

Autokorelacja błędów oszacowań w Badaniu Aktywności Ekonomicznej Ludności

Badanie Aktywności Ekonomicznej Ludności (BAEL) prowadzone przez GUS jest kwartalnym badaniem panelowym z rotacyjnym schematem doboru próby. Jego podstawy spełniają zalecenia Międzynarodowej Organizacji Pracy (International Labour Organization — ILO) oraz regulacje prawne Rady i Komisji Europejskiej. Głównym celem BAEL jest ocena bieżącej sytuacji na rynku pracy. Zastosowanie panelu rotacyjnego pozwala dodatkowo porównywać sytuację na rynku pracy kwartał do kwartału oraz rok do roku.

Próba kwartalna losowana do BAEL składa się z czterech prób elementarnych, z których dwie były badane w poprzednim kwartale, jedna jest wprowadzona po raz pierwszy oraz jedna była wprowadzona do badania przed rokiem (*Aktywność...*, 2013). Taki schemat doboru próby powoduje, że parametry populacji w kolejnych kwartałach szacowane są po części na podstawie tych samych jednostek. Błędy estymacji mogą zatem być ze sobą skorelowane.

Oszacowanie współczynników autokorelacji błędów za pomocą klasycznych estymatorów nie jest możliwe, gdyż błędy estymacji nie są obserwowalne. Problemem szacowania współczynników autokorelacji błędów w badaniach z panelem rotacyjnym zajmowali się m.in. Pfeffermann i in. (1997). Wprowadzili oni estymator, który następnie zastosowali do szacowania autokorelacji w błędach estymacji stopy bezrobocia i współczynnika aktywności ekonomicznej, otrzymanych na podstawie *Labour Force Survey* w Australii i Stanach Zjednoczonych. Aplikację tego estymatora można znaleźć także w pracach Yu'a i Mantela (1997), Bella i Carolana (1998), Griffithsa i Mansura (2001).

Wiedza na temat autokorelacji błędów w BAEL jest potrzebna m.in. do oszacowania trendu parametrów polskiego rynku pracy. Nieuwzględnienie autokorelacji w błędach oszacowania może powodować, że krzywa trendu będzie obciążona znacznymi wahaniami, charakterystycznymi dla procesów autoregresyjnych. Wahania te mogą być znaczne, gdy estymacja dotyczy małych domen, gdzie błędy oszacowań są stosunkowo duże. Wyniki badań przeprowadzonych przez wspomnianych autorów pokazują, że oszacowania krzywych trendu, które uwzględniają autokorelację w błędach są bardziej wygładzone niż te, które ją pomijają. Krzywe te są także bardziej stabilne, tzn. że nowe obserwacje w mniejszym stopniu zmieniają kształt krzywej trendu.

W artykule podjęto próbę oszacowania współczynnika autokorelacji w błędach estymacji stopy bezrobocia w woj. wielkopolskim dla trzech grup wieku z uwzględnieniem płci (6 domen). W tym celu dostosowano metodę szacowania współczynnika autokorelacji, zaproponowaną przez Pfeffermanna i in. (1997), do schematu losowania w BAEL.

SCHEMAT ROTACYJNY W BAEL

Zastosowany schemat losowania ma zasadniczy wpływ na korelację błędów oszacowań otrzymanych na podstawie badania z panelem rotacyjnym. Należy zatem opisać schemat rotacyjny z BAEL oraz jego wpływ na strukturę próby kwartalnej.

Losowana do badania próba kwartalna składa się z czterech prób elementarnych, przy czym jedna z nich jest wymieniana co kwartał. Próby elementarne losowane są niezależnie. Każda wylosowana próba elementarna badana jest według zasady: 2 kwartały w badaniu, 2 kwartały przerwy, znów 2 kwartały w badaniu i wymiana na nową próbę elementarną. Sytuacja osób wylosowanych do próby elementarnej badana jest za każdym razem „od nowa”, niezależnie od tego, który raz bierze udział w badaniu. Taki schemat losowania wprowadzono w czwartym kwartale 1999 r.

