

KAMIL WILAK

STRUKTURALNE MODELE SZEREGÓW CZASOWYCH W ESTYMACJI STOPY BEZROBOCIA W DEZAGREGACJI NA WOJEWÓDZTWA, PŁEĆ I WIEK

1. WSTĘP

Jednym z podstawowych wskaźników opisujących sytuację na rynku pracy jest stopa bezrobocia, określona jako stosunek liczby osób bezrobotnych do liczby osób aktywnych zawodowo. Głównymi źródłami wiedzy na temat stopy bezrobocia w Polsce są Badanie Aktywności Ekonomicznej Ludności (BAEL) oraz rejestr osób bezrobotnych. W przypadku BAEL, za bezrobotne uważa się te osoby w wieku 15–74 lat, które w badanym tygodniu nie pracowały, lecz aktywnie poszukiwały pracy i były gotowe podjąć ją w okresie 2 tygodni. Osobami aktywnymi zawodowo wg BAEL są wszystkie osoby pracujące lub bezrobotne (GUS 2014a). Natomiast w przypadku bezrobocia rejestrowanego, za bezrobotne uważa się osoby zarejestrowane w powiatowym urzędzie pracy, zaś osobami aktywnymi zawodowo są osoby bezrobotne i pracujące w jednostkach sektora publicznego lub prywatnego (GUS 2014b). Różnice w definicjach osób bezrobotnych i aktywnych zawodowo sprawiają, że stopa bezrobocia rejestrowanego może znacznie różnić się od stopy bezrobocia wg BAEL.

Do szacowania stopy bezrobocia na podstawie BAEL, Główny Urząd Statystyczny stosuje estymację bezpośrednią. Estymacja ta wykorzystuje tylko informacje z wylosowanej próby. Zaletą tego podejścia jest nieobciążoność estymatorów, natomiast wadą jest duża wariancja w przypadku małej liczby jednostek z badanej domeny, wylosowanych do próby. Estymatory o dużej wariancji uznawane są za mało precyzyjne. Domeny, w których liczba wylosowanych do badania jednostek jest tak mała, że estymacja bezpośrednia cechuje się zbyt małą precyzją, nazywane są małymi domenami (ang. *small domain*).

Oszacowania stopy bezrobocia są publikowane przez GUS dla województw w domenach określonych oddzielnie przez płeć, miejsce zamieszkania, poziom wykształcenia. Ze względu na zbyt dużą wariancję estymatorów bezpośrednich, GUS nie publikuje m.in. oszacowań dla województw w domenach określonych przez płeć i wiek łącznie. W niniejszej pracy rozważa się estymację stopy bezrobocia w przekroju województw dla trzech grup wieku (15–24 lat, 25–44 lat, 45–59/64 lat) z uwzględnieniem płci, co stanowi łącznie $3 \times 2 = 6$ domen dla województwa. W badaniu przepro-

wadzonym w niniejszym artykule ograniczono się do estymacji na podstawie danych z województwa wielkopolskiego.

Metodami szacowania w warunkach małej liczebności próby zajmuje się statystyka małych obszarów (ang. *small area estimation* – *SAE*), inaczej zwana statystyką małych domen lub estymacją dla małych domen. Sposobem na zwiększenie precyzji oszacowań w małych domenach jest zastosowanie bazującej na modelach statystycznych estymacji pośredniej, która wykorzystuje informacje spoza części próby, wylosowanej z rozpatrywanej domeny. Są to m.in. dane administracyjne, dane spoza badanej domeny, dane z wcześniej przeprowadzonych badań. Zmienne wykorzystane do zwiększenia precyzji estymacji badanego parametru nazywane są zmiennymi pomocniczymi. Estymacja pośrednia charakteryzuje się zazwyczaj mniejszą wariancją niż estymacja bezpośrednia, natomiast w przypadku źle dobranego modelu może cechować się znaczną obciążonością. Przegląd metod stosowanych w statystyce małych obszarów można znaleźć w pracach Rao (2003), Domańskiego, Pruski (2001). Zastosowanie metod statystyki małych obszarów w estymacji stopy bezrobocia na polskim rynku pracy można znaleźć w publikacji Gołaty (2004).

W literaturze światowej, w kontekście estymacji pośredniej charakterystyk rynku pracy, popularne jest podejście zwane pożyczaniem mocy w czasie (ang. *borrowing strength across time*). Polega ono na wykorzystaniu informacji z wcześniej przeprowadzonych badań, poprzez modelowanie szeregów czasowych składających się z oszacowań bezpośrednich. W tym celu stosowane są m.in. strukturalne modele szeregów, które można przedstawić w postaci dynamicznych modeli liniowych. Modele te zbudowane są z trendu, sezonowości i składnika systematycznego regresji liniowej. Zastosowanie takich modeli można znaleźć m.in. w pracach Brakela, Kriega (2008, 2009, 2010), Pfeiffermana i in. (2005), Pfeiffermana, Tillerera (2006). W artykule Wilaka (2013) zastosowano strukturalny model szeregu czasowego do danych z Badania Aktywności Ekonomicznej Ludności.

Celem niniejszego artykułu jest ocena jakości estymacji pośredniej stopy bezrobocia wg BAEL w przekroju województw dla sześciu domen określonych przez wiek i płeć, wykorzystującej strukturalne modele szeregów czasowych. Przedmiotem badania jest także korelacja pomiędzy stopą bezrobocia wg BAEL a stopą bezrobocia rejestrowanego. W niniejszej pracy rozważa się wykorzystanie tej drugiej jako zmiennej pomocniczej. Weryfikacji poddano hipotezę głoszącą, że estymacja pośrednia stopy bezrobocia w województwach dla małych domen określonych według płci i wieku, wykorzystująca strukturalne modele szeregów czasowych, cechuje się lepszą jakością niż estymacja bezpośrednia. Jakość estymatorów rozważa się pod kątem ich dokładności i precyzji. Z dwóch estymatorów dokładniejszy jest ten, który cechuje się mniejszą obciążonością, natomiast precyzyjniejszy jest ten, którego odchylenie standardowe jest mniejsze.

Dla weryfikacji postawionej hipotezy przeprowadzono eksperyment Monte Carlo z wykorzystaniem danych jednostkowych z Badania Aktywności Ekonomicznej Ludności z lat 2000–2009. Na podstawie tych danych stworzono pseudo-populację,

dla której szacowano stopę bezrobocia. Porównania jakości estymatorów pośrednich i bezpośrednich dokonano na podstawie ich obciążeń, odchyłeń standardowych, błędów średniokwadratowych i średnich błędów bezwzględnych, wyznaczonych na podstawie eksperymentu.

Metoda estymacji pośredniej wykorzystana w niniejszej pracy jest zaczerpnięta ze światowej literatury. Wkładem autora jest natomiast jej adaptacja do warunków polskich oraz ocena jakości estymacji pośredniej stopy bezrobocia w małych domenach określonych przez województwo, płeć i wiek, wykorzystującej tę metodę. Wartością dodaną jest również fakt, że autor pracuje na danych rzeczywistych z Badania Aktywności Ekonomicznej Ludności.

2. PRZEDSTAWIENIE MODELU

Model opisany w tej części pracy jest uproszczoną, jednowymiarową wersją modelu przedstawionego przez Brakela, Kriega (2010). Uproszczenie to wiąże się z pominięciem w modelu korelacji między składnikami trendów z różnych domen.

Niech $\theta_{d,t}$ oznacza parametr populacji w domenie d ($d = 1, \dots, D$) w czasie t ($t = 1, \dots, T$), a $Y_{d,t}$ wartość estymatora bezpośredniego tego parametru. Można wtedy zapisać¹:

$$Y_{d,t} = \theta_{d,t} + e_{d,t}, \quad (1)$$

gdzie $e_{d,t}$ jest błędem estymacji. Na potrzeby badania zakłada się, że próby w poszczególnych okresach losowane są niezależnie. Z nieobciążoności estymatora $Y_{d,t}$ wynika, że wartość oczekiwana błędu $e_{d,t}$ wynosi $E(e_{d,t}) = 0$. Wariancję $Var(e_{d,t})$ można oszacować, uwzględniając przy tym schemat losowania próby.

