Rozwój Regionalny i Polityka Regionalna 50: 159–174
https://doi.org/10.14746/rrpr.2020.50.09
2020

Tomasz M. Kossowski

Uniwersytet im. Adama Mickiewicza w Poznaniu
Wydział Geografii Społeczno-Ekonomicznej i Gospodarki Przestrzennej
Zakład Ekonometrii Przestrzennej
tkoss@amu.edu.pl

Wybrane zagadnienia modelowania matematyczno-statystycznego struktur i procesów przestrzennych

Zarys treści: Niniejsze opracowanie jest przeglądem wybranych zagadnień z zakresu modelowania matematyczno-statystycznego struktur i procesów przestrzennych. Ogólny charakter tego artykułu obejmuje dyskusję nad podstawowymi pojęciami, takimi jak: układ przestrzenny, struktura przestrzenna, proces przestrzenny, i ich wzajemnymi relacjami. Następnie definiowany jest w sposób ogólny (stochastyczny) proces przestrzenny i jego składniki, ze szczególnym uwzględnieniem reprezentacji struktury przestrzennej. Artykuł omawia sposób budowy modelu stochastycznego procesu przestrzennego, analizując jednocześnie najważniejsze problemy pojawiające się na etapie jego specyfikacji, estymacji i weryfikacji. Uwypuklono również wkład poznańskich geografów w rozwiązywanie problemów teoretycznych związanych z modelowaniem struktur i procesów przestrzennych.

Słowa kluczowe: model przestrzenny, układ przestrzenny, struktura przestrzenna, proces przestrzenny, estymacja.

Wprowadzenie

Modelowanie matematyczno-statystyczne struktur i procesów przestrzennych ma długą tradycję, sięgającą połowy ubiegłego stulecia. Pierwsze prace Morana (1950) oraz Geary'ego (1954) rozwiązywały problem pomiaru autokorelacji przestrzennej w danych zlokalizowanych geograficznie oraz wprowadzały modelowanie struktury przestrzennej interakcji za pomocą macierzy sąsiedztwa. Whittle (1954, 1963) zaproponował autoregresyjny model dla interakcji przestrzennej, który był następnie rozwijany do bardziej złożonych form. W późniejszych opracowaniach Cliffa i Orda (1970, 1972, 1973, 1981) analizowano modelowanie struktury przestrzennej poprzez macierze wag przestrzennych oraz metody pomiaru autokorelacji przestrzennej w danych zarówno ilościowych, jak i jakościowych. W pracach tych przedstawiono też metody wnioskowania statystycznego dotyczącego autokorelacji przestrzennej. Anselin (1988) w książce „Spatial Econometrics: Methods and Models" zebrał podstawy ekonometrycznego modelowania procesów przestrzennych wraz z metodami ich estymacji i testowania, a także walidacji modeli przestrzennych.

Wzrost mocy obliczeniowych komputerów, zwiększenie możliwości oprogramowania matematyczno-statystycznego, upowszechnienie się systemów informacji geograficznej wraz z pojawieniem się baz danych przestrzennych stały się nowym impulsem dla rozwoju modelowania struktur i procesów przestrzennych. W latach 90. XX w. rozwijano metody badań lokalnych procesów przestrzennych, wprowadzając lokalne miary autokorelacji przestrzennej (Anselin 1995) oraz modelowania procesu heterogeniczności przestrzennej (Fotheringham i in. 2002). Pod koniec lat 90. niektórzy badacze zwrócili się w kierunku poszukiwania nowych metod estymacji modeli procesu przestrzennego, alternatywnych do dotychczas wykorzystywanej metody największej wiarogodności. Przede wszystkim należy zwrócić tutaj uwagę na prace Kelejiana i Pruchy (1998, 1999, 2002), w których określono estymatory uogólnionej metody momentów oraz podwójnej metody najmniejszych kwadratów. W ostatnich latach XX w. i pierwszej dekadzie XXI w. próbowano również rozwiązać niektóre z problemów związanych z estymatorami metody największej wiarogodności (m.in. Pace, Barry 1997, 1999, Smirnow, Anselin 2001, 2009, Pace, LeSage 2004, Kossowski, Hauke 2010, Bivand i in. 2013).

Jednocześnie rozwijano metody badania lokalnej zależności przestrzennej dla danych jakościowych. Jedną z nich jest metoda Bootsa (2003, 2006) pozwalająca na wyznaczanie lokalnych klastrów dla zmiennych binarnych określonych na regularnej siatce kwadratów. Uogólnienie metody Bootsa na *k*-kolorowe mapy z nieregularnym podziałem przestrzeni przedstawili Bivand i in. (2017). Natomiast Anselin i Li (2019) zoperacjonalizowali lokalne statystyki *join-count*. Na gruncie polskiej geografii, globalne statystyki *join-count* zastosował Ratajczak (1980) w pracy dotyczącej modelowania wpływu czynników społeczno-ekonomicznych na kształtowanie się sieci transportowych.

Druga dekada XXI w. przyniosła także szereg nowych metod pozwalających modelować strukturę przestrzenną. Odrzucono przy tym wcześniejsze założenie, że macierz wag przestrzennych jest znana i z góry ustalona. Podjęto wiele prób estymacji nieznanej lub endogenicznej macierzy wag (m.in. Bhattacharjee i Jensen-Butler 2013, Kelejian, Piras 2014, Mur i in. 2019).

W niniejszej pracy omawiane są wybrane zagadnienia dotyczące modelowania matematyczno-statystycznego struktur i procesów przestrzennych. W pierwszej kolejności definiowane są pojęcia: układu przestrzennego, struktury przestrzennej i procesu przestrzennego. Przedstawione są także wzajemne zależności pomiędzy tymi pojęciami. Takie wprowadzenie jest konieczne, gdyż wiele prac geograficznych posługuje się definiowanymi tutaj pojęciami w sposób intuicyjny. W kolejnym punkcie pracy omawiane są poszczególne elementy procesu przestrzennego, przy czym szczególny nacisk jest położony na zagadnienie modelowania struktury przestrzennej i pojawiające się tutaj związane z tym trudności.

Następnie definiowana jest ogólna postać formalna modelu procesu przestrzennego, a po niej analizowane są, najczęściej stosowane w badaniach, modele przestrzennej regresji liniowej. Przedmiotem szerszej dyskusji jest metoda estymacji parametrów takich modeli, ponieważ od wyboru metody oszacowania zależy końcowa jakość modelu, a następnie zagadnienie jego weryfikacji. Całość opracowania zamyka podsumowanie.

