

Jacek Batóg

Instytut Ekonometrii i Statystyki
Uniwersytet Szczeciński

Identyfikacja i znaczenie obserwacji nietypowych w modelach konwergencji dochodowej

Streszczenie

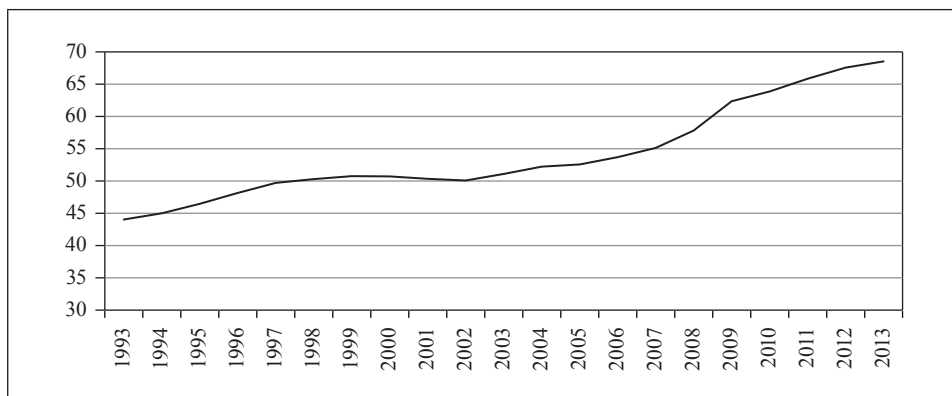
Badanie zjawiska konwergencji dochodowej znajduje szerokie odzwierciedlenie w dotychczasowym dorobku nauki i praktyce gospodarczej. Otrzymywane rezultaty charakteryzują się jednak stosunkowo dużym zróżnicowaniem. Wielu autorów wskazuje na silne uzależnienie uzyskiwanych wyników od zakresu czasowego i przekrojowego prowadzonych analiz oraz stosowanych metod badawczych. Mało uwagi poświęca się jednak roli obserwacji nietypowych, które mogą być wynikiem błędnego pomiaru, wystąpienia zdarzenia losowego, niestandardowych warunków lub działań o charakterze celowym. Weryfikacji poddana została hipoteza o istotnym wpływie tych obserwacji na uzyskiwane wyniki procesu estymacji. Głównym celem pracy było ustalenie, czy występowanie obserwacji uznanych za nietypowe istotnie zmienia jakość modeli oraz szybkość procesu konwergencji dochodowej.

Słowa kluczowe: obserwacje nietypowe, konwergencja dochodowa, modele ekonometryczne, Unia Europejska.

1. Wprowadzenie

Oceniając efekty rozwoju społeczno-gospodarczego danego kraju, dokonuje się często porównań z innymi obiektami. Jednym z prostszych sposobów stosowanych w praktyce jest odniesienie wartości określonego miernika do wartości przeciętnej charakteryzującej pewien obiekt wzorcowy. Typowym przykładem

takiego miernika jest zaprezentowany na rys. 1 udział PKB *per capita* Polski w wartości średniej dla 28 krajów Unii Europejskiej. Obserwacja kształtowania się tej proporcji w latach 1993–2013 prowadzi do co najmniej dwóch wniosków. Po pierwsze, widoczne jest zjawisko zbieżności produktu krajowego brutto przypadającego na 1 mieszkańca Polski do średniej unijnej. W okresie ostatnich 20 lat odnotowany został wzrost analizowanego udziału z 44,0% do 68,5%.



Rys. 1. Udział PKB *per capita* Polski w wartości średniej dla UE-28 w latach 1993–2013 (w %)

Źródło: obliczenia własne na podstawie [The Conference Board... 2014].

