

*Uniwersytet Mikołaja Kopernika w Toruniu
Wydział Nauk Ekonomicznych i Zarządzania*

Michał Bernard Pietrzak

PROBLEM IDENTYFIKACJI STRUKTURY DANYCH PRZESTRZENNYCH

Zarys treści. Tematem artykułu będzie zagadnienie prawidłowej identyfikacji struktury danych przestrzennych oraz ukazanie, jak pominięcie ważnego składnika struktury danych może prowadzić do błędnych wniosków i interpretacji. Rozpatrzony zostanie wpływ nieuwzględnienia autokorelacji przestrzennej na jakość prowadzonych badań. Dodatkowym celem artykułu jest podniesienie świadomości badacza o potencjalnych niebezpieczeństwach w trakcie realizacji analiz związanych z tematyką przestrzenną, które to następnie stają się podstawą procesu podejmowania decyzji.

Słowa kluczowe: wewnętrzna struktura danych przestrzennych, trend przestrzenny, autokorelacja przestrzenna.

WSTĘP

W wyniku ostatniej reformy administracyjnej wprowadzono nie tylko obowiązujący od dnia 1.01.1999 roku nowy podział administracyjny Polski, ale również dano samorządom lokalnym możliwość samodzielnego wpływu na politykę regionu. Oznaczało to decentralizację zarządzania państwem i miało przyczynić się do podniesienia jego jakości. Dodatkowo Polska od 1.05.2004 roku stała się członkiem Unii Europejskiej, co otworzyło możliwość korzystania z pomocy funduszy programów operacyjnych 2004–2006 oraz funduszy programów operacyjnych 2007–2013. Wszystko to przyczyniło się do wzrostu znaczenia prowadzonej polityki regionalnej, której jakość przekłada się bezpośrednio na społeczno-ekonomiczny rozwój regionu. Zwiększyła się również rola informacji, ponieważ rzetelne i obiektywne stanowią dla decydentów, związanych z zarządzaniem regionem, podstawę podejmowania trafnych decyzji. Użyteczne informacje są wynikiem analiz ekonomicznych, bazujących na danych

przestrzennych. W przypadku realizacji analiz przestrzennych zjawisk ekonomicznych proces wnioskowania może zostać zakłócony w wyniku niepoprawnej identyfikacji wewnętrznej struktury danych, co stanowi o ważności tej czynności badawczej. W związku z tym głównym celem artykułu jest rozpatrzenie zagadnienia prawidłowej identyfikacji wewnętrznej struktury danych przestrzennych, która powinna zapewnić sensowność otrzymanych wniosków zarówno w przypadku opisu badanego zjawiska, jak i w przypadku analizy istniejących zależności. Szczególna uwaga zostanie zwrócona na fakt, jak pominięcie ważnego składnika struktury danych, w postaci zjawiska autokorelacji przestrzennej, może prowadzić do błędnych interpretacji. W polskiej literaturze problematyka modelowania zjawisk przestrzennych nie była często podejmowana. Do najważniejszych pozycji należą: Bivand (1981), Czyż (1978), Kopczewska (2006), Szulc (2007), Zeliaś (1991).

1. IDENTYFIKACJA WEWNĘTRZNEJ STRUKTURY ZJAWISK PRZESTRZENNICH

W przypadku danych przestrzennych badane zjawisko możemy potraktować jako dwuwymiarowe pole losowe $X(u)$, nazywane dalej procesem przestrzennym, gdzie $u = (u_1, u_2)$ oznacza współrzędne na płaszczyźnie¹. W analizach empirycznych, oprócz opisu badanego zjawiska, istotne jest również wykrycie zależności przyczynowo-skutkowych. W ramach wykonania pełnej analizy przyczynowo-skutkowej należy rozpocząć badanie od identyfikacji wewnętrznej struktury wybranych procesów przestrzennych, w których wyróżnić można dwa składniki. Pierwszy składnik $f(u)$ wyraża niejednorodność zjawiska w średniej. Chodzi tu o niejednorodność systematyczną, która wyraża zmienność w wartości średniej procesu przestrzennego² i może być opisana modelem trendu przestrzennego. Drugi składnik, określony symbolem $e(u)$, wyraża zmienność zjawiska, mającą własność jednorodności³ w średniej. Składnik jednorodny w średniej opisuje

¹ Problem określenia pola losowego został rozpatrzony w pracy Szulc (2007), s. 88.

² Tamże, s. 102.

³ Zakłada się tutaj jednorodność w szerszym sensie (słabą jednorodność), w przypadku której spełnione są następujące założenia:

$$E(X(u)) = \mu,$$

$$\frac{1}{|N(h)|} \sum_{N(h)} ([X_i(u) - \mu][X_j(u) - \mu]) = K(u_i, u_j) = K(h),$$

$$h = u_i - u_j, u_i = (u_{1i}, u_{2i}), u_j = (u_{1j}, u_{2j}),$$

gdzie $X_i(u)$ wartości procesu, $N(h)$ zbiór par lokalizacji o różnicy h .

model szumu przestrzennego⁴ albo – w przypadku stwierdzenia autozależności – model przestrzennego procesu autoregresyjnego.

