

Kamil Wilak

Uniwersytet Ekonomiczny w Poznaniu

WYKORZYSTANIE DYNAMICZNYCH MODELI LINIOWYCH W ESTYMACJI POŚREDNIEJ

Streszczenie: W artykule przedstawiono metodę estymacji wykorzystującą dynamiczne modele liniowe, następnie zastosowano ją do szacowania stopy bezrobocia. Podjęto także próbę oceny tego podejścia pod kątem jakości oszacowań. W tym celu przeprowadzono badanie symulacyjne, którego zadaniem jest porównanie estymatorów bazujących na dynamicznym modelu liniowym z estymatorami bezpośrednimi. Wyniki tych badań pokazują, że zastosowanie dynamicznych modeli liniowych może w dużym stopniu obniżyć wariancję estymatorów bezpośrednich, zwiększając tym samym precyzję oszacowania.

Słowa kluczowe: estymacja pośrednia, pożyczanie mocy w czasie, dynamiczne modele liniowe.

1. Wstęp

Wśród technik estymacji można wyróżnić estymację bezpośrednią (*direct estimation*) i estymację pośrednią (*indirect estimation*).

Estymacja bezpośrednia bazuje wyłącznie na danych z wylosowanej próby. Za jej zaletę należy uznać fakt, że charakteryzuje się ona nieobciążonością (lub przynajmniej asymptotyczną nieobciążonością). Natomiast wadą estymacji bezpośredniej jest jej duża wariancja przy małej liczebności próby, co ma negatywny wpływ na precyzję oszacowań.

Estymacja pośrednia wykorzystuje informacje spoza badanej próby, np. dane z rejestrów administracyjnych, dane z badań specjalnych przeprowadzonych wcześniej, dane spoza badanej domeny. Estymatory pośrednie w ogólnym przypadku są obciążone, ale w porównaniu z estymatorami bezpośrednimi charakteryzują się małą wariancją.

Estymację pośrednią stosuje się szeroko w kontekście Statystyki Małych Obszarów (*Small Area Statistics*), a więc w warunkach, w których estymacja bezpośrednia obciążona jest zbyt dużą wariancją (a więc zbyt małą precyzją), aby jej wyniki mogły zostać podane opinii publicznej.

Uwzględnianie w estymacji pośredniej informacji spoza badanego obszaru (domeny) nazywa się „pożyczeniem mocy w przestrzeni” (*borrowing strength across space*), natomiast wykorzystywanie wcześniej przeprowadzonych badań nazywa się

„pożyczaniem mocy w czasie” (*borrowing strength across time*). W tej pracy zajęto się drugim spośród wyżej wymienionych podejść. Bazuje ono na założeniu, że estymowany parametr populacji jest realizacją procesu stochastycznego, który można opisać za pomocą modelu szeregu czasowego, por. [Scott, Smith 1974]. W literaturze wymieniane są dwie rodziny modeli szeregów czasowych, stosowane w estymacji pośredniej:

- mieszane modele liniowe (*mixed linear models*)¹,
- dynamiczne modele liniowe (*dynamic linear models*)².

Ideą wykorzystania mieszanych modeli liniowych w estymacji pośredniej zajmowali się m.in. Rao i Yu [1992], Rao [2003], Saei i Chambers [2003] czy Nissinen [2009]. Podejście to zastosowano m.in. w pracy Estebana i in. [2011] w celu estymacji ubóstwa.

Metody estymacji wykorzystujące dynamiczne modele liniowe pojawiają się w literaturze często w kontekście estymacji stopy bezrobocia, przykłady takich zastosowań znajdują się m.in. w pracach Pfeffermana i Tillera [2006], Brakela i Kriega [2008; 2009; 2010]. Podejście to jest wykorzystywane również w praktyce, m.in. przez urząd statystyczny w Stanach Zjednoczonych Ameryki Północnej do szacowania charakterystyk rynku pracy, por. [Pfeffermann i in. 1998].

W dalszej części pracy skupiono się na dynamicznych modelach liniowych, przedstawiono szczególny przypadek modelu z tej rodziny, a następnie zastosowano go do estymacji stopy bezrobocia. Celem artykułu jest weryfikacja hipotezy głoszącej, że (w warunkach małej liczebności próby³) estymacja pośrednia charakterystyk polskiego rynku pracy, wykorzystująca dynamiczne modele liniowe, charakteryzuje się lepszą jakością niż estymacja bezpośrednia. Jakość estymatorów rozważano pod kątem ich dokładności i precyzji. Z dwóch estymatorów dokładniejszy jest ten, który jest mniej obciążony, natomiast precyzyjniejszy jest ten, którego odchylenie standardowe jest mniejsze. Dla weryfikacji postawionej hipotezy przeprowadzono badanie symulacyjne na danych jednostkowych z Badania Aktywności Ekonomicznej Ludności. Porównania jakości estymatorów pośrednich i bezpośrednich dokonano na podstawie ich obciążeń, odchyień standardowych, błędów średniokwadratowych i średnich błędów względnych, które wyznaczono na podstawie symulacji.