Drugie spostrzeżenie odnosi się do charakteru wpływu światowych kryzysów gospodarczych na szybkość zbliżania się poziomu życia w Polsce do poziomu obserwowanego w Unii Europejskiej. Pierwszy kryzys, który miał miejsce w latach 2001–2003, spowodował nie tylko spowolnienie tej tendencji – widoczne było nawet zjawisko jej odwrócenia w latach 2001–2002. Drugi kryzys wiązał się ze znaczącym wzrostem zbieżności PKB *per capita* Polski z poziomem europejskim, obserwowanym zwłaszcza w latach 2008–2009, czyli w okresie największego spowolnienia i recesji światowej gospodarki. Odmienny sposób kształtowania się badanej proporcji podczas drugiego kryzysu wynikał z wysokiej odporności polskiej gospodarki na ogólnoświatowy szok ekonomiczny wywołany na rynku nieruchomości w Stanach Zjednoczonych.

Zjawisko polegające na wyrównywaniu się poziomów PKB przypadających na 1 mieszkańca, spowodowane tym, że kraje charakteryzujące się niższymi poziomami kapitału na 1 zatrudnionego wykazują tendencję do szybszego wzrostu PKB *per capita*, natomiast kraje o wysokim poziomie rozwoju, posiadające wysokie wartości kapitału na 1 zatrudnionego, odznaczają się niższym tempem

wzrostu gospodarczego ze względu na malejącą krańcową produktywność kapitału, określane jest mianem konwergencji dochodowej.

Poznanie przyczyn i mechanizmu kształtowania się tego zjawiska wymaga bardziej precyzyjnych narzędzi (miar) niż analizowana powyżej proporcja¹. Wśród tych narzędzi znaczącą rolę odgrywają modele konwergencji dochodowej typu β o postaci [Batóg i Batóg 2007, s. 132–133]:

$$\frac{1}{T} \ln(Y_{iT}) / (Y_{i0}) = y_i = \alpha_0 + \alpha_1 Y_{i0} + u_i, \quad (1)$$

gdzie:

$\frac{1}{T} \ln(Y_{iT}) / (Y_{i0})$ – średnioroczna stopa wzrostu realnego PKB *per capita* kraju *i*

po zlogarytmowaniu,

Y_{i0} – początkowy poziom realnego PKB *per capita* kraju *i* po zlogarytmowaniu,

Y_{iT} – poziom realnego PKB *per capita* kraju *i* w ostatnim okresie po zlogarytmowaniu.

2. Hipoteza badawcza i cel pracy

Jakkolwiek badanie zjawiska konwergencji dochodowej znajduje szerokie odzwierciedlenie w dotychczasowym dorobku nauki i praktyce gospodarczej, uzyskiwane rezultaty charakteryzują się stosunkowo dużym zróżnicowaniem. Jest ono skutkiem m.in. odmiennych definicji konwergencji dochodowej i w efekcie konstrukcji powiązanych z nimi metod badawczych, silnego uzależnienia uzyskiwanych wyników od zakresu czasowego lub przekrojowego prowadzonych analiz decydującego o jednorodności wykorzystywanych danych, zachodzących zmian strukturalnych i wpływu pojawiających się kryzysów gospodarczych. Szersze rozważania na ten temat zaprezentowano w pracy [Batóg 2010, s. 159–160]. Stosunkowo mało uwagi poświęca się jednak znaczeniu obserwacji nietypowych, a zwłaszcza ich roli w procesie estymacji parametrów strukturalnych modeli konwergencji realnej.

Obserwacja nietypowa określana jest często jako „obserwacja znacząco różniąca się od pozostałych” [Barnett i Lewis 1994; Maddala 2006, s. 125]. Wskazuje się też, że różnica między obserwacją nietypową a pozostałymi obserwacjami jest wynikiem odmiennego procesu generowania tej pierwszej [Hawkins

¹ „Jeżeli chcemy zrozumieć, czym powodowane są znaczące różnice w jakości życia między krajami, to musimy wyjaśnić mechanizm zróżnicowania długookresowych stóp wzrostu, którymi się one charakteryzują” [Barro i Sala-i-Martin 2004, s. 6].