Proces przestrzenny a układ i struktura przestrzenna

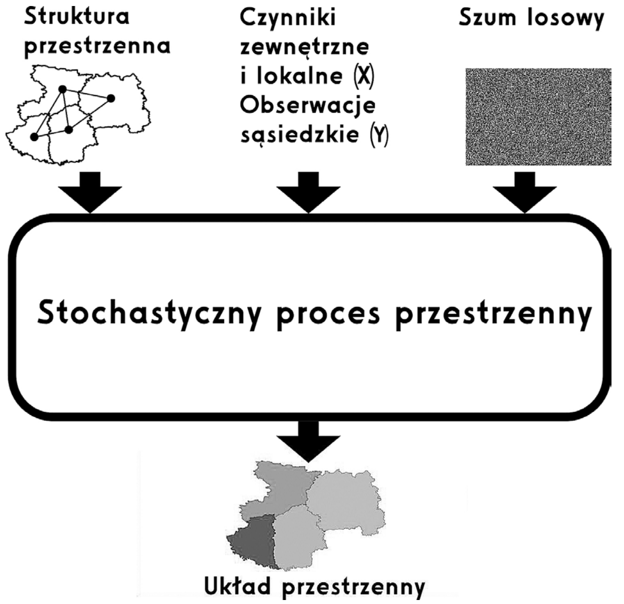
Obserwacja różnych zjawisk w przestrzeni geograficznej prowadzi niejednokrotnie do wniosku, że przebiegają one w różnym natężeniu lub sile w zależności od położenia w danym punkcie lub regionie. Przykładem jest różnicowanie się stopy bezrobocia w układzie powiatowym w Polsce. Układ przestrzenny, czyli rozmieszczenie wartości określonych cech w przestrzeni, jest na ogół wynikiem działania pewnego procesu przestrzennego. Zakłada się, że realizacje procesu przestrzennego są obserwowalne w przestrzeni i mierzalne. Pojęcia układu przestrzennego, struktury przestrzennej i procesu przestrzennego muszą więc być ściśle ze sobą powiązane.

Proces przestrzenny jest nierozerwalnie związany ze strukturą przestrzenną (Tiefelsdorf 2000, s. 1). Jak zauważa Haining (1990, s. 24), struktura przestrzenna powstaje w wyniku działania tych relacji przestrzennych, które wpływają na sposób zachowania się procesu przestrzennego. Relacją przestrzenną nazywamy podzbiór $R \subseteq S \times S$, gdzie S jest zbiorem obiektów w przestrzeni geograficznej, $S = \{s_i, i = 1, ..., n\}$. Taką relacją może być np. sąsiedztwo: dwa dowolne obiekty i, j albo są sąsiadujące (wówczas $\neg\, s_i R s_j$), albo nie są sąsiadujące (wówczas $s_i R s_j$). Strukturę przestrzenną definiuje się jako zbiór relacji pomiędzy punktami (obiektami) w przestrzeni geograficznej.

Procesem przestrzennym nazywamy prawidłowości, których działanie prowadzi do powstania układu przestrzennego określonego zjawiska w przestrzeni geograficznej. Prawidłowości te mogą zostać opisane w przypadku procesu deterministycznego za pomocą modeli matematycznych lub, w przypadku procesu stochastycznego, przy użyciu modeli statystycznych. Proces deterministyczny charakteryzuje się tym, że badana cecha może przyjąć w danej lokalizacji tylko jedną określoną wartość, natomiast proces stochastyczny pozwala na to, aby cecha przyjęła jedną z możliwych wartości zgodnie z jej rozkładem prawdopodobieństwa. Modele procesów stochastycznych z reguły zawierają w sobie, obok składnika losowego, również składnik deterministyczny.

Zasadnicze podejście do wyjaśniania układu przestrzennego jest następujące. Jak wspomniano powyżej, układ przestrzenny jest wynikiem działania pewnego procesu przestrzennego, którego ogólna postać jest znana. Problem sprowadza się zatem do oszacowania wartości parametrów modelu procesu przestrzennego przy założeniu, że obserwowane wartości cechy są realizacją tego procesu.

Modelowanie procesu przestrzennego wymaga uwzględnienia następujących elementów (por. ryc. 1): 1) wyjaśnianego układu przestrzennego zmiennej Y, 2)



Ryc. 1. Schemat stochastycznego procesu przestrzennego
Źródło: opracowanie własne.

czynników działających zewnętrznie lub lokalnie (heterogeniczność przestrzenna), reprezentowanych przez zmienne X, 3) wpływu obserwacji sąsiedzkich na układ przestrzenny zmiennej Y (zależność przestrzenna) oraz 4) szumu losowego. W przypadku deterministycznego procesu przestrzennego nie bierze się pod uwagę tego ostatniego elementu, gdyż wyklucza się losowość. Każdy z tych elementów zostanie szczegółowo omówiony w kolejnej części artykułu.

Ważnymi własnościami procesów przestrzennych są izotropowość i stacjonarność. Pierwsza z tych własności oznacza, że dany proces przestrzenny działa w każdym kierunku przestrzeni tak samo. Stacjonarność z kolei wymaga, aby

$$f(Y_1, Y_2, ..., Y_k) = f(Y_{1+l}, Y_{2+l}, ..., Y_{k+l}),$$

tzn. łączny rozkład zmiennych $Y_1, Y_2, ..., Y_k$ był taki sam jak rozkład zmiennych $Y_{1+l}, Y_{2+l}, ..., Y_{k+l}$ przesuniętych o l, przy czym to przesunięcie ma miejsce w przestrzeni. W praktyce oznacza to, że proces jest stacjonarny, gdy (Suchecki i in. 2010): 1) wartość oczekiwana zmiennych Y_i jest taka sama, 2) wariancja wszystkich zmiennych Y_i jest stała oraz 3) kowariancja zależy tylko od odległości między lokalizacjami. Niestacjonarność procesu występuje w powiązaniu ze zjawiskiem heterogeniczności przestrzennej. Związki pomiędzy stacjonarnością i niestacjonarnością oraz zależnością i heterogenicznością przestrzenną wraz z metodami ich wspólnego modelowania dyskutuje Fotheringham (2009).

Elementy procesu przestrzennego

Reprezentacja układu przestrzennego

Układ przestrzenny zjawiska może być, jak wcześniej wspomniano, przedstawiony na mapie. Jednakże dla celów modelowania matematyczno-statystycznego układ przestrzenny powinien zostać sformalizowany w następujący sposób. Niech S będzie zbiorem obiektów przestrzennych. Zbiór S może być zbiorem nieskończonym (punktów w przestrzeni geograficznej) albo skończonym (regiony geograficzne). Zbiór zmiennych losowych $\{Y(s): \in S, Y(s) \in \mathbf{R}\}$ nazywamy polem losowym, a wektor $\mathbf{y} = [Y(s): s \in S, Y(s) \in \mathbf{R}]$ nazywamy wektorem losowym. W najczęstszym przypadku, gdy zmienne losowe $Y(s)$ mają rozkład normalny, wówczas wektor $\mathbf{y}$ ma wielowymiarowy rozkład normalny o następujących własnościach. Wartość oczekiwana jest wektorem wartości oczekiwanych zmiennych $Y(s)$, tzn. $E(\mathbf{y}) = [\mathrm{E}(Y(s_1)), \mathrm{E}(Y(s_2)), \ldots, \mathrm{E}(Y(s_n)))$, natomiast macierz kowariancji jest określona

$$\Sigma = \begin{bmatrix} cov(Y(s_1), Y(s_1)) & cov(Y(s_1), Y(s_2)) & \cdots & cov(Y(s_1), Y(s_n)) \\ \cdots & \cdots & \cdots & \cdots \\ cov(Y(s_n), Y(s_1)) & cov(Y(s_n), Y(s_2)) & \cdots & cov(Y(s_n), Y(s_n)) \end{bmatrix}.$$

W ogólności zakłada się, że zmienne losowe $Y(s)$ są wzajemnie zależne. W przypadku, gdyby zmienne były parami niezależne, wówczas macierz $\sum$ miałaby elementy niezerowe tylko na diagonali. Realizację tego wektora losowego $\mathbf{y}$ nazywa się układem przestrzennym. Tak zdefiniowany wektor losowy wymaga oszacowania $\frac{n(n+1)}{2} + n$ parametrów (macierz kowariancji $\sum$ jest symetryczna). W praktyce, ze względu na fakt, że przeważnie dysponuje się jedną realizacją wektora losowego, przyjmuje się, iż wartości tego wektora są realizacją pewnej zmiennej losowej Y.

