

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 327

Taksonomia 22

**Klasyfikacja i analiza danych –
teoria i zastosowania**

Redaktorzy naukowci

Krzysztof Jajuga, Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2014

Redaktor Wydawnictwa: Barbara Majewska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

www.ibuk.pl, www.ebscohost.com,

w Dolnośląskiej Bibliotece Cyfrowej www.dbc.wroc.pl,

The Central and Eastern European Online Library www.ceeol.com,

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się

na stronie internetowej Wydawnictwa

www.wydawnictwo.ue.wroc.pl

Tytuł dofinansowany ze środków Narodowego Banku Polskiego

oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Kopiowanie i powielanie w jakiegokolwiek formie

wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu

Wrocław 2014

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	9
Eugeniusz Gatnar , Balance of payments statistics and external competitiveness of Poland.....	15
Andrzej Sokolowski, Magdalena Czaja , Efektywność metody k -średnich w zależności od separowalności grup.....	23
Barbara Pawelek, Józef Pocięcha, Adam Sagan , Wielosektorowa analiza ukrytych przejść w modelowaniu zagrożenia upadłością polskich przedsiębiorstw	30
Elżbieta Gołata , Zróżnicowanie procesu starzenia i struktur demograficznych w Poznaniu i aglomeracji poznańskiej na tle wybranych dużych miast Polski w latach 2002-2011.....	39
Aleksandra Łuczak, Feliks Wysocki , Ustalanie systemu wag dla cech w zagadnieniach porządkowania liniowego obiektów	49
Marek Walesiak , Wzmacnianie skali pomiaru dla danych porządkowych w statystycznej analizie wielowymiarowej	60
Paweł Lula , Identyfikacja słów i fraz kluczowych w tekstach polskojęzycznych za pomocą algorytmu <i>RAKE</i>	69
Mariusz Kubus , Propozycja modyfikacji metody złagodzonego LASSO.....	77
Andrzej Bąk, Tomasz Bartłomowicz , Wielomianowe modele logitowe wyborów dyskretnych i ich implementacja w pakiecie <i>DiscreteChoice</i> programu R.....	85
Justyna Brzezińska , Wykorzystanie modeli logarytmiczno-liniowych do analizy bezrobocia w Polsce w latach 2004-2012.....	95
Andrzej Bąk, Marcin Pelka, Aneta Rybicka , Zastosowanie pakietu <i>dcMNM</i> programu R w badaniach preferencji konsumentów wódki	104
Barbara Batóg, Jacek Batóg , Analiza stabilności klasyfikacji polskich województw według sektorowej wydajności pracy w latach 2002-2010	113
Małgorzata Markowska, Danuta Strahl , Klasyfikacja europejskiej przestrzeni regionalnej ze względu na filary inteligentnego rozwoju z wykorzystaniem referencyjnego systemu granicznego.....	121
Kamila Migdał-Najman, Krzysztof Najman , Formalna ocena jakości odwzorowania struktury grupowej na mapie Kohonena	131
Kamila Migdał-Najman, Krzysztof Najman , Graficzna ocena jakości odwzorowania struktury grupowej na mapie Kohonena	139
Beata Basiura, Anna Czapkiewicz , Badanie jakości klasyfikacji szeregów czasowych	148
Michał Trzęsiok , Wybrane metody identyfikacji obserwacji oddalonych.....	157

