

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 384

Taksonomia 24

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

Redaktorzy naukowci

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2015

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronie internetowej Wydawnictwa
www.pracnaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

Publikacja udostępniona na licencji Creative Commons
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2015

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp.....	9
Krzysztof Jajuga, Józef Pociecha, Marek Walesiak: 25 lat SKAD.....	15
Beata Basiura, Anna Czapkiewicz: Symulacyjne badanie wykorzystania entropii do badania jakości klasyfikacji.....	25
Andrzej Bąk: Zagadnienie wyboru optymalnej procedury porządkowania liniowego w pakiecie <code>pllord</code>	33
Justyna Brzezińska: Analiza klas ukrytych w badaniach sondażowych.....	42
Grażyna Dehnel: Rejestr podatkowy oraz rejestr ZUS jako źródło informacji dodatkowej dla statystyki gospodarczej – możliwości i ograniczenia ..	51
Sabina Denkowska: Wybrane metody oceny jakości dopasowania w <i>Propensity Score Matching</i>	60
Marta Dziechciarz-Duda, Klaudia Przybysz: Zastosowanie teorii zbiorów rozmytych do identyfikacji pozafiskalnych czynników ubóstwa.....	75
Iwona Foryś: Potencjał rynku mieszkaniowego w Polsce w latach dekonjunktury gospodarczej.....	84
Eugeniusz Gatnar: Statystyczna analiza konwergencji krajów Europy Środkowej i Wschodniej po 10 latach członkostwa w Unii Europejskiej.....	93
Ewa Genge: Zaufanie do instytucji publicznych i finansowych w polskim społeczeństwie – analiza empiryczna z wykorzystaniem ukrytych modeli Markowa.....	100
Alicja Grześkowiak: Wielowymiarowa analiza uwarunkowań zaangażowania Polaków w kształcenie ustawiczne o charakterze pozaformalnym.....	108
Monika Hamerska: Wykorzystanie metod porządkowania liniowego do tworzenia rankingu jednostek naukowych.....	117
Bartłomiej Jefmański: Zastosowanie modeli IRT w konstrukcji rozmytego systemu wag dla zmiennych w zagadnieniu porządkowania liniowego – na przykładzie metody TOPSIS.....	126
Tomasz Józefowski, Marcin Szymkowiak: Wykorzystanie uogólnionej miary odległości do porządkowania liniowego powiatów województwa podkarpackiego w świetle funkcjonowania specjalnej strefy ekonomicznej Euro-Park Mielec.....	135
Krzysztof Kompa: Zastosowanie testów parametrycznych i nieparametrycznych do oceny sytuacji na światowym rynku kapitałowym przed kryzysem i po jego wystąpieniu.....	144
Mariusz Kubus: Rekurencyjna eliminacja cech w metodach dyskryminacji....	154