Opis struktury przestrzennej

Najprostszym sposobem reprezentacji struktury przestrzennej jest macierz bezpośredniego sąsiedztwa **C**: każdy obiekt przestrzenny jest reprezentowany przez wiersz i kolumnę macierzy. Jeżeli s_iRs_j, to wówczas $c_{ij} = 1$, natomiast jeżeli $\neg s_iRs_j$, to wtedy $c_{ij} = 0$. Przyjmuje się, że $\neg s_iRs_j$, a zatem $c_{ii} = 0$. Tak zdefiniowana macierz wag zawiera informację o sąsiedztwie, ale nie uwzględnia faktu, że oddziaływania pomiędzy jednostkami przestrzennymi poprzez strukturę przestrzenną mogą mieć różną siłę i nie muszą być symetryczne. Stąd też w praktyce dokonuje się takiego przekształcenia macierzy bezpośredniego sąsiedztwa, aby zawierała również te dodatkowe informacje. Taką macierz nazywa się macierzą wag przestrzennych.

Obecnie wyróżnia się dwa podejścia do wyznaczania macierzy wag przestrzennych: 1) macierz **W** może zostać zdefiniowana egzogenicznie względem modelu procesu przestrzennego; wartości elementów są ustalane według pewnych schematów, które zakładają określony sposób oddziaływań poprzez strukturę przestrzenną, 2) macierz **W** jest nieznana i jest estymowana na podstawie danych. Podejście pierwsze – egzogeniczne – ciągle jest najpopularniejsze.

Ogólny sposób definiowania egzogenicznej macierzy wag przestrzennych podał Tiefelsdorf (2000), formułując poniższe wyrażenie:

$$\mathbf{W}_q = \frac{n}{\sum_{i=1}^{n} d_i^{q+1}} \mathbf{D}^q\mathbf{C},$$

gdzie **C** jest zero-jedynkową macierzą sąsiedztwa, **D** jest macierzą diagonalną, której elementy na diagonali d_i należą do wektora $\mathbf{d} = \mathbf{C} \cdot \mathbf{1}_n$, przy czym $\mathbf{1}_n$ jest wektorem jedynek długości n. Z powyższej definicji można wyprowadzić różne macierze wag w zależności od parametru q. Na ogół jednak badacze ograniczają się do kilku podstawowych schematów oddziaływania:

$$\mathbf{W}_q = \begin{cases} \dfrac{n}{\sum_{i=1}^{n} d_i} \mathbf{C},\ q = 0, \\ \dfrac{n}{\sum_{i=1}^{n} \sqrt{d_i}} \mathbf{D}^{-1/2}\mathbf{C},\ q = -1/2, \\ \mathbf{D}^{-1}\mathbf{C},\ q = -1. \end{cases}$$

Macierz dla $q = 0$ nosi nazwę macierzy typu **C** (C-coding) i jest macierzą symetryczną. W przypadku, gdy $q = -1/2$, wówczas macierz spłaszcza wariancję wag przypisanych do poszczególnych par obiektów i nazywana jest macierzą typu **S** (S-coding). Dla $q = -1$ otrzymuje się tak zwaną standaryzowaną wierszami macierz typu **W** (W-coding). Szerszy przegląd zastosowań macierzy typu **W** pozwala na przypuszczenie, że około 90% badań zakłada przebieg oddziaływań przestrzennych w sposób określony przez ten typ macierzy. Jej własnością jest to, że jest ona niesymetryczna co do rozkładu wartości, ale zachowuje symetrię co do położenia elementów niezerowych i zerowych. Wadą tak zdefiniowanej macierzy wag przestrzennych jest uzależnienie wartości jej elementów od liczby sąsiadów, których posiada jednostka przestrzenna. W takiej sytuacji macierz może przeszacowywać siłę interakcji dla jednostek przestrzennych położonych na granicy badanego obszaru, które w naturalny sposób mają mniejszą liczbę sąsiadów niż jednostki położone wewnątrz.

Innym rodzajem często wykorzystywanych macierzy wag przestrzennych są te skonstruowane o kryterium k-najbliższych sąsiadów (z reguły $k = 5$ lub $k = 6$). Macierze te również zostają przekształcone w wierszowo standaryzowane macierze wag, które nie zachowują symetrii położenia elementów niezerowych ani zerowych. Jednakże mogą one zostać poddane procedurze symetryzacji[1] (Cliff, Ord 1981, s. 18) w przypadku, gdy taka symetria jest wymagana.

[1] Zgodnie z formułą $w'_{ij} = w'_{ji} = \frac{1}{2}(w_{ij} + w_{ji})$.

W przypadku macierzy opartych na kryterium odległości przyjmuje się, że dwie jednostki przestrzenne są swoimi sąsiadami, gdy są położone w odległości nie większej niż *d*. Konstruuje się macierz odległości **D**, a potem zastępuje odległości większe niż *d* zerami. Taką macierz przekształca się następnie w macierz wag przestrzennych, zastępując odległości ich odwrotnościami. Macierz taka może być później poddana standaryzacji wierszowej w taki sposób, jak macierz **W**-coding.

Krytyka macierzy egzogenicznych sprowadza się zasadniczo do podważenia założenia, że siła oddziaływań przestrzennych pomiędzy jednostkami przestrzennymi jest z góry znana i może być oszacowana za pomocą prostych operacji arytmetycznych na macierzy bezpośrednich połączeń lub odległości. Istnieją jednakże prace, które uzasadniają zastosowanie odgórnie zdefiniowanych macierzy wag przestrzennych (np. Hays i in. 2010). Co więcej, wyrażany jest również pogląd (LeSage, Pace 2014), że wybór różnych macierzy egzogenicznych nie powoduje istotnych zmian w oszacowaniach parametrów modelu przestrzennego, o ile różnice pomiędzy macierzami nie są zbyt wielkie. Z drugiej strony, powstają też prace wskazujące na wpływ niepoprawnej specyfikacji macierzy wag przestrzennych (np. Florax, Rey 1995, Lee, Yu 2012) na jakość modelu. Mimo tych krytycznych uwag, macierze egzogeniczne wciąż są szeroko stosowane w modelowaniu struktury przestrzennej.