W modelu Brakela, Kriega (2010) zakłada się, że parametr $\theta_{d,t}$ można przedstawić za pomocą strukturalnego modelu szeregu czasowego:

$$\theta_{d,t} = L_{d,t} + S_{d,t} + x_{d,t}\beta_{d,t} + \varepsilon_{d,t}, \quad (2)$$

$$\varepsilon_{d,t} \sim NID(0, \sigma_{\varepsilon,d}^2). \quad (3)$$

gdzie $L_{d,t}$ oznacza trend stochastyczny, $S_{d,t}$ sezonowość stochastyczną, $x_{d,t}$ zmienną pomocniczą, β_t współczynnik regresji a $\varepsilon_{d,t}$ reszty losowe, które odzwierciedlają zmienność parametru populacji nietłumaczoną przez pozostałe składniki.

¹ W literaturze częściej używa się zapisu $\theta_{d,t} = Y_{d,t} + e_{d,t}$, natomiast na potrzeby wyprowadzenia estymatora pośredniego stosuje się zapis $Y_{d,t} = \theta_{d,t} + e_{d,t}$.

Pierwszy składnik modelu, trend stochastyczny $L_{d,t}$, jest postaci²:

$$L_{d,t} = L_{d,t-1} + R_{d,t}, \quad (4)$$

$$R_{d,t} = R_{d,t-1} + \eta_{R,d,t}, \quad \eta_{R,d,t} \sim NID(0, \sigma_{R,d}^2), \quad (5)$$

gdzie $R_{d,t}$ jest składnikiem losowym, który odzwierciedla nachylenie krzywej trendu, a $\eta_{R,d,t}$ jest składnikiem losowym odpowiedzialnym za zmiany tego nachylenia w czasie.

Drugi składnik modelu, sezonowość stochastyczna $S_{d,t}$, to kwartalna sezonowość trygonometryczna (*trigonometric seasonal*) postaci³:

$$S_{d,t} = S_{d,t,1} + S_{d,t,2}, \quad (6)$$

gdzie dla $j = 1, 2$:

$$S_{d,t,j} = \cos\left(\frac{j\pi}{2}\right) S_{d,t-1,j} + \sin\left(\frac{j\pi}{2}\right) S_{d,t-1,j}^* + \eta_{S,d,t,j}, \quad \eta_{S,d,t,j} \sim NID(0, \sigma_{S,d,j}^2), \quad (7)$$

$$S_{d,t,j}^* = -\sin\left(\frac{j\pi}{2}\right) S_{d,t-1,j} + \cos\left(\frac{j\pi}{2}\right) S_{d,t-1,j}^* + \eta_{S^*,d,t,j}, \quad \eta_{S^*,d,t,j} \sim NID(0, \sigma_{S^*,d,j}^2). \quad (8)$$

Składniki losowe $\eta_{S,d,t,j}$ i $\eta_{S^*,d,t,j}$ odpowiadają za zmiany w czasie wartości składnika sezonowości.

W trzecim składniku modelu, systematycznym składniku regresji liniowej $x_{d,t}\beta_{d,t}$, współczynnik regresji $\beta_{d,t}$ opisany jest są za pomocą błędzenia losowego:

$$\beta_{d,t} = \beta_{d,t-1} + \eta_{\beta,d,t}, \quad \eta_{\beta,d,t} \sim NID(0, \sigma_{\beta,d}^2). \quad (9)$$

Zakłada się brak zależności pomiędzy składnikami $\eta_{R,d,t}$, $\eta_{S,d,t,1}$, $\eta_{S^*,d,t,1}$, $\eta_{S,d,t,2}$, $\eta_{S^*,d,t,2}$, $\eta_{\beta,d,t}$, a także brak zależności pomiędzy składnikami losowymi z różnych domen.

Wstawiając równanie (2) do (1) otrzymuje się strukturalny model dla szeregu czasowego oszacowań bezpośrednich $Y_{d,1}, \dots, Y_{d,T}$ ($d = 1, \dots, D$):

$$Y_{d,t} = L_{d,t} + S_{d,t} + x_{d,t}\beta_{d,t} + \varepsilon_{d,t} + e_{d,t}. \quad (10)$$

Ideą estymacji pośredniej parametrów populacji $\theta_{d,t}$ ($d = 1, \dots, D$, $t = 1, \dots, T$) za pomocą strukturalnych modeli szeregów czasowych jest oczyszczenie oszacowań bez-

² W literaturze można znaleźć również inne propozycje trendu stochastycznego.

³ W literaturze można znaleźć również inne propozycje sezonowości stochastycznej.

pośrednich $Y_{d,t}$ z błędów oszacowań $e_{d,t}$. W celu estymacji parametrów modelu (10) zakłada się, że błędy oszacowań $e_{d,t}$ mają rozkład normalny. Wtedy suma składnika resztowego i błędu oszacowania $v_{d,t} = \varepsilon_{d,t} + e_{d,t}$ ma również rozkład normalny. Wartość oczekiwana składnika $v_{d,t}$ wynosi $E(v_{d,t}) = 0$, natomiast o wariancji tego składnika Brakel, Krieg (2010) zakładali, że jest równa:

$$\text{Var}(v_{d,t}) = \sigma_{v,d}^2 \text{Var}(Y_{d,t}). \quad (11)$$

gdzie $\sigma_{v,d}^2$ jest stałą, którą należy oszacować.

Oczyszczenie $Y_{d,t}$ z błędów oszacowań $e_{d,t}$, wiąże się z usunięciem składnika $v_{d,t}$. A więc oprócz błędów oszacowań $e_{d,t}$ usuwa się także część zmienności parametru populacji w postaci składnika resztowego $\varepsilon_{d,t}$. Stąd ważne jest, aby składnik $v_{d,t}$ był zdominowany przez $e_{d,t}$. Wówczas wariancja składnika $v_{d,t}$ będzie bliska wariancji $\text{Var}(e_{d,t})$ błędu estymatora bezpośredniego, co oznacza, że wartość $\sigma_{v,d}^2$ jest bliska 1.

Jeżeli przedstawimy strukturalny model szeregu czasowego opisanego równaniem (1)–(11) za pomocą notacji macierzowej otrzymamy klasyczną postać modelu przetrzeni stanów:

$$Y_{d,t} = \mathbf{Z}\boldsymbol{\alpha}_{d,t} + v_{d,t}, \quad (12)$$

$$\boldsymbol{\alpha}_{d,t} = \mathbf{T}\boldsymbol{\alpha}_{d,t-1} + \boldsymbol{\eta}_{d,t}, \quad (13)$$

gdzie \mathbf{Z} to wektor znanych współczynników regresji:

$$\mathbf{Z} = (1, 0, 1, 0, 1, 0, 1), \quad (14)$$

$\boldsymbol{\alpha}_t$ to wektor nieznanych parametrów stanu:

$$\boldsymbol{\alpha}_{d,t} = (L_{d,t}, R_{d,t}, S_{d,t,1}, S_{d,t,1}^*, S_{d,t,2}, S_{d,t,2}^*, \beta_{d,t})^T, \quad (15)$$

\mathbf{T} to macierz przejścia:

$$\mathbf{T} = \text{blockdiag}(\mathbf{T}^R, \mathbf{T}^{S_1}, \mathbf{T}^{S_2}, \mathbf{T}^\beta), \quad (16)$$

$$\mathbf{T}^R = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad (17)$$

$$\mathbf{T}^{S_j} = \begin{pmatrix} \cos(j\pi/2) & \sin(j\pi/2) \\ -\sin(j\pi/2) & \cos(j\pi/2) \end{pmatrix}, \quad (18)$$

$$\mathbf{T}^\beta = 1, \quad (19)$$

$\boldsymbol{\eta}_{d,t}$ to wektor składników losowych, odzwierciedlających zmiany w czasie wartości parametrów stanu:

$$\boldsymbol{\eta}_{d,t} = (0, \eta_{R,d,t}, \eta_{S,d,t,1}, \eta_{S^*,d,t,1}, \eta_{S,d,t,2}, \eta_{S^*,d,t,2}, \eta_{\beta,d,t}), \quad (20)$$

$$\boldsymbol{\eta}_{d,t} \sim N(\mathbf{0}, \mathbf{Q}_d), \quad (21)$$

$$\text{Cov}(\eta_{d,t}, \eta_{d,t'}) = \begin{cases} \mathbf{Q}_d & \text{gdy } t = t' \\ \mathbf{0} & \text{wpp} \end{cases}, \quad (22)$$

$$\mathbf{Q}_d = \text{diag}(0, \sigma_{R,d}^2, \sigma_{S,d,1}^2, \sigma_{S^*,d,1}^2, \sigma_{S,d,2}^2, \sigma_{S^*,d,2}^2, \sigma_{\beta,d}^2). \quad (23)$$