Marta Kuc: Wpływ sposobu definiowania macierzy wag przestrzennych na wynik porządkowania liniowego państw Unii Europejskiej pod względem poziomu życia ludności	163
Paweł Lula: Kontekstowy pomiar podobieństwa semantycznego	171
Iwona Markowicz: Model regresji Feldsteina-Horioki – wyniki badań dla Polski	182
Kamila Migdał-Najman: Ocena wpływu wartości stałej Minkowskiego na możliwość identyfikacji struktury grupowej danych o wysokim wymiarze	191
Małgorzata Misztal: O zastosowaniu kanonicznej analizy korespondencji w badaniach ekonomicznych.....	200
Krzysztof Najman: Zastosowanie przetwarzania równoległego w analizie skupień	209
Edward Nowak: Klasyfikacja danych a rachunkowość. Rozważania o relacjach	218
Marcin Pelka: Adaptacja metody <i>bagging</i> z zastosowaniem klasyfikacji pojęciowej danych symbolicznych.....	227
Józef Pocięcha, Mateusz Baryła, Barbara Pawelek: Porównanie skuteczności klasyfikacyjnej wybranych metod prognozowania bankructwa przedsiębiorstw przy losowym i nielosowym doborze prób	236
Agnieszka Przedborska, Małgorzata Misztal: Wybrane metody statystyki wielowymiarowej w ocenie jakości życia słuchaczy uniwersytetu trzeciego wieku	246
Wojciech Roszka: Konstrukcja syntetycznych zbiorów danych na potrzeby estymacji dla małych domen	254
Aneta Rybicka: Połączenie danych o preferencjach ujawnionych i wyrażonych	262
Elżbieta Sobczak: Poziom specjalizacji w sektorach intensywności technologicznej a efekty zmian liczby pracujących w województwach Polski	271
Andrzej Sokołowski, Grzegorz Harańczyk: Modyfikacja wykresu radarowego	280
Marcin Szymkowiak, Marek Witkowski: Wykorzystanie mediany do klasyfikacji banków spółdzielczych według stanu ich kondycji finansowej ..	287
Justyna Wilk, Michał B. Pietrzak, Roger S. Bivand, Tomasz Kossowski: Wpływ wyboru metody klasyfikacji na identyfikację zależności przestrzennych – zastosowanie testu <i>join-count</i>	296
Dorota Witkowska: Wykorzystanie drzew klasyfikacyjnych do analizy zróżnicowania płac w Niemczech	305
Artur Zaborski: Analiza niesymetrycznych danych preferencji z wykorzystaniem modelu punktu dominującego i modelu grawitacji.....	315

Summaries

Krzysztof Jajuga, Józef Pociecha, Marek Walesiak: XXV years of SKAD	24
Beata Basiura, Anna Czapkiewicz: Simulation study of the use of entropy to validation of clustering.....	32
Andrzej Bąk: Problem of choosing the optimal linear ordering procedure in the p_llord package.....	41
Justyna Brzezińska-Grabowska: Latent class analysis in survey research...	50
Grażyna Dehnel: Tax register and social security register as a source of additional information for business statistics – possibilities and limitations.....	59
Sabina Denkowska: Selected methods of assessing the quality of matching in Propensity Score Matching	74
Marta Dziechciarz-Duda, Klaudia Przybysz: Applying the fuzzy set theory to identify the non-monetary factors of poverty.....	83
Iwona Foryś: The potential of the housing market in Poland in the years of economic recessions.....	92
Eugeniusz Gatnar: Statistical analysis of the convergence of CEE countries after 10 years of their membership in the European Union.....	99
Ewa Genge: Trust to the public and financial institutions in the Polish society – an application of latent Markov models.....	107
Alicja Grześkowiak: Multivariate analysis of the determinants of Poles' involvement in non-formal lifelong learning	116
Monika Hamerska: The use of the methods of linear ordering for the creating of scientific units ranking.....	125
Bartłomiej Jefmański: The application of IRT models in the construction of a fuzzy system of weights for variables in the issue of linear ordering – on the basis of TOPSIS method	134
Tomasz Józefowski, Marcin Szymkowiak: GDM as a method of finding a linear ordering of districts of Podkarpackie Voivodeship in the light of the operation of the Euro-Park Mielec special economic zone	143
Krzysztof Kompa: Application of parametric and nonparametric tests to the evaluation of the situation on the world financial market in the pre- and post-crisis period.....	153
Mariusz Kubus: Recursive feature elimination in discrimination methods ...	162
Marta Kuc: The impact of the spatial weights matrix on the final shape of the European Union countries ranking due to the standard of living.....	170
Paweł Lula: The impact of context on semantic similarity.....	181
Iwona Markowicz: Feldstein-Horioka regression model – the results for Poland.....	190