Estymacja macierzy wag przestrzennych na podstawie danych doczekała się już kilku procedur. Do oszacowania wykorzystuje się topologię przestrzeni geograficznej oraz własności danych (Herrera i in. 2018). Wśród najstarszych podejść należy wspomnieć metodę maksymalizacji współczynnika autokorelacji przestrzennej *I* Morana (Kooijman 1976) oraz metodę Getisa i Aldstadta (2004) opartą na lokalnych statystykach przestrzennych. Folmer i Oud (2008) wykorzystują podejście oparte na zmiennych sztucznych, zbliżone do koncepcji wieloczynnikowego modelu zależności przestrzennej. Z kolei Bhattacharjee i Jensen-Butler (2013) proponują metodę estymacji symetrycznych macierzy wag przestrzennych dla modelu przestrzennego błędu losowego wraz z rozszerzeniem na inne modele przestrzenne oraz wykorzystują do analizy regionalnego rynku nieruchomości w Zjednoczonym Królestwie. Również Mur i in. (2019) zaprezentowali metodę estymacji nieznanej macierzy wag przestrzennych przy wykorzystaniu ogólnej metody momentów (GMM). Zakładają przy tym symetrię macierzy wag przestrzennych oraz autoregresyjny model procesu przestrzennego (SAR) lub przestrzenny proces średniej ruchomej. Poważną trudnością w tym podejściu jest jednakże konieczność rozwiązywania złożonych układów równań nieliniowych.

W ostatnim czasie podejmowane są również próby oszacowania endogenicznych macierzy wag przestrzennych. Specyficzną własnością tych macierzy jest to, że ich wartości zmieniają się długookresowo (ewoluują) endogenicznie wraz z modelem. Kelejian i Piras (2014) podali sposób oszacowania endogenicznej macierzy wag przestrzennych dla ogólnego panelowego modelu przestrzennego z opóźnioną zmienną zależną i zastosowali w praktyce dla modelu popytu na papierosy w Stanach Zjednoczonych. Qu i Lee (2015) oszacowali, wykorzystując dwuetapową metodę zmiennych instrumentalnych, metodę pseudonajwiększej wiarogodności oraz uogólnioną metodę momentów, endogeniczną macierz wag przestrzennych dla modelu przestrzennej zależności ze składnikiem regresyjnym (SAR). Złożoność procedur estymacji nieznanych lub endogenicznych macierzy wag przestrzennych nie pozwala jednakże, ze względu na objętość tego opracowania, na ich przedstawienie w tym miejscu.

Pozostałe składniki procesu przestrzennego

W modelowaniu procesu przestrzennego można brać pod uwagę czynniki oddziałujące na układ przestrzenny zarówno zewnętrznie, jak i lokalnie. Czynniki działające zewnętrznie mogą mieć charakter instytucjonalny lub wynikać z pewnych ogólnie działających praw (np. ekonomicznych, fizycznych). W modelu procesu zewnętrznego reprezentowane są przez zbiór zmiennych, których parametry są niezmiennicze ze względu na obserwacje.

Czynniki działające lokalnie (w każdej lokalizacji inaczej) to czynniki związane z heterogenicznością przestrzenną. Dotyczące pewnych, często niemierzalnych własności miejsc i powodują dodatkowe odchylenia układu przestrzennego od wartości wynikającej z procesu przestrzennego. Heterogeniczność przestrzenna skutkuje niestabilnością badanej relacji w przestrzeni geograficznej. W sensie formalnym może być modelowana jako heteroskedastyczny błąd losowy lub pod postacią zmiennych z parametrami zmieniającymi się w zależności od lokalizacji.

Szum losowy jest nieobserwowalny i musi być modelowany w postaci zmiennej losowej o pewnym ustalonym rozkładzie. Najczęściej jest to rozkład normalny $N(\mu, \sigma_i^2)$, w którym wartość oczekiwana wynosi $\mu = 0$. Przyjęcie takiego założenia pozwala na uwzględnienie heteroskedastyczności błędu losowego, tj. sytuacji, gdy jego wariancja zmienia się w zależności od miejsca obserwacji s_i. Na ogół jednak zakłada się, w celu uproszczenia modelu i eliminacji czynników działających heterogenicznie, że dla każdej lokalizacji s_i wariancja błędu losowego jest taka sama $\sigma_i^2 = \sigma^2$.

Przedstawione powyżej elementy procesu przestrzennego powinny zostać uwzględnione w jego matematyczno-statystycznej specyfikacji. W praktyce jednak zdarza się, że modeluje się uproszczone wersje procesów przestrzennych, co zostanie zaprezentowane w następnej części opracowania.

Model procesu przestrzennego

Wyjaśnianie układu przestrzennego pociąga za sobą konieczność budowy modelu procesu przestrzennego, generującego dany układ. W tym celu konieczne jest sformalizowanie modelu konceptualnego przedstawionego na rycinie 1 do postaci matematyczno-statystycznej. Wykorzystuje się wiele możliwych specyfikacji modeli przestrzennych, a próbę klasyfikacji tych modeli, stosowanych w analizie regionalnej, podjął na polskim gruncie Ratajczak (2008).

Specyfikacja modelu procesu przestrzennego

Niech punktem wyjścia będzie model postaci:

$$Y = f(Y, \mathbf{X}, \boldsymbol{\beta}, \mathbf{W}) + \varepsilon,$$

gdzie Y jest zmienną objaśnianą (reprezentującą układ przestrzenny), $\mathbf{X}$ są zmiennymi reprezentującymi czynniki zewnętrzne wpływające na układ przestrzenny lub zmiennymi opisującymi działanie czynników lokalnych (heterogeniczność), $\boldsymbol{\beta}$ są parametrami tych czynników, natomiast ε jest losowym szumem gaussowskim. Zakłada się, że $\mathbf{W}$ reprezentuje strukturę przestrzenną, opisując interakcje przestrzenne na trzy możliwe sposoby, jako: 1) autoregresyjny wpływ układu przestrzennego Y na samego siebie (wpływ Y z obserwacji sąsiedzkich), 2) wpływ zmiennych $\mathbf{X}$ z obserwacji w innych lokalizacjach na Y nazywany krzyżową regresją przestrzenną, 3) autokorelację błędu losowego. Tak określony model zawiera wszystkie elementy wymienione w poprzednich punktach pracy.

