functions and Cause-Specific Hazard Rates. *Journal of the American Statistical Association*, Vol. 89, No. 427.
- Bieszk-Stolorz, B. (2013). *Analiza historii zdarzeń w badaniu bezrobocia*. Szczecin: Volumina.pl Daniel Krzanowski.
- Bieszk-Stolorz, B., Markowicz, I. (2012a). *Modele regresji Coxa w analizie bezrobocia*. Warszawa: CeDeWu.
- Bieszk-Stolorz, B., Markowicz, I. (2012b). Wykorzystanie wielomianowego modelu logitowego do oceny szansy podjęcia pracy przez bezrobotnych. W: K. Jajuga, M. Walesiak (red.), *Taksonomia 19. Klasyfikacja i analiza danych — teoria i zastosowania. Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, nr 242.
- Bryant, J., Dignam, J.J. (2004). Semiparametric models for cumulative incidence functions. *Biometrics*, Vol. 60, No. 1.
- Crowder, M. (1994). Identifiability Crises in Competing Risks. *International Statistical Review*, Vol. 62, No. 3.
- Crowder, M. (1996). On assessing independence of competing risks when failure times are discrete. *Lifetime Data Analysis*, Vol. 2, No. 2.
- Crowder, M. (1997). A test for independence of competing risks with discrete failure times. *Lifetime Data Analysis*, Vol. 3, No. 3.
- Gooley, T.A., Leisenring, W., Crowley, J., Storer, B.E. (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*, Vol. 18, No. 6.
- Kalbfleisch, J.D., Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data. Second Edition*. Hoboken: John Wiley & Sons.
- Kaplan, E.L., Meier, P. (1958). Non-parametric estimation from incomplete observations. *Journal of American Statistical Association*, Vol. 53.
- Klein, J.P., Bajorunaite, R. (2004). *Inference for Competing Risks*. W: N. Balakrishnan, C.R. Rao (red.), *Handbook of Statistics: Advances in Survival Analysis*, Vol. 23. Elsevier.
- Klein, J.P., Moeschberger, M.L. (1984). Asymptotic bias of the product limit estimator under dependent competing risks. *Indian Journal of Productivity, Reliability and Quality Control*, Vol. 9.
- Klein, J.P., Moeschberger, M.L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data. Second Edition*. Nowy Jork: Springer-Verlag.
- Kleinbaum, D., Klein, M. (2005). *Survival Analysis. A Self-Learning Text*. Springer.
- Landmesser, J.M. (2008a). *Aktywność ekonomiczna ludności: klasyfikacja osób za pomocą wielomianowych modeli logitowych oraz jej związek z modelami hazardu dla czasów trwania*. W: K. Jajuga, M. Walesiak (red.), *Taksonomia 15. Klasyfikacja i analiza danych — teoria i zastosowania. Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, nr 7.
- Landmesser, J.M. (2008b). *Modele ryzyka konkurencyjnego dla czasu trwania czynności*. W: T. Trzaskalik (red.), *Modelowanie preferencji a ryzyko '08. Prace Naukowe Akademii Ekonomicznej w Katowicach*.
- Landmesser, J. (2013). *Wykorzystanie metod analizy czasu trwania do badania aktywności ekonomicznej ludności w Polsce*. Warszawa: Wydawnictwo SGGW.
- Markowicz, I. (2012). *Statystyczna analiza żywotności firm*. Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.
- Marubini, E., Valsecchi, M. (1995). *Analysing Survival Data from Clinical Trials and Observational Studies*. John Wiley & Sons.
- Matuszyk, A. (2015). *Zastosowanie analizy przetrwania w ocenie ryzyka kredytowego klientów indywidualnych*. Warszawa: CeDeWu.

- Pepe, M.S. (1991). Inference for Events with Dependent Risks in Multiple Endpoint Studies. *Journal of the American Statistical Association*, Vol. 86, No. 415.
- Pepe, M.S., Mori, M. (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine*, Vol. 12.
- Sączewska-Piotrowska, A. (2016). Dynamika ubóstwa w miejskich i wiejskich gospodarstwach domowych. *Wiadomości Statystyczne*, nr 7, s. 39—59.
- Sherif Bintu, N. (2008). *A comparison of Kaplan-Meier and cumulative incidence estimate in the presence or absence of competing risks in breast cancer data*. Master's Thesis, University of Pittsburgh.
- Wycinka, E. (2015). Modelowanie czasu do zaprzestania spłat rat kredytu lub wcześniejszej spłaty kredytu jako zdarzeń konkurujących. *Problemy Zarządzania*, Vol. 13, nr 3 (55), t. 2.

**Summary.** *The purpose of this article is to present selected methods of the survival analysis to evaluate the probability of leaving unemployment for the various types of competing risks. Complement to the unity of the Kaplan-Meier estimator, cumulative incidence function and cumulative conditional probability were used in the study. With these three estimators, the probability of deregistering caused by undertaking work, refusal and other causes were compared. The analysis was based on data from the Powiat Labour Office in Szczecin.*

**Keywords:** Kaplan-Meier estimator, cumulative incidence function, cumulative conditional probability, competing risks, unemployment.