

Marek Sobolewski

Politechnika Rzeszowska
e-mail: msobolew@prz.edu.pl

PORZĄDKOWANIE LINIOWE Z UWZGLĘDNIENIEM RELACJI PRZESTRZENNYCH

LINEAR ORDERING WITH REGARD TO SPATIAL RELATIONS

DOI: 10.15611/pn.2018.508.21

JEL Classification: C38, O18

Streszczenie: W ostatnich latach zaproponowano pewne modyfikacje metod porządkowania liniowego jednostek terytorialnych, mające na celu uwzględnienie występujących pomiędzy nimi relacji przestrzennych. Nowość polega na wprowadzeniu etapu wygładzania przestrzennego wartości zmiennych diagnostycznych. W większości opublikowanych do tej pory prac przyjęto zasadę, by korektę na relacje przestrzenne stosować tylko dla cech diagnostycznych wykazujących istotną statystycznie autokorelację przestrzenną. Jednak w takiej sytuacji zastosowanie wygładzania przestrzennego w niewielkim stopniu wpływa na wartości tej cechy, a więc i wyniki rankingu. Z drugiej strony można podać przykłady cech diagnostycznych, które nie wykazują autokorelacji przestrzennej, a z powodów merytorycznych powinny być korektom przestrzennym poddane. Wnioski są jednoznaczne – konieczność stosowania procedury wygładzania przestrzennego zmiennych diagnostycznych nie może być utożsamiana z występowaniem autokorelacji przestrzennej.

Słowa kluczowe: porządkowanie liniowe, autokorelacje przestrzenne, analizy regionalne.

Summary: In the recent years some modifications have been proposed to linear ordering methods for territorial units to take into account spatial relationships between them. The innovation involves introducing spatial smoothing of diagnostic variables. In most published works the principle has been adopted that the correction for spatial relationships should only be applied to diagnostic features showing statistically significant spatial autocorrelation. However, in this situation, the use of spatial smoothing has little effect on the value of this feature, and thus the results of the ranking. On the other hand, there are some examples of diagnostic features that do not show spatial autocorrelation, and for substantive reasons should be subjected to spatial corrections. The conclusions are clear – the need to use the spatial smoothing procedure of diagnostic variables cannot be equated with spatial autocorrelation.

Keywords: linear ordering, spatial autocorrelation, regional analysis.

1. Wstęp

Tworzenie rankingów jednostek terytorialnych jest jednym z najpopularniejszych obszarów zastosowania metod taksonomicznych, zwłaszcza zaś procedur porządkowania liniowego. Szczególnie cenny jest fakt, iż wiele rankingów jednostek terytorialnych jest publikowanych systematycznie¹.

W tym kontekście nieco zaskakujący jest fakt, iż metodologia zdecydowanej większości tak licznie publikowanych rankingów nie uwzględnia specyfiki powiązań przestrzennych analizowanych jednostek terytorialnych. Nawet na poziomie tak dużych jednostek jak państwa relacje przestrzenne są istotne dla aktualnego ich funkcjonowania i dla perspektyw rozwoju. Wszak poziom życia mieszkańców w takich obszarach jak np. bezpieczeństwo wewnętrzne zależy bezpośrednio od sąsiedztwa geograficznego.

Jednostki terytorialne oddziałują na siebie wzajemnie, ponieważ występuje pomiędzy nimi przepływ ludzi i towarów. Jest to zjawisko pożądane, gdy zwiększa się w ten sposób możliwości podjęcia pracy, edukacji czy korzystania z innych usług i dóbr materialnych. Jednak położenie przestrzenne może determinować też przepływy niepożądane, związane chociażby z przestępczością albo zanieczyszczeniem powietrza. Z powyższych rozważań wynika przesłanie, iż wartości niektórych zmiennych diagnostycznych dla jednostek terytorialnych powinny zostać poddane wygładzaniu przestrzennemu. Warto też nadmienić, że im bardziej szczegółowy podział administracyjny, tym większe znaczenie relacji przestrzennych i większa potrzeba wyrównania przestrzennego danych przed przystąpieniem do rzetelnych analiz.