Jednym z najczęściej wykorzystywanych modeli dla procesu przestrzennego jest model przestrzennej regresji liniowej. Ogólna postać tego modelu jest dobrze znana (Anselin 1988, Kossowski 2010) i wygląda następująco:

$$\begin{gathered}\mathbf{y} = \varrho \mathbf{W}_1\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} = \lambda\, \mathbf{W}_2\boldsymbol{\varepsilon} + \mathbf{u} \\ \mathbf{u} \sim \mathcal{N}(0, \boldsymbol{\Omega}) \\ \Omega_{ii} = h_i(\mathbf{z}, \boldsymbol{\alpha}),\ h_i > 0\end{gathered}$$

gdzie $\mathbf{y}$ jest wektorem reprezentującym objaśniany układ przestrzenny, jest k-elementowym wektorem parametrów zmiennych egzogenicznych (reprezentujących czynniki zewnętrzne), $\mathbf{X}$ jest macierzą zmiennych egzogenicznych, a są współczynnikami autoregresyjnymi opóźnionych przestrzennie: zmiennej zależnej i błędu losowego. Macierze wag przestrzennych $\mathbf{W}_1$, $\mathbf{W}_2$ są stopnia n i reprezentują strukturę przestrzenną oddziaływań. Macierz kowariancji $\boldsymbol{\Omega}$ niezależnego przestrzennie błędu losowego ma na głównej przekątnej wartości pewnej nieujemnej funkcji h_i, która zależy od wektora zmiennych (sztucznych) oraz wektora parametrów o długości p. Taka konstrukcja macierzy kowariancji pozwala na uchwycenie działania czynników występujących lokalnie (heterogeniczność przestrzenna), która może ujawnić się pod postacią heteroskedastyczności błędu losowego. Jeżeli $\boldsymbol{\alpha} = \mathbf{0}$, to otrzymujemy sytuację błędu homoskedastycznego, tzn. $h_i = \sigma^2$.

W badaniach ekonomiczno-przestrzennych najczęściej wykorzystuje się jeden z kilku wariantów modelu typu Cliffa-Orda. Są to: 1) klasyczny model regresji liniowej $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$, $\rho = \lambda = 0$, $\boldsymbol{\alpha} = 0$, który ma $k+1$ nieznanych parametrów oraz $p+2$ ograniczenia, 2) model typu SAC (*spatial autoregressive combined*): $\mathbf{y} = \boldsymbol{\rho}\mathbf{W}_1\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\alpha} = 0$, gdzie $\boldsymbol{\varepsilon} = \boldsymbol{\lambda}\mathbf{W}_2\boldsymbol{\varepsilon} + \mathbf{u}$, $\mathbf{u} \sim \mathcal{N}(0, \boldsymbol{\Omega})$, $\Omega_{ii} = \sigma^2$. Model ten ma $k+3$ nieznanych parametrów i p ograniczeń, a macierze wag przestrzennych $\mathbf{W}_1$, $\mathbf{W}_2$ mogą być równe. Model SAC może zostać dalej uproszczony do jednego z dwóch modeli: a) model opóźnienia przestrzennego SAR: $\mathbf{y} = \boldsymbol{\rho}\mathbf{W}_1\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varrho} = 0$, $\boldsymbol{\alpha}$

= 0 lub b) modelu błędu przestrzennego SEM: $\mathbf{y} + \mathbf{X\beta} + \boldsymbol{\varepsilon}$, $\lambda = 0$, $\boldsymbol{\alpha} = 0$, gdzie $\boldsymbol{\varepsilon} = \lambda \mathbf{W}_2 \boldsymbol{\varepsilon} + \mathbf{u}$, $\mathbf{u} \sim \mathcal{N}(0, \boldsymbol{\Omega})$, $\Omega_{ii} = \sigma^2$. Obydwa modele mają $k+2$ nieznanych parametrów i ograniczeń.

Wybór właściwej specyfikacji modelu procesu przestrzennego jest kolejnym problemem, który musi zostać rozwiązany przez badacza w trakcie modelowania. Z reguły przyjmuje się jedną z dwóch strategii modelowania (Suchecki i in. 2010). Pierwsza z nich to tak zwana strategia klasyczna (Anselin, Florax 1995) „od szczególnego do ogólnego", w której punktem wyjścia jest budowa klasycznego modelu nie uwzględniającego efektów przestrzennych (zależności i heterogeniczności). Następnie, po serii testów, badacz wprowadza do modelu kolejne składniki odpowiedzialne za ich modelowanie. Strategia Hendry'ego (Florax i in. 2003) prowadzi postępowanie w odwrotnym kierunku „od ogólnego do szczególnego". Badacz formułuje na początku jak najszerszy model procesu przestrzennego, z którego następnie eliminuje zbędne składniki. Strategia ta wymaga od danych i zmiennych dodatkowego zbioru sześciu założeń (Florax i in. 2005).

Estymacja modelu przestrzennego

Estymacja modelu procesu przestrzennego polega na oszacowaniu wartości wektora jego parametrów Wektor ten ma $k+p+3$ nieznane parametry. W przypadku modelu klasycznej regresji liniowej, wektor parametrów ulega redukcji i wymaga oszacowania nieznanych parametrów. Estymacja modelu regresji liniowej odbywa się przy wykorzystaniu klasycznej metody najmniejszych kwadratów:

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

W przypadku modelu przedstawionego w punkcie 2) wraz z podpunktami a) i b), zastosowanie metody najmniejszych kwadratów do oszacowania wektora parametrów prowadzi do uzyskania wadliwych wyników. Oszacowania parametrów są albo obciążone i niezgodne w przypadku modelu a), albo nieefektywne, ale nieobciążone jak w przypadku b) (Anselin 1988).

W odniesieniu do modeli procesu przestrzennego typu SAC aktualnie istnieje kilka metod estymacji ich parametrów. Wyróżnić można następujące (choć nie wszystkie) podejścia do estymacji modeli procesu przestrzennego: 1) metoda największej wiarogodności (*maximum likelihood*) (Cliff, Ord 1981, Anselin 1988, Tiefelsdorf 2000, Suchecki i in. 2010), 2) podwójna metoda najmniejszych kwadratów (2SLS, GS2SLS) zaproponowana przez Kelejiana i Pruchę (1998, 2002) oraz Lee (2007), 3) uogólniona metoda momentów (GMM) wprowadzona również przez Kelejiana i Pruchę (1999, 2007), 4) metoda zmiennych instrumentalnych (IV) zasugerowana przez Haininga (1978) i Bivanda (1984), a sformalizowana przez Anselina (1988).

Spośród powyżej wymienionych najczęściej stosowana jest metoda największej wiarogodności. Wynika to z dominacji poglądu (Suchecki i in. 2010), że jest to metoda najlepsza. Jest ona jednocześnie metodą o dużej złożoności i wymaga zastosowania technik numerycznych.

Niech dany będzie model typu SAC z heteroskedastycznym błędem losowym:

$$\mathbf{y} = \boldsymbol{\rho}\mathbf{W}_1\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\ \boldsymbol{\alpha} = 0,\ \text{gdzie } \boldsymbol{\varepsilon} = \boldsymbol{\lambda}\mathbf{W}_2\boldsymbol{\varepsilon} + \mathbf{u},\ \mathbf{u} \sim \mathcal{N}\,(0, \boldsymbol{\Omega})\ \Omega_{ii} = \sigma^2.$$

Model ten można zapisać następująco:

$$\mathbf{Ay} = \mathbf{X}\boldsymbol{\beta} + \mathbf{B}^{-1}\mathbf{u},$$

gdzie $\mathbf{A} = \mathbf{I} - \varrho\mathbf{W}_1$, $\mathbf{B} = \mathbf{I} - \lambda\mathbf{W}_2$. Korzystając z faktu, że macierz jest diagonalna, można heteroskedastyczny błąd losowy przekształcić do sytuacji homoskedastycznej $v = \boldsymbol{\Omega}^{-1/2}\mathbf{u}$, $\mathbf{u} = \boldsymbol{\Omega}^{1/2}\mathbf{v}$. Stąd:

$$\mathbf{Ay} = \mathbf{X}\boldsymbol{\beta} + \mathbf{B}^{-1}\,\boldsymbol{\Omega}^{1/2}\mathbf{v},$$

a zatem:

$$\mathbf{v} = \boldsymbol{\Omega}^{-1/2}\,\mathbf{B}(\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta}).$$