Kamila Migdal-Najman: The assessment of impact value of Minkowski's constant for the possibility of group structure identification in high dimensional data.....	199
Małgorzata Misztal: On the use of canonical correspondence analysis in economic research.....	208
Krzysztof Najman: The application of the parallel computing in cluster analysis.....	217
Edward Nowak: Data classification and accounting. A study of correlations	226
Marcin Pelka: The adaptation of bagging with the application of conceptual clustering of symbolic data.....	235
Józef Pociecha, Mateusz Baryła, Barbara Pawelek: Comparison of classification accuracy of selected bankruptcy prediction methods in the case of random and non-random sampling technique.....	244
Agnieszka Przedborska, Małgorzata Misztal: Selected multivariate statistical analysis methods in the evaluation of the quality of life of the members of the University of the Third Age.....	253
Wojciech Roszka: Construction of synthetic data sets for small area estimation.....	261
Aneta Rybicka: Combining revealed and stated preference data.....	270
Elżbieta Sobczak: Specialization in sectors of technical advancement vs. effects of workforce number changes in Poland's voivodships.....	279
Andrzej Sokółowski, Grzegorz Harańczyk: Modification of radar plot.....	286
Marcin Szymkowiak, Marek Witkowski: Classification of cooperative banks according to their financial situation using the median.....	295
Justyna Wilk, Michał B. Pietrzak, Roger S. Bivand, Tomasz Kossowski: The influence of classification method selection on the identification of spatial dependence – an application of join-count test.....	304
Dorota Witkowska: Application of classification trees to analyze wages disparities in Germany.....	314
Artur Zaborski: Asymmetric preference data analysis by using the dominance point model and the gravity model.....	323

Marta Kuc

Politechnika Gdańska

e-mail: marta.kuc@zie.pg.gda.pl

WPLYW SPOSOBU DEFINIOWANIA MACIERZY WAG PRZESTRZENNYCH NA WYNIK PORZĄDKOWANIA LINIOWEGO PAŃSTW UNII EUROPEJSKIEJ POD WZGLĘDEM POZIOMU ŻYCIA LUDNOŚCI

Streszczenie: W artykule dokonano analizy wpływu sposobu definiowania macierzy wag przestrzennych na wynik porządkowania liniowego na przykładzie rankingu państw Unii Europejskiej ze względu na poziom życia. Ustalenie struktury zależności przestrzennych jest istotnym zagadnieniem badawczym, gdyż umożliwia uwzględnienie w analizach niemierzalnych efektów przestrzennych, a ponadto pozwala uniknąć błędów poznawczych i podnosi wiarygodność analiz. Za aproksymację poziomu życia przyjęto przestrzenny taksonomiczny miernik rozwoju [Antczak 2013]. W opracowaniu przyjęto kilka wariantów macierzy wag przestrzennych, m.in.: macierz kontyngencji rzędu I i II, macierz odległości geograficznej oraz macierz odległości ekonomicznej. Wyniki przeprowadzonej analizy wskazują, że ostateczny kształt rankingu jest zależny od sposobu konstrukcji macierzy wag przestrzennych.

Słowa kluczowe: macierz wag przestrzennych, przestrzenny taksonomiczny miernik rozwoju, porządkowanie liniowe, poziom życia.

DOI: 10.15611/pn.2015.384.17

1. Wstęp

Konstrukcję przestrzennego taksonomicznego miernika rozwoju zaproponowała Antczak [2013], jako próbę uwzględnienia zależności przestrzennych w formule taksonomicznego miernika rozwoju. Postępowanie takie wydaje się zasadne, gdyż:

- zgodnie z I prawem geografii Toblera: „wszystko jest powiązane ze sobą, ale bliższe obiekty są bardziej zależne od siebie niż odległe” [Tobler 1970],
- wykorzystując dane regionalne odnoszące się do obiektów geograficznych, należy przynajmniej rozważyć możliwość, że obserwacje mogą nie być nieza-

leżne w wyniku wzajemnych połączeń pomiędzy sąsiadującymi obiektami [Buccellato 2007],

- państwa, regiony i województwa podlegają wpływom innych sąsiadujących obiektów przestrzennych oraz zależą od zachodzących w nich procesów ekonomicznych, społecznych i politycznych [Paelinck, Klaassen 1983],
- lepiej wykorzystać najprostszą macierz sąsiedztwa geograficznego, niż z góry zakładać niezależność [Griffith 1996].