TABL. 1. KARTA ROTACJI W BAEL

Kwartały	<i>t</i>	Numery prób												
		38	39	40	41	42	43	44	45	46	47	48	49	50
2009														
I	1	x	x	—	—	x	x	—	—	—	—	—	—	—
II	2	—	x	x	—	—	x	x	—	—	—	—	—	—
III	3	—	—	x	x	—	—	x	x	—	—	—	—	—
IV	4	—	—	—	x	x	—	—	x	x	—	—	—	—
2010														
I	5	—	—	—	—	x	x	—	—	x	x	—	—	—
II	6	—	—	—	—	—	x	x	—	—	x	x	—	—
III	7	—	—	—	—	—	—	x	x	—	—	x	x	—
IV	8	—	—	—	—	—	—	—	x	x	—	—	x	x

Źródło: *Aktywność...* (2013).

Z karty rotacji (tabl. 1) można wyczytać, że każdy kwartał *t*:

- ma dwie wspólne próby elementarne z kwartałem *t*-1 i *t*-4,
- ma jedną wspólną próbę elementarną z kwartałem *t*-3 i *t*-5,
- z pozostałymi kwartałami nie ma wspólnych prób elementarnych.

Teoretyczne frakcje osób zbadanych w kwartałach $t-1$, $t-2$, $t-3$, $t-4$, $t-5$, $t-6$, $t-7$, ... w łącznej próbie z kwartału t wynoszą odpowiednio: 50%, 0%, 25%, 50%, 25%, 0%, 0%, ... W praktyce frakcje te są mniejsze, ponieważ z różnych względów nie wszystkie osoby wylosowane do próby biorą udział w badaniu tyle razy, ile wynika to ze schematu rotacyjnego. Należy także pamiętać, że jeżeli domeny badawcze zdefiniowane są przez wiek, to z kwartału na kwartał część osób ze „skrajnych” roczników przechodzi z „młodszej” do „starszej” domeny¹, co wpływa na wielkość tych frakcji.

Niech $f_{1,l}$ oznacza średnią ważoną empiryczną frakcję osób badanych w kwartale t , które zostały przebadane w kwartale $t-l$, zdefiniowaną w następujący sposób:

$$f_{1,l} = \frac{1}{T-l} \sum_{t=1+l}^T \frac{\sum_{i \in S_t \cap S_{t-l}} w_i}{\sum_{i \in S_t} w_i}$$

gdzie:

S_t — próba z kwartału t ,

w_i — waga osoby i stosowana w BAEL, będąca odwrotnością prawdopodobieństwa wylosowania do próby.

Można zauważyć, że opóźnienie l ma ujemny wpływ na stosunek średniej frakcji empirycznej do frakcji teoretycznej (tabl. 2). Można również dostrzec, że stosunek ten jest tym większy, im „starsza” jest domena.

TABL. 2. ŚREDNIE WAŻONE EMPIRYCZNE FRAKCJE $f_{1,l}$ OSÓB BADANYCH W KWARTALE t , KTÓRE PRZEBADANO W KWARTAŁACH $t-l$ W %

Domeny	$f_{1,l}$			
	$l=1$	$l=3$	$l=4$	$l=5$
1 — mężczyźni 15—24 lata	46,5 (0,929)	22,3 (0,890)	42,2 (0,844)	20,2 (0,809)
2 — mężczyźni 25—44	46,9 (0,938)	23,0 (0,921)	43,7 (0,874)	21,0 (0,841)
3 — mężczyźni 45—64	47,7 (0,955)	23,5 (0,940)	45,4 (0,908)	22,2 (0,886)
4 — kobiety 15—24	46,5 (0,931)	22,2 (0,887)	42,1 (0,842)	20,0 (0,802)
5 — kobiety 25—44	47,0 (0,941)	22,9 (0,917)	43,7 (0,873)	21,0 (0,840)
6 — kobiety 45—59 lat	47,5 (0,949)	23,5 (0,939)	44,7 (0,894)	21,5 (0,861)

U w a g a. W nawiasie podano stosunek średniej frakcji empirycznej do frakcji teoretycznej.
Ź r ó d ł o: opracowanie własne na podstawie BAEL 2000—2010, GUS.

¹ Z dwóch domen „starsza” jest ta, która jest określona przez starszą grupę wiekową.