Błąd losowy v jest nieobserwowalny, dlatego konieczne jest przekształcenie, które pozwoli wyznaczyć łączny rozkład Y na podstawie łącznego rozkładu $\mathbf{v}$. Jakobian takiego przekształcenia jest równy:

$$J = \det\,(\boldsymbol{\Omega}^{-1/2}\mathbf{BA}) = \det\,(\boldsymbol{\Omega}^{-1/2})\,\det\,(\mathbf{B})\,\det\,(\mathbf{A}).$$

Przy założeniu łącznego rozkładu normalnego $\mathbf{v}$ logarytm funkcji wiarogodności dla obserwacji $\mathbf{y}$ jest postaci:

$$L\,(\theta) = \left(-\frac{n}{2}\right)\ln(\pi) - \frac{1}{2}\ln|\boldsymbol{\Omega}| + \ln|\mathbf{B}| + \ln|\mathbf{A}| - \left(\frac{1}{2}\right)(\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta})'\mathbf{B}'\boldsymbol{\Omega}^{-1}\mathbf{B}(\mathbf{Ay} - \mathbf{X}\boldsymbol{\beta})$$

Następnie poszukuje się maksimum funkcji względem wektora parametrów. Szczególnie istotnym problemem jest obliczenie wartości logarytmów wyznaczników $\ln|\boldsymbol{\Omega}|$, $\ln|\mathbf{B}|$, $\ln|\mathbf{A}|$, gdzie $\mathbf{A} = \mathbf{I} - \varrho\mathbf{W}_1$, $\mathbf{B} = \mathbf{I} - \lambda\mathbf{W}_2$. Wynika to z faktu, że macierze reprezentujące strukturę przestrzenną interakcji $\mathbf{W}_1$, $\mathbf{W}_2$ są macierzami rzadkimi. Istnieje kilka rozwiązań tego problemu, których porównanie zostało przedstawione w pracy Bivanda i in. (2013). Autorzy porównali ze sobą numeryczną efektywność wyliczania tych wyznaczników w trzech grupach metod. Do pierwszej z nich należy zaliczyć metody wykorzystujące wartości własne macierzy, takie jak metoda Orda (1975) oraz metoda Griffitha i Sone (1995). Druga grupa metod obejmuje rozkłady macierzy na trójkątne oraz rozkład Cholesky'ego zastosowane do macierzy rzadkich (Pace, Barry 1997). Ostatnią porównywaną grupą metod są aproksymacje Czebyszewa (Pace, LeSage 2004), Monte Carlo (Barry, Pace 1999) oraz metoda Smirnova-Anselina (2009) bazująca na obliczaniu śladów macierzy.

Bivand i in. (2013) w konkluzji swoich badań stwierdzają, że w przypadku małego zbioru danych użyteczne są metody z pierwszej grupy, bez względu na to, czy macierz wag przestrzennych **W** jest symetryczna. Dla większych i rzadkich macierzy wag przestrzennych, ale zachowujących warunek symetrii, badacze mogą wybierać pomiędzy metodami Cholesky'ego, aproksymacjami Czebyszewa i Monte Carlo oraz metodą Smirnova-Anselina. W przypadku regularnych układów przestrzennych i zero-jedynkowych macierzy wag przestrzennych najwygodniejsza jest metoda Griffitha i Sone (1995), natomiast gdy wykorzystywana jest duża, rzadka i asymetryczna macierz wag, wtedy należy wybrać metodę rozkładu na macierze trójkątne lub aproksymację Monte Carlo.

Dalsza procedura estymacji metodą największej wiarogodności polega na wyliczeniu pochodnych cząstkowych funkcji $L(\theta)$ względem nieznanych parametrów z wektora θ i przyrównaniu ich do zera. Otrzymany w ten sposób układ równań nieliniowych jest rozwiązywany w procedurze iteracyjnej.

Podwójna metoda najmniejszych kwadratów (2SLS, GS2SLS) zaproponowana przez Kelejiana i Pruchę (1998, 2002) wymaga przyjęcia sześciu założeń dotyczących możliwych wartości parametrów ρ, λ, braku losowości macierzy wag przestrzennych, nieosobliwości macierzy **A** i **B**, braku losowości i ograniczoności zmiennych **X**, dodatkowych restrykcji na błędzie losowym **u** i wykorzystywanego zbioru instrumentów. Metoda ta składa się z trzech etapów: 1) wstępnego oszacowania parametrów $\boldsymbol{\alpha}$ za pomocą podwójnej metody najmniejszych kwadratów, 2) estymacji parametru λ poprzez metodę GMM, 3) końcowej estymacji całego modelu, przy czym w etapie 1) dopuszcza się zastosowanie metody zmiennych instrumentalnych. Podejście zaproponowane przez Kelejiana i Pruchę jest prostsze z punktu widzenia implementacji numerycznej, ale minusem jest konieczność przyjmowania kolejnych założeń.

Uogólniona metoda momentów (GMM) to kolejna metoda estymacji modeli procesu przestrzennego wprowadzona przez Kelejiana i Pruchę. W pracy z 1999 r. wyprowadzili oni estymator parametru w modelu przestrzennym typu SAR, który obliczeniowo jest prostszy do uzyskania niż estymator metody największej wiarogodności (ML). Co więcej, przy pewnym zbiorze dodatkowych założeń udowodnili, że estymator ten jest również zgodny. Rozszerzeniem wyników ich badań jest praca z 2010 r., w której wyznaczyli estymator GMM dla parametru w modelu typu SAC oraz ustalili estymatory metody zmiennych instrumentalnych (IV) dla parametrów regresyjnych $\boldsymbol{\beta}$. Walde i in. (2008) dokonali numerycznej efektywności metody największej wiarogodności i uogólnionej metody momentów, natomiast Lee (2007) dodatkowo włącza do analizy estymatory uzyskane z podwójnej metody najmniejszych kwadratów. Procedurę estymacji za pomocą uogólnionej metody momentów przystępnie przedstawiono w pracy Sucheckiego i in. (2010).

Metoda zmiennych instrumentalnych do estymacji parametrów modelu procesu przestrzennego została przedstawiona przez Anselina (1988). Autor w sposób ogólny określa metodę estymacji dla modelu typu SAR i SAC. Mimo względnej prostoty obliczeniowej metoda ta nie zdobyła szerszego uznania w badaniach geograficznych i regionalnych, dlatego też nie będzie w niniejszej pracy szerzej dyskutowana.

Wybór metody estymacji wektora parametrów modelu procesu przestrzennego nie jest zadaniem łatwym i zależy od doświadczenia badacza oraz możliwości, zwłaszcza numerycznych, przeprowadzenia oszacowania. Stąd też ciągle najpopularniejsza jest metoda największej wiarogodności, która została zaimplementowana w popularnych pakietach, takich jak R, czy specjalnym programie dedykowanym dla analiz przestrzennych GeoDa.