1980]. Precyzyjna definicja obserwacji nietypowej jest jednak uzależniona od struktury analizowanych danych oraz stosowanej metody identyfikacji [Ben-Gal 2005].

Obserwacja nietypowa może być wynikiem błędnego pomiaru, błędnego określenia badanej populacji, niejednorodności zbiorowości statystycznej, z której pobrano próbę, występowania obserwacji ekstremalnych lub zdarzeń losowych, nieprawidłowych założeń dotyczących rozkładów generowanych danych oraz działań o charakterze celowym (zob. m.in. [Walfish 2006]).

Obserwacje nietypowe często zmieniają zależność między badanymi zmiennymi w wyniku pojawiania się błędów specyfikacji modelu. Z drugiej strony dostarczają one informacji o danym procesie oraz o jego zakłóceniach. Pozwala to na liczne zastosowania praktyczne analizy obserwacji nietypowych ukierunkowane np. na wykrywanie oszustw, badanie nietypowych zachowań wyborców, wspomaganie leczenia klinicznego, wykrywanie włamań do sieci, prognozowanie anomalii pogodowych lub analizę gospodarności [Ruts i Rousseeuw 1996, Fawcett i Provost 1997, Penny i Jolliffe 2001, Acuna i Rodriguez 2004, Hozer 1993].

W przeprowadzonym badaniu weryfikacji poddana zostanie hipoteza o istotnym wpływie obserwacji nietypowych na uzyskiwane wyniki procesu estymacji parametrów strukturalnych modelu konwergencji absolutnej typu β . Głównym celem pracy jest natomiast identyfikacja obserwacji nietypowych za pomocą zróżnicowanych metod, a także ustalenie, czy ich występowanie istotnie zmienia jakość budowanych modeli oraz ocenianą szybkość procesu konwergencji dochodowej.

3. Rodzaje i metody identyfikacji obserwacji nietypowych

Prawidłowa interpretacja wpływu obserwacji nietypowych na wyniki prowadzonych analiz wymaga precyzyjnego rozróżnienia, z jakim rodzajem obserwacji nietypowej mamy do czynienia, oraz zastosowania odpowiedniej metody jej identyfikacji. Wśród klasyfikacji obserwacji nietypowych dominuje podejście, zgodnie z którym wyróżnia się cztery rodzaje tych obserwacji. Należą do nich: *univariate outlier*, czyli obserwacja o wartościach jednej ze zmiennych znacznie odbiegających od reszty obserwacji, która niekoniecznie powoduje jednak negatywne zjawiska podczas estymacji, *regression (vertical) outlier*, czyli obserwacja o znacznie odbiegającej od reszty obserwacji wartości zmiennej zależnej dla danej wartości zmiennej niezależnej, która zwiększa błędy szacunku parametrów, nie zawsze jednak powoduje zmianę nachylenia, może natomiast mieć wpływ na wyraz wolny i obniżać stopień dopasowania modelu, *leverage*, czyli obserwacja

charakteryzująca się wartościami zmiennej niezależnej znacznie odbiegającymi od jej średniej, której występowanie zmniejsza błędy szacunku parametrów i zwiększa stopień dopasowania modelu, oraz *influence*, czyli obserwacja powodująca istotną zmianę wartości ocen parametrów strukturalnych po jej usunięciu, która jednocześnie nie musi generować dużych reszt regresji.

