

ANALIZA WSKAŹNIKA ZATRUDNIENIA W KRAJACH NOWEJ UE Z WYKORZYSTANIEM FPCA

Jolanta Wojnar (ORCID: 0000-0001-6962-4610)

Wydział Ekonomii
Uniwersytet Rzeszowski
e-mail: jwojnar@ur.edu.pl

Wojciech Zieliński (ORCID: 0000-0003-0749-8764)

Wydział Zastosowań Informatyki i Matematyki
SGGW w Warszawie
e-mail: wojciech_zielinski@sggw.pl

Streszczenie. W pracy porównano kraje nowej Unii Europejskiej ze względu na wskaźnik zatrudnienia w latach 2004-2016. Zastosowano analizę funkcjonalnych składowych głównych (ang. *Functional Principal Components Analysis*). Metoda ta polega na transformacji pierwotnych zmiennych funkcyjnych w zbiór nowych wzajemnie ortogonalnych zmiennych, zwanych funkcjonalnymi składowymi głównymi. Możliwości wizualizacyjne pozwoliły na porównanie krajów i wyodrębnienie grup o zbliżonym poziomie wskaźnika zatrudnienia w całym badanym okresie.

Słowa kluczowe: analiza funkcjonalnych składowych głównych, wskaźnik zatrudnienia

JEL classification: C83, C99

WSTĘP

Obserwowane zmiany na rynku pracy są pochodną zmian zachodzących w całej gospodarce narodowej. W latach 2004-2008 sytuacja na rynku pracy we wszystkich krajach nowej UE kształtowała się pod wpływem procesów makroekonomicznych, których rozwój, w pierwszej części tego okresu, był ściśle związany z przystąpieniem do Unii Europejskiej. Był to czas wyso-

kiej dynamiki wzrostu gospodarczego, wzrostu wpływów do budżetu funduszy europejskich oraz niskiej i relatywnie stabilnej stopy inflacji. W tym okresie we wszystkich krajach sytuacja na rynku pracy z każdym rokiem była coraz lepsza. Odsetek osób zatrudnionych dynamicznie wzrastał. Załamanie gospodarcze w 2008 roku, które z sektora finansów przeniosło się na całą gospodarkę objęło swym zasięgiem niemal wszystkie kraje. Światowy kryzys zmienił oblicze europejskiego rynku pracy niwecząc wcześniejsze osiągnięcia wielu krajów w walce z bezrobociem. Poszczególne kraje wykazały jednak różną odporność na wpływ kryzysu.

Celem pracy jest porównanie krajów nowej Unii Europejskiej pod kątem wskaźnika zatrudnienia w latach 2004-2016. Ważnym zagadnieniem jest rozpatrywanie wskaźnika zatrudnienia w badanych krajach w całym analizowanym okresie łącznie, a nie oddzielnie dla każdego roku. Można to było osiągnąć po przedstawieniu oryginalnych szeregów czasowych oddzielnie dla każdego kraju w postaci funkcji ciągłych określonych na ustalonym przedziale czasowym zwanych danymi funkcjonalnymi [Jacques, Preda 2014; Górecki i in. 2014]. Do syntetycznego scharakteryzowania zmian odsetka zatrudnionych w badanych krajach zastosowano analizę składowych głównych dla danych funkcjonalnych (analizę funkcjonalnych składowych głównych). Jest to analiza podobna do znanej klasycznej analizy składowych głównych dla obiektów obserwowanych wielocechowo. Analiza funkcjonalnych składowych głównych jest stosowna w przypadku, gdy obiekty opisane są pewnymi funkcjami. Ta technika statystyczna została wprowadzona przez Ramseya i Silvermana (2002, 2005) (patrz również Ramsay J. O., Dalzell C. J. (1991), Horváth L., Kokoszka P. (2012)). Pozwala na dokonanie rzutu wysokowymiarowych danych na przestrzeń o dużo mniejszym wymiarze jednocześnie zachowując maksymalnie dużo informacji dotyczącej zmienności danych. W pracy ten rodzaj analizy został zastosowany do badania zatrudnienia jako funkcji czasu.