Metoda estymacji pośredniej wykorzystana w niniejszej pracy jest zaczerpnięta ze światowej literatury, natomiast *novum* jest zastosowanie tej metody do szacowania stopy bezrobocia na podstawie danych z polskiego rynku pracy.

¹ Mieszane modele liniowe mają szersze zastosowanie aniżeli tylko w kontekście szeregów czasowych.

² Dynamiczne modele liniowe są szczególnym przypadkiem modeli przestrzeni stanów (*state space models*).

³ Mała liczebność próby to taka, dla której estymacja bezpośrednia cechuje się względnie dużym błędem oszacowania.

2. Specyfikacja modelu

Model przedstawiony w tej części pracy jest jednowymiarową wersją modelu zaproponowanego przez Brakela i Kriega [2010], bez składnika regresyjnego.

Niech θ_t oznacza parametr populacji w czasie t , a Y_t niech oznacza wartość estymatora bezpośredniego tego parametru. Można wtedy zapisać:

$$Y_t = \theta_t + e_t, \quad (1)$$

gdzie e_t jest błędem estymacji. Niech estymator Y_t będzie asymptotycznie nieobciążony, a próba, na podstawie której obliczamy wartość tego estymatora, będzie losowana niezależnie. W związku z powyższym wartość oczekiwaną i kowariancję błędu estymacji e_t można opisać za pomocą układu równań:

$$E(e_t) \approx 0 \quad (2)$$

$$\text{Cov}(e_t, e_{t'}) = \begin{cases} \text{Var}(Y_t), & \text{gd } t = t' \\ 0, & \text{wpp} \end{cases}. \quad (3)$$

Niech parametr θ_t można przedstawić za pomocą strukturalnego modelu szeregu czasowego (*structural time series model*):

$$\theta_t = L_t + S_t + \varepsilon_t \quad (4)$$

$$\varepsilon_t \sim NID(0, \sigma_\varepsilon^2), \quad (5)$$

gdzie L_t to trend stochastyczny w postaci trendu lokalnie liniowego (*local linear trend*):

$$L_t = L_{t-1} + R_t \quad (6)$$

$$R_t = R_{t-1} + \eta_{R,t} \quad (7)$$

$$\eta_{R,t} \sim NID(0, \sigma_R^2), \quad (8)$$

a S_t to sezonowość stochastyczna w postaci sezonowości trygonometrycznej (*trigonometric seasonality*):

$$S_t = \sum_{j=1}^6 S_{t,j}, \quad (9)$$

gdzie dla $j = 1, 2, \dots, 6$:

$$S_{t,j} = \cos\left(\frac{j\pi}{6}\right) S_{t-1,j} + \sin\left(\frac{j\pi}{6}\right) S_{t-1,j}^* + \omega_{t,j} \quad (10)$$

$$S_{t,j}^* = -\sin\left(\frac{j\pi}{6}\right) S_{t-1,j} + \cos\left(\frac{j\pi}{6}\right) S_{t-1,j}^* + \omega_{t,j}^* \quad (11)$$

$$\omega_{t,j} \sim N(0, \sigma_{\omega,j}^2), \quad \omega_{t,j}^* \sim N(0, \sigma_{\omega,j}^2) \quad (12)$$

$$Cov(\omega_{t,j}, \omega_{t,j'}) = Cov(\omega_{t,j}^*, \omega_{t,j'}^*) = \begin{cases} \sigma_{\omega,j}^2, & \text{gdy } t = t', j = j' \\ 0, & \text{wpp} \end{cases}. \quad (13)$$

Podobne modele w swoich pracach stosowali m.in. Pfeffermann i Tiller [2006], Brakel i Krieg [2008; 2009; 2010]. Więcej na temat strukturalnych modeli liniowych można znaleźć w pracach Harveya [1989], Petrisa i in. [2007].

Po wstawieniu równania (4) do równania (1) otrzymano:

$$Y_t = L_t + S_t + \varepsilon_t + e_t = L_t + S_t + v_t, \quad (14)$$

gdzie $v_t = \varepsilon_t + e_t$. Następnie zastosowano podejście, które zaproponowali Brakel i Krieg [2010], a mianowicie o składniku v_t założono, że jest zdominowany przez błąd estymacji e_t , i opisano jego wartość oczekiwaną i kowariancję za pomocą układu równań:

$$E(v_t) = 0 \quad (15)$$

$$Cov(v_t, v_{t'}) = \begin{cases} \sigma_v^2 Var(Y_t), & \text{gdy } t = t' \\ 0, & \text{wpp} \end{cases}, \quad (16)$$

gdzie wartość parametru σ_v^2 powinna być bliska 1. O składniku v_t zakłada się dodatkowo, że ma rozkład normalny.