Weryfikacja modelu procesu przestrzennego

Weryfikacja i testowanie uzyskanych w wyniku estymacji modeli procesu przestrzennego składa się z kilku etapów i jest bardziej złożone niż w przypadku modeli klasycznych. Suchecki in. (2010) wyróżniają sześć grup problemów związanych z procedurą testowania modeli przestrzennych. Pierwsza z nich dotyczy badania stopnia dopasowania, a także testowania parametrów strukturalnych oraz całego modelu. W badaniu dopasowania zastosowanie ma współczynnik pseudo-R2, natomiast w przypadku istotności parametrów modelu należy posłużyć się testem t (bądź z w odniesieniu do unormowanych estymatorów metody ML). Normalność rozkładu reszt sprawdzana jest testem Jarque'a-Bery lub Kołmogorowa-Smirnowa. Do porównania modeli wykorzystuje się kryteria informacyjne Akaike'a AIC oraz logarytmu wiarogodności loglike.

Druga grupa testów dotyczy weryfikacji zależności przestrzennych. Wykorzystywane są w tym celu, zarówno ex ante, jak i ex post, testy mnożników Lagrange'a *LM* w wersji klasycznej i w wersji odpornej. Uzupełnieniem mogą być testy ilorazu wiarogodności *LR* oraz test Walda. Porównanie wartości testów *LM*, *LR* i Walda pozwala wnioskować o poprawności specyfikacji modelu. Trzeci testowany problem dotyczy heteroskedastyczności składnika losowego, która wskazuje na występowanie procesu heterogeniczności przestrzennej. Używa się tutaj testów White w przypadku, gdy postać heteroskedastyczności jest nieznana, albo Breuscha-Pagana dla pewnej zakładanej funkcji heteroskedastyczności.

Jednoczesne występowanie heterogeniczności i autokorelacji przestrzennej może być testowane na dwa sposoby (Suchecki i in. 2010): a) testowanie autokorelacji przy wystąpieniu heteroskedastyczności lub 2) testowanie heteroskedastyczności przy występowaniu autokorelacji przestrzennej. W przypadku pierwszym używa się zmodyfikowanych testów mnożnika Lagrange'a, natomiast w drugim stosujemy *Joint LM test* lub przestrzenną wersję testu Breuscha-Pagana. Ostatni testowany problem to badanie niestabilności parametrów strukturalnych modelu w przypadku występowania zależności przestrzennej. Niestabilność ta sprawdzana jest za pomocą testów Chowa.

Poprawna weryfikacja oszacowanego modelu procesu przestrzennego wymaga użycia wielu narzędzi sprawdzających. Trzeba zwrócić uwagę na fakt, że testy te mają ograniczone zastosowanie. Niektóre z nich mogą być użyte tylko w przypadku konkretnej postaci modelu. Inne z kolei, np. testy heteroskedastyczności, mogą postulować tylko konkretną postać funkcji. To wszystko nakłada na badacza konieczność zachowania szczególnej ostrożności w formułowaniu końcowej decyzji dotyczącej przydatności oszacowanego modelu procesu przestrzennego.

Zakończenie

Rozwijane od lat 50. ubiegłego stulecia metody modelowania matematyczno-statystycznego struktur i procesów przestrzennych mają już uznaną pozycję na gruncie nauk ekonomiczno-przestrzennych i regionalnych. Istnieje bogata literatura, głównie zagraniczna, obejmująca problematykę modelowania przestrzennego całościowo, oraz wiele artykułów koncentrujących się na konkretnych zagadnieniach. Siłą rzeczy, w niniejszym przeglądzie tylko część z tych zagadnień mogła zostać omówiona, niemniej jednak należą one do kluczowych. Dotyczy to zarówno rozważań teoretycznych, jak i zastosowań praktycznych.

Prace polskie w tym zakresie nie są zbyt liczne i z reguły koncentrują się na zastosowaniu metod modelowania przestrzennego do wybranych problemów. Niemniej jednak wśród polskich badaczy można zauważyć zwiększone zainteresowanie modelowaniem przestrzennym w ostatnich latach. Zespół badaczy z Zakładu Ekonometrii Przestrzennej UAM w Poznaniu, wraz z partnerami zagranicznymi, stara się wnieść pewien wkład w rozwój metod modelowania struktur i procesów przestrzennych. Prace teoretyczne członków tej grupy zostały przytoczone powyżej przy okazji omówienia wybranych zagadnień. Natomiast w ujęciu praktycznym modelowanie przestrzenne jest wykorzystywane np. w badaniach nad dochodami własnymi gmin w Polsce (Kossowski, Motek 2009), czy układem przestrzennym stopy bezrobocia (Kossowski 2009).

Powyższe rozważania pozwalają na stwierdzenie, że modelowanie przestrzenne jest procesem złożonym i wymaga od badacza podejmowania wielu decyzji w jego trakcie. Dotyczą one ustalenia reprezentacji struktury przestrzennej interakcji w formie konkretnej macierzy wag przestrzennych, przez wybór właściwej specyfikacji modelu, metody estymacji i w końcu sposobów jego testowania. W każdym z tych punktów badacz ma wpływ na końcowy wynik modelowania, a tym samym na jakość i przydatność formułowanego modelu. Z drugiej strony, modelowanie struktur i procesów przestrzennych wielokrotnie potwierdziło swoją przydatność jako narzędzie wykorzystywane przy wyjaśnianiu zjawisk zachodzących w przestrzeni geograficznej.

Literatura

Anselin L. 1988. Spatial Econometrics: Methods and Models. Kluwer, Dordrecht.

Barry R., Pace R. 1999. Monte Carlo Estimates of the Log Determinant of Large Sparse Matrices. Linear Algebra and Its Applications, 289: 41–54.

Anselin L. 1995. Local Indicators of Spatial Association-LISA. Geographical Analysis, 27: 93–115.

Anselin L., Florax R.J.G.M (red.) 1995. New Directions in Spatial Econometrics. Springer-Verlag, Berlin.

Anselin, L., Li, X. 2019. Operational local join count statistics for cluster detection. J. Geogr. Syst., 21: 189–210.

Bhattacharjee A., Jensen-Butler C. 2013. Estimation of the spatial weights matrix under structural constrains. Regional Science and Urban Economics, 43: 617–634.

Bivand R.S. 1984. Regression Modelling with Spatial Dependence: An Application of Some Class Selection and Estimation Methods. Geographical Analysis, 16: 25–37.

Bivand R., Hauke J., Kossowski T. 2013. Computing the Jacobian in Spatial Autoregressive Models: An Illustrated Comparison of Available Methods. Geographical Analysis, 45, 2: 150–179

Bivand R.S., Wilk J., Kossowski T. 2017. Spatial association of population pyramids across Europe: The application of symbolic data, cluster analysis and join-count tests. Spatial Statistics, 21: 339–361.

Cliff A.D., Ord J.K. 1970. Spatial Autocorrelation: A Review of Existing and New Measures with Applications. Economic Geography, 46: 269–272.