Pomysł uwzględnienia relacji przestrzennych w procedurach tworzenia rankingów jednostek terytorialnych przedstawiono niemal jednocześnie w trzech pracach [Antczak 2013; Pietrzak 2014; Sobolewski i in. 2014]. Wspólną cechą tych prac jest propozycja wprowadzenia do algorytmu porządkowania liniowego korekty wartości wybranych zmiennych diagnostycznych względem relacji przestrzennych². Natomiast pewne różnice przedstawionych algorytmów dotyczą szczegółów obliczeniowych. Interesującego porównania tych metod dokonano w pracy [Kuc 2017]. Korektę wartości zmiennych na relacje przestrzenne można (i trzeba) przeprowadzać nie tylko w analizach taksonomicznych [Markowska, Sobolewski 2014].

Jako że mamy do czynienia z nowym podejściem do popularnej procedury porządkowania jednostek terytorialnych, które zapewne będzie zastępowało lub przynajmniej współwystępowało z podejściem klasycznym, ważne jest, by na eta-

¹ Są to liczne i powszechnie znane wskaźniki publikowane corocznie, takie jak HDI, GII – na poziomie międzynarodowym czy na przykład wskaźniki atrakcyjności inwestycyjnej, poziomu życia, wykorzystania środków unijnych (i wiele innych) – publikowane w przekroju województw, powiatów czy gmin.

² Dla procedury tej proponujemy określenie „wygładzanie przestrzenne”, które będzie używane w dalszej części artykułu.

pie formułowania zasad wygładzania przestrzennego nie pojawiły się pewne błędy metodologiczne. Głównym celem poniższej pracy jest wskazanie, iż sformułowana w pracach [Antczak 2013, s. 42-43; Pietrzak 2014, s. 188] i powielana w innych opracowaniach idea, że wygładzaniu przestrzennemu powinny być poddane tylko zmienne wykazujące autokorelację przestrzenną, jest niepotrzebna. Ponadto tworzenie statystycznych kryteriów doboru zmiennych podlegających wygładzaniu przestrzennemu jest tu chyba niepotrzebną formalizacją.

2. Formalizacja procedury wygładzania przestrzennego

Formalizacja procedury wygładzania przestrzennego może wyglądać następująco. Definiujemy macierz podobieństwa jednostek terytorialnych $[w_{ij}]$ (może to być na przykład sąsiedztwo 1-go lub wyższego stopnia, długość wspólnej granicy, odwrotność odległości pomiędzy obiektami), a następnie dokonujemy normalizacji macierzy \mathbf{W} wierszami do jedności. Wartości zmiennej diagnostycznej z dla obiektu o numerze i korygujemy, stosując poniższą formułę wygładzania przestrzennego:

$$z_i^* = \alpha \cdot z_i + (1 - \alpha) \cdot \sum_{i \neq j} w_{ij} \cdot z_j. \quad (1)$$

Wagę α określamy arbitralnie, przy czym dla każdej zmiennej macierz \mathbf{W} i parametr wygładzania α mogą być inne. Wartość $\alpha = 1$ oznacza brak wygładzania przestrzennego, wartość $\alpha = 0$ to uwzględnienie tylko sąsiednich wartości, bez wartości dla „wygładzanego” obiektu.