Model opisany równaniami (1)-(16) można zapisać w postaci macierzowej:

$$Y_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + v_t \quad (17)$$

$$\boldsymbol{\alpha}_t = \mathbf{T} \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \quad (18)$$

gdzie:

$$\mathbf{Z}_t = (1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0) \quad (19)$$

$$\boldsymbol{\alpha}_t = (L_t, R_t, S_{t,1}, S_{t,1}^*, \dots, S_{t,6}, S_{t,6}^*)^T \quad (20)$$

$$\mathbf{T} = \text{blockdiag}(\mathbf{T}^R, \mathbf{T}^{S_1}, \dots, \mathbf{T}^{S_6}) \quad (21)$$

$$\mathbf{T}^R = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad (22)$$

$$\mathbf{T}^{S_j} = \begin{pmatrix} \cos(j\pi/6) & \sin(j\pi/6) \\ -\sin(j\pi/6) & \cos(j\pi/6) \end{pmatrix} \quad (23)$$

$$\boldsymbol{\eta}_t = (0, \eta_{R,t}, \omega_{t,1}, \omega_{t,1}^*, \dots, \omega_{t,6}, \omega_{t,6}^*) \quad (24)$$

$$\boldsymbol{\eta}_t \sim N(\mathbf{0}, \text{diag}(0, \sigma_R^2, \sigma_{\omega,1}^2, \dots, \sigma_{\omega,1}^2)) \quad (25)$$

$$\text{Cov}(\boldsymbol{\eta}_t, \boldsymbol{\eta}_{t'}) = \begin{cases} \text{diag}(0, \sigma_R^2, \sigma_{\omega,1}^2, \dots, \sigma_{\omega,1}^2), & \text{gdy } t = t' \\ 0, & \text{wpp.} \end{cases} \quad (26)$$

Model opisany za pomocą układu równań postaci (17)-(18) należy do rodziny dynamicznych modeli liniowych. Równania (17) i (18) nazywane są odpowiednio równaniem pomiarowym (*measurement equation*) i równaniem przejścia (*transition equation*).

Do szacowania wektora $\boldsymbol{\alpha}_t$ można użyć metody zaproponowanej przez Kalmana [1960], zwanej filtrem Kalmana (*Kalman filter*). Rekurencyjna procedura szacowania wektora $\boldsymbol{\alpha}_t$ na podstawie informacji dostępnych w okresie t nazywana jest filtrowaniem (*filtering*). Procedura szacowania wektora $\boldsymbol{\alpha}_t$ z uwzględnieniem informacji dostępnych po okresie t nazywana jest wygładzaniem (*smoothing*). Zastosowanie wygładzania w okresie T wiąże się z modyfikacją wcześniejszych oszacowań $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_{T-1}$. Przed użyciem filtra Kalmana należy oszacować hiperparametry⁴ $\sigma_v^2, \sigma_R^2, \sigma_{\omega,1}^2$, można to zrobić za pomocą metody największej wiarygodności.

Oszacowanie parametru θ_t otrzymane za pomocą dynamicznego modelu liniowego opisanego równaniami (17) i (18) jest postaci:

$$Y_t^F = \mathbf{Z}_t \hat{\boldsymbol{\alpha}}_t^F \quad \text{lub} \quad Y_t^S = \mathbf{Z}_t \hat{\boldsymbol{\alpha}}_t^S, \quad (27)$$

gdzie $\hat{\boldsymbol{\alpha}}_t^F$ i $\hat{\boldsymbol{\alpha}}_t^S$ są oszacowaniami wektora $\boldsymbol{\alpha}_t$ uzyskanymi odpowiednio poprzez filtrowanie i wygładzanie. W aktualnym okresie T do szacowania parametru populacji θ_T stosuje się filtrowanie, natomiast wygładzanie można zastosować w celu poprawienia wcześniejszych oszacowań dla okresów $1, \dots, T-1$ lub do szacowania trendu czy sezonowości w szeregu $\theta_1, \dots, \theta_T$.

3. Badanie symulacyjne

W symulacji wykorzystano dane jednostkowe z Badania Aktywności Ekonomicznej Ludności (BAEL) z lat 2000-2005. BAEL jest badaniem reprezentacyjnym, prowadzonym przez Główny Urząd Statystyczny (GUS), w celu uzyskania informacji o sytuacji na polskim rynku pracy. Badaniem tym objęte są osoby w wieku 15 lat i więcej. BAEL przeprowadzany jest kwartalnie, w sposób ciągły, tzn. w każdym z 13 tygodni kwartału losowana jest 1/13 próby kwartalnej. Więcej na temat BAEL można znaleźć w [GUS 2012].