Grażyna Dehnel, Tomasz Klimanek , Taksonomiczne aspekty estymacji pośredniej uwzględniającej autokorelację przestrzenną w statystyce gospodarczej.....	167
Michał Bernard Pietrzak, Justyna Wilk , Odległość ekonomiczna w modelowaniu zjawisk przestrzennych z wykorzystaniem modelu grawitacji.....	177
Maciej Beręsewicz , Próba zastosowania różnych miar odległości w uogólnionym estymatorze Petersena.....	186
Marcin Szymkowiak, Tomasz Józefowski , Konstrukcja i praktyczne wykorzystanie estymatorów typu SPREE na przykładzie dwuwymiarowych tabel kontyngencji.....	195
Marcin Pelka , Klasyfikacja pojęciowa danych symbolicznych w podejściu wielomodelowym.....	202
Małgorzata Machowska-Szewczyk , Ocena klas w rozmytej klasyfikacji obiektów symbolicznych.....	210
Justyna Wilk , Problem wyboru liczby klas w taksonomicznej analizie danych symbolicznych.....	220
Andrzej Dudek , Metody analizy skupień w klasyfikacji markerów map Google.....	229
Ewa Roszkowska , Ocena ofert negocjacyjnych w słabo ustrukturyzowanych problemach negocjacyjnych z wykorzystaniem rozmytej procedury SAW.....	237
Marcin Szymkowiak, Marek Witkowski , Zastosowanie analizy korespondencji do badania kondycji finansowej banków spółdzielczych.....	248
Bartłomiej Jefmański , Budowa rozmytych indeksów satysfakcji klientów z zastosowaniem programu R.....	257
Karolina Bartos , Odkrywanie wzorców zachowań konsumentów za pomocą analizy koszykowej danych transakcyjnych.....	266
Joanna Trzęsiok , Taksonomiczna analiza krajów pod względem dzietności kobiet oraz innych czynników demograficznych.....	275
Beata Bal-Domańska , Próba identyfikacji większych skupisk regionalnych oraz ich konwergencja.....	285
Beata Bieszk-Stolorz, Iwona Markowicz , Wpływ zasiłku na proces poszukiwania pracy.....	294
Marta Dziechciarz-Duda, Klaudia Przybysz , Wykształcenie a potrzeby rynku pracy. Klasyfikacja absolwentów wyższych uczelni.....	303
Tomasz Klimanek , Problem pomiaru procesu dezagrarnizacji wsi polskiej w świetle wielowymiarowych metod statystycznych.....	313
Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska , Wybrane metody analizy danych wzdluznych.....	321
Artur Zaborski , Zastosowanie miar odległości dla danych porządkowych do agregacji preferencji indywidualnych.....	330
Mariola Chrzanowska, Nina Drejerska, Iwona Pomianek , Zastosowanie analizy korespondencji do badania sytuacji mieszkańców strefy podmiejskiej Warszawy na rynku pracy.....	338

Katarzyna Wawrzyniak , Klasyfikacja województw według stopnia realizacji priorytetów Strategii Rozwoju Kraju 2007-2015 z wykorzystaniem wartości centrum wierszowego	346
---	-----

Summaries

Eugeniusz Gatnar , Statystyka bilansu płatniczego a konkurencyjność gospodarki Polski	22
Andrzej Sokółowski, Magdalena Czaja , Cluster separability and the effectiveness of k -means method	29
Barbara Pawelek, Józef Pocięcha, Adam Sagan , Multisectoral analysis of latent transitions in bankruptcy prediction models.....	38
Elżbieta Golata , Differences in the process of aging and demographic structures in Poznań and the agglomeration compared to selected Polish cities in the years 2002-2011	48
Aleksandra Łuczak, Feliks Wysocki , Determination of weights for features in problems of linear ordering of objects	59
Marek Walesiak , Reinforcing measurement scale for ordinal data in multivariate statistical analysis	68
Paweł Lula , Automatic identification of keywords and keyphrases in documents written in Polish.....	76
Mariusz Kubus , The proposition of modification of the relaxed LASSO method.....	84
Andrzej Bąk, Tomasz Bartłomowicz , Microeconomic multinomial logit models and their implementation in the <code>DiscreteChoice</code> R package .	94
Justyna Brzezińska , The analysis of unemployment data in Poland in 2004-2012 with application of log-linear models	103
Andrzej Bąk, Marcin Pelka, Aneta Rybicka , Application of the MMLM package of R software for vodka consumers preference analysis.....	112
Barbara Batóg, Jacek Batóg , Analysis of the stability of classification of Polish voivodeships in 2002-2010 according to the sectoral labour productivity	120
Małgorzata Markowska, Danuta Strahl , Classification of the European regional space in terms of smart growth pillars using the reference limit system.....	130
Kamila Migdał Najman, Krzysztof Najman , Formal quality assessment of group structure mapping on the Kohonen's map	138
Kamila Migdał Najman, Krzysztof Najman , Graphical quality assessment of group structure mapping on the Kohonen's map	147
Beata Basiura, Anna Czapkiewicz , Validation of time series clustering	156
Michał Trzęsiok , Selected methods for outlier detection.....	166