Ideą modeli przestrzeni stanów jest założenie, że obserwowana zmienna $Y_{d,t}$ jest zależna od nieobserwowanego wektora parametrów stanu $\boldsymbol{\alpha}_{d,t}$. Równanie (12), nazywane równaniem pomiarowym, określa zależność między zmienną $Y_{d,t}$ a wektorem stanów $\boldsymbol{\alpha}_{d,t}$, zaś wektor \mathbf{Z} w tym równaniu składa się ze współczynników określających tę zależność. Równanie (13), nazywane równaniem przejścia opisuje natomiast zmiany w czasie nieobserwowanego wektora parametrów stanu $\boldsymbol{\alpha}_{d,t}$. Równania (12)–(19) opisują szczególny przypadek modelu z rodziny przestrzeni stanów, w którym równanie pomiarowe i równanie przejścia są liniowe a rozkład składników losowych jest normalny. Taki model nazywany jest dynamicznym modelem liniowym.

W celu oszacowania wektora $\boldsymbol{\alpha}_{d,t}$ zastosować można metodę zaproponowaną przez Kalmana (1960), zwaną filtrem Kalmana. Rekurencyjna procedura szacowania wektora $\boldsymbol{\alpha}_{d,t}$ opierająca się na informacjach dostępnych w okresie t nazywana jest filtrowaniem. Natomiast estymacja wektora $\boldsymbol{\alpha}_{d,t}$ z uwzględnieniem danych dostępnych po okresie t nazywana jest wygładzaniem. Filtrowanie stosuje się do szacowania w obecnym okresie T , natomiast wygładzanie wykorzystuje się do poprawy oszacowań z wcześniejszych okresów $1, \dots, T-1$.

Dla użycia filtru Kalmana potrzebna jest znajomość hiperparametrów⁴ $\sigma_{v,d}^2, \sigma_{R,d}^2, \sigma_{S,d,1}^2, \dots, \sigma_{S^*,d,1}^2, \sigma_{\beta,d}^2$ ($d = 1, \dots, D$). W praktyce najczęściej ich wartości nie są znane, należy więc je oszacować. Dzięki założeniu o normalności reszt w przedstawionym wyżej modelu, hiperparametry można oszacować za pomocą metody największej wiarygodności. Należy także przyjąć wartości początkowe wektora $\boldsymbol{\alpha}_{d,0}$ ⁵.

⁴ W literaturze anglojęzycznej pojęcie hiperparametr (ang. *hyperparameter*) stosuje się dla parametrów rozkładu *a priori*, w celu odróżnienia ich od parametrów modelu. W przypadku modelu zastosowanego w artykule hiperparametry są wariancjami reszt parametrów stanów mających normalny rozkład *a priori*.

⁵ W obliczeniach przeprowadzonych w dalszej części opracowania przyjęto domyślne wartości paczki dlm w pakiecie statystycznym R: $\boldsymbol{\alpha}_{d,0} = (0,0,0,0,0,0,0)$, $d = 1, \dots, D$.

Więcej na temat szacowania parametrów i hiperparametrów dynamicznych modeli liniowych można znaleźć w pracach Harvey'a (1989), Durbina, Koopmana (2001), Petrisa i in. (2007).

Oszacowanie parametru $\theta_{d,t}$ otrzymane za pomocą dynamicznego modelu liniowego opisanego równaniami (12)–(13) jest postaci:

$$Y_{d,t}^F = \mathbf{Z}\hat{\boldsymbol{\alpha}}_{d,t}^F \text{ lub } Y_{d,t}^S = \mathbf{Z}\hat{\boldsymbol{\alpha}}_{d,t}^S, \quad (24)$$

gdzie $\hat{\boldsymbol{\alpha}}_{d,t}^F$ i $\hat{\boldsymbol{\alpha}}_{d,t}^S$ są oszacowaniami wektora $\boldsymbol{\alpha}_{d,t}$ uzyskane odpowiednio poprzez filtrowanie (F – filtered) i wygładzanie (S – smoothing).

3. KONSTRUKCJA PSEUDOPOPULACJI

W dalszej części opracowania podjęto próbę oceny jakości estymacji pośredniej stopy bezrobocia z wykorzystaniem dynamicznych modeli liniowych. Oceny jakości dokonano na podstawie eksperymentu Monte Carlo. W tym celu wykorzystano dane jednostkowe z Badania Aktywności Ekonomicznej Ludności z lat 2000–2009.

Na potrzeby eksperymentu utworzono pseudo-populację odpowiadającą ludności w wieku 15+ zamieszkującej województwo wielkopolskie. Wagi osób biorących udział w BAEL z woj. wielkopolskiego zmodyfikowano w taki sposób, aby dla każdego kwartału t ($t = 1, \dots, 40$) i dla każdej domeny d ($d = 1, \dots, 6$), określonej przez płeć i wiek, sumowały się do liczebności populacji z woj. wielkopolskiego w roku odpowiadającym kwartałowi t i domenie d . Niech $w_{i,d,t}$ oznacza wagę i -tej osoby z domeny d w kwartale t . Zmodyfikowana waga i -tej osoby wyznaczona jest w następujący sposób:

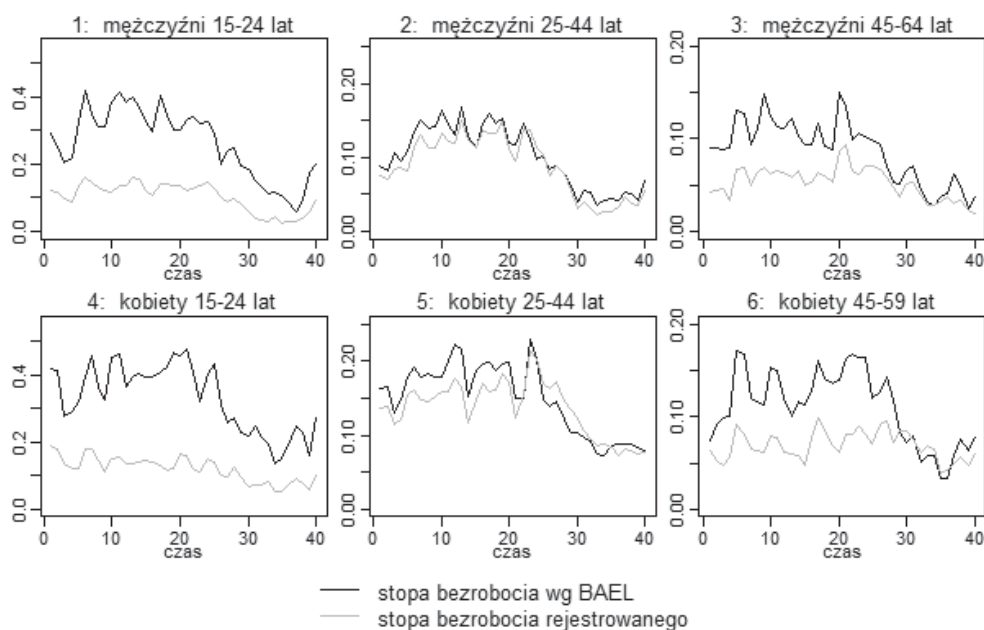
$$w_{i,d,t}^* = \left\lceil w_{i,d,t} \frac{N_{d,t}}{\sum_{i=1}^{n_{d,t}} w_{i,d,t}} \right\rceil, \quad (25)$$

gdzie $w_{i,d,t}$ to waga i -tej osoby, $N_{d,t}$ to liczebność populacji⁶, $n_{d,t}$ to liczebność próby wylosowanej do BAEL w domenie d i kwartale t , zaś operator $\lceil \cdot \rceil$ oznacza zaokrąglenie do całości. Następnie każdą jednostkę i z domeny d i kwartału t zduplikowano $w_{i,d,t}^*$ razy, zgodnie z zasadą, że osoba o wadze w reprezentuje w osób.