Dla każdego miesiąca t ($t = 1, 2, \dots, 72$) jako populację przyjęto część próby kwartalnej wylosowanej do Badania Aktywności Ekonomicznej Ludności, która została zbadana w miesiącu t . Wielkość tak zdefiniowanej populacji waha się między 13450 a 20239. Ze względu na status na rynku pracy naszą populację można podzielić na cztery grupy:

- osoby pracujące w pełnym wymiarze czasu,

⁴ Przez pojęcie hiperparametr (*hyperparameter*) rozumiemy parametr rozkładu przyjętego *a priori*.

- osoby pracujące w niepełnym wymiarze czasu,
- osoby bezrobotne,
- osoby biernie zawodowo.

Pierwsze trzy grupy łącznie tworzą grupę osób aktywnych zawodowo.

Szacowanym w symulacji parametrem populacji jest miesięczna stopa bezrobocia, którą definiuje się jako iloraz wartości globalnych liczby osób bezrobotnych i aktywnych zawodowo:

$$\theta_t = \frac{N_t^U}{N_t^{EF} + N_t^{EP} + N_t^U}, \quad (28)$$

gdzie N_t^{EF} , N_t^{EP} i N_t^U to odpowiednio wartości globalne liczby osób zatrudnionych w pełnym wymiarze czasu, zatrudnionych w niepełnym wymiarze czasu i bezrobotnych w miesiącu t .

Do bezpośredniej estymacji stopy bezrobocia θ_t zastosowano estymator ilorazu wartości globalnych liczby osób bezrobotnych i aktywnych zawodowo, który można sprowadzić do postaci:

$$Y_t = \frac{n_t^U}{n_t^{EF} + n_t^{EP} + n_t^U}, \quad (29)$$

gdzie n_t^{EF} , n_t^{EP} i n_t^U to odpowiednio liczba osób zatrudnionych w pełnym wymiarze czasu, zatrudnionych w niepełnym wymiarze czasu i bezrobotnych w próbie wylosowanej dla miesiąca t . Wariancja tego estymatora jest w przybliżeniu równa⁵:

$$Var(Y_t) \approx \frac{(N_t - n_t)n_t}{N_t(n_t - 1)} \cdot \frac{n_t^U (n_t^{EF} + n_t^{EP})}{(n_t^U + n_t^{EF} + n_t^{EP})^3}, \quad (30)$$

gdzie N_t i n_t to odpowiednio wielkość populacji i próby w miesiącu t .

Symulacja przebiega według następującej procedury:

1) Dla każdego miesiąca z lat 2000-2005 za pomocą losowania prostego niezależnego pobrano próbę o rozmiarze $10\% \cdot N$ (gdzie N jest liczebnością populacji), a następnie na jej podstawie z wykorzystaniem estymatora bezpośredniego danego wzorem (29) oszacowano miesięczną stopę bezrobocia. W ten sposób otrzymano szereg czasowy $\{Y_t\}_{t=1}^{72}$ estymatorów bezpośrednich.

2) Na podstawie szeregu $\{Y_t\}_{t=1}^{60}$ (od stycznia 2000 do grudnia 2004 r.) oszacowano hiperparametry σ_v^2 , σ_R^2 , $\sigma_{\omega,1}^2$ dynamicznego modelu liniowego, opisanego w poprzedniej części. Następnie dla szeregu $\{Y_t\}_{t=1}^{72}$ (od stycznia 2000 do grudnia 2005 r.) za pomocą filtrowania oszacowano wektory α_t ($t = 1, \dots, 72$) dynamicznego modelu liniowego, otrzymując w ten sposób szereg czasowy $\{Y_t^F\}_{t=1}^{72}$ estymatorów pośrednich.

3) Na podstawie szeregu $\{Y_t\}_{t=1}^{72}$ (od stycznia 2000 do grudnia 2005) oszacowano hiperparametry σ_v^2 , σ_R^2 , $\sigma_{\omega,1}^2$ dynamicznego modelu liniowego, a następnie za