Grażyna Dehnel, Tomasz Klimanek , Taxonomic aspects of indirect estimation accounting for spatial correlation in enterprise statistics	176
Michał Bernard Pietrzak, Justyna Wilk , Economic distance in modeling spatial phenomena with the application of gravity model.....	185
Maciej Beręsewicz , An attempt to use different distance measures in the Generalized Petersen estimator	194
Marcin Szymkowiak, Tomasz Józefowski , Construction and practical using of SPREE estimators for two-dimensional contingency tables.....	201
Marcin Pelka , The ensemble conceptual clustering for symbolic data.....	209
Małgorzata Machowska-Szewczyk , Evaluation of clusters obtained by fuzzy classification methods for symbolic objects.....	219
Justyna Wilk , Problem of determining the number of clusters in taxonomic analysis of symbolic data	228
Andrzej Dudek , Clustering techniques for Google maps markers.....	236
Ewa Roszkowska , The evaluation of negotiation offers in ill structure negotiation problems with the application of fuzzy SAW procedure	247
Marcin Szymkowiak, Marek Witkowski , The use of correspondence analysis in analysing the financial situation of cooperative banks.....	256
Bartłomiej Jefmański , The construction of fuzzy customer satisfaction indexes using R program.....	265
Karolina Bartos , Discovering patterns of consumer behaviour by market basket analysis of the transactional data.....	274
Joanna Trzęsiok , Cluster analysis of countries with respect to fertility rate and other demographic factors	284
Beata Bal-Domańska , An attempt to identify major regional clusters and their convergence	293
Beata Bieszk-Stolorz, Iwona Markowicz , The influence of benefit on the job finding process	302
Marta Dziechciarz-Duda, Klaudia Przybysz , Education and labor market needs. Classification of university graduates	312
Tomasz Klimanek , The problem of measuring deagrarianisation process in rural areas in Poland using multivariate statistical methods.....	320
Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska , Selected methods for an analysis of longitudinal data.....	329
Artur Zaborski , The application of distance measures for ordinal data for aggregation individual preferences	337
Mariola Chrzanowska, Nina Drejerska, Iwona Pomianek , Application of correspondence analysis to examine the situation of the inhabitants of Warsaw suburban area in the labour market	345
Katarzyna Wawrzyniak , Classification of voivodeships according to the level of the realization of priorities of <i>the National Development Strategy 2007-2015</i> with using the values of centroid of the rows	355

Beata Bal-Domańska

Uniwersytet Ekonomiczny we Wrocławiu

PRÓBA IDENTYFIKACJI WIĘKSZYCH SKUPISK REGIONALNYCH ORAZ ICH KONWERCENCJA

Streszczenie: Celem artykułu jest identyfikacja oraz ocena znaczenia relacji przestrzennych dla estymacji modeli konwergencji dochodowej na poziomie regionów NUTS-2 Unii Europejskiej. Ponadto podjęto próbę wskazania właściwej definicji macierzy odległości oraz w dalszej kolejności istnienia efektu dyfuzji. Analizy dokonano dla danych z lat 2000-2011. Do analizy wykorzystano narzędzia statystyki (statystyki *I* Morana i *join-count*) i ekonometrii przestrzennej (modele opóźnienia i błędu przestrzennego).

Słowa kluczowe: ekonometria przestrzenna, macierz odległości, konwergencja, regiony NUTS-2 UE.

1. Wstęp

Dysproporcje w rozwoju gospodarczym są jednym z kluczowych problemów polityki regionalnej i spójności Unii Europejskiej. W przekroju jednostek terytorialnych widoczne są tendencje do powstawania klas silnych centrów rozwoju i obszarów peryferyjnych. Istnienie skupisk jednostek terytorialnych, w których wartości określonej zmiennej determinują i są determinowane przez jej realizacje w innych lokalizacjach, prowadzi do procesów (dodatniej) autokorelacji przestrzennej.

Zależności przestrzenne mogą być wynikiem istnienia powiązań między sąsiadującymi jednostkami, które przekraczają wyznaczone granice administracyjne. Może to wynikać np. z dojazdów do pracy, powiązań między firmami a dostawcami i podmiotami współpracującymi. W analizach prowadzonych na poziomie dużych jednostek terytorialnych, jak regiony czy kraje, powiązania mogą wynikać z otwartości rynków na współpracę, z ustroju politycznego, ale także szoków, jak wojny itp.

W literaturze przedmiotu prezentowane są liczne wyniki badań regionalnych prowadzonych na podstawie szeregów przekrojowych [Markowska, Strahl 2012] czy panelowych (np. dla państw [Ciołek 2004], regionów [Bal-Domańska 2013]). W badaniach tych często pomija się wpływ aspektów przestrzennych na relacje

łącznie analizowane zjawiska i w konsekwencji oszacowania parametrów strukturalnych modeli. Z metodologicznego punktu widzenia istnienie zależności przestrzennych (nieuwzględnionych w strukturze modelu) może skutkować niską jakością oszacowań parametrów strukturalnych modeli.

Problem relacji przestrzennych i ich wpływ na oszacowania parametrów strukturalnych modeli zostanie omówiony na przykładzie modeli rozwoju regionalnego i konwergencji. Zagadnienia te były tematem wielu analiz przeprowadzonych przez autorkę oraz innych badaczy. W niniejszym artykule podjęto próbę weryfikacji wpływ aspektów przestrzennych na wyniki analizy konwergencji dochodowej regionów szczebla NUTS-2 państw Unii Europejskiej w latach 2000-2011.