Celem niniejszej pracy jest ocena wrażliwości przestrzennego taksonomicznego miernika rozwoju na sposób definiowania macierzy wag przestrzennych. Aby sprawdzić, jak zmieniają się wyniki porządkowania liniowego państw ze względu na poziom życia ludności, zastosowano kilka macierzy wag przestrzennych: macierz bezpośredniego sąsiedztwa, macierz sąsiedztwa II rzędu, macierze odległości geograficznej oraz ekonomicznej.

Badaniem objęto 24 państwa członkowskie Unii Europejskiej (z wyjątkiem Cypru, Malty i Luksemburga) w roku 2012. Materiał statystyczny został zaczerpnięty z bazy Passport GMID Euromonitoru (<http://www.euromonitor.com/passport>).

2. Przestrzenny taksonomiczny miernik rozwoju

Idea przestrzennego taksonomicznego miernika rozwoju polega na modyfikacji taksonomicznego miernika rozwoju Hellwiga [1968] przez dodanie do tegoż miernika macierzy wag przestrzennych [Antczak 2013]. Modyfikacja następuje już na etapie normalizacji. W przypadku gdy zmienne wykazujące charakter przestrzenny przekształca się zgodnie z formułą:

$$z_{ij}^* = \frac{x_{ij}^* - \bar{x}_j^*}{s_j^*},$$

gdzie: x_{ij}^* – macierz zmiennych diagnostycznych wykazujących charakter przestrzenny, zaś \mathbf{W} oznacza przyjętą macierz wag przestrzennych.

z_{ij}^* – znormalizowana macierz zmiennych diagnostycznych wykazujących charakter przestrzenny.

W niniejszym badaniu za zmienną o charakterze przestrzennym uważa się taką, dla której wartość statystyki Morana I jest statystycznie istotna przy poziomie istotności $\alpha = 0,05$. Zmienne niewykazujące charakteru przestrzennego są normalizowane w sposób klasyczny, nieuwzględniający macierzy wag przestrzennych [Hellwig 1968]. Po znormalizowaniu zmiennych można przejść do wyznaczenia odległości obiektów od wzorca [Antczak 2013]:

$$d_{spi}^* = \begin{cases} \sqrt{\sum_{j=1}^m (z_{ij}^* - \varphi_j^*)^2}, & \text{gdyn } j\text{-ta zmienna ma charakter przestrzenny} \\ \sqrt{\sum_{j=1}^m (z_{ij} - \varphi_j)^2}, & \text{w przeciwnym przypadku} \end{cases},$$

gdzie: z_{ij} – znormalizowana macierz zmiennych diagnostycznych niewykazujących charakteru przestrzennego.

Ostateczny przestrzenny miernik rozwoju powstaje jako [Antczak 2013]:

$$\mu_{spi}^* = 1 - \frac{d_{spi}^*}{d_{spi-}^*},$$

gdzie: d_{spi-}^* – graniczna odległość danego i -tego obiektu od wzorca, liczona jako:

$$d_{spi-}^* = \bar{d}^* - 2s_d^*, \text{ gdzie: } \bar{d}^* \text{ – średnia arytmetyczna wektora } d_{sp}^*, s_d^* \text{ – odchylenie standardowe wektora } d_{sp}^*.$$

Podobnie jak w przypadku miary zaproponowanej przez Hellwiga przestrzenny taksonomiczny miernik rozwoju zazwyczaj przyjmuje wartości z przedziału $\langle 0,1 \rangle$. Zatem im wyższa wartość miernika, tym lepiej z punktu widzenia analizowanego zjawiska.