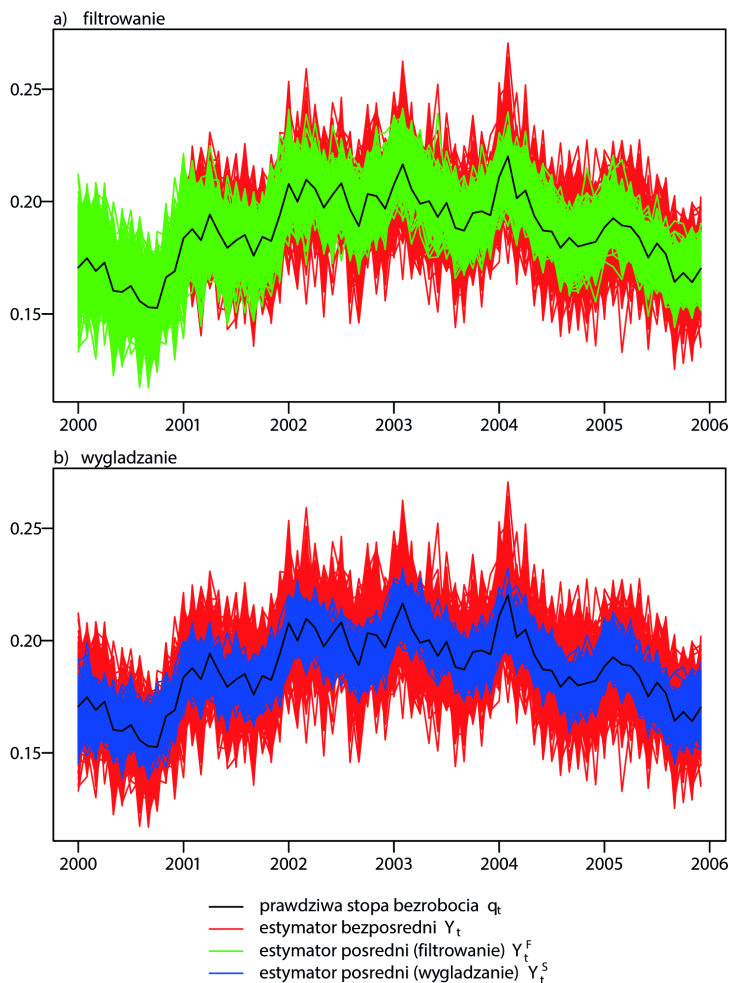
⁵ Wzór został wprowadzony na podstawie [Bracha 1996].

pomocą wygładzania oszacowano wektory α_t ($t = 1, \dots, 72$), otrzymując szereg czasowy $\{Y_t^S\}_{t=1}^{72}$ estymatorów pośrednich.

Powyższą procedurę powtórzono 500 razy.

4. Wyniki symulacji

W wyniku symulacji dla każdego miesiąca z lat 2000-2005 otrzymano 500 ocen estymatorów bezpośrednich Y_t oraz po 500 ocen estymatorów pośrednich obliczonych za pomocą filtrowania Y_t^F i wygładzania Y_t^S . Wartości te zestawiono na rys. 1.



Rys. 1. Oceny estymatorów bezpośrednich Y_t i pośrednich Y_t^F , Y_t^S otrzymanych w wyniku symulacji

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

Wyznaczono następnie miary jakości ocen estymatorów: obciążenie (B), odchylenie standardowe (SD), błąd średniokwadratowy (*Mean Squared Error* – MSE) i średni bezwzględny błąd procentowy (*Mean Absolute Percentage Error* – $MAPE$). Dla estymatora bezpośredniego Y_t obliczono je za pomocą wzorów:

$$B(Y_t) = m(Y_t) - \theta_t, \quad m(Y_t) = \frac{1}{500} \sum_{i=1}^{500} Y_{t,i} \quad (31)$$

$$SD(Y_t) = \sqrt{\frac{1}{N} \sum_{i=1}^{500} (Y_{t,i} - m(Y_t))^2} \quad (32)$$

$$MSE(Y_t) = B^2(Y_t) + SD^2(Y_t) \quad (33)$$

$$MAPE(Y_t) = \frac{1}{500} \sum_{i=1}^{500} \left| \frac{\theta_t - Y_{t,i}}{\theta_t} \right|, \quad (34)$$

gdzie $Y_{t,i}$ to ocena estymatora bezpośredniego otrzymana w i -tej iteracji. Dla estymatorów pośrednich wzory są analogiczne. Na rysunku 2 zestawiono wartości tych miar, w tab. 1 zaś podano ich podstawowe charakterystyki, a na rys. 3 przedstawiono część z tych charakterystyk za pomocą wykresów pudełkowych.

Dla pierwszych 12 miesięcy wartości Y_t i Y_t^F pokrywają się, a więc miary jakości w tych miesiącach dla tych estymatorów także będą równe, w związku z powyższym porównania jakości estymatorów dokonamy, pomijając ten okres.