Osoby, które biorą udział w badaniu w danym kwartale charakteryzują się często takim samym statusem na rynku pracy, jak w poprzednich kwartałach, w których były badane. Niech $f_{2,l}$ oznacza średnią ważoną empiryczną frakcję osób, które w kwartałach t i $t-l$ miały taki sam status na rynku pracy, w grupie osób badanych zarówno w kwartale t , jak i w kwartale $t-l$, zdefiniowaną w następujący sposób:

$$f_{2,l} = \frac{1}{T-l} \sum_{t=1+l}^T \frac{\sum_{i \in S_t^P \cap S_{t-l}^P} w_i + \sum_{i \in S_t^B \cap S_{t-l}^B} w_i + \sum_{i \in S_t^N \cap S_{t-l}^N} w_i}{\sum_{i \in S_t \cap S_{t-l}} w_i}$$

gdzie:

- S_t — próba z kwartału t ,
 S_t^P , S_t^B i S_t^N — odpowiednio osoby pracujące, bezrobotne i nieaktywne zawodowo z próby z kwartału t ,
 w_i — waga osoby i .

Średnie frakcje $f_{2,l}$ maleją wraz z opóźnieniem l oraz rosną wraz z „wiekiem” badanej domeny (tabl. 3). Pierwsza z tych zależności wynika z wpływu długości okresu na prawdopodobieństwo zmiany statusu na rynku pracy. Z kolei druga zależność wynika m.in. z mobilności zawodowej czy ochrony pracy, które różnią się w poszczególnych grupach wiekowych.

TABL. 3. ŚREDNIE WAŻONE FRAKcje $f_{2,l}$ OSÓB, KTÓRE W KWARTAŁACH t I $t-l$ MIAŁY TAKI SAM STATUS NA RYNKU PRACY, W GRUPIE OSÓB BADANYCH ZARÓWNO W KWARTALE t , JAK I W KWARTALE $t-l$ W %

Domeny	$f_{2,l}$			
	$l=1$	$l=3$	$l=4$	$l=5$
1 — mężczyźni 15—24 lata	94,3	88,1	84,7	81,9
2 — mężczyźni 25—44	96,9	93,5	92,3	91,3
3 — mężczyźni 45—64	97,6	94,7	93,7	92,6
4 — kobiety 15—24	94,8	87,9	85,4	81,8
5 — kobiety 25—44	95,8	91,3	89,4	87,8
6 — kobiety 45—59 lat	97,9	94,0	93,0	92,1

Źródło: jak przy tabl. 2.

Zaobserwowane zależności pozwalają wysunąć następujące hipotezy:

- współczynniki autokorelacji z opóźnieniem 1 i 4 są wyższe niż z opóźnieniem 3 i 5;
- współczynniki autokorelacji z opóźnieniem 1 i 3 są wyższe niż odpowiednio z opóźnieniem 4 i 5;
- „wiek” domeny ma dodatni wpływ na siłę autokorelacji.

METODA ESTYMACJI WSPÓŁCZYNNIKA AUTOKORELACJI W BŁĘDACH OSZACOWANIA

Metodę estymacji współczynnika autokorelacji błędów oszacowania, zaproponowaną przez Pfeffermanna i in. (1997), dostosowano do schematu losowania w BAEL.

Niech Y_t oznacza badany parametr populacji w kwartale t , np. stopę bezrobocia. Niech y_t oznacza jego zgodny estymator postaci:

$$y_t = \frac{1}{M} \sum_{k=1}^M y_t^{(k)}$$

gdzie:

- $y_t^{(k)}$ ($k = 1, \dots, M$) — bezpośredni estymator parametru Y_t otrzymany na podstawie próby elementarnej, która w kwartale t badana była k -ty raz,
- M — liczba oznaczająca, ile razy według schematu rotacyjnego są badane próby elementarne. W przypadku BAEL liczba M jest równa 4.

Niech $e_t = y_t - Y_t = \frac{1}{M} \sum_{k=1}^M e_t^{(k)}$ oznacza błąd oszacowania parametru Y_t , gdzie $e_t^{(k)} = y_t^{(k)} - Y_t$ to błąd oszacowania parametru Y_t na podstawie próby elementarnej, która w kwartale t brała udział w badaniu po raz k -ty. Współczynnik autokorelacji w szeregu błędów e_t ($t = 1, \dots, T$) definiujemy jako $\rho_l = \text{corr}(e_t, e_{t-l}) = \frac{\text{cov}(e_t, e_{t-l})}{\text{var}(e_t)}$, gdzie $l = 1, 2, \dots$, o ile szereg ten jest stacjonarny.