Ogólnie model konwergencji absolutnej można opisać jako wpływ początkowej wartości poziomu rozwoju gospodarczego na tempo wzrostu gospodarczego:

$$\frac{1}{T}(\ln y_{i0+T} - \ln y_{i0}) = \alpha + \theta \ln(y_{i0}) + \varepsilon_{it},$$

gdzie: $\theta = -\frac{(1 - e^{-\beta^k T})}{T}$ – parametr określający szybkość zbieżności do stanu równowagi długookresowej (odległość pokonywana w ciągu roku), T – liczba lat¹.

Uzyskanie ujemnej, istotnej oceny parametru stojącego przy początkowym poziomie dochodów θ oznacza potwierdzenie istnienia konwergencji. Wartość tej oceny informuje o kierunku zależności między początkowym poziomem rozwoju a tempem wzrostu gospodarczego. Parametr β^k informuje o szybkości konwergencji, czyli o jaki procent odległości w kierunku stanu równowagi długookresowej gospodarka pokonuje w ciągu 1 okresu.

2. Podstawy teoretyczne analiz przestrzennych

Jakość i dokładność oszacowań parametrów strukturalnych modelu zależy od stopnia spełnienia założeń wykorzystanej metody estymacji. Jedną z najpopularniejszych metod szacowania modeli regresji jest klasyczna metoda najmniejszych kwadratów (KMNK). Najefektywniejszy liniowy estymator nieobciążony KMNK BLUE (*best linear unbiased estimator*) powinien spełniać określone założenia co do rozkładu składnika losowego, w tym założenie o niezależności obserwacji.

Jednym z kluczowych postulatów KMNK jest założenie sferyczności składnika losowego. Jednym z przejawów niesferyczności składnika losowego jest korelacja (powiązania) i -tego składnika losowego (gdzie $i = 1, 2, \dots, L$ oznacza liczbę obserwacji wykorzystanych do oszacowania modelu) z wartościami składnika losowego dla innych obserwacji. Niespełnienie tego założenie powoduje, że estymator para-

¹ Często jest definiowany jako $\theta = -(1 - e^{-\beta^k T})$, stąd $\beta = -\frac{\ln(1 + \theta)}{T}$ [Arbia 2006, s. 8-14].

metrów strukturalnych jest nadal zgodny i nieobciążony, ale staje się mało efektywny, a błędy standardowe są obciążone. Stosowanie standardowych statystyk testowych w modelu z niesferycznym składnikiem losowym może skutkować błędnymi wnioskami.

Identyfikacja zjawiska autokorelacji w szeregach przekrojowych nie jest łatwa ze względu na brak naturalnego uporządkowania obserwacji. Elementem utrudniającym ocenę powiązań między obiektami badania jest możliwość wystąpienia powiązań wielokierunkowych, gdy wpływ może następować zarówno z obiektu i do $i + 1$, jak i z $i + 1$ do i . W przypadku jednakowej siły oddziaływania mówimy o warunkach izotropowości, w przypadku zróżnicowania oddziaływań w zależności od kierunku – o anizotropowości.

Przyjmuje się, że autokorelacja w szeregach danych opisujących obiekty (gospodarstwa domowe, regiony, państwa) jest wynikiem powiązania z sobą wybranych obiektów. W szczególności dotyczy to obiektów umieszczonych w pewnym sąsiedztwie. Zakłada się, że siła oddziaływania między jednostkami maleje wraz ze wzrostem odległości (zgodnie z pierwszym prawem Toblera). Z dodatnią autokorelacją przestrzenną mamy do czynienia w sytuacji, gdy obiekty sąsiednie mają zbliżone wartości. Autokorelacja ujemna to zdecydowanie różne wartości w obiektach położonych w swoim otoczeniu. Najczęściej występującym przypadkiem autokorelacji w badaniach regionalnych jest autokorelacja dodatnia przejawiająca się powstawaniem klastrów, czyli obiektów o podobnych wartościach zmiennej (wysokich lub niskich).

Kluczowym pojęciem dla autokorelacji przestrzennej jest pojęcie sąsiedztwa i definicja macierzy je opisującej. Jej poprawne skonstruowanie wymaga nie tylko rozpoznania sieci i kierunku powiązań, ale także zasięgu oddziaływania. W ekonometrycznych analizach przestrzennych możemy spotkać dwa typy macierzy sąsiedztwa wynikające z odległości fizycznej i ekonomicznej między obiektami [Arbia 2006; Suchecki (red.) 2010; Pietrzak 2010]. Istnieje kilka możliwych wariantów konstrukcji macierzy zależności przestrzennych. Wykorzystanie różnych definicji odległości może prowadzić do odmiennych rezultatów [Młodak 2012].