Idea uwzględnienia zależności przestrzennych w konstrukcji zmiennej syntetycznej wydaje się zasadna, należy jednak zbadać, czy i jak silnie wyniki otrzymanej analizy są zależne od sposobu definiowania macierzy wag przestrzennych.

3. Poziom życia ludności

W niniejszym opracowaniu wpływ sposobu definiowania macierzy wag przestrzennych na rezultaty porządkowania liniowego obiektów zostanie przedstawiony na przykładzie rankingu państw ze względu na poziom życia ludności. Poziom życia rozumiany jest jako „stopień zaspokojenia potrzeb ludności wynikający z konsumpcji wytworzonych przez człowieka dóbr materialnych i usług oraz wykorzystania walorów środowiska naturalnego i społecznego [Bywalec, Wydimus 1992]. Przyjęcie tak szerokiej definicji poziomu życia wymaga uwzględniania czynników o charakterze zarówno pieniężnym, jak i niepieniężnym. Za aproksymację poziomu życia ludności przyjęto przestrzenny taksonomiczny miernik rozwoju, który poprzez agregację wskaźników cząstkowych będzie w stanie uwzględnić złożoność analizowanego zjawiska, a poprzez uwzględnienie macierzy wag przestrzennych pozwoli na uwzględnienie interakcji pomiędzy obiektami.

Analiza literatury związanej z poziomem życia pozwoliła przygotować zbiór 110 potencjalnych cech diagnostycznych. Ze zbioru zmiennych diagnostycznych

usunięto te cechy diagnostyczne, które nie spełniały warunków formalno-statystycznych [Zeliaś 2004]. Ostatecznie, przestrzenny taksonomiczny miernik rozwoju powstał na podstawie 40 zmiennych podzielonych na 10 kategorii:

1. Populacja – współczynnik płodności, współczynnik zależności demograficznej osób starszych.

2. Rynek pracy i bezpieczeństwo pracy – stopa bezrobocia zarejestrowanego, stopa zatrudnienia, liczba wypadków śmiertelnych w miejscu pracy na 100 tys. osób.

3. Ochrona zdrowia i opieka społeczna – liczba lekarzy przypadających na 100 tys. osób, liczba pielęgniarek przypadających na 100 tys. osób, liczba łóżek szpitalnych przypadających na 100 tys. osób, liczba zgonów na gruźlicę na 100 tys. osób, liczba zgonów na cukrzycę na 100 tys. osób, liczba zgonów noworodków na 1000 urodzeń żywych, liczba nowo zarejestrowanych przypadków AIDS na 100 tys. osób, odsetek ludzi otyłych (BMI>30) w populacji 18+.

4. Edukacja – liczba studentów na 1000 osób, liczba nauczycieli akademickich przypadających na 1 studenta.

5. Kultura, rekreacja i czas wolny – liczba hoteli na 1000 osób, średnioroczna liczba wyjść do kina na osobę.

6. Warunki mieszkaniowe – liczba nowo wybudowanych mieszkań przypadających na 1000 gospodarstw domowych.

7. Komunikacja i łączność – liczba nowo zarejestrowanych samochodów osobowych na 1000 osób, transport lotniczy w pasażerokilometrach na osobę, transport kolejowy w pasażerokilometrach na osobę; gęstość sieci drogowej w km na km² powierzchni ogólnej kraju, odsetek dróg o twardej powierzchni, odsetek autostrad w ogólnej długości dróg, abonenci telefonii komórkowej na 1000 osób, odsetek osób korzystających z Internetu.

8. Bezpieczeństwo społeczne – współczynnik postrzeganej korupcji, liczba zabójstw na 100 tys. osób, liczba przestępstw związanych z narkotykami na 100 tys. osób, liczba samobójstw na 100 tys. osób, liczba rozwodów na 1000 osób.

9. Dochody i wydatki ludności – oszczędności jako procent dochodu rozporządzalnego, podatki i składki na ubezpieczenie społeczne jako procent dochodu brutto, stopa inflacji, stawka za godzinę pracy w przemyśle w euro (ceny stałe z 2012 r.).