Niech $y_{t-l}^{(t,k)}$ i $e_{t-l}^{(t,k)}$ oznaczają odpowiednio oszacowanie i błąd oszacowania w kwartale $t-l$ na podstawie próby elementarnej, której k -te badanie przypada w kwartale t . Dalsze rozważania opierają się na założeniu, że błędy oszacowań otrzymanych na podstawie tej samej próby elementarnej, ale z różnych kwartałów są ze sobą skorelowane. Z kolei wartość współczynnika kowariancji $cov(e_t^{(t,k)}, e_{t-l}^{(t,k)})$ zależy od opóźnienia l między kwartałami oraz od liczby k , nie zależy natomiast od kwartału t . Takie założenia, przy uwzględnieniu schematu rotacyjnego w BAEL, można zapisać następująco:

$$cov(e_t^{(t,k)}, e_{t-l}^{(t,k)}) = \begin{cases} \gamma_l^{(k)} & \text{dla } (l, k) \in \{(0,1), (0,2), (0,3), (0,4), (1,2), (1,4), (3,3), (4,3), (4,4), (5,4)\} \\ 0 & \text{w pozostałych przypadkach} \end{cases}$$

Parametry populacji Y_t nie są znane, a więc błędy oszacowania $e_{t-l}^{(t,k)}$ nie są obserwowalne. Zatem do oszacowania $\gamma_l^{(k)}$ nie można zastosować klasycznego estymatora kowariancji. Pfeffermann i in. (1997) zaproponowali, aby w tym celu rozważyć tzw. pseudobłędy panelowe (*pseudo panel-survey errors*) zdefiniowane jako $e_{p,t-l}^{(t,k)} = y_{t-l}^{(t,k)} - y_{t-l} = e_{t-l}^{(t,k)} - e_{t-l}$.

Niech $C_l^{(k)} = cov(e_{p,t}^{(t,k)}, e_{p,t-l}^{(t,k)})$ oznacza współczynnik autokowariancji tych błędów. Przy założeniu, że estymator y_t jest nieobciążony, można pokazać, że:

$$C_l^{(k)} = \left(\frac{M-1}{M}\right)^2 \gamma_l^{(k)} + \left(\frac{1}{M}\right)^2 \sum_{m \neq k} \gamma_l^{(m)}$$

Na podstawie tej zależności Pfeffermann i in. (1997) wyprowadzili wzór na współczynnik autokorelacji w błędach oszacowania:

$$\rho_l = \frac{\sum_{k=1}^M c_l^{(k)}}{\sum_{k=1}^M c_0^{(k)}}$$

Po uwzględnieniu schematu rotacyjnego z BAEL wzór ten przyjmuje następującą postać:

$$p_l = \begin{cases} \frac{C_1^{(2)} + C_1^{(4)}}{\sum_{k=1}^4 C_0^{(k)}} & l = 1 \\ \frac{C_3^{(3)}}{\sum_{k=1}^4 C_0^{(k)}} & l = 3 \\ \frac{C_4^{(3)} + C_4^{(4)}}{\sum_{k=1}^4 C_0^{(k)}} & l = 4 \\ \frac{C_5^{(4)}}{\sum_{k=1}^4 C_0^{(k)}} & l = 5 \\ 0 & \text{w pozostałych przypadkach} \end{cases}$$

Błędy pseudopanelowe $e_{t-l}^{(t,k)}$ są obserwowalne, można zatem oszacować współczynniki autokorelacji $C_l^{(k)}$ w następujący sposób:

$$\hat{C}_l^{(k)} = \frac{1}{T-k} \sum_{t=l+1}^T (e_{p,t}^{(t,k)} - e_{p.}^{(0,k)})(e_{p,t-l}^{(t,k)} - e_{p.}^{(l,k)})$$

gdzie $e_{p.}^{(l,k)} = \frac{1}{T} \sum_{t=l+1}^{T+l} e_{p,t-l}^{(t,k)}$

Można zauważyć, że w przypadku schematu rotacyjnego zastosowanego w BAEL:

$$e_{p.}^{(l,k)} = \begin{cases} e_{p.}^{(0,1)} & (l,k) \in \{(0,1), (1,2), (4,3), (5,4)\} \\ e_{p.}^{(0,2)} & (l,k) \in \{(0,2), (3,3), (4,4)\} \\ e_{p.}^{(0,3)} & (l,k) \in \{(0,3), (1,4)\} \\ e_{p.}^{(0,4)} & (l,k) = (0,4) \end{cases}$$