METODYKA

Analizowane zjawisko opisane jest pewnym procesem stochastycznym $X(t)$ obserwowanym na przedziale $[0, T]$. Zakładamy, że dla tego procesu istnieje operator kowariancji \mathcal{C} , tzn. zakładamy istnienie kowariancji

$$c(s, t) = E(X(s) - EX(s))(X(t) - EX(t)), \quad \text{dla } s, t \in T.$$

Oznacza to, że proces stochastyczny jest klasy $L^2[0, T]$, czyli funkcji określonych na przedziale $[0, T]$ całkowalnych z kwadratem. Z ogólnej teorii wiadomo, że proces $X(t)$ można przedstawić w następującej postaci

$$X(t) = \sum_{j=1}^{\infty} \varepsilon_j c_j(t),$$

gdzie $\mathbf{c}_j(t)$ jest bazą w przestrzeni $L^2[0, T]$ oraz ε_j są zmiennymi losowymi.

Zadanie polega na takim wyborze bazy $\mathbf{b}_j(t)$ i zmiennych losowych η_j , by

$$\int_0^T \mathbf{b}_j^2(t) dt = 1, \quad \int_0^T \mathbf{b}_j(t) \mathbf{b}_k(t) dt = 0, \quad \text{dla wszystkich } j \neq k$$

oraz $D^2\eta_1 \geq D^2\eta_2 > \dots$. Elementy bazy nazywamy funkcjonalnymi składowymi głównymi.

Niech $X_1(t), \dots, X_n(t)$ będą obserwacjami procesu stochastycznego $X(t)$ dokonanymi w punktach czasowych $t_1 < t_2 < \dots < t_p$ na n obiektach, tzn. dysponujemy $n \times p$ macierzą obserwacji

$$\mathbf{X} = \begin{bmatrix} X_1(t_1) & X_1(t_2) & \dots & X_1(t_p) \\ X_2(t_1) & X_2(t_2) & \dots & X_2(t_p) \\ \dots & \dots & \dots & \dots \\ X_n(t_1) & X_n(t_2) & \dots & X_n(t_p) \end{bmatrix}.$$

Na podstawie tych obserwacji wyznaczmy funkcjonalne składowe główne. Ponieważ mamy do dyspozycji skończoną liczbę obserwacji, więc nie jest możliwe wyznaczenie całej bazy (formalnie elementów bazy jest nieskończenie wiele), więc musimy ograniczyć się do skończonej liczby elementów bazy. Naturalnym ograniczeniem liczby wyznaczanych składowych jest liczba p punktów pomiarowych.

Metodą najmniejszych kwadratów dopasowujemy każdą z obserwacji do bazy:

$$\hat{X}_i(t) = \sum_{j=1}^p \hat{a}_{ij} b_j(t) \quad \text{dla } i = 1, 2, \dots, n$$

i wyznaczamy próbkowy operator kowariancji

$$c(s, t) = \frac{1}{n} \sum_{i,j=1}^n (\hat{X}_i(s) - \bar{X}_i)(\hat{X}_j(t) - \bar{X}_j),$$

Numerycznie rozwiązujemy zagadnienie własne dostając próbkowe funkcjonalne składowe główne $\hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_p$ oraz próbkowe wariancje funkcjonalnych składowych głównych $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p$. Do wybrania „efektywnej” liczby funkcjonalnych składowych wykorzystuje się wykres osypiska, czyli wykres skumulowanego udziału kolejnych składowych głównych w wyjaśnianej ogólnej zmienności:

$$\frac{\sum_{i=1}^m \hat{\lambda}_i}{\sum_{i=1}^p \hat{\lambda}_i}.$$

Dla porównania analizowanych obiektów wykorzystuje się wykresy typu biplot. Na biplotcie r -tej oraz s -tej składowej głównej umieszcza się punkty