Cliff A.D., Ord J.K. 1972. Testing for Spatial Autocorrelation Among Regression Residuals. Geographical Analysis, 4: 267–284.

Cliff A.D., Ord J.K. 1973. Spatial Autocorrelation. Pion, London.

Cliff A.D., Ord J.K. 1981. Spatial Processes: Models and Applications. Pion, London.

Florax R.J., Rey S. 1995. The impact of misspecified spatial interaction in linear regression models. [W:] L. Anselin, R.J. Florax (red.), New Directions in Spatial Econometrics, s. 111–135.

Florax R.J.G.M., Folmer H., Rey S.J. 2003. Specification searches in spatial econometrics: the relevance of Hendry's methodology. Regional Science and Urban Economics, 33: 557–579.

Florax R.J.G.M., Folmer H., Rey S.J. 2005. A comment on specification searches in spatial econometrics: the relevance of Hendry's methodology: A reply. Regional Science and Urban Economics, 36: 557–579.

Folmer H., Oud J.H. 2008. How to get more rid of W: a latent variables approach to modelling spatially lagged variables. Environment and Planning, A 40: 2526–2538.

Fotheringham A.S. 2009. The Problem of Spatial Auctocorrelation and Local Spatial Statistics. Geographical Analysis, 41: 398–403.

Fotheringham A.S., Brundson C., Charlton M. 2002. Geographically Weighted Regression. Wiley & Sons, Chichester.

Geary R. 1954. The Contiguity Ratio and Statistical Mapping. The Incorporated Statistician, 5: 115–145.

Getis A., Aldstadt J. 2004. Constructing the Spatial Weights Matrix Using a Local Statistic. Geographical Analysis, 36: 90–104.

Griffith D.A., Sone A. 1995. "Trade-offs Associated with Normalizing Constant Computational Simplifications for Estimating Spatial Statistical Models. Journal of Statistical Computation and Simulation, 51: 165–83.

Haining R.P. 1978. Estimating Spatial Interaction Models. Environment and Planning, A, 10: 305–320.

Haining R. 1990. Spatial Data Analysis in the Social and Environmental Sciences. Cambridge University Press, Cambridge.

Hays J.C., Kachi A., Franzese R.J. 2010. A spatial model incorporating dynamic, endogenous network interdependence: A political science application. Statistical Methodology, 7: 406–428.

Herrera M., Mur J., Ruiz M. 2018. A Comparison Study on Criteria to Select the Most Adequate Weighting Matrix. Entropy, 21, 2: 160.

Kelejian H.H., Prucha I. 2002. 2SLS and OLS in spatial autoregressive model with equal weights. Regional Science and Urban Economics, 32: 691–707.

Kelejian H.H., Prucha I. 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. Journal of Real Estate Finance and Economics, 17, 1: 99–121.

Kelejian H.H., Prucha I. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review, 40, 2: 509–533.

Kelejian H.H., Prucha I. 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics, 157: 53–67.

Kelejian H.H., Piras G. 2014. Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. Regional Science and Urban Economics, 46: 140–149.

Kooijman S. 1976. Some remarks on the statistical analysis of grids especially with respect to ecology. Annals of System Research, 5: 113–132.

Kossowski T. 2009. Metody i modele ekonometrii przestrzennej. [W:] Z. Zwoliński (red.), GIS – platforma integracyjna geografii. Bogucki Wydawnictwo Naukowe, Poznań, s. 145–165.

Kossowski T., Motek P. 2009. Spatial modelling of the local public finance in Poland. [W:] T. Markowski, M. Turała (red.), Theoretical and practical aspects of urban and regional development. Studia Regionalia, 24: 152–167.
Kossowski T.M., Hauke J. 2011. The method of computing the Log-Jacobian of the variable transformation for spatial models-test and comments. Acta Universitatis Lodziensis, Folia Oeconomica, 252: 161–173.
Lee L.-F. 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics, 137: 489–514.
Lee L.F., Yu J. 2012. QML estimation of spatial dynamic panel data models with time-varying spatial weights matrices. Spatial Economic Analysis, 7: 31–74.
LeSage J.P., Pace R.K. 2014. The biggest myth in spatial econometrics. Econometrics, 2, 217–249.
Moran P.A.P. 1950. Notes on continuous stochastic phenomena. Biometrika, 37: 17–23.
Mur J., Hauke J., Kossowski T. 2019. Searching for the spatial weighting matrix. A GMM approach to W. Manuscript submitted to Sankhya A.
Ord J. 1975. Estimation Methods for Models of Spatial Interaction. Journal of the American Statistical Association, 70: 120–26.
Qu X., Lee L.F. 2015. Estimating a spatial autoregressive model with an endogenous spatial weighting matrix. Journal of Econometrics: 209–232.
Pace R., Barry R. 1997. Sparse Spatial Autoregressions. Statistics and Probability Letters, 33: 291–97.
Pace R., LeSage J. 2004. Chebyshev Approximation of Log-Determinants of Spatial Weight Matrices. Computational Statistics and Data Analysis 45: 179–96.
Ratajczak W. 1980. Analiza i modele wpływu czynników społeczno-gospodarczych na kształtowanie się sieci transportowych. PAN, seria Geografia, 5. PWN, Warszawa.
Ratajczak W. 2008. Modele ekonometrii przestrzennej w analizie regionalnej. [W:] T. Stryjakiewicz, T. Czyż (red.), O nowy kształt badań w geografii i gospodarce przestrzennej. Biuletyn KPZK PAN, 237: 186–202.
Ripley B. 1981. Spatial Statistics. Wiley, New York.
Smirnov O., Anselin L. 2001. Fast Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models: A Characteristic Polynomial Approach. Computational Statistics and Data Analysis, 35: 301–319.
Smirnov O., Anselin L. 2009. An O(N) Parallel Method of Computing the Log-Jacobian of the Variable Transformation for Models with Spatial Interaction on A Lattice. Computational Statistics and Data Analysis, 53: 2980–2988.
Tiefelsdorf M. 2000. Modeling spatial processes: the identification and analysis of spatial relationships in regression residuals by means of Moran's I. Springer, Berlin–Heidelberg.
Walde J., Larch M., Tappeiner G. 2008, Performance contest between MLE and GMM for huge spatial autoregressive models. Journal of Statistical Computation and Simulation, 78, 2: 151–166.
Whittle P. 1954. On Stationary Process in the Plane. Biometrika, 41: 434–449.
Whittle P. 1963. Stochastic Processes in Several Dimensions. Bulletin of the International Statistical Institute, 40: 974–994.

Mathematical and statistical modeling of spatial structures and processes: selected issues

Abstract: This study is a review of selected issues in mathematical and statistical modeling of spatial structures and processes. The review includes a discussion of basic concepts such as spatial pattern, spatial structure and spatial process and their relationships. Then, a general (stochastic) spatial process and its components are defined with a special focus on the problem of the spatial structure representation. The article discusses a procedure of constructing a stochastic spatial process model, and analyses the most important problems that arise during the specification, estimation and validation of the model. The Polish contribution to solving theoretical questions related to the modeling of spatial structures and processes was also emphasized.

Key words: spatial model, spatial pattern, spatial structure, spatial process, estimation