Do określenia sąsiedztwa obiektów wykorzystywanych jest kilka podejść. Do najpopularniejszych – stosowanych w niniejszym opracowaniu – należą macierze:

1. sąsiedztwa n -tego rzędu, gdzie n to liczba granic, które trzeba przekroczyć od jednego obiektu do drugiego,
2. k najbliższych sąsiadów – definiowana na podstawie odległości od środków ciężkości regionu, określonych jako współrzędne geograficzne,
3. sąsiadów w promieniu d – których środki ciężkości znajdują się w linii prostej w odległości mniejszej niż d .

Wzięcie pod uwagę macierzy relacji przestrzennych uwzględniających odległość d lub liczbę sąsiadów k , umożliwia identyfikację efektu dyfuzji, czyli znaczenia dla interakcji przestrzennych coraz to dalszych sąsiadów.

Kolejnym ważnym krokiem analizy jest modelowy opis interakcji przestrzennych. Przestrzenne zależności można uwzględnić w modelu na trzy sposoby, jako:

- autokorelację przestrzenną składnika losowego (SEM – *Spatial Error Model*) – gdy w modelu pominięto zmienne przestrzennie autokorelowane, a składnik przestrzenny jest częścią elementu losowego,
- autoregresję przestrzenną (SAR/SLM – *Spatial Lag Model*) – gdy wartości zmiennej endogenicznej Y z jednostki s wpływają na kształtowanie się zmiennej w jednostce i oraz innych lokalizacjach,
- krzyżową regresję przestrzenną – gdy na wartości zmiennej Y w jednostce i wpływają wartości zmiennych objaśniających X z innych jednostek.

Powstaje pytanie, która konstrukcja modelowa jest właściwa. Anselin wskazuje, że w sytuacji, gdy rozważamy model równowagi procesów przestrzennych lub społecznych, w których na kształtowanie się zmiennej objaśnianej mają wpływ także procesy zachodzące w sąsiednich jednostkach, właściwy jest model autoregresji przestrzennej. Jeżeli natomiast autokorelacja nie wynika z przestrzennych lub społecznych interakcji, a jest po prostu wynikiem niesferyczności składnika losowego, wtedy właściwy jest model z przestrzenną autokorelacją składnika losowego. Do wyboru właściwej konstrukcji modelu możliwe jest wykorzystanie testów statystycznych opartych na mnożnikach Lagrange’a LM (*Lagrange Multiplier*) [Anselin, Bera 1998]. Do ostatecznego wyboru modelu przydatne są także statystyki oparte na funkcji wiarygodności. W niniejszym badaniu wykorzystano kryterium informacyjne Akaike (AIC) [Akaike 1974] oraz bayesowskie kryterium Schwarza (BIC) [Schwarz 1978].

Po określeniu schematu interakcji przestrzennych kolejnym krokiem analizy jest weryfikacja istnienia lub braku zależności przestrzennych między obiektami. Najpopularniejszym testem pozwalającym ocenić siłę zależności przestrzennej jest statystyka I Morana [Moran 1947; Cliff, Ord 1981]. Statystyka ta istnieje w dwóch wariantach – globalna, do określenia ogólnego podobieństwa regionów, oraz lokalna, która określa, czy i -ty obiekt (region) otoczony jest obiektami o podobnych lub różnych wartościach. Globalną statystykę I Morana zapisać można jako [Arbia 2006]:

$$I = \frac{L}{\sum_{i=1}^L \sum_{j=1}^L w_{ij}} \frac{\sum_{i=1}^L \sum_{j=1}^L w_{ij} \hat{e}_i \hat{e}_j}{\sum_{i=1}^L \hat{e}_i^2},$$

gdzie: w_{ij} – elementy macierzy odległości, \hat{e}_i – reszty modelu oszacowanego KMNK.

W przypadku braku autokorelacji przestrzennej statystyka I Morana ma tendencję do przyjmowania wartości $I \approx -\frac{1}{n-1}$. Do oceny zależności przestrzennych wykorzystany może być także test *join-count* [Kopczewska 2006], który pozwala na weryfikację, czy zależność przestrzenna dotyczy reszt dodatnich, czy ujemnych. Ideą testu jest weryfikacja prawdopodobieństwa stykania się obiektów o podobnych wartościach. Hipoteza zerowa w tym teście zakłada, że rozkład zdarzeń w przestrzeni jest losowy i autokorelacja nie występuje.