Symbol $e_{p.}^{(0,k)}$ oznacza średni błąd oszacowań $y_t^{(k)}$ ($t = 1, \dots, T$) otrzymanych na podstawie prób elementarnych badanych k -ty raz.

Estymator współczynnika ρ_l autokorelacji błędów w BAEL przyjmuje zatem postać:

$$\hat{\rho}_l = \begin{cases} \frac{\hat{C}_1^{(2)} + \hat{C}_1^{(4)}}{\sum_{k=1}^4 \hat{C}_0^{(k)}} & l = 1 \\ \frac{\hat{C}_3^{(3)}}{\sum_{k=1}^4 \hat{C}_0^{(k)}} & l = 3 \\ \frac{\hat{C}_4^{(3)} + \hat{C}_4^{(4)}}{\sum_{k=1}^4 \hat{C}_0^{(k)}} & l = 4 \\ \frac{\hat{C}_5^{(4)}}{\sum_{k=1}^4 \hat{C}_0^{(k)}} & l = 5 \\ 0 & \text{w pozostałych przypadkach} \end{cases}$$

OSZACOWANIA WSPÓŁCZYNNIKÓW AUTOKORELACJI W BAEL

W tej części artykułu przedstawiono wyniki estymacji współczynnika autokorelacji w błędach oszacowania stopy bezrobocia (tabl. 4) w woj. wielkopolskim dla trzech grup wieku: 15—24 lata, 25—44, 45—59/64 lata, z uwzględnieniem płci. Oszacowania dokonano na podstawie danych z lat 2000—2010.

TABL. 4. OSZACOWANIA WSPÓŁCZYNNIKÓW AUTOKORELACJI W BŁĘDACH ESTYMACJI STOPY BEZROBOCIA WYZNACZONE NA PODSTAWIE BAEL Z LAT 2000—2010

Domeny	Oszacowanie współczynników autokorelacji $\hat{\rho}_l$			
	$l = 1$	$l = 3$	$l = 4$	$l = 5$
1 — mężczyźni 15—24 lata	0,226	0,018	0,165	0,064
2 — mężczyźni 25—44	0,354	0,128	0,283	0,109
3 — mężczyźni 45—64	0,387	0,133	0,294	0,140
4 — kobiety 15—24	0,253	0,108	0,222	0,071
5 — kobiety 25—44	0,265	0,132	0,270	0,010
6 — kobiety 45—59 lat	0,409	0,221	0,326	0,122

Źródło: jak przy tabl. 2.

Oszacowania współczynników autokorelacji błędów potwierdzają przypuszczenia wysunięte na podstawie analizy schematu rotacyjnego i jego wpływu na strukturę próby kwartalnej. Współczynniki autokorelacji z opóźnieniem 1 i 4 są większe niż z opóźnieniem 3 i 5, co bezpośrednio wynika ze schematu losowania. Współczynniki autokorelacji z opóźnieniem 1 i 3 są większe niż odpowiednio z opóźnieniem 4 i 5, co jest skutkiem ujemnego wpływu czasu na średnie frakcje $f_{1,l}$ i $f_{2,l}$. W większości przypadków siła autokorelacji jest tym większa, im „starsza” jest domena, co jest związane ze zmiennością statusu na rynku pracy w poszczególnych grupach wiekowych.

Wnioski

Zastosowanie metody szacowania współczynników autokorelacji, zaproponowanej przez Pfeffermanna i in. (1997), dostosowanej do BAEL, pozwoliło na oszacowanie współczynników autokorelacji w błędach oszacowania stopy bezrobocia.