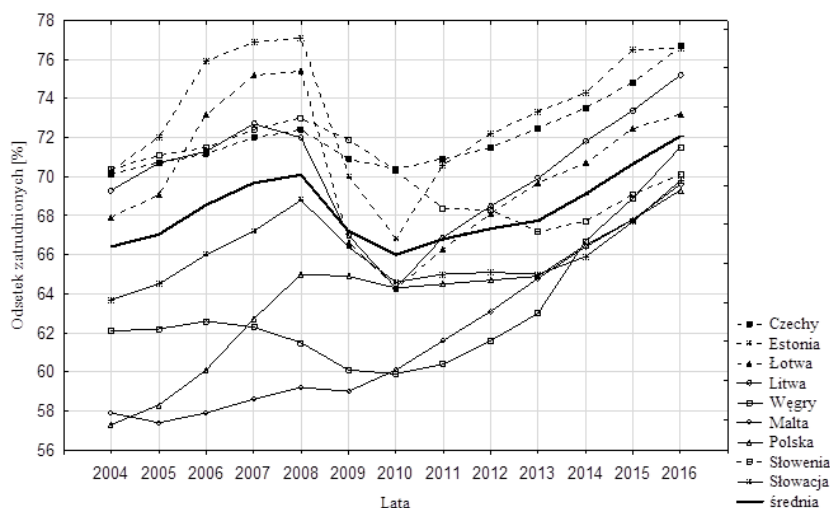
$$\left\{ \left(\int_0^T \hat{\mathbf{b}}_r(t) (\hat{X}_i(t) - \bar{X}_i(t)) dt, \int_0^T \hat{\mathbf{b}}_s(t) (\hat{X}_i(t) - \bar{X}_i(t)) dt \right) : i = 1, \dots, n \right\}.$$

ANALIZA WSKAŹNIKA ZATRUDNIENIA

Analizowany jest wskaźnik zatrudnienia, obliczany jako udział pracujących w ogólnej liczbie ludności w wieku 20-64 lata, w nowych krajach Unii Europejskiej, tzn. w krajach które 1 maja 2004 roku przystąpiły do Unii Europejskiej. Są to: Cypr, Czechy, Estonia, Łotwa, Litwa, Węgry, Malta, Polska, Słowenia i Słowacja ($n = 10$). Wskaźnik zatrudnienia obserwowany był w latach 2004-2016, tzn. w ciągu trzynastu lat ($p = 13$).

Analizę wskaźnika zatrudnienia rozpoczynamy od naszkicowania jego przebiegu w badanych latach (Rys. 1). Wykres jest pomocny przy wyborze typu bazy w przestrzeni $L^2[0, T]$. Do dalszych analiz wybrano bazę funkcji B -sklejanych.

Rysunek 1. Wskaźnik zatrudnienia



Źródło: Eurostat (2018)

Do wyznaczenia funkcjonalnych składowych głównych zastosowano autorski* program napisany w języku R (w programie wykorzystano pakiety `fda` oraz `fda.usc`). Początkowo wyznaczono dziesięć składowych głównych, których procenty wyjaśnionej zmienności podane są w Tabeli 1 oraz graficznie na wykresie osypiska (Rys. 2).

W ustaleniu liczby składowych głównych „wystarczająco” opisujących zmienność analizowanego zjawiska pomocny jest wykres osypiska (Rys. 2), na którym przedstawione jest tempo spadku wartości własnych, czyli procentu wyjaśnionej wariancji. Kolejne funkcjonalne składowe główne wyjaśniają coraz