1.1. IDENTYFIKACJA WŁASNOŚCI NIEJEDNORODNOŚCI W ŚREDNIEJ

Niejednorodność systematyczną można wyrazić w postaci trendu przestrzennego, który definiowany jest jako ogólna tendencja przestrzenna⁵. W modelowaniu trendu przestrzennego wykorzystywane są funkcje wielomianowe. Ze względu na postać analityczną wyróżnić można przestrzenny trend liniowy, trend kwadratowy, trend sześcienny oraz trendy wyższych rzędów.

W przypadku danych przestrzennych dominujące są dwie przestrzenne tendencje rozwoju zjawiska: systematyczny rozwój zjawiska w określonym kierunku, nazywany dalej trendem kierunkowym, oraz rozwój zjawiska wokół dominującego centrum, określony w pracy jako trend centralny. Trend kierunkowy opisywany jest modelem trendu liniowego, a w przypadku większej złożoności systematycznych zmian modelem trendu sześciennego lub wyższych rzędów. Natomiast trend centralny poprawnie opisuje model trendu kwadratowego, który może być zastąpiony modelami trendu wyższych rzędów w miarę wzrostu systematycznej złożoności.

Model trendu liniowego można określić za pomocą wzoru

$$X_i(u) = \alpha_0 + \alpha_1 u_{1i} + \alpha_2 u_{2i} + e_i(u), \quad (1)$$

gdzie $\alpha_0, \alpha_1, \alpha_2$ to parametry strukturalne, u_{1i}, u_{2i} to współrzędne na płaszczyźnie, a $e_i(u)$ stanowi proces jednorodny w szerszym sensie.

Model trendu liniowego, jako model trendu kierunkowego, służy do określenia kierunku rozwoju przestrzennego badanego zjawiska. O kierunku i sile rozwoju świadczą znaki i wartości ocen parametrów strukturalnych stojących przy współrzędnych geograficznych. Na wykresie 1 przedstawiono kształtowanie się stopy bezrobocia z podziałem na powiaty w roku 2007. Natomiast na wykresie 2 zaprezentowane zostało dopasowanie modelu przestrzennego trendu liniowego, gdzie widoczna jest przestrzenna tendencja wzrostu stopy bezrobocia wraz z przechodzeniem w kierunku północno-wschodniej części Polski. Przy czym

⁴ Szum przestrzenny (biały szum) stanowi jednorodne w szerszym sensie pole losowe. Dodatkowo czyni się założenia o zerowej wartości średniej oraz funkcji kowariancji postaci

$$K(u, u') = \begin{cases} \delta^2, & \text{gdy } u = u' \\ 0, & \text{gdy } u \neq u' \end{cases}, \text{ co znaleźć można w pracy Szulc (2007), s. 30.}$$

⁵ Problematyka związana z trendem przestrzennym przedstawiona została kompleksowo w pracach: Czyż (1978), Szulc (2007).

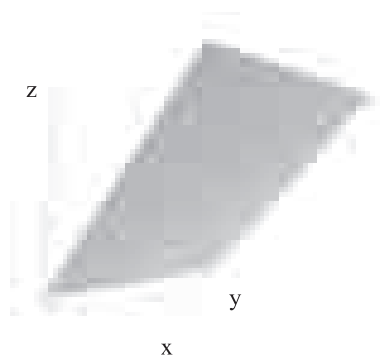
znacznie silniejszy wzrost stopy bezrobocia występuje przy założeniu kierunku zachód-wschód.

Model trendu liniowego uwzględnia główny kierunek przestrzennego rozwoju. Systematyczne zmiany najczęściej jednak charakteryzują się większą złożonością i w przypadku stopy bezrobocia, oprócz ogólnej tendencji kierunkowej, występuje też tendencja wyższej stopy bezrobocia w powiatach przygranicznych województw: zachodniopomorskiego, dolnośląskiego i opolskiego.



Wykres 1. Wartości stopy bezrobocia⁶ w powiatach w 2007 roku

Źródło: opracowanie własne.

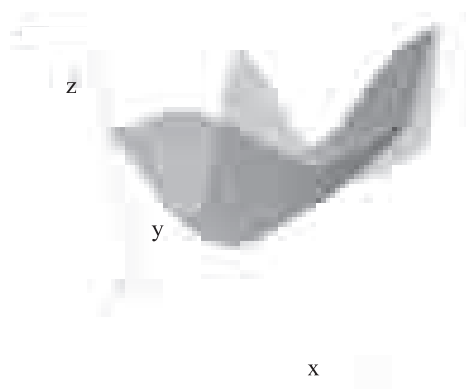


Wykres 2. Dopasowanie modelu trendu liniowego

Źródło: opracowanie własne.

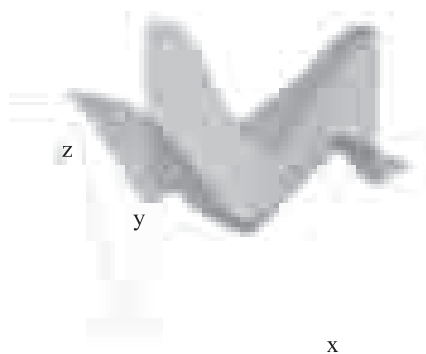
⁶ Oś x , y stanowią dwuwymiarowy układ przestrzenny, oś z opisuje wartości badanego zjawiska.