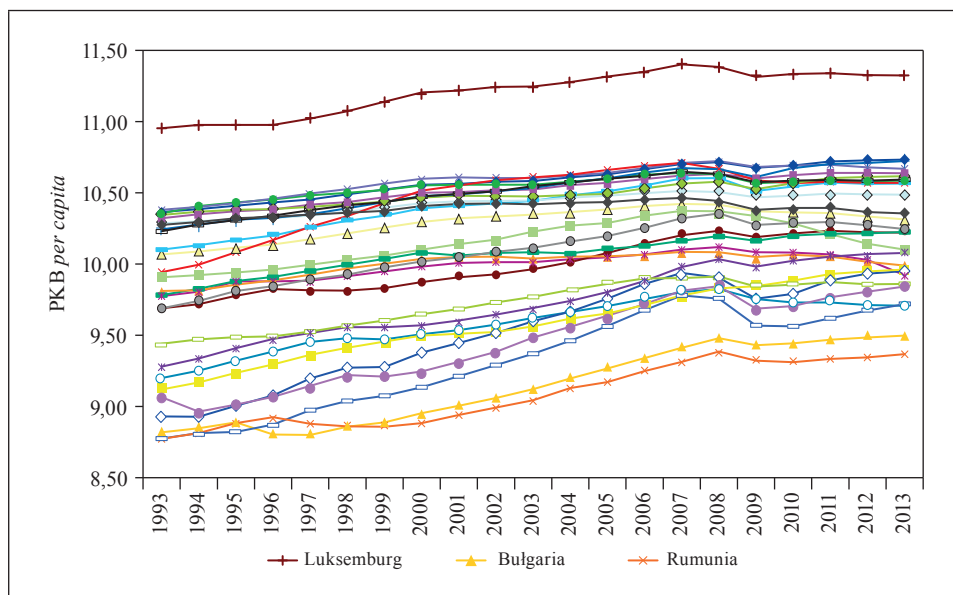
Do najczęściej stosowanych metod identyfikacji obserwacji nietypowych zaliczane są metody wykorzystujące reszty regresji standaryzowane lub studentyzowane albo wartości h_i (*hat values*) będące miarą dźwigni (zob. m.in. [Williams i in. 2002, Ben-Gal 2005]), metoda DFBETAs (*difference of betas*), pozwalająca ocenić różnicę między wartościami ocen uzyskiwanymi dla regresji przy pełnej liczbie obserwacji oraz dla regresji z usuniętą wartością nietypową, jak również metoda wykorzystująca odległość Cooka, która umożliwia ocenę wpływu danej obserwacji nietypowej na wszystkie oceny parametrów strukturalnych jednocześnie.

4. Uzyskane wyniki

W badaniu występowania obserwacji nietypowych oraz w ocenie ich znaczenia dla rezultatów analizy zjawiska konwergencji dochodowej w krajach Unii Europejskiej wykorzystane zostały dane pochodzące z *The Conference Board Total Economy Database* [2014]. Zaprezentowane w tej bazie danych szeregi czasowe PKB *per capita* podawane są w USD według poziomu cen z 2007 r., które są dodatkowo modyfikowane za pomocą EKS PPP z 2005 r.