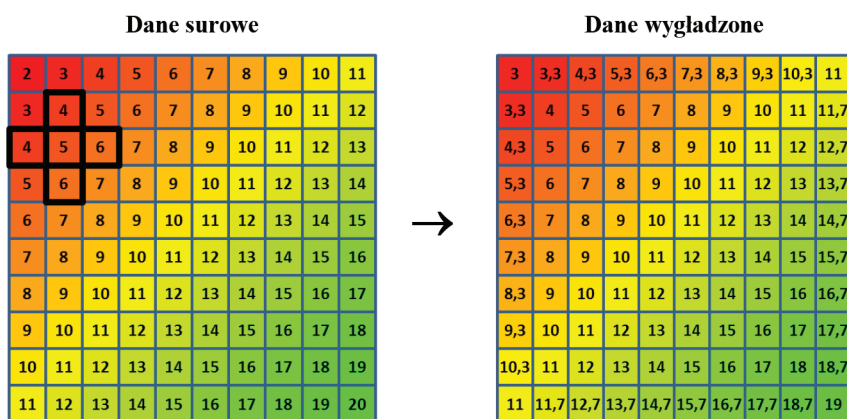
W procedurze wygładzania przestrzennego jest wiele elementów, które mają charakter arbitralny – wybór macierzy podobieństwa \mathbf{W} , wybór stałej wygładzania α . Taką koncepcję wygładzania przestrzennego przedstawiono w pracy [Sobolewski i in. 2014]. Procedury opisane w dwóch pozostałych wzmiankowanych wyżej pracach [Antczak 2013; Pietrzak 2014] są bardzo podobne.

3. Autokorelacja przestrzenna a wygładzanie przestrzenne

Jak już wcześniej wspomniano, w literaturze pojawiła się sugestia – niestety bardzo wyraźnie sformułowana – by wygładzania przestrzennego dokonywać tylko dla zmiennych wykazujących autokorelacje przestrzenne. W kolejnych trzech (pod punktach poniższej pracy zaprezentowano argumenty przeciwko tej koncepcji, znacząco ograniczającej zakres i użyteczność metody porządkowania liniowego z korektą przestrzenną.

3.1. Co daje wygładzanie przestrzenne zmiennych wykazujących wysoką autokorelację przestrzenną?

Poniższy przykład oparty jest na danych umownych – rozważamy 100 obiektów z sąsiedztwem „poziomym i pionowym”, które tworzą „mapę” pokazaną na rys. 1. Dane te wykazują bardzo wysoką autokorelację przestrzenną (współczynnik Morana $I = 0,95$), więc przyjmując kontestowane kryterium wygładzania przestrzennego, cecha ta powinna być tej procedurze bezwzględnie poddana.



Rys. 1. Dane przestrzenne charakteryzujące się bardzo silną autokorelacją przestrzenną ($I = 0,95$) – przed i po wygładzeniu przestrzennym

Źródło: opracowanie własne.