Do weryfikacji wpływu struktury przestrzennej na wyniki analizy konwergencji w badaniu wykorzystano następującą procedurę:

1. Określenie obiektu badania i powiązań między nimi.
2. Zdefiniowanie macierzy odległości – przyjęto standaryzowaną wierszami macierz wag według kryterium wspólnej granicy (**W**).
3. Oszacowanie modelu KMNK (bez uwzględnienia relacji przestrzennych).
4. Ocena wystąpienia autokorelacji przestrzennej reszt z wykorzystaniem testu *I* Morana oraz *join-count*.
5. Wybór typu relacji przestrzennej – modele błędu lub opóźnienia przestrzennego – z wykorzystaniem testu LM (*Lagrange Multiplier*).
6. Oszacowanie parametrów modeli przestrzennych:
 - model opóźnienia przestrzennego (SAR/SLM):

$$\frac{1}{T}(\ln y_{i0+T} - \ln y_{i0}) = \alpha - \pi \ln(y_{i0}) + \delta W\left(\frac{1}{T}(\ln y_{i0+T} - \ln y_{i0})\right) + \varepsilon \quad \varepsilon \sim IID,$$

- model błędu przestrzennego (SEM):

$$\frac{1}{T}(\ln y_{i0+T} - \ln y_{i0}) = \alpha - \pi \ln(y_{i0}) + \varepsilon \quad \varepsilon = \lambda W\varepsilon + \zeta.$$

Modele opóźnienia przestrzennego szacowane są metodą największej wiarygodności – możliwość ekonomicznej interpretacji parametrów. Model błędu przestrzennego oszacowano metodą GLS, a δ przez optymalizację. Wszystkie obliczenia wykonano w programie R.

7. Oszacowanie modeli przestrzennych dla różnych typów macierzy.

Oprócz zdefiniowanej w 2. punkcie procedury standaryzowanej macierzy wag według kryterium wspólnej granicy (**W**) rozważono następujące macierze wag:

k – macierz wag według kryterium najbliższych 10 sąsiadów,

d – macierz wag według kryterium sąsiadów w promieniu 200 km.

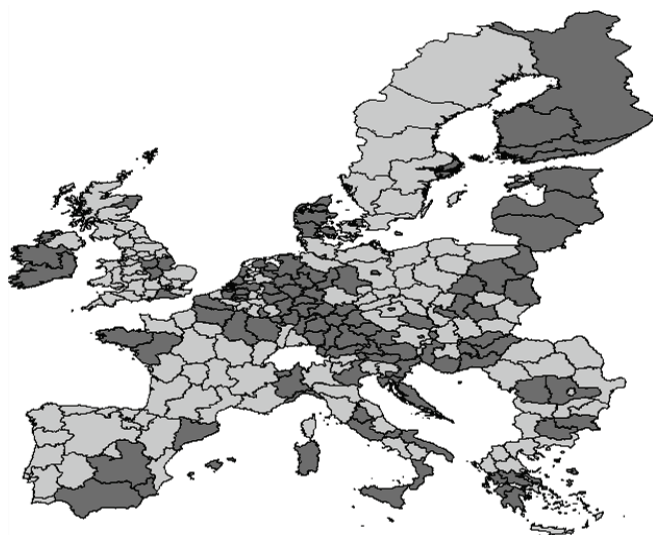
W badaniu uwzględniono 261 regionów Unii Europejskiej szczebla NUTS-2. W badaniu pominięto oddalone od centralnej części kontynentu wyspy lub państwa wyspiarskie (Kypros, Malta, zamorskie regiony Francji: Guadeloupe, Martinique, Guyane, Réunion, trzy regiony hiszpańskie: Ciudad Autónoma de Ceuta, Ciudad Autónoma de Melilla i Canarias oraz dwa portugalskie Região Autónoma dos Açores i Região Autónoma da Madeira). Badanie dotyczyło lat 2000-2011.

3. Wyniki estymacji modeli konwergencji regionalnej w zakresie PKB

Analizę konwergencji rozpoczęto od oszacowania modelu obrazującego wpływ początkowego poziomu PKB (w przeliczeniu na 1 mieszkańca według parytetu siły nabywczej) na tempo wzrostu gospodarczego w przekroju regionów NUTS-2 Unii Europejskiej z wykorzystaniem klasycznej metody najmniejszych kwadratów.

Oszacowania podano w tab. 2. Uzyskana ujemna ocena parametru strukturalnego wskazuje na zachodzenie procesów beta konwergencji i doganianie przez regiony o niskim poziomie rozwoju w 2000 roku poziomu rozwoju regionów bogatych.