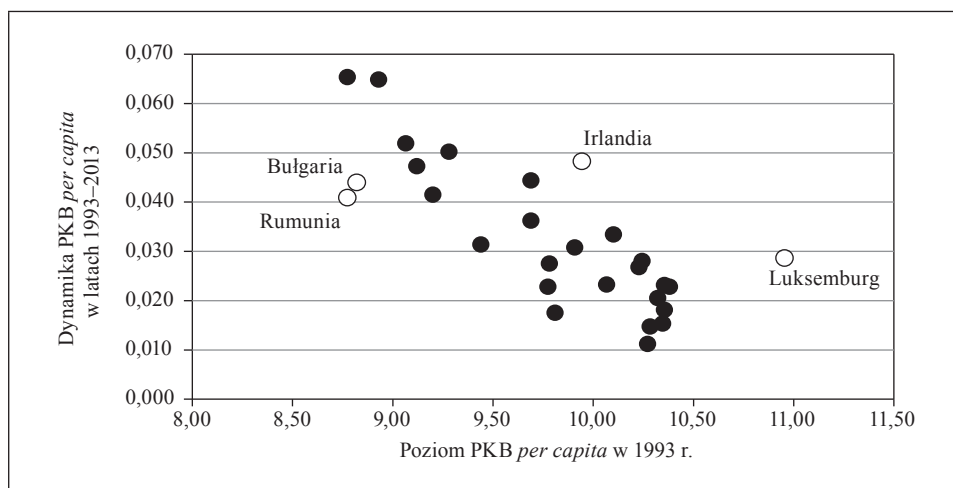
Na rys. 2 przedstawiono kształtowanie się produktów krajowych brutto na 1 mieszkańca 28 krajów Unii Europejskiej w latach 1993–2014. Wyróżniono trzy obiekty charakteryzujące się odmiennym poziomem analizowanej zmiennej. Pierwszym z nich jest Luksemburg odznaczający się znacznie wyższymi wartościami PKB *per capita* w porównaniu z innymi krajami. Dwa pozostałe to Bułgaria i Rumunia o nieco niższych w porównaniu z innymi krajami UE-28 wartościach tej zmiennej w całym badanym okresie. W grupie badanych krajów występuje jeszcze jeden obiekt – Irlandia – który również może być uznany za nietypowy w zakresie kształtowania się dynamiki PKB *per capita*, ponieważ w jego przypadku obserwowany był bardzo silny wzrost tej zmiennej (awans z 13 pozycji w 1993 r. na 2 w 2008 r.). Odmienność wymienionych krajów od pozostałych krajów UE-28 jest bardziej zauważalna na rys. 3, który prezentuje zależność początkowych poziomów i średnich stóp wzrostu realnego PKB *per capita*.

Zauważyć można również znaczący wpływ kryzysu gospodarczego na zjawisko konwergencji dochodowej obserwowane w Unii Europejskiej od 1993 r. W 2009 r. odnotowany został wzrost zróżnicowania analizowanej zmiennej, co z pewnością wpłynęło na spadek szybkości zbieżności dochodowej.



Rys. 2. PKB *per capita* w krajach Unii Europejskiej (logarytmy, USD, poziom cen z 2007 r., aktualizacja EKS PPP z 2005 r.)

Źródło: opracowanie własne na podstawie [The Conference Board... 2014].



Rys. 3. Wykres korelacyjny początkowych poziomów i średnich stóp wzrostu realnego PKB *per capita* w krajach Unii Europejskiej

Źródło: opracowanie własne na podstawie [The Conference Board... 2014].

Zależność ta została również scharakteryzowana poprzez oszacowanie parametrów strukturalnych modelu (1):

$$\hat{y}_i = 0,2224 - 0,0193 Y_{i0}, \quad R^2 = 0,61.$$

(0,029) (0,003)

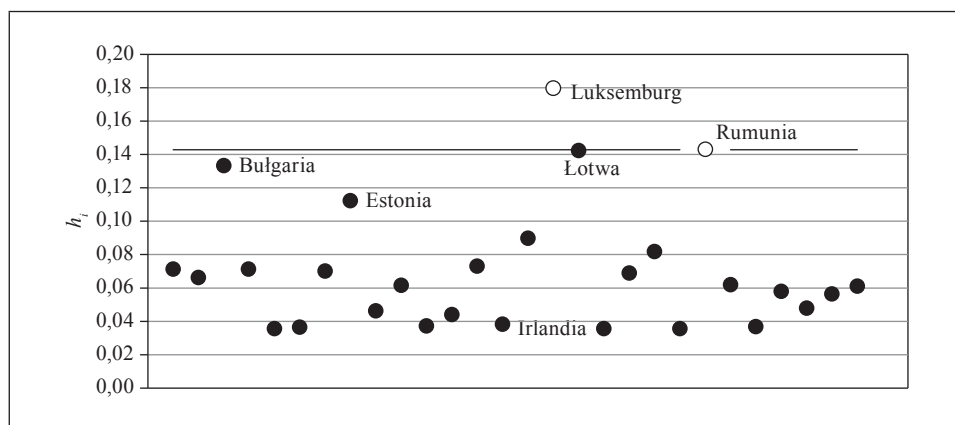
W nawiasach podano średnie błędy szacunku, które pozwalają stwierdzić statystyczną istotność parametrów strukturalnych. Uwagę zwraca jednak niezbyt wysokie dopasowanie modelu do danych rzeczywistych. Wykorzystując wzory (2) i (3), można obliczyć odpowiednio szybkość konwergencji dochodowej β oraz połowę okresu konwergencji $T_{\frac{1}{2}}$ w przypadku uwzględnienia wszystkich obiektów.