10. Środowisko naturalne – emisja pyłów w mikrogramach na km² powierzchni ogólnej kraju, procent powierzchni chronionej kraju, emisja tlenu węgla w kg na osobę, lesistość kraju, dzienna emisja zanieczyszczeń wód w kg na 1000 osób.

4. Analiza empiryczna

W celu sprawdzenia wrażliwości przestrzennego taksonomicznego miernika rozwoju na sposób definiowania macierzy wag przestrzennych wyznaczono syntetyczne mierniki poziomu życia na podstawie pięciu różnych macierzy wag przestrzennych:

- Wariant I – bezpośrednie sąsiedztwo (kryterium wspólnej granicy).
- Wariant II – sąsiedztwo II rzędu.
- Wariant III – odległość pomiędzy stolicami.
- Wariant IV – długość wspólnej granicy.
- Wariant V – odległość ekonomiczna (PKB *per capita* według parytetu siły nabywczej).

W literaturze można znaleźć wiele rozmaitych sposobów definiowania macierzy wag przestrzennych, a ich przegląd znajduje się m.in. w pracy Getisa i Aldstadta [2004]. Liczba i rodzaj zmiennych wykazujących charakter przestrzenny różniły się w zależności od zastosowanego wariantu macierzy wag (tab. 1). Niezależnie od wybranej macierzy wag przestrzennych charakter przestrzenny wykazywały m.in. następujące zmienne: stopa bezrobocia, liczba zgonów na gruźlicę na 100 tys. osób, średnioroczna liczba wyjść do kina na osobę, emisja CO w kg na osobę, oszczędności jako procent dochodów rozporządzalnych ludności. Analizując zawartość tab. 1, można zauważyć, że przyjmując macierz odległości ekonomicznej (wariant V), otrzymano znacznie niższą niż w pozostałych przypadkach liczbę zmiennych, dla których statystyka I Morana była statystycznie istotna.

Tabela 1. Liczba zmiennych o charakterze przestrzennym i nieprzestrzennym w zależności od wariantu macierzy wag przestrzennych

	Wariant I	Wariant II	Wariant III	Wariant IV	Wariant V
Zmienne wykazujące charakter przestrzenny	22	23	23	21	14
Zmienne niewykazujące charakteru przestrzennego	19	18	18	20	27

Źródło: opracowanie własne w programie STATA.

Wyniki porządkowania liniowego państw członkowskich Unii Europejskiej zostały zaprezentowane w tab. 2. W czołówce państw ze względu na poziom życia ludności znalazły się Wielka Brytania, Irlandia, Szwecja oraz Holandia. Najniższy poziom życia został natomiast odnotowany w Bułgarii, Rumunii, na Litwie oraz Łotwie. Otrzymane wyniki zdają się być merytorycznie i logicznie poprawne. W celu sprawdzenia, czy istnieje zgodność uporządkowań, dokonano analizy z wykorzystaniem współczynnika tau-Kendalla (tab. 3). Współczynnik tau-Kendalla dla każdej z analizowanych par przyjmuje stosunkowo wysokie wartości oraz jest statystycznie istotny. Analizując zawartość tab. 2, zauważyć można występowanie znacznych rozbieżności w porządkowaniu liniowym obiektów, szczególnie w przypadku wariantu III, opartego na odległości geograficznej pomiędzy stolicami poszczególnych państw. W wyniku zastosowania wariantu III najwyższy poziom życia został przypisany Francji, która w przypadku innych uporządkowań zajmowała miejsce 6 (wariant IV i V) bądź nawet 11 (wariant I). Zastosowanie macierzy wag opartej na odległości geograficznej spowodowało, że Irlandia znalazła się na