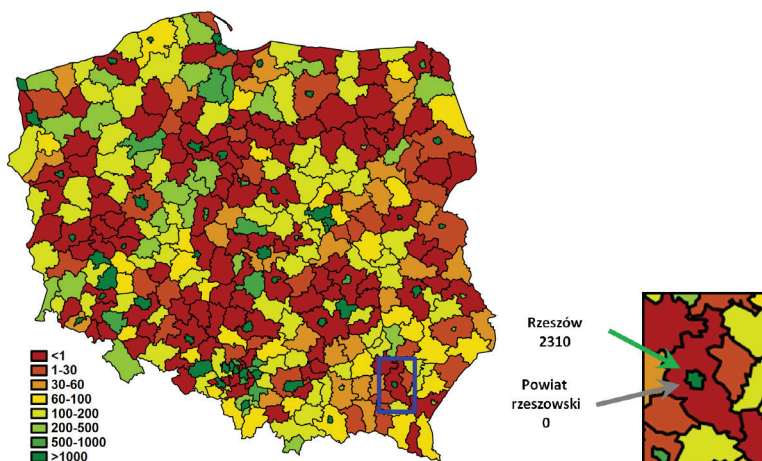
Po zastosowaniu do danych z rys. 1 procedury wygładzania przestrzennego według wzoru (1), nawet z maksymalną siłą wygładzania (a więc dla $\alpha = 0$, co oznacza, że wartość cechy dla każdego obiektu jest zastępowana średnią z obiektów z nim sąsiadujących), uzyskamy układ niemal identyczny z wyjściowym. Jedyne dla części obiektów skrajnych, którym „brakuje” jednego lub dwóch sąsiadów, wartości wygładzone będą się minimalnie różniły od wartości pierwotnych. Oczywiście przedstawiony przykład jest bardzo wyidealizowany, w przypadku innych danych, wykazujących istotne statystycznie korelacje przestrzenne, lecz o mniejszej sile, efekt wygładzania może być nieco większy. Jednakże zawsze pozostaje pytanie, po co poddawać dodatkowej procedurze wygładzania przestrzennego dane będące już przestrzennie wygładzone? Jeżeli już za konieczne uznamy istnienie obiektywnego kryterium doboru zmiennych poddawanych wygładzaniu przestrzennemu, to wtedy należałoby raczej wziąć pod uwagę przypadki ujemnej autokorelacji przestrzennej. Jak jednak pokazano w kolejnym (pod)punkcie, tak przeformułowane kryterium nie jest zadowalające, bowiem także (a może przede wszystkim) zmienne nie wykazujące autokorelacji przestrzennej powinny być poddane wygładzaniu przestrzennemu.

3.2. Wygładzanie przestrzenne w przypadku braku autokorelacji

W poprzednim (pod)punkcie wskazano, iż dla cech o dodatniej autokorelacji przestrzennej wygładzanie przestrzenne oznaczać będzie jedynie niewielkie korekty wartości wyjściowych. Natomiast w przypadku braku autokorelacji przestrzennej (lub autokorelacji ujemnej) wygładzanie będzie znacząco zmieniało wyjściowe wartości. Co więcej, jak wynika z dalszych rozważań, łatwo znaleźć przykłady takich cech, dla których operacja wygładzania przestrzennego jest uzasadniona merytorycznie.

W rankingach jakości życia wykorzystuje się wskaźniki dotyczące dostępności dóbr kulturalnych³ (kin, teatrów, muzeów), które w naturalny sposób koncentrują się w większych ośrodkach. Wobec dużej mobilności współczesnego społeczeństwa ta koncentracja nie jest problemem dla mieszkańców mniejszych ośrodków – szczególnie tych położonych blisko wielkich miast.

Poszukując przykładu, który spełniałby wskazane powyżej kryteria, zwrócono uwagę na wskaźnik liczby seansów w kinach na 10 tys. mieszkańców w przekroju powiatów (rys. 2). Cecha ta nie wykazuje autokorelacji przestrzennej w sposób niemal wzorcowy, ponieważ współczynnik Morana $I = 0,00$. A więc w myśl kontestowanego w tej pracy kryterium nie powinna być wygładzana przestrzennie.



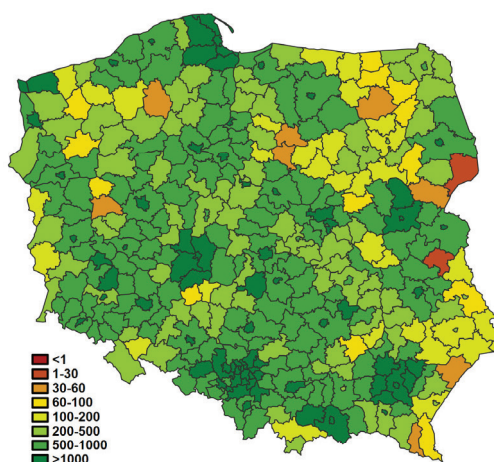
Rys. 2. Rozkład przestrzenny wskaźnika liczby seansów w kinach na 10 tys. mieszkańców – cecha nie wykazuje autokorelacji ($I = 0,00$), ale powinna być poddana wygładzaniu przestrzennemu

Źródło: opracowanie własne na podstawie danych GUS.