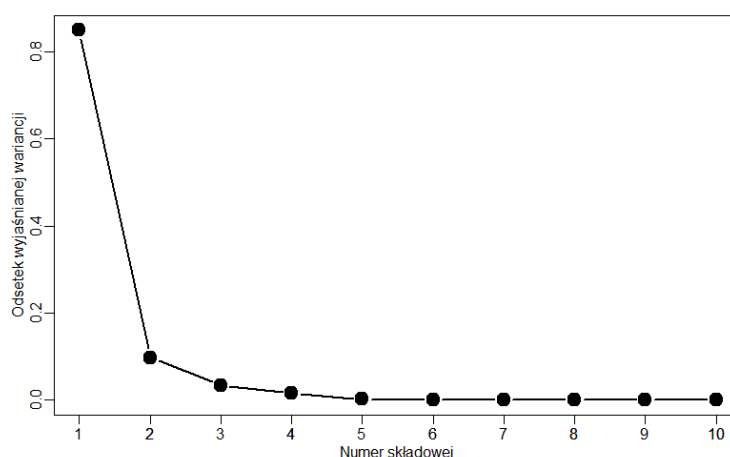
* Autorzy dziękują drowi Stanisławowi Jaworskiemu za istotną pomoc w napisaniu programu

Tabela 1. Udziały składowych głównych

składowa	1	2	3	4	5	6	7	8	9	10
procent	84.89	9.75	3.47	1.52	0.20	0.09	0.04	0.02	0.01	0.00

Źródło: obliczenia własne

Rysunek 2. Wykres osypiska

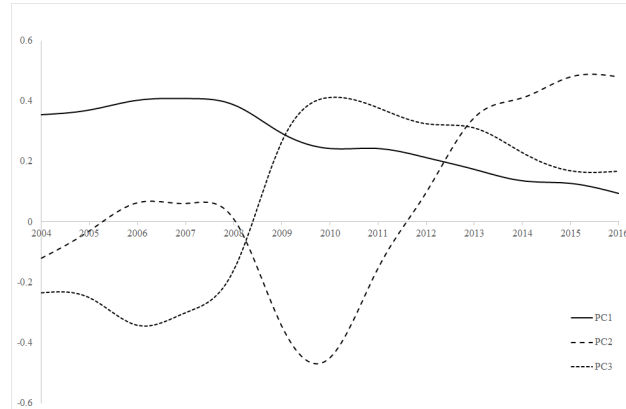


Źródło: obliczenia własne

mniejszą część zmienności, stąd spadki wartości na wykresie są coraz mniejsze. Pierwsza składowa główna wyjaśnia najwięcej zmienności, druga jest ortogonalna do pierwszej i wyjaśnia maksymalnie dużo z tego co pozostało. Wkład każdej kolejnej składowej w ogólna zmienność początkowych zmiennych jest coraz mniejszy i może być pominięty w dalszej analizie, z uwagi na to, że wyjaśnia znikomy zakres zmienności i nie wnosi istotnych informacji. Do dalszych analiz wybrano pierwsze trzy funkcjonalne składowe główne wyjaśniające łącznie 98.11% całkowitej zmienności zatrudnienia w analizowanych krajach w badanych latach. Wybrane funkcjonalne składowe główne przedstawione są na rysunku 3.

Pierwsza składowa wyjaśniająca prawie 85% całkowitej zmienności odssetka zatrudnionych w badanej grupie państw w analizowanym okresie jest uśrednieniem zatrudnienia w latach 2004-2016. Nieco większe wagi przypisane są wskaźnikowi zatrudnienia w okresie przed kryzysowym. Większa wartość składowej może być interpretowana jako ponadprzeciętne zatrudnienie. Dokładniej mówiąc, większą wartość tej składowej ma to państwo, w którym wskaźnik zatrudnienia był wyższy niż jego średnia wartość w analizowanych krajach. Wartości tej składowej głównej dla poszczególnych krajów można odczytać na osi poziomej biplotu dla pierwszych dwóch składowych głównych (Rys. 4) lub biplotu dla pierwszej i trzeciej składowej głównej (Rys. 5). Analiza wartości pierwszej funkcjonalnej składowej głównej wskazuje, że do