Tabela 2. Rankingi państw ze względu na poziom życia ludności w roku 2012

Kod ISO	Wariant I	Wariant II	Wariant III	Wariant IV	Wariant V
AT	12	7	2	7	8
BE	5	11	6	8	7
BG	24	23	21	24	23
DK	7	8	5	11	10
EE	23	17	23	19	11
FI	1	6	14	3	2
FR	11	3	1	6	6
GR	18	19	19	18	15
ES	13	10	12	12	13
NL	6	5	4	5	4
IE	4	1	17	2	5
LT	20	20	16	23	19
LV	21	21	20	20	22
DE	8	12	9	10	14
PL	15	16	15	17	17
PT	16	13	10	15	20
CZ	14	14	13	13	12
RO	19	24	24	22	21
SK	9	15	11	14	9
SI	17	18	18	16	16
SE	2	4	8	4	3
HU	22	22	22	21	24
UK	3	2	3	1	1
IT	10	9	7	9	18

Źródło: opracowanie własne.

Tabela 3. Ocena podobieństw uporządkowania z wykorzystaniem współczynnika tau-Kendalla

	Wariant I	Wariant II	Wariant III	Wariant IV	Wariant V
Wariant I	1,000	0,696	0,536	0,775	0,688
Wariant II		1,000	0,652	0,848	0,674
Wariant III			1,000	0,587	0,456
Wariant IV				1,000	0,710
Wariant V					1,000

Źródło: opracowanie własne.

miejscu 17, podczas gdy w innych rankingach zajmowała pozycje 1-5. W uporządkowaniu stworzonym według wariantu III otrzymano znacznie niższą pozycję Szwecji oraz Finlandii (odpowiednio o 4 i 8 pozycji) w stosunku do pozostałych rankingów. Wydaje się zatem, że taka macierz wag przestrzennych prowadzi do zaniżenia wartości miernika syntetycznego w obiektach najdalej wysuniętych. Dość duże rozbieżności w pozycjach poszczególnych państw w stosunku do innych rankingów odnotowano również w wyniku zastosowania wariantu V. W wyniku zastosowania macierzy wag opartej na odległości ekonomicznej Estonia została

uplasowana na 11 miejscu, podczas gdy dla wariantu I i III zajmowała pozycję 23 lub 19 dla wariantu IV. Korzystniej wypadła również Słowacja, która zajęła 9 lokatę (również w wariacie I), podczas gdy w wyniku zastosowania macierzy sąsiedztwa II rzędu uplasowała się na 15 pozycji. Spadek pozycji w wyniku zastosowania wariantu V zanotowano natomiast we Włoszech (9 pozycji w dół w stosunku do wariantu II i IV) oraz Niemiec (6 pozycji w dół w stosunku do wariantu I).

Przeprowadzona analiza wykazała występowanie zależności przestrzennych dla większości zmiennych uwzględnionych w badaniu. Jest to zatem uzasadnienie dla stosowania taksonomicznego miernika rozwoju w porządkowaniu i grupowaniu państw pod względem poziomu życia ludności. Ważnym problemem badawczym jest dobór odpowiedniej macierzy wag przestrzennych, gdyż jak wynika z niniejszego badania ma ona wpływ na ostateczny ranking obiektów. Najniższą zgodność wyników z pozostałymi uporządkowaniami zaobserwowano dla wariantu III (odległość od stolic). Taki sposób definiowania macierzy wag przestrzennych przyczynia się do osłabiania pozycji obiektów wysuniętych. Wysoką zgodność uporządkowań otrzymano w wyniku zastosowania klasycznie pojmowanego sąsiedztwa (istnienie wspólnych granic), sąsiedztwa II rzędu oraz długości wspólnej granicy. Zgodnie ze wspomnianym już wcześniej I prawem geografii Toblera obiekty bliższe sobie silniej na siebie oddziałują, również im większa powierzchnia styczności (długość wspólnej granicy), tym zachodzące interakcje przestrzenne są silniejsze. Wydaje się zatem, że należy preferować macierze wag przestrzennych oparte bądź na długości wspólnej granicy bądź na najbliższym sąsiedztwie.