³ Jeżeli kogoś nie przekonuje ciężar gatunkowy zaproponowanego przykładu, może zamiast wskaźników dostępności dóbr kulturalnych rozważyć wskaźniki związane z opieką zdrowotną – szczególnie tą wysokospecjalistyczną, która także w oczywisty sposób koncentruje się w dużych ośrodkach.

Rozważmy dwa przykładowe powiaty wyróżnione na rys. 2. Czy rzeczywiście mieszkańcy powiatu rzeszowskiego mają nieskończenie wiele razy gorszy dostęp do kin niż mieszkańcy Rzeszowa? Wszak w całym powiecie rzeszowskim nie ma ani jednego czynnego kina! Tyle że czas dojazdu z terenu powiatu do Rzeszowa nie przekracza 30 minut i nie jest to znaczące utrudnienie w dobie powszechnej motoryzacji. Wygładzenie przestrzenne tych danych jest konieczne, bo oryginalny rozkład wartości, choć jest prawdziwy, nie pokazuje rzeczywistego obrazu zjawiska! Zresztą danych tego typu, choć – podkreślmy to raz jeszcze: prawdziwych – nie powinno się poddawać żadnym analizom, dopóki nie uwzględni się ich przestrzennej specyfiki i nie wygładzi przestrzennie.

Jak zmieni się rozkład wartości wskaźnika liczby seansów po przeprowadzeniu wygładzania przestrzennego? Odpowiedź na to pytanie znaleźć można na kartogramie zamieszczonym na rys. 3. Dokładniejszego omówienia wymaga sposób przeprowadzenia wygładzania przestrzennego w rozważanym przypadku. Automatyczne zastosowanie wzoru (1) miałyby się z celem z uwagi na specyficzny charakter analizowanego wskaźnika. Podstawą obliczeń będzie macierz odległości pomiędzy powiatami⁴. Przyjęte zostanie założenie, iż wzajemne oddziaływania dotyczyć będą powiatów odległych nie więcej niż o 50 km. Dla określonego powiatu „przeszukiwane” będą więc powiaty odległe odeń nie więcej niż o 50 km, z których wybierany będzie ten o największym wskaźniku liczby seansów kinowych (najbardziej atrakcyjny cel wyjazdu). Następnie wyliczana będzie średnia ważona wartości dla



Rys. 3. Rozkład przestrzenny wskaźnika liczby seansów w kinach na 10 tys. mieszkańców po przeprowadzeniu wygładzania przestrzennego

Źródło: opracowanie własne na podstawie danych GUS.

⁴ Odległości między powiatami były wyznaczone jako odległości ich środków ciężkości.

rozważanego powiatu i maksimum z otoczenia. Dodać należy, że udział wartości z otoczenia będzie malał wraz ze wzrostem odległości pomiędzy rozważanym powiatem a powiatem realizującym maksimum wskaźnika w jego otoczeniu.

Jak widać, po przeprowadzaniu wygładzania przestrzennego nastąpiły duże zmiany w poziomie i rozkładzie wskaźnika dostępności seansów kinowych (por. rys. 2 i 3). W zmodyfikowanej wersji „zyskały” przede wszystkim powiaty stanowiące sąsiedztwo miast na prawach powiatów. Tak zmodyfikowany wskaźnik pokazuje zupełnie inną hierarchię powiatów, na pewno bliższą rzeczywistym odczuciom mieszkańców niż wartości oryginalne pokazane na rys. 2. Po przeprowadzeniu wygładzania przestrzennego wskaźnik dostępności do kin dla żadnego powiatu nie wynosi zero, co jest chyba zgodne ze stanem faktycznym. Tymczasem wśród danych wyjściowych takich powiatów było aż 118, co stanowiło niemal jedną trzecią całej zbiorowości.