$$\beta = -\frac{1}{T} \cdot \ln(1 + \alpha_1 \cdot T), \quad (2)$$

$$T_{\frac{1}{2}} = \frac{\ln 2}{\beta}. \quad (3)$$

Uzyskane wyniki wskazują na zbieżność poziomów PKB *per capita* w krajach Unii Europejskiej w tempie 2,28% rocznie oraz połowę okresu konwergencji równą 30,38 lat.

Nietypowość dwóch obserwacji została potwierdzona za pomocą wartości *hat values* h_i (zob. rys. 4) tylko dla dwóch krajów: Luksemburga i Rumunii. Podobne rezultaty uzyskano na podstawie reszt standaryzowanych i studentyzowanych.



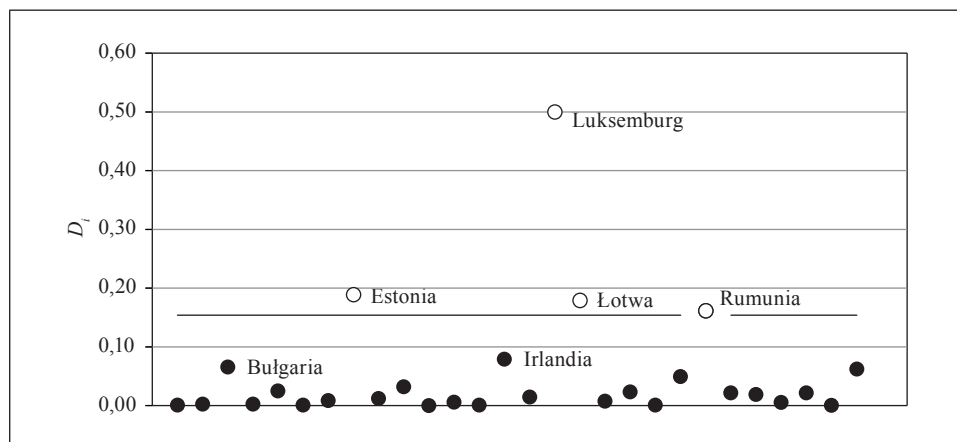
Rys. 4. Wartości *hat values* h_i dla krajów Unii Europejskiej

Źródło: opracowanie własne na podstawie [The Conference Board... 2014].

Za pomocą metody wykorzystującej odległości Cooka jako nietypowe wskazano dodatkowo Estonię i Łotwę (zob. rys. 5), przy czym identyczny wynik

uzyskano, stosując metodę DFBETAs. Jak widać, zastosowanie żadnej z miar i metod nie potwierdziło nietypowości Bułgarii i Irlandii.

W dalszej części badania oszacowane zostały parametry strukturalne modeli, w których dokonywano kolejnej eliminacji obiektów uznanych wstępnie za nietypowe na podstawie obserwacji dynamiki PKB *per capita* (zob. tabela 1).



Rys. 5. Obserwacje nietypowe wskazane za pomocą odległości Cooka

Źródło: opracowanie własne na podstawie [The Conference Board... 2014].