Analiza reszt modelu wykazała istnienie autokorelacji przestrzennej. Wartość statystyki *I* Morana wyniosła 0,2013. Rozkład reszt przedstawiony został na rys. 1. Wyraźnie zarysowują się klastry regionów o dodatnich lub ujemnych wartościach reszt modelu. Wyniki testu *join-count* pozwoliły przy poziomie istotności 0,01 odrzucić hipotezę zerową o braku autokorelacji przestrzennej dla każdej grupy reszt dodatniej i ujemnej. Potwierdzając tym samym tendencje do tworzenia skupisk regionów, dla których reszty przyjęły zarówno wartość dodatnią, jak i ujemną. Wartości statystyk testowych wynosiły odpowiednio dla reszt ujemnych – 44,2, dla reszt dodatnich – 32,9.



Rys. 1. Rozkład wartości reszt modelu konwergencji dla regionów NUTS-2 Unii Europejskiej (jasny kolor reszty ujemne; ciemny kolor reszty dodatnie)

Źródło: opracowanie własne.

Wyniki analizy przestrzennego rozkładu reszt potwierdziły istnienie zależności, dlatego w kolejnym kroku podjęto próbę wprowadzenia interakcji przestrzennych w strukturę modelu konwergencji. O wyborze właściwej definicji zależności przestrzennej zdecydowano na podstawie wyników testu LM oraz w dalszej kolejności – wartości kryterium informacyjnego AIC. Wyniki testu LM, zarówno w wersji podstawowej, jak i odpornej, wskazały model błędu przestrzennego jako właściwy do opisu konwergencji (por. tab. 1).

Oceny parametru strukturalnego stojącego przy logarytmie początkowego poziomu PKB, uzyskane w każdym z modeli, nie różniły się znacząco (tab. 2).

Wszystkie miały ujemne znaki i kształtowały się od $-0,226$ w modelu opóźnienia przestrzennego do $-0,24$ w podstawowym oszacowanym KMNK. Tym samym parametr zbieżności kształtował się od $0,0213$ do $0,023$. Wartości kryterium informacyjnego AIC potwierdziły wybór modelu błędu przestrzennego jako właściwego do opisu procesów konwergencji absolutnej.

Tabela 1. Wartości testu LM i LM odpornego

Wyszczególnienie	LM		LM odporny (<i>robust</i>)	
	wartość statystyki	<i>p-value</i>	wartość statystyki	<i>p-value</i>
Model opóźnienia przestrzennego	10,2987	0,00133	0,0323	0,8573
Model błędu przestrzennego	20,5854	5,703e-06	10,319	0,00132

Źródło: opracowanie własne.

Niewielkie różnice w oszacowaniach sugerują, iż interakcje przestrzenne nie wpływały w sposób znaczący na oszacowania ocen parametrów strukturalnych i pomimo istnienia w przekroju regionów klastrów o niskich i wysokich wartościach reszt, oceny parametrów strukturalnych pozostały na zbliżonym poziomie w modelach bez uwzględnienia struktur zależności przestrzennych i z ich uwzględnieniem.

Tabela 2. Oszacowania modelu beta konwergencji absolutnej dla regionów NUTS-2 w latach 2000-2011

Wyszczególnienie	Podstawowy (KMNK)	Opóźnienia przestrzennego (SLM)	Błąd przestrzennego (SEM)
Wyraz wolny	2,604 (0,155)***	2,422 (0,165)***	2,577 (0,162)***
lnPKB2000	$-0,240$ (0,016)***	$-0,226$ (0,016)***	$-0,237$ (0,017)***
β / <i>Half-life</i> (lata)	0,023 / 30,3	0,0213 / 32,5	0,0226 / 30,7
AIC	-350,818	-357,911	-365,688

***statystycznie istotne przy poziomie 0,001.

Źródło: opracowanie własne.

Tabela 3. Oszacowania modelu beta konwergencji absolutnej dla regionów NUTS-2 w latach 2000-2011

Typ macierzy wag	AIC	BIC
W – standaryzowana macierz wag według kryterium wspólnej granicy	-365,6882	-351,4301
k – macierz wag według kryterium najbliższych 10 sąsiadów	-364,1032	-349,8451
d – macierz wag według kryterium sąsiadów w promieniu 200 km	-363,7816	-349,5235

Źródło: opracowanie własne.

Dodatkowo oszacowano modelu błędu przestrzennego przy uwzględnieniu różnych typów macierzy odległości (tab. 3). Testy oparte na kryteriach informacyjnych AIC i BIC wskazały jako najlepsze oszacowania z wykorzystaniem standaryzowanej macierzy wag według kryterium wspólnej granicy pierwszego rzędu. W dalszej kolejności kryteria wskazały na macierz wag według kryterium najbliższych 10 sąsiadów, najgorzej wypadły oszacowania dla kryterium sąsiadów w promieniu 200 km.