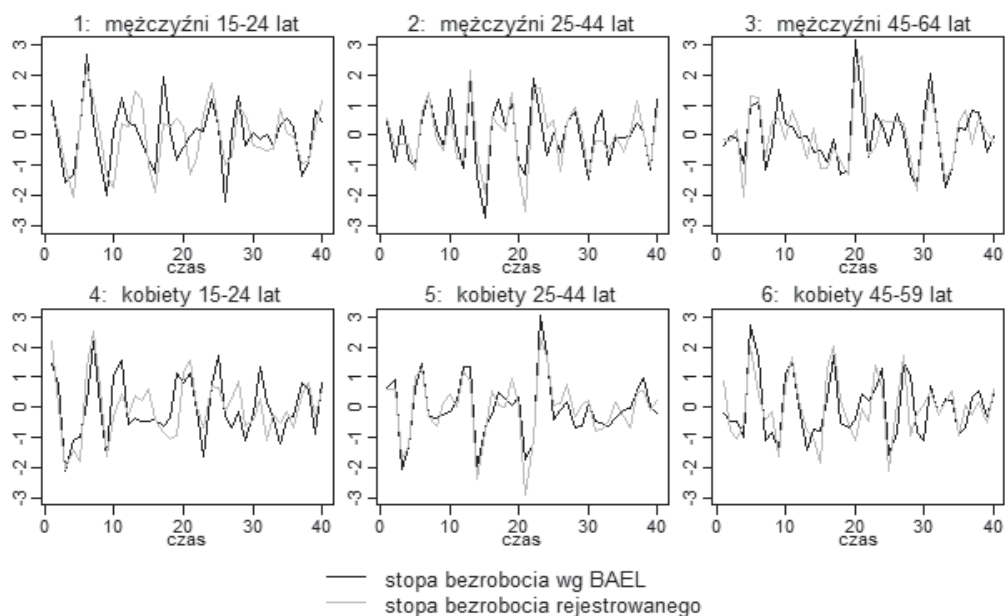
⁶ jako liczebności $N_{d,t}$ przyjęto oszacowania publikowane przez Główny Urząd Statystyczny.

4. STOPA BEZROBOCIA REJESTROWANEGO JAKO ZMIENNA POMOCNICZA

Konstrukcja modelu statystycznego wiąże się z wyborem zmiennej objaśniającej (pomocniczej). Zasadniczym wymogiem w tej kwestii jest silne skorelowanie ze zmienną objaśnianą. Naturalnym kandydatem na zmienną pomocniczą do szacowania stopy bezrobocia wg BAEL jest stopa bezrobocia rejestrowanego. Badając relację między tymi zmiennymi w skonstruowanej pseudo-populacji, zauważyć można duże podobieństwo zmienności w przekroju domen (por. rys. 1). Model strukturalny, przedstawiony w drugiej części tego opracowania, zawiera trend i sezonowość. Wprowadzenie stopy bezrobocia rejestrowanego do tego modelu jako zmiennej pomocniczej wymaga oczyszczenia jej z trendu i wahań sezonowych. W tym celu wykorzystano model opisany równaniami (12)–(23) z pominięciem składnika systematycznego regresji liniowej. Po oczyszczeniu z trendu i sezonowości, a następnie po znormalizowaniu, relacje między stopą bezrobocia i stopą bezrobocia rejestrowanego wskazują na silną korelację (rys. 2, rys. 3, tab. 1). Szczególnie silna zależność występuje w domenach kobiet w wieku 25–44 lat i mężczyzn w wieku 25–44 lat.

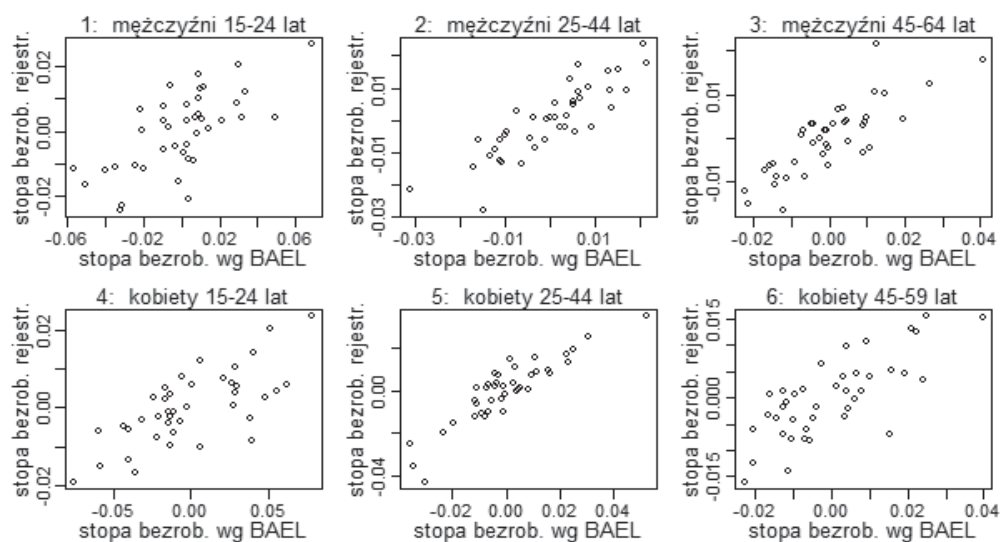


Rysunek 1. Stopa bezrobocia wg BAEL i stopa bezrobocia rejestrowanego w pseudo-populacji
Źródło: opracowanie własne na podstawie danych z Badania Aktywności Ekonomicznej Ludności.



Rysunek 2. Stopa bezrobocia wg BAEL i stopa bezrobocia rejestrowanego w pseudo-populacji, po oczyszczeniu z trendu i sezonowości oraz znormalizowaniu

Źródło: opracowanie własne na podstawie danych z Badania Aktywności Ekonomicznej Ludności.



Rysunek 3. Wykres korelacyjny stopy bezrobocia wg BAEL i stopy bezrobocia rejestrowanego w pseudo-populacji, po oczyszczeniu z trendu i sezonowości

Źródło: opracowanie własne na podstawie danych z Badania Aktywności Ekonomicznej Ludności.

Tabela 1.

Współczynniki korelacji liniowej między oczyszczonymi z trendu i sezonowości stopami bezrobocia wg BAEL i bezrobocia rejestrowanego

Domena	Współczynnik korelacji liniowej
1: mężczyźni 15–24 lat	0,69
2: mężczyźni 25–44 lat	0,86
3: mężczyźni 45–64 lat	0,82
4: kobiety 15–24 lat	0,72
5: kobiety 25–44 lat	0,90
6: kobiety 45–59 lat	0,76

Źródło: opracowanie własne na podstawie Badania Aktywności Ekonomicznej Ludności.

5. EKSPERYMENT MONTE CARLO

Weryfikując przydatność strukturalnych modeli szeregów czasowych do szacowania stopy bezrobocia, przeprowadzono eksperyment Monte Carlo. W eksperymencie tym szacowanymi parametrami są kwartalne stopy bezrobocia w pseudo-populacji w następujących sześciu domenach:

1. mężczyźni w wieku 15–24 lat,
2. mężczyźni w wieku 25–44 lat,
3. mężczyźni w wieku 45–64 lat,
4. kobiety w wieku 15–24 lat,
5. kobiety w wieku 25–44 lat,
6. kobiety w wieku 45–59 lat.

Jako zmienną pomocniczą w estymacji pośredniej stopy bezrobocia wg BAEL wykorzystano stopę bezrobocia rejestrowanego w pseudo-populacji.