Tabela 1. Wyniki estymacji parametrów strukturalnych modeli oraz szybkość konwergencji dochodowej absolutnej typu β w krajach Unii Europejskiej w latach 1993–2013

Liczba krajów	R_{sk}^2	$\hat{\alpha}_1$	$t(\hat{\alpha}_1)$	$\hat{\beta}$	$T_{\frac{1}{2}}$
Wszystkie kraje Unii Europejskiej					
$n = 28$	0,60	-0,0193	-6,42	2,28%	30,38
Kraje Unii Europejskiej bez Luksemburga					
$n = 27$	0,67	-0,0220	-7,29	2,67%	25,94
Kraje Unii Europejskiej bez Luksemburga i Irlandii					
$n = 26$	0,73	-0,0225	-8,20	2,75%	25,22
Kraje Unii Europejskiej bez Luksemburga, Irlandii i Rumunii					
$n = 25$	0,76	-0,0244	-8,87	3,04%	22,77
Kraje Unii Europejskiej bez Luksemburga, Irlandii, Rumunii i Bułgarii					
$n = 24$	0,79	-0,0263	-9,37	3,35%	20,68

Źródło: obliczenia własne na podstawie [The Conference Board... 2014].

Uzyskane rezultaty pozwalają stwierdzić, że pominięcie Luksemburga, Irlandii, Bułgarii i Rumunii powoduje znaczącą zmianę wyników estymacji, a w konsekwencji również szybkości konwergencji – połowa okresu konwergencji zmniejsza się z ponad 30 lat do niecałych 21 lat. Jednocześnie potwierdzają się wyniki identyfikacji obserwacji nietypowych, zgodnie z którymi Irlandia nie została uznana za obserwację tego typu. Warto jednak zauważyć, że Bułgaria, która również nie była wskazana jako obserwacja nietypowa, charakteryzuje się silnym wpływem na obserwowaną szybkość konwergencji. W przypadku usunięcia dwóch obiektów, które zostały zidentyfikowane jako nietypowe przy wykorzystaniu odległości Cooka, czyli Estonii i Łotwy, obserwowany był natomiast spadek szybkości konwergencji (ocena parametru α_1 w obu przypadkach kształtowała się na poziomie $-0,0178$) oraz pogorszenie stopnia dopasowania modelu do rzeczywistych danych.

5. Podsumowanie

W okresie ostatnich dwudziestu lat można zaobserwować zjawisko zbieżności dochodowej krajów Unii Europejskiej. Proces ten został jednak spowolniony przez ogólnoświatowy kryzys gospodarczy zapoczątkowany w gospodarce amerykańskiej w 2007 r. Wpływ poszczególnych krajów na szybkość konwergencji dochodowej jest zróżnicowany², a kluczową rolę w tym zakresie odgrywają obiekty uznane za nietypowe. Należy jednak zauważyć, że obserwacje uznane za nietypowe na podstawie kształtowania się dynamiki PKB *per capita* nie pokrywają się w pełni z obserwacjami wskazanymi jako nietypowe za pomocą zróżnicowanych miar i metod ich identyfikacji. Warto też podkreślić, że poszczególne podejścia stosowane w identyfikacji obserwacji nietypowych generują zróżnicowane zbiory tych obserwacji. W dużym stopniu jest to zapewne rezultatem odmiennej konstrukcji stosowanych miar.

Modele z usuniętymi obserwacjami nietypowymi charakteryzują się wyższą jakością, przy czym jest to na ogół widoczne w przypadku obserwacji uznanych za nietypowe z tytułu kształtowania się dynamiki PKB *per capita*. Redukcja zbioru obserwacji o obserwacje nietypowe pozwala na precyzyjniejsze wyodrębnienie analizowanej prawidłowości dzięki eliminacji wpływu wartości różniących się istotnie od pozostałych. Nie można jednak usuwać ich w sposób automatyczny, ponieważ powoduje to zmianę obserwowanej relacji między zmiennymi. Decydujące znaczenie powinna mieć w tym przypadku analiza merytoryczna

² Dokonanie pomiaru wkładu poszczególnych krajów w proces wyrównywania się dochodów w grupie obserwowanych obiektów nie jest możliwe za pomocą modelu (1). Można w tym celu wykorzystać koncepcję pionowej konwergencji krańcowej opisaną w pracy [Batóg 2013].

oparta np. na obserwacji dynamicznych właściwości badanych zjawisk lub znaczenia poszczególnych obiektów dla formułowanych wniosków.