Wraz ze wzrostem stopnia modelu trendu uwzględniane jest coraz większe zróżnicowanie badanego szeregu. Widoczne jest to na wykresach 3 oraz 4, na których oprócz wymienionych powiatów przygranicznych został również uwzględniony pas wyższych wartości stopy bezrobocia w województwach warmińsko-mazurskim i mazowieckim oraz pas utworzony od terenów byłego województwa radomskiego, przechodzący przez województwo świętokrzyskie i sąsiadujące województwa małopolskie i podkarpackie.



Wykres 3. Dopasowanie modelu trendu sześciennego

Źródło: opracowanie własne.



Wykres 4. Dopasowanie modelu trendu czwartego stopnia

Źródło: opracowanie własne.

W przypadku dopasowania kilku modeli trendu przestrzennego pojawia się problem wyboru właściwego modelu. Podobnie jak w przypadku analizy

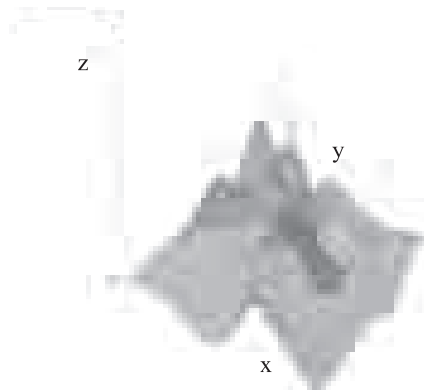
szeregów czasowych, stosowany jest test F^7 , w którym w hipotezie alternatywnej zakłada się istotny spadek wariancji resztowej. Ostatecznie jednak wybór stopnia modelu trendu należy do badacza, który powinien ustalić granicę, po przekroczeniu której opisywane jest dowolne zróżnicowanie procesu, a nie, zgodnie z założeniem, jedynie zróżnicowanie systematyczne.

Model trendu kwadratowego⁸ opisany jest wzorem:

$$X_i(u) = \alpha_0 + \alpha_1 u_{1i} + \alpha_2 u_{2i} + \alpha_3 u_{1i}^2 + \alpha_4 u_{1i} u_{2i} + \alpha_5 u_{2i}^2 + e_i(u), \quad (2)$$

gdzie $\alpha_0, \alpha_1, \dots, \alpha_5$ to parametry strukturalne, u_{1i}, u_{2i} to współrzędne na płaszczyźnie, a $e_i(u)$ stanowi proces jednorodny w szerszym sensie.

Model trendu kwadratowego poprawnie opisuje tendencję przestrzenną, określoną przez autora jako trend centralny. Na wykresie 5 zaprezentowano kształtowanie się dochodów własnych na osobę dla gmin województwa kujawsko-pomorskiego w roku 2006.



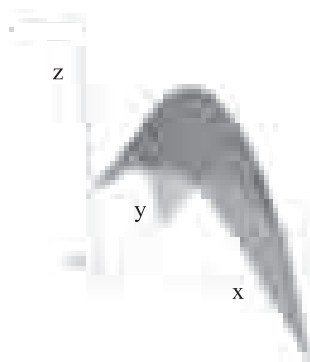
Wykres 5. Wartości dochodów własnych gmin na osobę w województwie kujawsko-pomorskim w roku 2006

Źródło: opracowanie własne.

Natomiast na wykresie 6 zostało przedstawione dopasowanie przestrzennego trendu kwadratowego. Widoczna jest tu przestrzenna tendencja w postaci dominującego centrum tworzonego przez dwa, położone blisko siebie, największe miasta województwa, Bydgoszcz i Toruń. Biorąc pod uwagę niewielką odległość między tymi miastami, przypuszczać można, że w przyszłości stworzą one aglomerację bydgosko-toruńską.

⁷ Zastosowanie testu F dla trendu przestrzennego znaleźć można w pracy Czyż (1978), s. 32.

⁸ Analogicznie tworzone są modele trendu wyższych rzędów. Należy jednak pamiętać o założeniu, że łączny stopień iloczynu u_{1i}, u_{2i} nie może być większy niż stopień zakładanego trendu.



Wykres 6. Dopasowanie modelu trendu kwadratowego

Źródło: opracowanie własne.

1.2. IDENTYFIKACJA WŁASNOŚCI AUTOKORELACJI PRZESTRZENNEJ

Składnik $e(u)$ wyraża zmienność zjawiska, którą charakteryzuje własność jednorodności w średniej. Po eliminacji trendu przestrzennego zjawisko przestrzenne powinno charakteryzować się jednorodnością w szerszym sensie. Własność ta nie wyklucza jednak występowania autozależności przestrzennych⁹. Zjawisko autokorelacji przestrzennej oznacza istnienie zależności między wartościami procesu w określonej jednostce a wartościami tego procesu w jednostkach sąsiadujących. Dlatego kolejnym krokiem powinna być identyfikacja własności autokorelacji przestrzennej. Pominięcie zjawiska autokorelacji przestrzennej może stać się przyczyną błędnego wnioskowania, co zostanie pokazane w dalszej części pracy. Również pominięcie trendu przestrzennego w modelu regresji powoduje wzrost autozależności w resztach modelu.