W obliczeniach stosowano losowanie proste bez zwracania. Do bezpośredniej estymacji stopy bezrobocia wykorzystano estymator postaci:

$$Y_{d,t} = \frac{n_{d,t}^B}{n_{d,t}^A}, \quad (26)$$

gdzie $n_{d,t}^B$ i $n_{d,t}^A$ to odpowiednio liczba osób bezrobotnych i liczba osób aktywnych zawodowo w wylosowanej próbie w kwartale t w domenie d . Oszacowanie wariancji tego estymatora, przy losowaniu prostym bez zwracania, jest dane wzorem:

$$\text{Var}(Y_{d,t}) = \frac{(N_{d,t} - n_{d,t})n_{d,t}}{N_{d,t}(n_{d,t} - 1)} \frac{n_{d,t}^B(n_{d,t}^A - n_{d,t}^B)}{(n_{d,t}^A)^3}. \quad (27)$$

Eksperyment przeprowadzono według następującej procedury:

1. Dla każdego kwartału t z lat 2000–2009 ($t = 1, \dots, 40$) za pomocą niezależnego losowania prostego bez zwracania pobrano próbę o rozmiarze 4000⁷.
2. Na podstawie wylosowanej próby z wykorzystaniem estymatora bezpośredniego danego wzorem (26) oszacowano stopę bezrobocia dla 6 domen. W ten sposób otrzymano 6 szeregów czasowych $\{Y_{d,t}\}_{t=1}^{40}$ ($d = 1, \dots, 6$) ocen estymatorów bezpośrednich.
3. Na podstawie szeregów $\{Y_{d,t}\}_{t=1}^{40}$ ($d = 1, \dots, 6$) oszacowano hiperparametry $\sigma_{v,d}^2, \sigma_{R,d}^2, \sigma_{S,d,1}^2, \dots, \sigma_{S,d,1}^2, \sigma_{\beta,d}^2$ ($d = 1, \dots, 6$) dynamicznego modelu liniowego (12)–(23).
4. Następnie dla szeregu $\{Y_{d,t}\}_{t=1}^{40}$ ($d = 1, \dots, 6$) za pomocą wygładzania oszacowano wektory parametrów stanów \mathbf{a}_t ($t = 1, \dots, 40$) dynamicznego modelu liniowego.
5. Podstawiając oszacowania parametrów stanów \mathbf{a}_t do estymatora pośredniego danego równaniem (24) otrzymano szeregi czasowe $\{Y_{d,t}^S\}_{t=1}^{40}$ ($d = 1, \dots, 6$) ocen estymatorów pośrednich.

Powyższą procedurę powtórzono 500 razy, w wyniku czego dla każdej domeny d ($d = 1, \dots, 6$) i dla każdego kwartału t ($t = 1, \dots, 40$) otrzymano po 500 ocen estymatorów bezpośrednich $Y_{d,t}$ oraz pośrednich $Y_{d,t}^S$ (rys. 4). Na ich podstawie wyznaczono miary jakości estymatorów: obciążenie (B), odchylenie standardowe (S), pierwiastek błędu średniokwadratowego ($RMSE$) i średni bezwzględny błąd (MAE). Dla estymatora bezpośredniego $Y_{d,t}$ obliczono je za pomocą wzorów:

$$B(Y_{d,t}) = m(Y_{d,t}) - \theta_{d,t}, \quad m(Y_{d,t}) = \frac{1}{500} \sum_{i=1}^{500} Y_{d,t,i}, \quad (28)$$

$$S(Y_{d,t}) = \sqrt{\frac{1}{500} \sum_{i=1}^{500} (Y_{d,t,i} - m(Y_{d,t}))^2}, \quad (29)$$

$$RMSE(Y_{d,t}) = \sqrt{B^2(Y_{d,t}) + S^2(Y_{d,t})}, \quad (30)$$

$$MAE(Y_{d,t}) = \frac{1}{500} \sum_{i=1}^{500} |\theta_{d,t} - Y_{d,t,i}|, \quad (31)$$

gdzie $Y_{d,t,i}$ to ocena estymatora bezpośredniego otrzymana w i -tej iteracji. Dla estymatorów pośrednich wzory są analogiczne.

⁷ Zaokrąglona do tysięcy przeciętna wielkość próby losowanej do BAEL z woj. wielkopolskiego

W celu porównań między domenami, interpretacji poddano stosunki tych miar do „prawdziwej” wartości stopy bezrobocia: $\frac{B(Y_{d,t})}{\theta_{d,t}}, \frac{S(Y_{d,t})}{\theta_{d,t}}, \frac{RMSE(Y_{d,t})}{\theta_{d,t}}, \frac{MAE(Y_{d,t})}{\theta_{d,t}}$.

6. WYNIKI BADANIA

Estymacja pośrednia cechują się dużym obciążeniem, sięgającym nawet do 25% wartości stopy bezrobocia w domenach mężczyzn w wieku 15–24 lat i 45–64 lat oraz kobiet w wieku 15–24 lat i 45–59 lat (domeny 1, 3, 4 i 6) (por. rys 5 i 6). W tych domenach obciążenie stanowi przeciętnie około 7% wartości stopy bezrobocia. Natomiast w przypadku domen mężczyzn w wieku 25–44 lat oraz kobiet w wieku 25–44 lat (domeny 2 i 4) maksymalne obciążenie wyników estymacji pośredniej wynosi 15% wartości stopy bezrobocia, a średnio stanowi około 4% (por. tab. 2 i 3).

We wszystkich domenach estymacja pośrednia cechuje się mniejszym odchyleniem standardowym niż estymacja bezpośrednia, przeciętnie od 1,73 razy w domenach 2, 6 do 1,9 razy w domenach 3 i 5 (por. rys. 7, tab. 4).

Estymacja pośrednia w większości przypadków cechuje się mniejszym błędem średniokwadratowym niż estymacja bezpośrednia (por. rys. 8). Odmienną sytuację zaobserwowano w domenach mężczyzn w wieku 15–24 lat i kobiet w wieku 15–24 lat (domeny 1 i 4) dla dwunastu i dziewięciu kwartałów, w domenach mężczyzn w wieku 44–64 lat i kobiet w wieku 44–59 lat (domeny 3 i 6) dla pięciu kwartałów, zaś w domenach mężczyzn w wieku 25–44 lat i kobiet w wieku 25–44 lat (domeny 2 i 5) tylko dla trzech kwartałów. Przeciętnie pierwiastek błędu średniokwadratowego w estymacji pośredniej jest mniejszy niż w estymacji bezpośredniej od 1,29 razy w domenie pierwszej do 1,59 razy w domenie piątej (por. tab. 5).

W większości przypadków estymacja pośrednia cechuje się mniejszym średnim błędem bezwzględnym niż estymacja bezpośrednia (por. rys. 9). Większy średni błąd bezwzględny w estymacji pośredniej w domenach mężczyzn w wieku 15–24 lat i kobiet w wieku 15–24 lat (domeny 1 i 4) występuje odpowiednio w piętnastu i dziewięciu kwartałach, w domenach mężczyzn w wieku 45–64 lat oraz kobiet w wieku 45–59 lat (domeny 3 i 6) w sześciu kwartałach, w domenach mężczyzn w wieku 25–44 lat i kobiet w wieku 25–44 lat (domeny 2 i 5) w czterech i pięciu kwartałach. Przeciętny stosunek średniego błędu bezwzględnego do prawdziwej stopy bezrobocia w estymacji bezpośredniej wynosi od 9% w domenie piątej do 14% w domenach trzeciej i szóstej. Natomiast w estymacji pośredniej stosunek ten wynosi od 6% w domenie piątej do 11% w domenach trzeciej i szóstej (por. tab. 6). Średni błąd bezwzględny w estymacji pośredniej jest mniejszy niż w estymacji bezpośredniej przeciętnie od 1,26 razy w domenie pierwszej do 1,57 razy w domenie piątej.