Literatura

- Acuna E., Rodriguez C.A. [2004], *A Meta Analysis Study of Outlier Detection Methods in Classification*, Proceedings IPSI 2004, Venice, <http://academic.uprm.edu/eacuna/paperout.pdf> (dostęp: 26.04.2014).
- Barnett V., Lewis T. [1994], *Outliers in Statistical Data*, John Wiley, Chichester.
- Barro R.J., Sala-i-Martin X. [2004], *Economic Growth*, 2nd ed., MIT Press, Cambridge, Massachusetts.
- Batóg J. [2010], *Konwergencja dochodowa w krajach Unii Europejskiej. Analiza ekonomiczna*, Rozprawy i Studia, t. DCCCLIV (780), Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin.
- Batóg J. [2013], *Analiza krańcowej pionowej konwergencji dochodowej typu β w krajach Unii Europejskiej w latach 1993–2010*, Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania, nr 31, *Metody ilościowe w ekonomii*, t. 1, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin.
- Batóg J., Batóg B. [2006], *Income Convergence in the European Countries: Empirical Analysis*, „Folia Oeconomica Stetinensia”, nr 5(13).
- Ben-Gal I. [2005], *Outlier Detection [w:] Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*, red. O. Maimon, L. Rokach, Kluwer Academic Publishers, Boston.
- The Conference Board Total Economy Database* [2014], January, <http://www.conference-board.org/data/economydatabase/> (dostęp: 22.04.2014).
- Fawcett T., Provost F. [1997], *Adaptive Fraud Detection*, „Data Mining and Knowledge Discovery”, vol. 1, nr 3.
- Hawkins D. [1980], *Identification of Outliers*, Chapman and Hall, London.
- Hozer J. [1993], *Mikroekonometria*, PWE, Warszawa.
- Maddala G.S. [2006], *Econometrics*, Wydawnictwo Naukowe PWN, Warszawa.
- Penny K.I., Jolliffe I.T. [2001], *A Comparison of Multivariate Outlier Detection Methods for Clinical Laboratory Safety Data*, „The Statistician”, vol. 50, nr 3, <http://dx.doi.org/10.1111/1467-9884.00279>.
- Ruts I., Rousseeuw P.J. [1996], *Computing Depth Contours of Bivariate Point Clouds*, „Computational Statistics and Data Analysis”, vol. 23, nr 1, [http://dx.doi.org/10.1016/s0167-9473\(96\)00027-8](http://dx.doi.org/10.1016/s0167-9473(96)00027-8).
- Walfish S. [2006], *A Review of Statistical Outlier Methods*, „Pharmaceutical Technology”, November 2.
- Williams G.J., Baxter R.A., He H.X., Hawkins S., Gu L. [2002], *A Comparative Study of RNN for Outlier Detection in Data Mining*, Proceedings of the 2nd IEEE International Conference on Data Mining (ICDM'02), Maebashi City, Japan.

Identification and Significance of Outliers in Models of Income Convergence

(Abstract)

The research of income convergence found remunerative findings in the existing literature and economic practice. The results obtained, however, show comparatively large differentiation. Many authors underline the strong dependence of the results obtained from the time and spatial character of the sample as well as the type of methods applied. Little attention is placed on the role of non-typical observations (outliers) which can occur as a result of incorrect measurement, random error, non-standard circumstances or intentional impact. The hypothesis verified was that outliers exert an essential influence on estimation results. The main objective of the analyses provided was to determine if the occurrence of such observations significantly changes the quality of the models built and the speed of the process of income convergence.

Keywords: outliers, income convergence, econometric models, European Union.