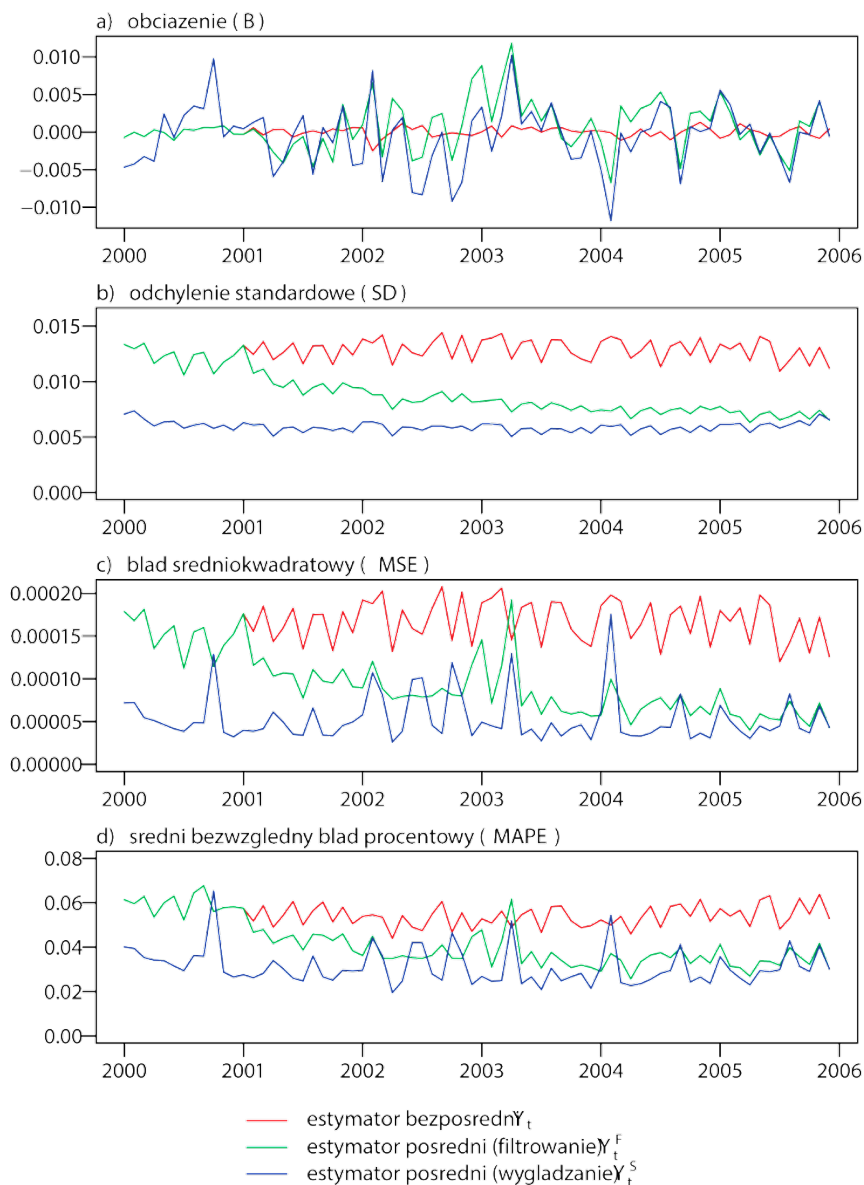
Obciążenie $B(Y_t^F)$ i $B(Y_t^S)$ estymatorów pośrednich przyjmuje wartości z przedziału $[-0.118, 0.0118]$, a więc jest znacznie bardziej rozproszone wokół 0 niż obciążenie $B(Y_t)$ estymatora bezpośredniego, które zawiera się w przedziale $[-0.0025, 0.0013]$.

Wartości estymatorów Y_t^S charakteryzują się najmniejszą zmiennością, ich odchylenie standardowe $SD(Y_t^S)$ należy do przedziału $[0.0050, 0.0071]$. Wartości estymatorów Y_t charakteryzują się największą zmiennością, ich odchylenie standardowe $SD(Y_t)$ waha się w przedziale $[0.0109, 0.0144]$. Natomiast odchylenie standardowe $SD(Y_t^F)$ ma tendencję malejącą w czasie, ich maksymalna i minimalna wartość to odpowiednio 0.0133 i 0.0063.

Błąd średniokwadratowy, we wszystkich miesiącach z wyjątkiem kwietnia 2003 r., przyjmuje największą wartość w przypadku $MSE(Y_t)$. W większości okresów, bo w 51 na 60 miesięcy, błąd średniokwadratowy $MSE(Y_t^S)$ jest mniejszy niż $MSE(Y_t^F)$.

We wszystkich miesiącach, z wyjątkiem czerwca i sierpnia 2003, średni bezwzględny błąd procentowy $MAPE(Y_t)$ jest największy, przyjmuje wartości z przedziału $[0.440, 0.0639]$. W większości okresów, bo w 50 na 60 miesięcy, średni bez-

względny błąd procentowy $MAPE(Y_t^S)$ jest mniejszy niż $MAPE(Y_t^F)$, zawierają się one odpowiednio w przedziałach $[0.0195, 0.0544]$ i $[0.0257, 0.0617]$.