Rysunek 3. Składowe główne



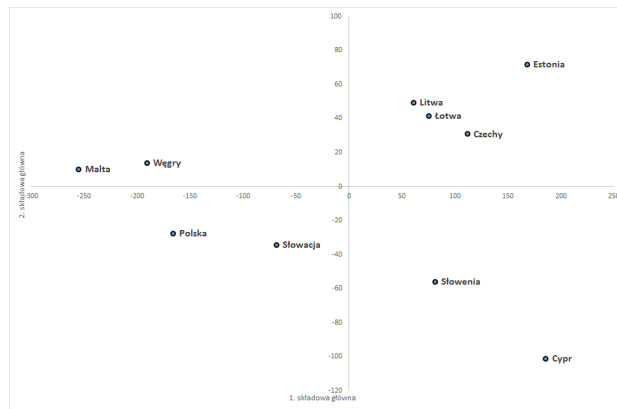
Źródło: obliczenia własne

krajów o ponadprzeciętnym wskaźniku zatrudnienia (oś pozioma) należą Litwa, Łotwa, Estonia oraz Czechy, Słowenia i Cypr, zaś do krajów o wskaźniku zatrudnienia niższym niż średni należą Słowacja, Węgry, Malta oraz Polska.

Druga składowa wyjaśniająca niemal 10% zmienności wskaźnika zatrudnienia w analizowanych państwach w badanym okresie jest porównaniem poziomu tego wskaźnika w latach 2008-2011 a jego poziomu w latach 2013-2016. Dodatnia wartość tej składowej może być interpretowana jako wzrost wskaźnika zatrudnienia w latach 2013-2016. Ujemna wartość tej składowej może być interpretowana w ten sposób, że państwo nie „poradziło” sobie z odrobieniem pokryzysowego spadku poziomu zatrudnienia. Analiza wykresu biplotu dla pierwszych dwóch składowych głównych (Rys. 4) wskazuje, że do krajów, które nadrobiły pokryzysowy spadek zatrudnienia (oś pionowa) należą kraje nadbałtyckie oraz Czechy, Węgry i Malta, zaś pozostałe wciąż odczuwają skutki kryzysu (w tym Polska). Należy pamiętać, że ten wniosek jest względny, tzn. odnosi się do sytuacji w danym kraju na tle pozostałych z badanej grupy państw.

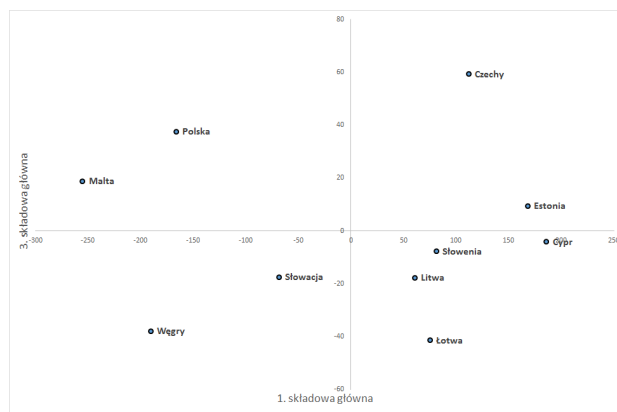
Trzecia składowa wyjaśniająca niecałe 3.5% ogólnej zmienności, porównuje wskaźnik zatrudnienia na początku i na końcu badanego okresu. Dodatnia wartość tej składowej oznacza, że zmiany wskaźnika zatrudnienia w początkowym okresie tj. latach 2004-2010 były wyższe niż w końcowym. Analiza wykresu biplotu dla drugiej i trzeciej składowej głównej (Rys. 6) wskazuje, że Polska wraz z Estonią, Malta i Czechami należy do krajów, w których zatrudnienie w latach 2010-2016 wzrastało zdecydowanie wolniej niż na początku badanego okresu (kraje te zmniejszały tempo wzrostu wskaźnika). W pozostałych krajach nowej UE większe tempo zmian zatrudnienia było na końcu badanego okresu.