Jedną z miar autokorelacji przestrzennej jest statystyka globalna I Morana¹⁰. Wartości statystyki I Morana istotnie większe od zera wskazują na istnienie dodatniej autokorelacji przestrzennej, a wartości istotnie ujemne na autokorelację ujemną. Brak istotnej różnicy od zera świadczy o niewystępowaniu autozależności przestrzennych. Miara ta określona jest wzorem¹¹:

⁹ Własność autokorelacji przestrzennej została rozważona w pracach: Bivand (1981); Cliff, Ord (1981); Czyż (1978); Szulc (2007); Zeliaś (1991).

¹⁰ Miary autokorelacji przestrzennej omówiono w pracach: Moran (1948); Cliff, Ord (1981).

¹¹ Statystyka I Morana ma rozkład normalny. Wyprowadzenie wzorów dla teoretycznej wartości oczekiwanej $E(I)$ i wariancji $Var(I)$ oraz problematyka związana z testowaniem statystyki została poruszona w pracy Cliff, Ord (1981).

$$I = \frac{\sum_i \sum_j w_{ij} (x_i(u) - \bar{x}(u))(x_j(u) - \bar{x}(u))}{\sum_i \sum_j w_{ij} (x_i(u) - \bar{x}(u))^2}, \quad (3)$$

gdzie N jest liczbą jednostek przestrzennych, $x_i(u) = x(u_{1i}, u_{2i})$ oznacza wartość badanej zmiennej w i -tym regionie, $\bar{x}(u)$ jest średnią ze wszystkich regionów, natomiast w_{ij} jest elementem ustalonej wcześniej macierzy wag.

W przypadku istnienia autozależności przestrzennych zmienność składnika $e_i(u)$ można opisać za pomocą modelu autoregresyjnego procesu przestrzennego pierwszego rzędu¹², co określamy wzorem:

$$e_i(u) = qWe_i(u) + \eta_i(u), \quad (4)$$

gdzie q jest parametrem autoregresji, W ¹³ stanowi macierz wag przestrzennych, a $\eta_i(u)$ jest szumem przestrzennym.

W przypadku właściwego ustalenia struktury zarówno objaśnianego procesu przestrzennego, jak i procesów objaśniających można przejść do tworzenia zgodnego modelu przyczynowo-skutkowego¹⁴. W wyniku uwzględnienia właściwej struktury wszystkich procesów przestrzennych uzyskiwany jest pełny model zgodny, określony wzorem¹⁵:

$$Y_i(u) = \alpha_0 + qWY_i(u) + f_i(u) + \alpha_1 X_{1i}(u) + \alpha_2 WX_{1i}(u) + \dots + \alpha_{2n-1} X_{ni}(u) + \alpha_{2n} WX_{ni}(u) + \eta_i(u), \quad (5)$$

gdzie q , α_0 , α_1 , ..., α_{2n} są parametrami strukturalnymi, $f_i(u)$ stanowi trend przestrzenny, W jest macierzą wag przestrzennych, a $\eta_i(u)$ jest szumem przestrzennym. Składniki nieistotne mogą następnie ulec redukcji, w wyniku czego otrzymany jest końcowy model zgodny.

¹² Patrz Bivand (1981), s. 13. W przypadku procesów stochastycznych zakłada się istnienie rzędu autoregresji dowolnego rzędu. Dla potrzeb analiz przestrzennych proces autoregresyjny pierwszego rzędu wydaje się w wystarczającym stopniu uwzględniać autozależności przestrzenne.

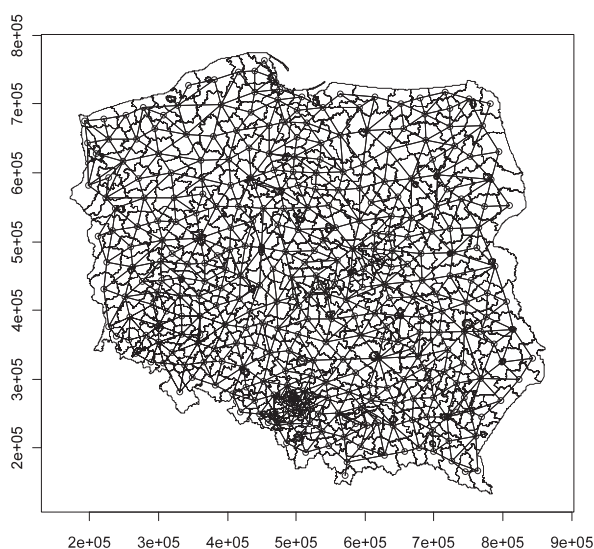
¹³ W badaniach wykorzystuje się najbardziej intuicyjne kryterium sąsiedztwa zdefiniowane na podstawie posiadania wspólnej granicy. Tak określona macierz wag przestrzennych nazywana jest macierzą sąsiedztwa $W = (w_{i,j})_{i,j=1,\dots,n}$, która ma zera na przekątnej, natomiast poza przekątną wartości macierzy są równe jeden, gdy obiekty i oraz j graniczą ze sobą, oraz wartość zero w przeciwnym wypadku. Dodatkowo standaryzuje się macierz w ten sposób, by suma elementów dla każdego z wierszy równa była jeden.