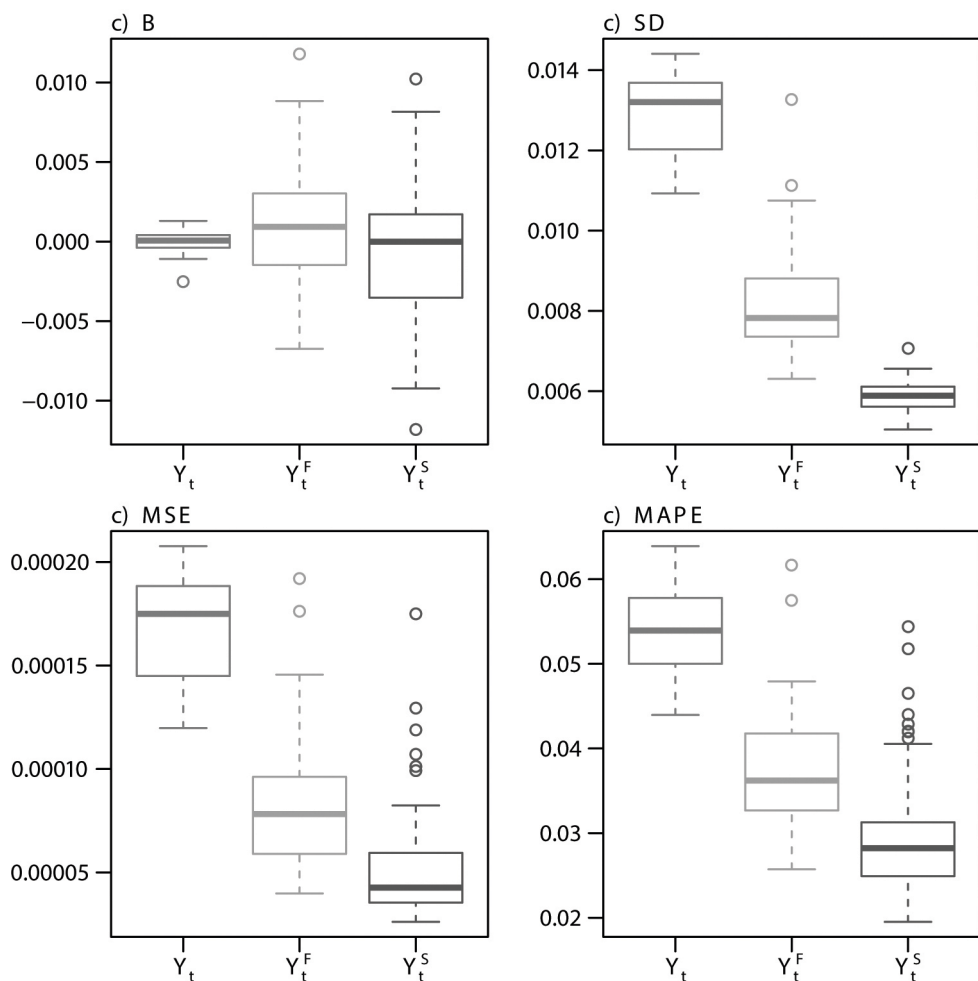
Rys. 2. Miary jakości ocen estymatorów bezpośrednich Y_t i pośrednich Y_t^F , Y_t^S otrzymanych w wyniku symulacji

Źródło: opracowanie własne na podstawie wyników symulacji.

Tabela 1. Charakterystyki miar jakości ocen estymatorów bezpośrednich Y_t i pośrednich Y_t^F , Y_t^S otrzymanych w wyniku symulacji

Charakterystyki statystyczne	Miary jakości ocen estymatora											
	bezpośredniego				pośredniego (filtrowanie)				pośredniego (wygładzanie)			
	B	STD	MSE	MAPE	B	STD	MSE	MAPE	B	STD	MSE	MAPE
Minimum	-0,0025	0,0109	0,0001	0,0440	-0,0067	0,0063	0,0000	0,0257	-0,0118	0,0050	0,0000	0,0195
Kwartył pierwszy	-0,0004	0,0120	0,0001	0,0500	-0,0014	0,0074	0,0001	0,0327	-0,0035	0,0056	0,0000	0,0249
Mediana	-0,0001	0,0132	0,0002	0,0539	0,0009	0,0078	0,0001	0,0362	-0,0000	0,0059	0,0000	0,0282
Kwartył trzeci	0,0004	0,0137	0,0002	0,0577	0,0030	0,0088	0,0001	0,0418	0,0016	0,0061	0,0001	0,0312
Maksimum	0,0013	0,0144	0,0002	0,0639	0,0118	0,0133	0,0002	0,0617	0,0102	0,0071	0,0002	0,0544
Średnia	0,0000	0,0129	0,0002	0,0540	0,0008	0,0082	0,0001	0,0377	-0,0008	0,0059	0,0001	0,0301
Odczylenie standardowe	0,0007	0,0009	0,0000	0,0048	0,0037	0,0013	0,0000	0,0068	0,0042	0,0004	0,0000	0,0075

Źródło: opracowanie własne na podstawie wyników symulacji.



Rys. 3. Wykres pudełkowy dla miar jakości ocen estymatorów bezpośrednich Y_t i pośrednich Y_t^F , Y_t^S otrzymanych w wyniku symulacji

Źródło: opracowanie własne na podstawie wyników symulacji.