Tabela 2.
Podstawowe statystyki stosunku obciążenia wyznaczonego na podstawie eksperymentu do „prawdziwej” stopy bezrobocia

Estymator	Bezpośredni						Pośredni					
	1 M 15-24	2 M 25-44	3 M 45-64	4 K 15-24	5 K 25-44	6 K 45-59	1 M 15-24	2 M 25-44	3 M 45-64	4 K 15-24	5 K 25-44	6 K 45-59
Minimum	-0,03	-0,01	-0,02	-0,01	-0,01	-0,02	-0,17	-0,15	-0,25	-0,17	-0,11	-0,25
Kwartyl 1.	0,00	0,00	-0,01	0,00	0,00	0,00	-0,08	-0,05	-0,08	-0,06	-0,03	-0,06
Mediana	0,00	0,00	0,00	0,00	0,00	0,00	-0,02	0,00	-0,01	0,00	-0,01	-0,03
Kwartyl 3.	0,00	0,00	0,00	0,00	0,00	0,01	0,06	0,03	0,04	0,06	0,03	0,04
Maximum	0,01	0,01	0,03	0,01	0,01	0,02	0,23	0,09	0,16	0,25	0,12	0,20
Średnia	0,00	0,00	0,00	0,00	0,00	0,00	-0,01	-0,01	-0,02	0,00	-0,01	-0,01
Odch. stand.	0,01	0,00	0,01	0,01	0,01	0,01	0,09	0,05	0,09	0,09	0,05	0,09

Źródło: opracowanie własne.

Tabela 3.
Podstawowe statystyki stosunku wartości bezwzględnej obciążenia wyznaczonego na podstawie eksperymentu do „prawdziwej” stopy bezrobocia

Estymator	Bezpośredni						Pośredni					
	1 M 15-24	2 M 25-44	3 M 45-64	4 K 15-24	5 K 25-44	6 K 45-59	1 M 15-24	2 M 25-44	3 M 45-64	4 K 15-24	5 K 25-44	6 K 45-59
Minimum	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Kwartyl 1.	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,02	0,02	0,03	0,02	0,03
Mediana	0,00	0,00	0,00	0,00	0,00	0,01	0,06	0,04	0,06	0,06	0,03	0,06
Kwartyl 3.	0,01	0,01	0,01	0,01	0,01	0,01	0,10	0,07	0,10	0,10	0,05	0,09
Maximum	0,03	0,01	0,03	0,01	0,01	0,02	0,23	0,15	0,25	0,25	0,12	0,25
Średnia	0,01	0,00	0,01	0,00	0,00	0,01	0,07	0,04	0,07	0,07	0,04	0,07
Odch. stand.	0,01	0,00	0,01	0,00	0,00	0,01	0,05	0,03	0,06	0,05	0,03	0,05

Źródło: opracowanie własne.

Tabela 4.
Podstawowe statystyki stosunku odchylenia standardowego wyznaczonego na podstawie eksperymentu do „prawdziwej” stopy bezrobocia

Estymator	Bezpośredni						Pośredni					
	1 M 15-24	2 M 25-44	3 M 45-64	4 K 15-24	5 K 25-44	6 K 45-59	1 M 15-24	2 M 25-44	3 M 45-64	4 K 15-24	5 K 25-44	6 K 45-59
Minimum	0,08	0,08	0,12	0,08	0,08	0,13	0,04	0,04	0,05	0,04	0,04	0,06
Kwartyl 1.	0,10	0,10	0,15	0,10	0,09	0,14	0,05	0,05	0,07	0,05	0,04	0,08
Mediana	0,11	0,11	0,16	0,11	0,10	0,16	0,07	0,07	0,08	0,06	0,05	0,09
Kwartyl 3.	0,17	0,15	0,20	0,17	0,13	0,19	0,10	0,09	0,12	0,09	0,08	0,13
Maximum	0,36	0,21	0,31	0,25	0,16	0,31	0,36	0,16	0,26	0,17	0,11	0,26
Średnia	0,14	0,13	0,18	0,13	0,11	0,18	0,09	0,08	0,10	0,08	0,06	0,11
Odch. stand.	0,06	0,04	0,05	0,05	0,02	0,05	0,06	0,03	0,05	0,03	0,02	0,05

Źródło: opracowanie własne.

Tabela 5.
Podstawowe statystyki stosunku pierwiastka błędu średniokwadratowego wyznaczonego na podstawie eksperymentu do „prawdziwej” stopy bezrobocia

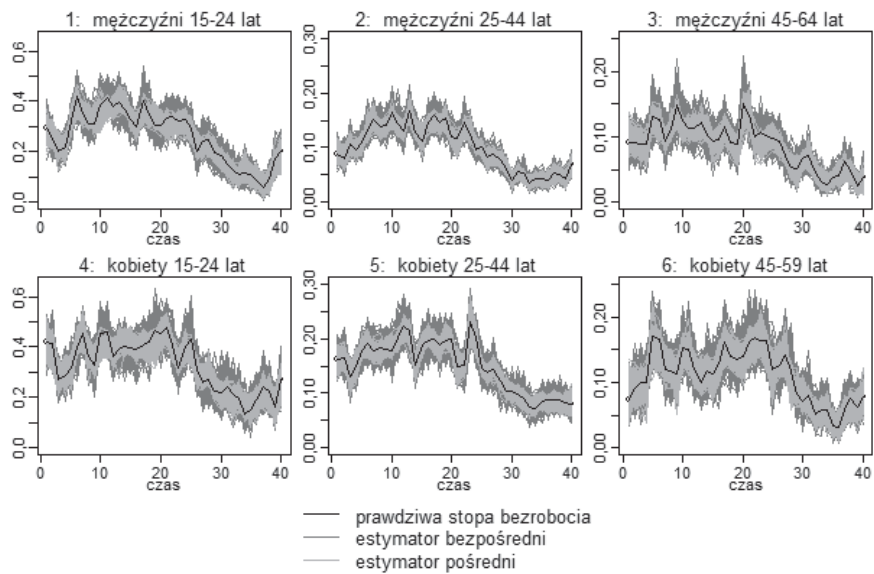
Estymator	Bezpośredni						Pośredni					
	1 M 15-24	2 M 25-44	3 M 45-64	4 K 15-24	5 K 25-44	6 K 45-59	1 M 15-24	2 M 25-44	3 M 45-64	4 K 15-24	5 K 25-44	6 K 45-59
Minimum	0,08	0,08	0,13	0,08	0,08	0,13	0,05	0,05	0,06	0,05	0,04	0,07
Kwartył 1.	0,10	0,10	0,15	0,10	0,09	0,14	0,08	0,06	0,08	0,07	0,05	0,09
Mediana	0,11	0,11	0,16	0,11	0,10	0,16	0,11	0,09	0,11	0,10	0,06	0,12
Kwartył 3.	0,17	0,15	0,20	0,17	0,13	0,19	0,15	0,12	0,15	0,14	0,09	0,15
Maximum	0,36	0,21	0,31	0,25	0,16	0,31	0,43	0,18	0,30	0,30	0,15	0,32
Średnia	0,14	0,13	0,18	0,13	0,11	0,18	0,12	0,09	0,13	0,11	0,07	0,13
Odc. stand.	0,06	0,04	0,05	0,05	0,02	0,05	0,07	0,04	0,06	0,05	0,03	0,06

Źródło: opracowanie własne.

Tabela 6.
Podstawowe statystyki stosunku średniego błędu bezwzględnego wyznaczonego na podstawie eksperymentu do „prawdziwej” stopy bezrobocia