¹⁴ Autor opiera się na pojęciach modelowania zgodnego, jak i modelu zgodnego wprowadzonych przez Profesora Zygmunta Zielińskiego. Koncepcja ta w przypadku przestrzennych pól losowych została rozważona w pracy Szulc (2007).

¹⁵ Można rozważać sytuację, gdy zjawisko autokorelacji opisuje model autoregresyjny wyższego rzędu niż pierwszy. Należałoby wtedy odpowiednio rozbudować model przyczynowo-skutkowy.

2. POTENCJALNE SKUTKI NIEUWZGLĘDNIENIA ZJAWISKA AUTOKORELACJI PRZESTRZENNEJ

Wnioskowanie statystyczne daje możliwość wydawania sądów o populacji na podstawie próby. Musi jednak zachodzić warunek niezależności obserwacji w próbie. Występowanie zjawiska autokorelacji prowadzi do obciążenia estymatorów, co z kolei przekłada się na jakość testów statystycznych i jakość wnioskowania statystycznego. W celu pokazania potencjalnych skutków nieuwzględnienia zjawiska autokorelacji przestrzennej wykonano trzy symulacje. Celem pierwszej symulacji była ocena jakości testu istotności dla wartości oczekiwanej w warunkach istnienia autokorelacji przestrzennej.



Wykres 7. Układ przestrzenny z uwzględnieniem sąsiedztwa I rzędu

Źródło: opracowanie własne.

W celu zwiększenia realności wykorzystania testu symulacja została przeprowadzona w układzie przestrzennym Polski w podziale na powiaty. Wykres 7 przedstawia układ przestrzenny wykorzystany w ramach zrealizowanej symulacji, wraz z uwzględnieniem sąsiedztwa pierwszego rzędu. Symulacja polegała na generowaniu kolejno szumu przestrzennego o określonej średniej oraz trzech kolejnych procesów autoregresyjnych o parametrach $\alpha = 0,5$; $\alpha = 0,7$; $\alpha = 0,9$ ¹⁶.

¹⁶ W przypadku wszystkich symulacji szumu przestrzennego, jak i kolejnych procesów autoregresyjnych generowano 50 000 razy.

Efektom generowania kolejnych realizacji procesów autoregresyjnych było uzyskanie struktury przestrzennej charakteryzującej się własnością autokorelacji.

W celu symulacji procesu autoregresyjnego skorzystano z zależności¹⁷:

$$X_i(u) = qWX_i(u) + e_i(u) \Leftrightarrow (I - qW)X_i(u) = e_i(u)$$

$$X_i(u) = (I - qW)^{-1}e_i(u) \Leftrightarrow X_i(u) = e_i(u) + qWe_i(u) + q^2W^2e_i(u) + \dots, \quad (6)$$

gdzie q jest parametrem autoregresji, W stanowi macierz wag przestrzennych, a $e_i(u)$ jest szumem przestrzennym.

W wyniku przeprowadzonej symulacji otrzymano wartości krytyczne testu¹⁸, którymi były odpowiednie kwantyle uzyskanych statystyk, oraz policzono wartości prawdopodobieństwa popełnienia błędu pierwszego rodzaju na podstawie liczby odrzuceń. Tabela 1 przedstawia wyniki przeprowadzonej symulacji w przypadku szumu przestrzennego. Uzyskane wartości krytyczne, jak i błąd pierwszego rodzaju można uznać za prawidłowe. Oznacza to, że wnioskowanie statystyczne na podstawie tego testu, przy faktycznym braku autozależności przestrzennych, prowadzi do właściwych wniosków¹⁹.

Tabela 1. Wyniki symulacji przy założeniu generowania szumu przestrzennego

Poziom istotności	0,01	0,05	0,1
Wartości krytyczne	2,33	1,64	1,27
Błąd pierwszego rodzaju	0,0104	0,0496	0,0994

Źródło: opracowanie własne.

Wyniki symulacji, w przypadku generowania przestrzennych procesów autoregresyjnych, przedstawione zostały w tabeli 2. Wraz ze wzrastającym poziomem autokorelacji przestrzennej rosną wartości krytyczne. Oznacza to, że stosując wartości krytyczne z rozkładu normalnego i nie uwzględniając zjawiska autokorelacji, zwiększa się znacznie prawdopodobieństwo odrzucenia hipotezy zerowej. Przejawia się to w znacznie zwiększonym poziomie błędu pierwszego rodzaju, który rośnie wraz ze wzrostem autokorelacji przestrzennej. Prowadzi to do wniosku, że w przypadku istnienia zjawiska autokorelacji jakość testu znacznie spada.

¹⁷ Problem odwrócenia macierzy we wzorze (6) rozważono w pracy Bivand (1981), s. 13.