5. Zakończenie

Wyniki badania symulacyjnego, przeprowadzonego na danych z polskiego rynku pracy, pokazują, że podejście w estymacji pośredniej, zwane „pożyczaniem mocy w czasie”, może zwiększyć jakość oszacowania stopy bezrobocia. Zastosowanie dynamicznych modeli liniowych w dużym stopniu zmniejsza odchylenie standardowe estymatorów (średnio o ok. 37% – filtrowanie, 54% – wygładzanie), a więc tym samym zwiększa ich precyzję. Z drugiej zaś strony, estymacja pośrednia cechuje się

dużym obciążeniem, a więc małą dokładnością, w porównaniu z estymacją bezpośrednią. Jednak w łącznej ocenie we wszystkich okresach, z wyjątkiem dwóch miesięcy, estymatory oparte na dynamicznych modelach liniowych w porównaniu z estymatorami bezpośrednimi cechują się mniejszymi błędami średniokwadratowymi (średnio o ok. 50 i 68%) i średnimi bezwzględnymi błędami procentowymi (średnio o ok. 30 i 44%).

Oceny estymatorów otrzymane za pomocą wygładzania cechują się lepszymi własnościami statystycznymi niż oszacowania otrzymane za pomocą filtrowania, a mianowicie charakteryzują się mniejszym błędem standardowym, mniejszym błędem średniokwadratowym oraz mniejszym średnim bezwzględnym błędem procentowym. Powyższa obserwacja jest zgodna z oczekiwaniami, ponieważ metoda wygładzania wykorzystuje większą liczbę informacji. Należy jednak pamiętać, że metody te wykorzystywane są dla różnych celów, jak już wcześniej wspomniano, filtrowanie stosuje się w przypadku szacowania w bieżącym okresie, natomiast wygładzanie wykorzystuje się w celu poprawy oszacowań z wcześniejszych okresów lub w celu szacowania trendu.

Literatura

- Bracha C., *Teoretyczne podstawy metody reprezentacyjnej*, Warszawa 1996.
- Brakel J., Krieg S., *Estimation of the Monthly Unemployment Rate through Structural Time Series modeling in a Rotating Panel Design*, Statistics Netherlands, Hague 2008.
- Brakel J., Krieg S., *Estimation of the Monthly Unemployment rate for Six Domains through Structural Time Series Modeling with Cointegrated Trends*, Statistics Netherlands, Hague 2010.
- Brakel J., Krieg S., *Structural Time Series Modeling of the Monthly Unemployment Rate in a Rotating Panel*, Statistics Netherlands, Hague 2009.
- Esteban M.D., Morales D., Perez A., Santamaria L., *Small area Estimation of Poverty Proportions under Area-Level Time Models*, "Computational Statistics and Data Analysis" 2011, nr 56.
- Główny Urząd Statystyczny, *Aktywność ekonomiczna ludności Polski, IV kwartał 2012*, Warszawa 2012.
- Harvey A.C., *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge 1989.
- Kalman R.E., *A new approach to linear filtering and prediction problems*, "Journal of Basic Engineering" 1960, nr 82.
- Nissinen K., *Small Area Estimation with Linear Mixed Models from Unit-Level Panel and Rotating Panel Data*, Jyväskylä 2009.
- Petris G., Petrone S., Campagnoli P., *Dynamic Linear Models with R*, Springer 2007.
- Pfeffermann D., Feder M., Signorelli D., *Estimation of autocorrelations of survey errors with application to trend estimation in small areas*, "Journal of the American Statistical Association" 1998, nr 3.
- Pfeffermann D., Tiller R., *Small area estimation with state-space models subject to benchmark constraints*, "Journal of the American Statistical Association" 2006, nr 476.
- Rao J.N.K., *Small Area Estimation*, New Jersey 2003.
- Rao J.N.K., Yu M., *Small area estimation by combining time series and cross-sectional data*, "The Canadian Journal of Statistics" 1992, nr 4.
- Saei A., Chambers R., *Small Area Estimation under Linear and Generalized Linear Mixed Models with Time and Area Effects*, Southampton Statistical Sciences Research Institute, University of Southampton 2003.

Scott A.J., Smith T.M.F., *Analysis of repeated surveys using time series methods*, "American Statistical Association" 1974, nr 347.

APPLICATION OF DYNAMIC LINEAR MODELS IN INDIRECT ESTIMATION

Summary: In this paper we describe a method of estimation which uses dynamic linear models and then we use this method for estimating unemployment rate. We attempt also to evaluate this approach in respect of the quality of assessment. In this aim we do simulation study which purpose is to compare estimators based on dynamic linear models to direct estimators. The results of the survey show that the use of time series models may greatly reduce variance of direct estimators, and thereby increase the precision of assessment.

Keywords: indirect estimation, borrowing strength across time, dynamic linear models.