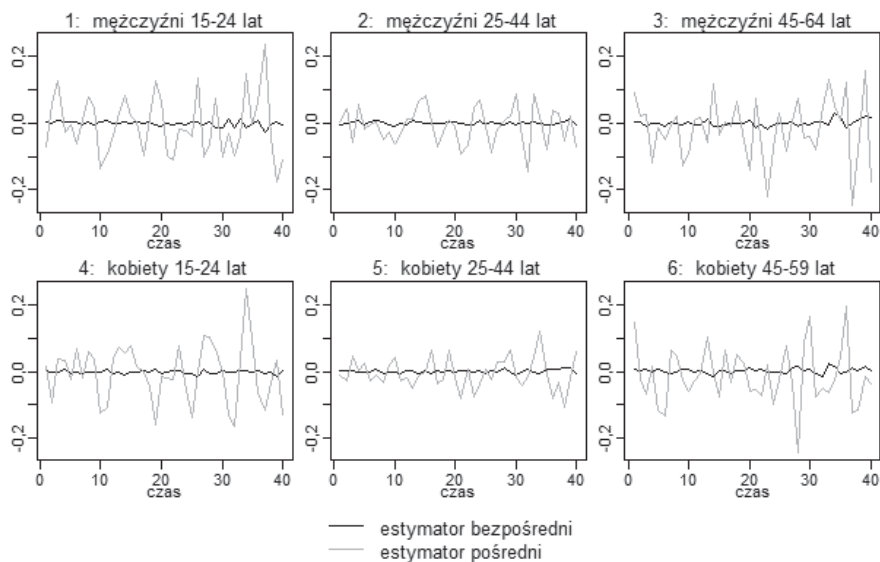
Estymator	Bezpośredni						Pośredni					
	1 M 15-24	2 M 25-44	3 M 45-64	4 K 15-24	5 K 25-44	6 K 45-59	1 M 15-24	2 M 25-44	3 M 45-64	4 K 15-24	5 K 25-44	6 K 45-59
Minimum	0,07	0,06	0,10	0,06	0,06	0,10	0,04	0,04	0,05	0,04	0,03	0,06
Kwartyl 1.	0,08	0,08	0,12	0,08	0,07	0,11	0,06	0,05	0,07	0,06	0,04	0,08
Mediana	0,09	0,09	0,13	0,09	0,08	0,13	0,09	0,07	0,09	0,08	0,05	0,10
Kwartyl 3.	0,13	0,12	0,15	0,14	0,10	0,16	0,13	0,10	0,12	0,12	0,07	0,13
Maximum	0,29	0,17	0,25	0,20	0,13	0,25	0,36	0,15	0,25	0,26	0,13	0,27
Średnia	0,12	0,10	0,14	0,11	0,09	0,14	0,10	0,08	0,11	0,09	0,06	0,11
Odc. stand.	0,05	0,03	0,04	0,04	0,02	0,04	0,06	0,03	0,05	0,05	0,02	0,05

Źródło: opracowanie własne.



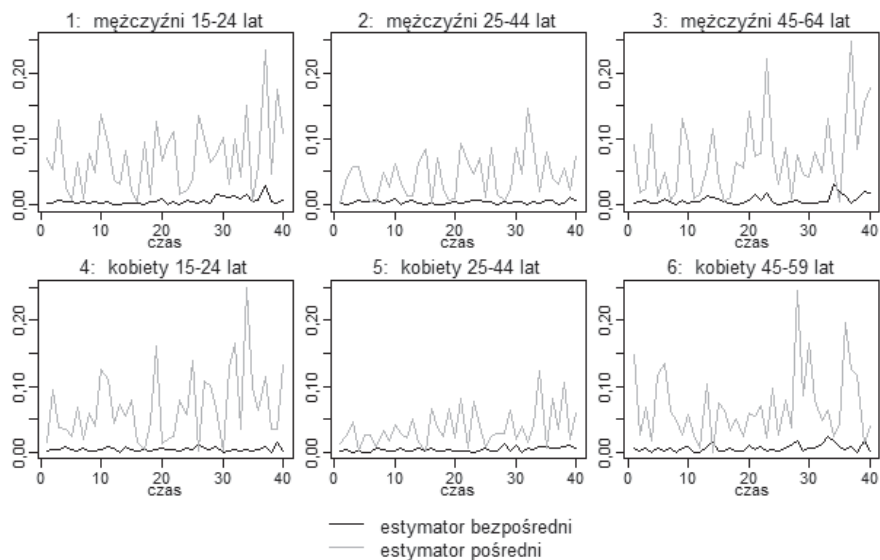
Rysunek 4. Oceny estymatorów otrzymane w wyniku eksperymentu

Źródło: opracowanie własne na podstawie danych z Badania Aktywności Ekonomicznej Ludności.



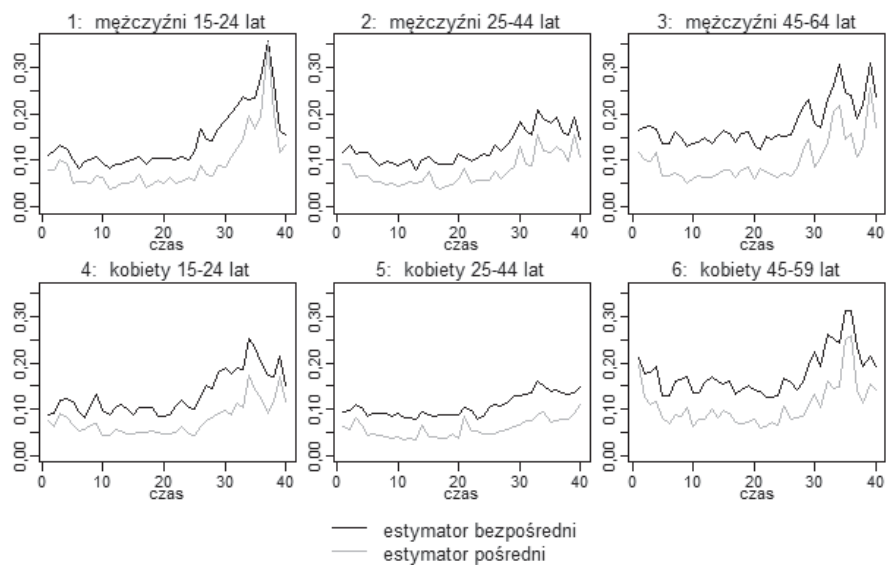
Rysunek 5. Stosunek obciążenia wyznaczonego na podstawie eksperymentu do „prawdziwej” stopy bezrobocia

Źródło: opracowanie własne na podstawie danych z Badania Aktywności Ekonomicznej Ludności.



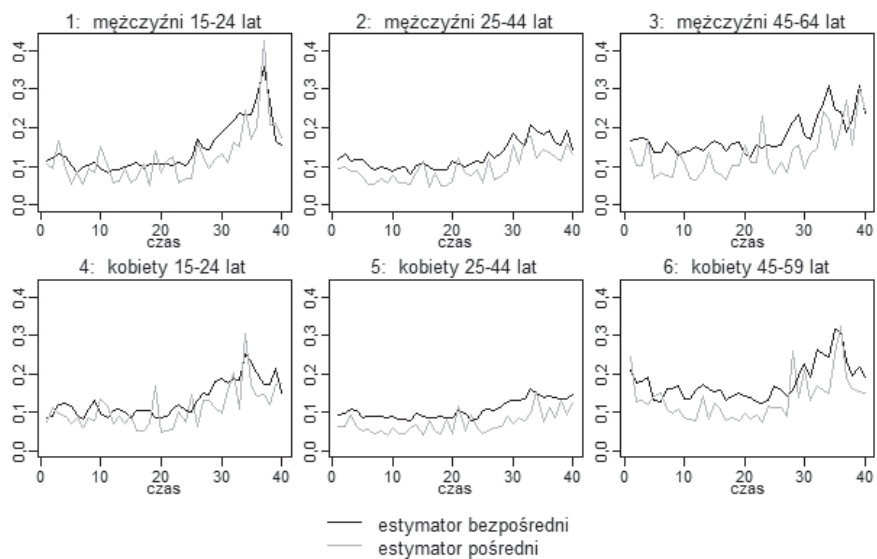
Rysunek 6. Stosunek wartości bezwzględnej obciążenia wyznaczonego na podstawie eksperymentu do „prawdziwej” stopy bezrobocia

Źródło: opracowanie własne na podstawie danych z Badania Aktywności Ekonomicznej Ludności.



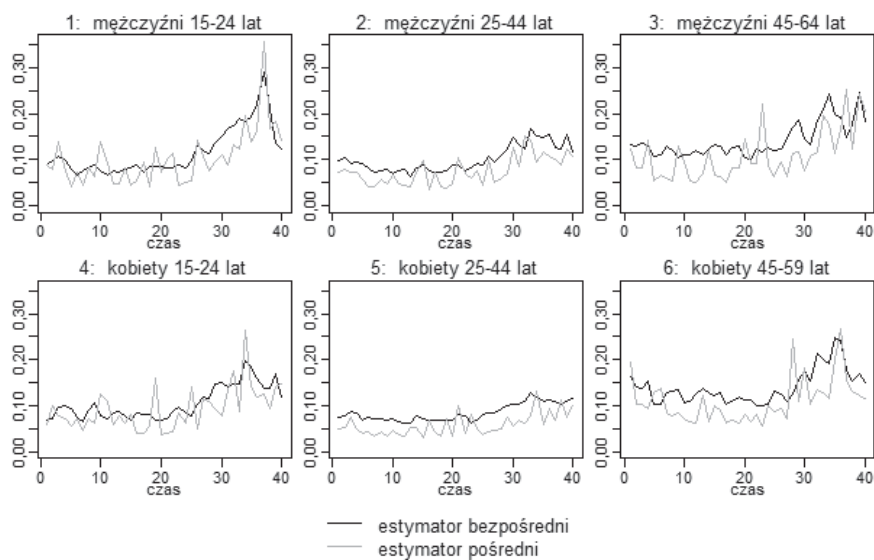
Rysunek 7. Stosunek odchylenia standardowego wyznaczonego na podstawie eksperymentu do „prawdziwej” stopy bezrobocia

Źródło: opracowanie własne na podstawie danych z Badania Aktywności Ekonomicznej Ludności.