Wyniki analizy pokazują, że schemat losowania zastosowany w BAEL powoduje, że błędy oszacowań cech rynku pracy wykazują znaczną autokorelację. Wysokość współczynników autokorelacji dla poszczególnych opóźnień jest ściśle związana z zastosowanym schematem rotacyjnym. Siła autokorelacji zależy również od zmienności statusu na rynku pracy wśród osób badanej domeny. Im ta zmienność jest mniejsza, tym współczynniki autokorelacji przyjmują większą wartość.

Autokorelacja błędów nie ma wpływu na jakość estymacji bezpośredniej wykorzystywanej w BAEL do bieżącej oceny sytuacji na rynku pracy, natomiast jest ona istotna w kontekście obserwacji długofalowych zmian. Na podstawie wyników badań opisanych w literaturze światowej można wysunąć hipotezę głoszącą, że uwzględnienie autokorelacji błędów poprawia jakość estymacji krzywych trendu dla charakterystyk rynku pracy otrzymanych na podstawie BAEL. Powyższa hipoteza będzie weryfikowana przez autora w kolejnych pracach.

mgr Kamil Wilak — *Uniwersytet Ekonomiczny w Poznaniu*

LITERATURA

- Aktywność Ekonomiczna Ludności Polski; III kwartał 2012* (2013), GUS
- Bell P. A., Carolan A. M. (1998), *Trend estimation for small areas from a counting surveys with controlled sample overlap*, „Working papers in econometrics and applied statistics”, Australian Bureau of Statistics, No. 98/1
- Griffiths R., Mansur K. (2001), *Current Population Survey Sampling Error Autocorrelations*, U.S. Census Bureau
- Pfeffermann D., Feder M., Signorelli D. (1997), *Estimation of Autocorrelations of Survey Errors with Application to Trend Estimation in Small Areas*, „Journal of Business & Economic Statistics”, No. 16/3
- Yu M., Mantel H. (1997), *Trend estimation for the Canadian Labour Force Survey*, Statistics Canada

SUMMARY

Rotating panel used in the Labour Force Survey (LFS) causes correlation possibility of estimations of labor markets errors. Knowledge of autocorrelation is important in the context of the trend estimation of labor market parameters. Dismissal of autocorrelation can result in the trend curve it will be fraught with volatility, characteristic of auto-regression processes. Estimation errors are not observable, thus it is not possible to estimate the autocorrelation coefficients by conventional estimators. This paper describes the adaptation of methods for estimating the errors of autocorrelation coefficients (proposed by Pfeffermann et al.), The rotational scheme in LFS. Then, this method was used to estimate the autocorrelation coefficients in error estimation of the unemployment rate in the province. Greater Poland for six domains defined by gender and age. Rotating panel used in the LFS (Labour Force Survey) causes estimation errors labor markets may be correlated. Knowledge of autocorrelation is important in the context of the trend estimation parameters labor market. Dismissal of autocorrelation can result in the trend curve it will be fraught with volatility, characteristic of auto-regression processes. Estimation errors are not observable, thus it is not possible to estimate the autocorrelation coefficients by conventional estimators. This paper describes the adaptation of methods for estimating the errors of autocorrelation coefficients (proposed by Pfeffermann and others), to the rotational scheme in LFS. Then, this method was used to estimate the autocorrelation coefficients in error estimation of the unemployment rate in the Wielkopolskie voivodship for six domains defined by gender and age.

РЕЗЮМЕ

Оборотная панель используемая в Обследовании экономической активности населения (ВАЕЛ) является причиной корреляции ошибок оценки характеристик рынка труда. Знания по теме автокорреляции являются важными в отношении к оценке тренда параметров на рынке труда. Неучтение автокорреляции может привести к тому, что кривая тренда будет подвергаться колебаниям, которые являются характеристическими для процессов авторегрессии. Ошибки оценивания не наблюдаются, таким образом невозможным оказывается оценка коэффициентов их автокорреляции с использованием обычных оценок. В статье характеризуется приспособление метода оценивания коэффициентов автокорреляции ошибок (предложенного Пфефферманном и другими), к оборотной схеме в ВАЕЛ. Затем этот метод был использован для оценки коэффициентов автокорреляции в ошибках оценивания нормы безработицы в великопольском воеводстве для шести домен определенных полом и возрастом.