Wyniki potwierdzają istotność na poziomie regionalnym relacji przestrzennych z najbliższymi sąsiadami. Ta struktura przestrzenna pozwoliła w największym stopniu na poprawę zawartości informacyjnej modelu.

4. Podsumowanie

W procesie estymacji modeli konwergencji na poziomie regionalnym wystąpił problem autokorelacji przestrzennej reszt (zarówno dla reszt dodatnich, jak i ujemnych).

Wprowadzenie interakcji przestrzennych do modeli konwergencji w niewielkim stopniu wpłynęło na szybkość dochodzenia do stanu równowagi długookresowej. W modelu podstawowym (KMNK) okres potrzebny na zmniejszenie o połowę dystansu do stanu równowagi długookresowej oszacowano w latach na poziomie 30,3, w modelu błędu przestrzennego był niewiele dłuższy i wynosił 30,7.

Najkorzystniejsze rezultaty estymacji uzyskano dla macierzy uwzględniającej relacje przestrzenne jedynie między najbliższymi sąsiadami, bezpośrednio graniczącymi ze sobą.

Nie bez znaczenia dla wyników oszacowań jest wielkość regionów. Regiony szczebla NUTS-2, będące obiektem analizy, obejmują relatywnie duże obszary o zróżnicowanej strukturze zagospodarowania przestrzennego i gospodarczego, które często stanowią mało spójne wewnętrznie gospodarki. Powodować to może rozmycie się relacji przestrzennych, gdyż największe oddziaływania przestrzenne są we wnętrzu regionów. Warto w dalszym etapie badania zweryfikować znaczenie interakcji przestrzennych na poziomie podregionów NUTS-3, które są jednostkami mniejszymi.

Literatura

- Akaike H. (1974), *A new look at the statistical model identification*, „IEEE Transactions on Automatic Control”, 19(6), s. 716-723.
- Anselin L., Bera A. (1998), *Spatial dependence in linear regression models with an introduction to spatial econometrics*, [w:] A. Ullah, D.E. Giles (red.), *Handbook of Applied Economic Statistics*, Marcel Dekker, New York, s. 237-289.
- Arbia G. (2006), *Spatial econometrics*, Springer, Berlin – Heidelberg.
- Bal-Domańska B. (2013), *Procesy konwergencji wydajności pracy w regionach Unii Europejskiej*, „Wiadomości Statystyczne” 2, s. 1-14.

- Ciołek D. (2004), *Szacowanie regresji wzrostu i konwergencji dla danych panelowych*, [w:] A. Welfe (red.), *Metody ilościowe w naukach ekonomicznych*, t. 4, SGH w Warszawie, Oficyna Wydawnicza, Warszawa.
- Cliff A.D., Ord J.K. (1981), *Spatial Processes: Models and Applications*, Pion, London.
- Kopczewska K. (2006), *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu.PL, Warszawa.
- Markowska M., Strahl D. (2012), *Evaluation of the European Union regions convergence regarding innovation*, „Argumenta Oeconomica”, Wrocław, 1 (28), s. 41-67.
- Młodak A. (2012), *Sąsiedztwo obszarów przestrzennych w ujęciu fizycznym oraz społeczno-ekonomicznym – podejście taksonomiczne*, [w:] K. Jajuga, M. Walesiak, *Klasyfikacja i analiza danych – teoria i zastosowania*, Prace Naukowe Uniwersytetu Ekonomicznego nr 242, Taksonomia 19. Uniwersytet Ekonomiczny, Wrocław, s. 76-85.
- Moran P.A.P. (1947), *The Interpretation of Statistical Maps*, „Journal of the Royal Statistical Society”, B10, s. 243-251.
- Pietrzak M. (2010), *Application of economic distance for the purposes of a spatial analysis of the unemployment rate for Poland*, „Oeconomia Copernicana”, vol. 1, s. 79-98.
- Suchecky B. (red.) (2010), *Ekonometria przestrzenna*, C.H. Beck, Warszawa.
- Schwarz G. (1978), *Estimating the dimension of a model*, „Annals of Statistics” 6, s. 46-464.

AN ATTEMPT TO IDENTIFY MAJOR REGIONAL CLUSTERS AND THEIR CONVERGENCE

Summary: The objective of the article is to identify and assess the significance of spatial relations for the estimation of convergence models at the European Union NUTS-2 level regions. It was attempted to indicate the correct definition of the distance matrix and next the occurrence of diffusion effect. The analysis was performed based on data covering the period 2000-2011. Statistical (*I* Moran statistics and *joint*-count) and spatial econometrics (spatial lag and spatial error models) tools were applied in the course of the conducted analysis.

Keywords: spatial econometrics, distance matrix, convergence, European Union NUTS-2 level regions.