¹⁸ Założono w symulacji test prawostronny. Wyniki dla testu lewostronnego lub dwustronnego można wywnioskować na podstawie wyników dla testu prawostronnego.

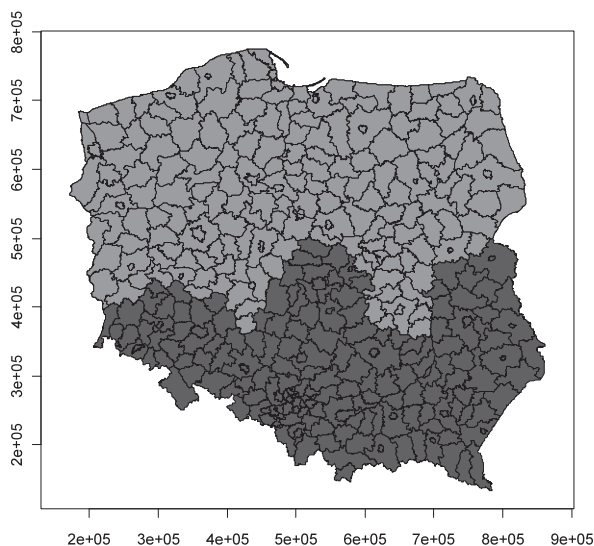
¹⁹ Badacz powinien jednak pamiętać, że może popełnić błąd pierwszego rodzaju z prawdopodobieństwem równym założonemu poziomowi istotności α .

Tabela 2. Wyniki symulacji przy założeniu generowania autoregresyjnych procesów przestrzennych

Proces autoregresyjny $\alpha = 0,5$			
Poziom istotności	0,01	0,05	0,1
Wartości krytyczne	4,15	2,96	2,31
Błąd pierwszego rodzaju	0,0969	0,1804	0,2378
Proces autoregresyjny $\alpha = 0,7$			
Poziom istotności	0,01	0,05	0,1
Wartości krytyczne	5,5	3,82	3
Błąd pierwszego rodzaju	0,1564	0,2358	0,2892
Proces autoregresyjny $\alpha = 0,9$			
Poziom istotności	0,01	0,05	0,1
Wartości krytyczne	6,6	4,71	3,61
Błąd pierwszego rodzaju	0,2119	0,2881	0,3317

Źródło: opracowanie własne.

Zbadana została również jakość testu na istotność różnicy dwóch wartości oczekiwanych. W tym przypadku posłużono się także układem administracyjnym Polski w podziale na powiaty. Wyodrębniono dwa obszary: północny, który utworzyły powiaty województw: zachodniopomorskiego, pomorskiego, warmińsko-mazurskiego, lubuskiego, wielkopolskiego, kujawsko-pomorskiego,



Wykres 8. Podział na dwa obszary przestrzenne

Źródło: opracowanie własne.

mazowieckiego i podlaskiego, oraz południowy, powstały przez zsumowanie województw: dolnośląskiego, opolskiego, łódzkiego, śląskiego, świętokrzyskiego, lubelskiego, małopolskiego oraz podkarpackiego. Test ten można wykorzystać do wykazania istotnej różnicy lub jej braku w przypadku dwóch wartości oczekiwanych wybranej cechy, przy założeniu, że cecha pochodzi z rozkładów obserwowanych na dwóch obszarach. Wykres 8 (zob. s. 93) przedstawia dwa ustalone obszary, na podstawie których przeprowadzono kolejną symulację.

Dla obydwu obszarów generowano szumy przestrzenne oraz przestrzenne procesy autoregresyjne, przy założeniu tych samych współczynników autoregresji $\alpha_{1,2} = 0,5$; $\alpha_{1,2} = 0,7$; $\alpha_{1,2} = 0,9$. Następnie wyznaczono wartości krytyczne jako kwantyle dla uzyskanych statystyk testu. Policzono również prawdopodobieństwo popełnienia błędu pierwszego rodzaju na podstawie liczby odrzuceń. Otrzymane wyniki dla szumu przestrzennego zamieszczone zostały w tabeli 3. Zawarte w niej wartości krytyczne oraz prawdopodobieństwa popełnienia błędu pierwszego rodzaju są prawidłowe, co oznacza poprawność testu w przypadku braku zależności przestrzennych.

Tabela 3. Wyniki symulacji przy założeniu generowania szumu przestrzennego

Poziom istotności	0,01	0,05	0,1
Wartości krytyczne	2,34	1,65	1,29
Błąd pierwszego rodzaju	0,0104	0,0516	0,1016

Źródło: opracowanie własne.

Tabela 4. Wyniki symulacji przy założeniu generowania autoregresyjnych procesów przestrzennych

Proces autoregresyjny $\alpha = 0,5$			
Poziom istotności	0,01	0,05	0,1
Wartości krytyczne	3,61	3	2,31
Błąd pierwszego rodzaju	0,0979	0,1798	0,2382
Proces autoregresyjny $\alpha = 0,7$			
Poziom istotności	0,01	0,05	0,1
Wartości krytyczne	4,71	3,82	2,96
Błąd pierwszego rodzaju	0,1589	0,2398	0,2909
Proces autoregresyjny $\alpha = 0,9$			
Poziom istotności	0,01	0,05	0,1
Wartości krytyczne	6,6	5,4	4,15
Błąd pierwszego rodzaju	0,2068	0,2805	0,3246

Źródło: opracowanie własne.