5. Zakończenie

Wyniki przeprowadzonej analizy wskazują, że przestrzenny taksonomiczny miernik rozwoju jest wrażliwy na sposób definiowania macierzy wag przestrzennych. Pomimo faktu, że współczynnik korelacji tau-Kendalla przyjmował stosunkowo wysokie wartości, różnice w pozycjach poszczególnych państw sięgały nawet 16 pozycji (Irlandia 1 miejsce w wariacie II, 17 miejsce w wariacie III).

Najniższą zgodnością cechowały się wyniki uzyskane w rezultacie wykorzystania macierzy opartej na odległości pomiędzy stolicami państw. Tak skonstruowana macierz zaniżała wartości zmiennej syntetycznej dla skrajnych obiektów.

Florax i Rey [1995] sugerują, że w ekonometrii przestrzennej powinno się preferować mniej skomplikowane macierze wag przestrzennych, gdyż korzystniejsze jest stosowanie macierzy niedoszacowanej niż przeszacowanej. Wydaje się, że i w przypadku taksonomicznego przestrzennego miernika rozwoju takie rozwiązanie będzie najkorzystniejsze.

Literatura

- Antczak E., 2013, *Przestrzenny taksonomiczny miernik rozwoju*, Wiadomości Statystyczne, 2013/7, s. 37-53.
- Buccellato T., 2007, *Convergence across Russian Regions: A spatial econometrics approach*, Centre for the Study of Economic and Social Change in Europe, Economics Working Paper nr 72, Londyn, s. 1-29.
- Bywalec Cz., Wydymus S., 1992, *Poziom życia ludności Polski w porównaniu z krajami Europejskiej Wspólnoty Gospodarczej*, Ekonomista, 5-6/1992, s. 669-687.
- Florax R.J.G.M., Rey S., 1995, *The impact of misspecified spatial interaction in linear regression models*, New Directions in Spatial Econometrics. Advances in Spatial Sciences, Springer, s. 111-135.
- Getis A., Aldstadt J., 2004, *Constructing the spatial weight matrix using a local statistics*, Geographical Analysis 36, s. 90-104.
- Griffith D.A., 1996, *Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data*, The Canadian Geographer, vol. 40/1996, s. 351-367.
- Hellwig Z., 1968, *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*”, Przegląd Statystyczny nr 4, s. 307-326.
- Paelinck J.H.P., Klaassen L.H., 1983, *Ekonometria przestrzenna*, PWN, Warszawa, s. 14.
- Tobler W.R., 1970, *A computer movie simulating urban growth in the Detroit region*, Economic Geography, nr 46, Clark University, Clark, s. 234-240.
- Zeliaś A. (red.), 2004, *Poziom życia w Polsce i krajach Unii Europejskiej*, Wydawnictwo Naukowe PWE, Warszawa 2004, s. 53.

THE IMPACT OF THE SPATIAL WEIGHTS MATRIX ON THE FINAL SHAPE OF THE EUROPEAN UNION COUNTRIES RANKING DUE TO THE STANDARD OF LIVING

Summary: This article analyzes the impact of the spatial weights matrix on the result of the European Union countries ranking due to the standard of living. The proper structure of spatial relationship is an important research issue, since it allows to extend the standard of living analysis of the immeasurable spatial effects. Moreover, it avoids cognitive errors and increases the reliability of the analyzes. The taxonomy spatial measure of development constructed by E. Antczak was used as an approximation of the standard of living. In this paper, several variants of the spatial weights matrix were used, for example: spatial contiguity matrix, shared-boundary weight matrix, geographical distance matrix and economic distance matrix. The results of the analysis indicate that the final shape of the ranking is dependent on the method of constructing the spatial weights matrix.

Keywords: spatial weights matrix, taxonomy spatial measure of development, linear ordering, standard of living.