3.3. Dlaczego kryterium $I \neq 0$ nie może być poprawne – argumenty dedukcyjne

Istnieją jeszcze inne argumenty za nieprzyjmowaniem kryterium istnienia autokorelacji przestrzennej jako podstawy wykonywania wygładzania przestrzennego.

Po pierwsze, wartość współczynnika autokorelacji I jest liczona „globalnie”, czyli na podstawie całej zbiorowości jednostek terytorialnych, natomiast procedura wygładzania przestrzennego zaproponowana w pracach dotyczących porządkowania liniowego działła „lokalnie” – czyli uzależnia wartość tylko od jednostek znajdujących się w pobliżu (to oczywiście zależy od postaci macierzy \mathbf{W}). Na tle całej zbiorowości przypadki wymagające wygładzenia mogą zostać nie zauważone.

Drugie spostrzeżenie jest następujące – otóż istotność współczynnika korelacji zależy od wielkości rozważanego zbioru danych. Współczynnik $I = 0,20$ może być nieistotny dla – przykładowo – powiatów woj. podkarpackiego, a istotny dla wszystkich powiatów w Polsce. Tymczasem relacje danego powiatu z woj. podkarpackiego z jego sąsiadami się nie zmieniają – niezależnie od tego, czy rozpatrujemy tylko jedno województwo, czy cały kraj⁵.

4. Propozycja rozszerzenia pojęcia wygładzania przestrzennego

Niniejszy punkt zawiera próbę formalizacji rozważań prowadzonych w (pod)punkcie 3.2. Wnikliwy czytelnik mógłby podważyć słuszność przeprowadzenia wygładzania przestrzennego dla danych prezentowanych na rys. 2 za pomocą formuły (1). Dla wskaźnika dostępności dóbr kulturalnych, obiekty „lepsze” oddziałują pozytywnie na „gorsze”, ale nie na odwrót. Mieszkańcy powiatu rzeszowskiego (przykładowo)

⁵ Gwoli ścisłości, relacje przestrzenne zmieniają się dla powiatów leżących na granicy województwa, ale dla większości pozostaną takie same, niezależnie od zakresu przestrzennego analizy.

mają całkiem dobry dostęp do kin, mieszkając w niedużej odległości od Rzeszowa, a więc skorygowany wskaźnik dla powiatu rzeszowskiego powinien być znacząco wyższy od zera. Jednak zastosowanie średniej ważonej macierzą podobieństwa dla wszystkich powiatów spowodowałoby obniżenie wartości wskaźnika dla Rzeszowa, a przecież rzeszowianin nie traci możliwości korzystania z usług kinowych z powodu przyjazdu (ograniczonej liczby) mieszkańców ościennych powiatów.

Tak więc formuła średniej ważonej (1) nie zawsze będzie odzwierciedlać złożone relacje przestrzenne. Dla danych dotyczących dostępności usług kulturalnych lepiej sprawdzi się wzięcie maksimum po powiatach z pewnego otoczenia, ewentualnie ważonego odległością i uśrednienie z wartością oryginalną. Dla omawianego w (pod)punkcie 3.2 wskaźnika wygładzanie przestrzenne opierać się zatem będzie na następującej formule:

$$z_i^* = (1 - \alpha) \cdot z_i + \alpha \cdot \max_{d_{ij} \leq 50} z_j. \quad (2)$$

W powyższym wzorze d_{ij} oznacza odległość (wyrażoną w kilometrach) pomiędzy powiatami, a wartość współczynnika wygładzania α zależy od odległości rozważanego powiatu od powiatu posiadającego optymalną wartość wygładzanej zmiennej w otoczeniu o promieniu 50 km. Zakładamy tu, że oddziaływanie jest jednokierunkowe – „lepsze” powiaty wpływają na „gorsze”, ale nie na odwrót. Zakres wahań parametru α trzeba ustalić arbitralnie. W procedurze zastosowanej w części 3.2 wahał się on od 0,25 do 0,50, w zależności od odległości pomiędzy rozważanym powiatem a powiatem stanowiącym dlań „lokalny” wzorzec.