Otrzymane wyniki jakości testu dla danych generowanych przez dwa przestrzenne procesy autoregresyjne zostały zawarte w tabeli 4 (zob. s. 94). Widoczne jest, że wraz ze wzrastającym poziomem autokorelacji przestrzennej rosną również wartości krytyczne testu. Fakt ten powinien przejawiać się w częstszym odrzucaniu hipotezy zerowej, niż to wynika z przyjętego poziomu istotności α . Prawdopodobieństwo popełnienia błędu pierwszego rodzaju znacznie wzrosło, co – podobnie jak dla testu na istotność wartości oczekiwanej – świadczy o pogarszaniu się jakości testu.

Problemy przedstawione wcześniej dotyczyły wnioskowania co do podstawowych charakterystyk badanego zjawiska. Kolejnym krokiem będzie rozpatrzenie jakości estymacji modelu trendu przestrzennego, w przypadku nieuwzględnienia zjawiska przestrzennej autokorelacji. Pojawiają się tutaj dwa aspekty. Aspekt pierwszy dotyczy istotności parametrów modelu. Istotna autokorelacja przestrzenna powoduje obciążenie wariancji składnika resztowego i w konsekwencji prowadzi do nieefektywnej estymacji parametrów strukturalnych²⁰. Oznacza to możliwość wystąpienia sytuacji pozornej istotności parametru w modelu lub sytuacji nieistotności parametru, podczas gdy jest on statystycznie istotny. Aspekt drugi dotyczy dopasowania modelu trendu przestrzennego, które przy nieuwzględnieniu autokorelacji przestrzennej pozornie rośnie. Ponownie wygenerowano szum przestrzenny oraz przestrzenne procesy autoregresyjne z parametrami $\alpha = 0,5$; $\alpha = 0,7$; $\alpha = 0,9$ ²¹. Następnie szacowano kolejno przestrzenny model trendu liniowego, trendu kwadratowego oraz trendu sześciennego. Na podstawie otrzymanych wyników policzono kwantyle dla uzyskanych współczynników determinacji oraz prawdopodobieństwo istotności parametrów strukturalnych.

Tabela 5. Wartości kwantyli współczynnika determinacji dla trendu liniowego

Kwantyle*	Szum przestrzenny	Proces autoregresyjny (0,5)	Proces autoregresyjny (0,7)	Proces autoregresyjny (0,9)
0,9	0,0129	0,038	0,062	0,094
0,95	0,0154	0,048	0,079	0,118
0,99	0,0242	0,073	0,118	0,173

* Kolejne kwantyle oznaczają wartość, poniżej której znajduje się odpowiednio 90%, 95% oraz 99% wszystkich uzyskanych wartości współczynnika determinacji.

Źródło: opracowanie własne.

²⁰ Problem ten został poruszony w pracy Czyż (1978), s. 39.

²¹ W przypadku wszystkich symulacji szumu przestrzennego, jak i kolejnych procesów autoregresyjnych generowano 50 000 razy.

Tabela 6. Wartości kwantyli współczynnika determinacji dla trendu kwadratowego

Kwantyle	Szum przestrzenny	Proces autoregresyjny (0,5)	Proces autoregresyjny (0,7)	Proces autoregresyjny (0,9)
0,9	0,024	0,073	0,094	0,134
0,95	0,029	0,086	0,118	0,183
0,99	0,0398	0,114	0,173	0,234

Źródło: opracowanie własne.

Tabela 7. Wartości kwantyli współczynnika determinacji dla estymacji trendu sześciennego

Kwantyle	Szum przestrzenny	Proces autoregresyjny (0,5)	Proces autoregresyjny (0,7)	Proces autoregresyjny (0,9)
0,9	0,038	0,11	0,17	0,22
0,95	0,043	0,12	0,19	0,25
0,99	0,056	0,15	0,23	0,31

Źródło: opracowanie własne.

Tabele 5, 6 oraz 7 przedstawiają wyliczone kwantyle dla zbiorów wartości współczynników determinacji, uzyskanych podczas estymacji kolejnych modeli trendu przestrzennego. We wszystkich trzech tabelach widoczny jest wzrost wartości kwantyli współczynnika determinacji w miarę wzrostu siły autozależności przestrzennych. Ponieważ w strukturze symulowanych procesów nie ma trendu przestrzennego, uzyskany wynik świadczy o możliwości pozornego dopasowania modelu. Błąd ten wynika z nieuwzględnienia zjawiska autokorelacji przestrzennej.

Tabela 8. Prawdopodobieństwo istotności parametrów dla trendu liniowego

Parametry strukturalne	Szum przestrzenny	Proces autoregresyjny (0,5)	Proces autoregresyjny (0,7)	Proces autoregresyjny (0,9)
α_1	0,051	0,2742	0,3969	0,4997
α_2	0,0499	0,2741	0,3961	0,4998
Łącznie	0,0975	0,4692	0,6311	0,7456

Źródło: opracowanie własne.