Rysunek 4. Bi-plot pierwszej i drugiej składowej



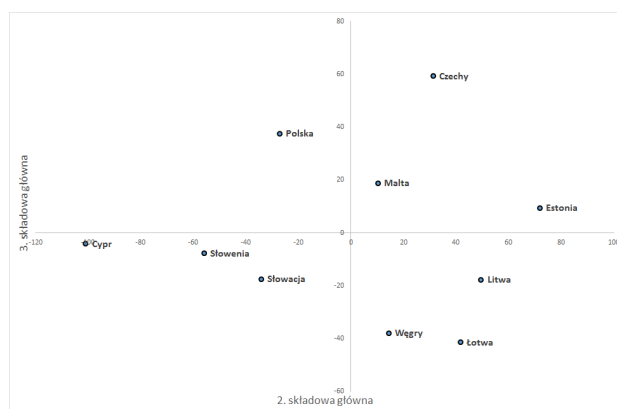
Źródło: obliczenia własne

Rysunek 5. Bi-plot pierwszej i trzeciej składowej



Źródło: obliczenia własne

Rysunek 6. Bi-plot drugiej i trzeciej składowej



Źródło: obliczenia własne

WNIOSKI

Przystąpienie do Unii Europejskiej zapewniło Polsce i pozostałym krajom UE-10 poprawę sytuacji na rynku pracy. Wskaźnik zatrudnienia w Polsce pomimo wyraźnego wzrostu w ostatnich latach nadal jest jednym z najniższych wskaźników wykorzystania zasobów pracy w krajach nowej Unii. Podobna sytuacja jest na Węgrzech natomiast nieco gorsza na Malcie. Wśród krajów które najlepiej wykorzystały swoje szanse na szybszy rozwój i pozytywne zmiany na rynku pracy po integracji z Unią Europejską należy wymienić kraje nadbałtyckie (Litwę, Łotwę i Estonię) oraz Czechy i Cypr.

Uzyskane wyniki pokazują dużą użyteczność opisanej metodyki zarówno w wizualizacji jak również we wnioskowaniu na podstawie szeregów czasowych z innej perspektywy niż było to czynione do tej pory przy użyciu powszechnie wykorzystywanych metod. Klasyczne metody statystycznej analizy wielowymiarowej (grupowania obiektów, analiza czynnikowa, analiza składowych głównych, itp.), mimo że są prostsze nie prezentują pełnego obrazu sytuacji. W prezentowanej pracy porównano kraje nowej Unii Europejskiej pod kątem jednej cechy (wskaźnika zatrudnienia) w wielu punktach czasowych. W takich analizach istnieje potrzeba badania wielu cech obserwowanych równocześnie w wielu punktach czasowych, co skłania do przeniesienia proponowanej metody na przypadek wielowymiarowy - analiza składowych głównych dla wielowymiarowych danych funkcjonalnych

BIBLIOGRAFIA

- Eurostat (2018) <http://ec.europa.eu/eurostat/data/database> (stan na dzień 11.01.2018).
- Górecki T., Krzyśko M., Waszak Ł., Wołyński W. (2014) Methods of Reducing Dimension for Functional Data. *Statistics in Transition - new series*, 15 (2), 231-242.
- Horváth L., Kokoszka P. (2012) *Inference for Functional Data with Applications*. Springer, New York.
- Jacques J., Preda C. (2014) Model-based Clustering for Multivariate Functional Data. *Computational Statistics & Data Analysis*, 71, 92-106.
- Ramsay J. O., Dalzell C. J. (1991) Some Tools for Functional Data Analysis. *Journal of the Royal Statistical Society, Ser. B*, 53, 539-572.
- Ramsay J. O., Silverman B. W. (2002) *Applied Functional Data Analysis: methods and case studies*, Springer, New York.
- Ramsay J. O., Silverman B. W. (2005) *Functional Data Analysis*, Springer, New York.

AN ANALYSIS OF EMPLOYMENT RATE IN NEW UE COUNTRIES USING FPCA

Abstract. In the paper new countries of UE are compared with respect to employment rate in 2004-2016. Functional principal components analysis was applied. On the basis of results of this analysis the groups of countries of similar employment rate were determined.

Keywords: Functional principal components analysis, employment rate