Na zakończenie tych rozważań warto wspomnieć, iż poza relacjami przestrzennymi w procedurze wygładzania przestrzennego warto wprowadzić pewien dodatkowy czynnik ważący, albowiem wzajemne oddziaływanie jednostek terytorialnych zależy w oczywisty sposób od ich wielkości (na przykład liczby mieszkańców). Nieporównywalna jest siła oddziaływania liczącej 1,7 mln mieszkańców Warszawy na każdy z powiatów ościennych, które liczą kilkanaście razy mniej mieszkańców, i oddziaływania w przeciwnym kierunku.

5. Zakończenie

W artykule przedstawiono argumenty za tym, że stosowanie procedury wygładzania przestrzennego zmiennej diagnostycznej jako etapu porządkowania liniowego nie powinno być uzależnione od występowania autokorelacji przestrzennej.

W przypadku dodatniej autokorelacji przestrzennej wygładzanie przestrzenne w niewielkim stopniu zmienia wartości danej cechy, więc takie ograniczenie stosowania tej procedury doprowadziłoby do wniosku, iż jest ona zbyteczna. Z kolei z prostego przykładu zamieszczonego w (pod)punkcie 3.2 wynika, że brak korelacji przestrzennej nie musi oznaczać braku oddziaływania pomiędzy sąsiadującymi ze sobą obszarami.

Wydaje się, że nie ma sensu na siłę poszukiwać statystycznych („obiektywnych”) kryteriów selekcjonowania zmiennych podlegających wygładzaniu przestrzennemu, gdyż, jak wykazano w niniejszej pracy, takie kryteria mogą mocno deformować wyniki. Niestety, analiza merytoryczna jest trudniejsza i nie ma uroku „automatyzmu”, trudno jednak akceptować reguły błędne tylko dlatego, że są wygodniejsze.

Warto też promować elastyczne podejście do procedury wygładzania danych przestrzennych. Nie zawsze formuła średniej ważonej (do tego sprowadza się wyznaczenie oryginalnych danych przez elementy macierzy \mathbf{W}) prowadzi do rozsądnych wyników. Ponadto do tej pory w pracach, w których zamieszczano wyniki rankingów uwzględniających relacje przestrzenne podczas wygładzania przestrzennego, nie brano pod uwagę potencjału ludnościowego, a wydaje się, że warto dokonać takiej modyfikacji prowadzonych wyliczeń.

Literatura

- Antczak E., 2013, *Przestrenny taksonomiczny miernik rozwoju*, Wiadomości Statystyczne, nr 7, s. 37-53.
- Kuc M., 2016, *Przemiany w przestrzennym zróżnicowaniu poziomu życia ludności w państwach Unii Europejskiej*, Wydawnictwo UMK w Toruniu, Toruń.
- Kuc M., 2017, *The Taxonomy Spatial Measure of Development in the Standard of Living Analysis*, Folia Oeconomica, nr 1 (327), s. 167-185.
- Markowska M., Sobolewski M., 2014, *Regionalna ruchoma średnia przestrzenna*, XXXIII konferencja naukowa: Multivariate Statistical Analysis: Uniwersytet Łódzki, Łódź: 17-19 listopada 2014.
- Pietrzak M.B., 2014, *Taksonomiczny miernik rozwoju (TMR) z uwzględnieniem zależności przestrzennych*, Przegląd Statystyczny, z. 2, s. 181-201.
- Sobolewski M., Migala-Warchoł A., Mentel G., 2014, *Ranking poziomu życia w powiatach w latach 2003-2012 z uwzględnieniem korelacji przestrzennych*, Acta Universitatis Lodziensis, Folia Oeconomica, nr 6 (308), s. 147-159.