Tabela 9. Prawdopodobieństwo istotności parametrów dla trendu kwadratowego²²

Parametry strukturalne	Szum przestrzenny	Proces autoregresyjny (0,5)	Proces autoregresyjny (0,7)	Proces autoregresyjny (0,9)
α_3	0,0520	0,2642	0,3860	0,4872
α_4	0,0510	0,2676	0,3792	0,4831
α_5	0,0494	0,2659	0,387	0,4797

Źródło: opracowanie własne.

²² W tabelach 9 i 10 podane zostały wyniki jedynie dla parametrów stojących przy iloczynach najwyższych stopni, dla trendu kwadratowego stopnia drugiego, a dla sześciennego trzeciego.

Tabele 8, 9 oraz 10 zawierają otrzymane wartości prawdopodobieństwa istotności parametrów strukturalnych. Analiza wyników wskazuje na fakt, że w przypadku istnienia autokorelacji znacznie zwiększa się prawdopodobieństwo istotności parametrów strukturalnych. W tabeli 8 przedstawiono dodatkowo w ostatnim wierszu prawdopodobieństwo w przypadku istotności choć jednego z dwóch parametrów²³, prawdopodobieństwo uznania modelu za istotny wzrasta prawie dwukrotnie. Uwzględnienie ostatniego wiersza w tabeli 8 powoduje prawie dwukrotny wzrost prawdopodobieństwa do poziomu 46%, 63% oraz 74%, w zależności od siły autozależności przestrzennych.

Tabela 10. Prawdopodobieństwo istotności parametrów dla estymacji trendu sześciennego

Parametry strukturalne	Szum przestrzenny	Proces autoregresyjny (0,5)	Proces autoregresyjny (0,7)	Proces autoregresyjny (0,9)
α_4	0,0507	0,2371	0,3390	0,4311
α_5	0,0478	0,2366	0,3333	0,4391
α_6	0,0508	0,2526	0,3578	0,4675
α_7	0,0495	0,2542	0,3579	0,4629

Źródło: opracowanie własne.

PODSUMOWANIE

Celem artykułu było przedstawienie problemu identyfikacji wewnętrznej struktury procesów przestrzennych. Ustalenie wewnętrznej struktury zapewnia poprawność uzyskiwanych informacji oraz powinno stanowić podstawę każdej analizy ekonomicznej dla danych przestrzennych. W artykule została pokazana możliwość opisu tej struktury poprzez wykorzystanie modelu trendu przestrzennego oraz przestrzennych modeli autoregresyjnych. Za pomocą przeprowadzonych symulacji wykazano, że nieuwzględnienie zjawiska autokorelacji przestrzennej znacznie obniża wartość poznawczą prowadzonych analiz. Istotna autokorelacja przestrzenna nie tylko pogarsza jakość testów w przypadku wnioskowania co do charakterystyk zjawiska, ale również wyklucza poprawne modelowanie ekonometryczne. Skutkiem zjawiska autokorelacji jest obciążenie estymatorów. Przekłada się to na jakość testów statystycznych, a w przypadku przestrzennych modeli ekonometrycznych na jakość estymacji i weryfikacji. Ostatecznie problem

²³ Zakłada się, że do uznania istotności trendu przestrzennego wystarczy istotność jednego z parametrów stojących przy iloczynach najwyższych stopni.

poprawnej identyfikacji struktury zjawisk przestrzennych, jako niezbędnej czynności badawczej w każdej ekonomicznej analizie przestrzennej, przekłada się na jakość uzyskiwanych informacji, które stają się następnie podstawą w podejmowaniu decyzji.

LITERATURA

- Bivand R. (1981), *Modelowanie geograficznych układów czasoprzestrzennych*, PWN, Warszawa–Poznań.
- Cliff A., Ord J. K. (1981), *Spatial process: Models and Applications*, Pion, London.
- Czyż T. (1978), *Metody generalizacji układów przestrzennych*, PWN, Warszawa–Poznań.
- Kopczewska K. (2006), *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, Wydawnictwo CeDeWu, Warszawa.
- Moran P. (1948), *The interpretation of statistical map*, „Journal of the Royal Statistical Society, Series B”, 10, 243–251.
- Szulc E. (2007), *Ekonometryczna analiza wielowymiarowych procesów gospodarczych*, Wydawnictwo UMK, Toruń.
- Zeliaś A. (red.) (1991), *Ekonometria przestrzenna*, Wydawnictwo Ekonomiczne, Warszawa.

THE PROBLEM OF IDENTIFICATION THE STRUCTURE OF SPATIAL DATA

A b s t r a c t. The article presents the problem of a proper identification of the structure of spatial data. It will be shown how the phenomenon of spatial autocorrelation causes mistakes in the economic analysis. Also, the impact of non-consideration of the spatial autocorrelation on the quality of the research conducted will be analysed. An additional purpose of the article is to increase the researcher's awareness of potential dangers related to analyses of spatial data which, later, form the basis for the decision-making process.

Key words: the structure of spatial data, the spatial trend, the spatial autocorrelation.