Rysunek 8. Stosunek pierwiastka błędów średniokwadratowych wyznaczonego na podstawie eksperymentu do „prawdziwej” stopy bezrobocia

Źródło: opracowanie własne na podstawie danych z Badania Aktywności Ekonomicznej Ludności.



Rysunek 9. Stosunek średniego błędów bezwzględnych wyznaczonego na podstawie eksperymentu do „prawdziwej” stopy bezrobocia

Źródło: opracowanie własne na podstawie danych z Badania Aktywności Ekonomicznej Ludności.

7. WNIOSKI I DALSZE KIERUNKI BADAŃ

W ogólnym przypadku estymacja pośrednia cechuje się mniejszą dokładnością, ale większą precyzją niż estymacja bezpośrednia. W większości kwartałów jakość estymacji mierzona za pomocą błędu średniokwadratowego i średniego błędu bezwzględnego jest większa w przypadku estymacji pośredniej. Duże obciążenia estymatorów pośrednich w niektórych kwartałach spowodowane były nie do końca dobrze dopasowanym modelem. Oczyszczona z trendu i sezonowości stopa bezrobocia rejestrowanego będąca zmienną pomocniczą, w niektórych kwartałach przebiegała znacząco odmiennie niż oczyszczona z trendu i sezonowości stopa bezrobocia, co powodowało, że systematyczny składnik regresji liniowej w tych kwartałach pogarszał dopasowanie modelu.

Najmniejsze błędy estymacji pośredniej można zaobserwować w domenach mężczyzn w wieku 25–44 lat i kobiet w wieku 25–44 lat (domeny 2 i 5), gdzie zmienna pomocnicza była najsilniej skorelowana z „prawdziwą” stopą bezrobocia. Zaś w domenach mężczyzn 15–24 lat i kobiet w wieku 15–24 lat (domeny 1 i 4) zmienna pomocnicza jest najslabiej skorelowana z „prawdziwą” stopą bezrobocia, co przekłada się na większe błędy estymacji pośredniej.

Oszacowania otrzymane metodą zastosowaną w niniejszym artykule charakteryzują się niespójnością, a mianowicie średnia ważona oszacowań z poszczególnych domen nie jest równa oszacowaniu dla całego województwa. W celu otrzymania spójnych oszacowań w literaturze stosuje się techniki zwane „benchmarkingiem”. Zastosowanie „benchmarkingu” można znaleźć m.in. w pracy Pfeffermanna i in. (2005), Pfeffermanna, Tillera (2006). Jednym z dalszych kierunków badań autora jest zastosowanie tych technik w celu otrzymania spójnych oszacowań charakterystyk polskiego rynku pracy za pomocą estymacji pośredniej.

Uniwersytet Ekonomiczny w Poznaniu

LITERATURA

- Brakel J., Krieg S., (2008), *Estimation of the Monthly Unemployment Rate through Structural Time Series Modeling in a Rotating Panel Design*, Statistics Netherlands, Hague.
- Brakel J., Krieg S., (2009), *Structural Time Series Modeling of the Monthly Unemployment Rate in a Rotating Panel*, Statistics Netherlands, Hague.
- Brakel J., Krieg S., (2010), *Estimation of the Monthly Unemployment Rate for Six Domains through Structural Time Series Modeling with Cointegrated Trends*, Statistics Netherlands, Hague.
- Domański C., Pruska K., (2001), *Metody statystyki małych obszarów*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Durbin J., Koopman S. J., (2001), *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.

- GUS (2014a), *Aktywność ekonomiczna ludności Polski, IV kwartał 2013*, Główny Urząd Statystyczny, Warszawa.
- GUS (2014b), *Bezrobocie rejestrowane, I–IV kwartał 2013*, Główny Urząd Statystyczny, Warszawa.
- Gołata E., (2004), *Estymacja pośrednia bezrobocia na lokalnym rynku pracy*, *Prace habilitacyjne*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań.
- Harvey A. C., (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.
- Kalman R. E., (1960), A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering*, 82 (Series D), 35–45.
- Petris G., Petrone S., Campagnoli P., (2007), *Dynamic Linear Models with R*, Springer, New York.
- Pfeffermann D., Tiller R., Brown S., (2005), *Small Area Estimation with Stochastic Benchmark Constraints: Theory and Practical Application in US Labor Statistics*, *Statistical Office of the European Communities (Eurostat) – Working papers and studies*, Luxemburg.
- Pfeffermann D., Tiller R., (2006), Small Area Estimation with State Space Models Subject to Benchmark Constraints, *Journal of the American Statistical Association*, 476 (101), 1387–1397.
- Rao J. N. K., (2003), *Small Area Estimation*, John Wiley & Sons, Hoboken.
- Wilak K., (2013), Wykorzystanie dynamicznych modeli liniowych w estymacji pośredniej, *Ekonometria*, 2 (40), 126–138.

STRUKTURALNE MODELE SZEREGÓW CZASOWYCH W ESTYMACJI STOPY BEZROBOCIA W DEZAGREGACJI NA WOJEWÓDZTWA, PŁEĆ I WIEK

Streszczenie

Informacje publikowane przez Główny Urząd Statystyczny na podstawie Badania Aktywności Ekonomicznej Ludności cechują się dużym poziomem agregacji. Oszacowania w przekroju województw dla małych grup określonych przez cechy demograficzne nie są publikowane ze względu na zbyt małą precyzję estymacji bezpośredniej, spowodowaną małą liczebnością próby. Sposobem na zwiększenie precyzji oszacowania jest zastosowanie estymacji pośredniej. W literaturze popularne jest podejście, w którym do estymacji pośredniej charakterystyk rynku pracy stosuje się strukturalne modele szeregów czasowych. W niniejszym artykule została podjęta próba oceny wykorzystania tej metody w kontekście zwiększenia precyzji estymacji stopy bezrobocia w dezagregacji na województwa, płeć i wiek. Ocena ta została dokonana na podstawie eksperymentu Monte Carlo z wykorzystaniem danych jednostkowych z Badania Aktywności Ekonomicznej Ludności z lat 2000–2009. Wyniki tego badania pokazują, że zastosowany estymator pośredni w większości przypadków cechuje się lepszą jakością niż estymacja bezpośrednia.

Słowa kluczowe: statystyka małych obszarów, estymacja pośrednia, dynamiczne modele liniowe, stopa bezrobocia

STRUCTURAL TIME SERIES MODELS IN UNEMPLOYMENT RATE ESTIMATION
IN DISAGGREGATION ON VOIVODESHIP, SEX AND AGE

A b s t r a c t

Central Statistical Office in Poland publishes information on labour market derived from Labour Force Survey at high level of aggregation. Estimates for small demographic domains on voivodeship level are not published due to insufficient precision of direct estimates, caused by small sample size. One of possible approaches to the problem is to apply small area estimation. Taking into account that LFS is panel research of households structural time series models can be used in order to borrow strength in time. The aim of the article is to evaluate this method in the context of unemployment rate estimation on voivodeship level including sex and age domains. Monte Carlo simulation study will be applied in order to assess results of estimation and compare to direct estimation. Data obtained from the Labour Force Survey in Poland between 2000–2009 will be used. Results of the study indicates that temporal small area estimation have better quality of estimates compared to direct estimation.

Keywords: Small Area Estimation, direct estimation, dynamic linear models, unemployment rate

