

Uniwersytet Mikołaja Kopernika w Toruniu
Katedra Ekonometrii i Statystyki

Elżbieta Szulc

EKONOMETRYCZNA IDENTYFIKACJA STRUKTUR PROCESÓW PRZESTRZENNYCH WOBEC PROBLEMU AGREGACJI DANYCH

Zarys treści. Artykuł dotyczy odkrywania zależności między ekonomicznymi procesami przestrzennymi, gdy są one mierzone na różnych poziomach agregacji danych. Rozważania nawiązują do badań potwierdzających efektywność tzw. quasi-zgodnego modelu przestrzennego jako narzędzia pomiaru rzeczywistych zależności między procesami, pod warunkiem, że modelowy opis struktur poszczególnych procesów jest prawidłowy i wystarczający. Poszukuje się opisu powiązań przestrzennych adekwatnego do wyrażenia autozależności w badanych procesach. Proponuje się wykorzystanie odległości ekonomicznej, mierzącej podobieństwo regionów na podstawie wartości analizowanych procesów.

Słowa kluczowe: przestrzeny model quasi-zgodny, autozależności, macierz sąsiedztwa, odległość ekonomiczna.

WSTĘP

Estymacja modeli przestrzennych przeprowadzana jest dla określonych jednostek przestrzennych, takich jak: kraje, regiony, województwa, obszary spisowe itp. Dane, które są zazwyczaj przestrzennie skorelowane, zebrane dla tych jednostek tworzą pewne konfiguracje determinujące kwantyfikację (auto)zależności przestrzennych. Konfiguracje danych zagregowanych upraszczają powiązania mniejszych jednostek przestrzennych, mogą zatem upraszczać opis własności

przestrzennych procesów ekonomicznych¹ oraz zależności między nimi. Jest to stwierdzenie nawiązujące do znanego w statystyce i ekonometrii przestrzennej problemu różnic w zależnościach między procesami, obserwowanych przy przechodzeniu z jednego poziomu agregacji danych do innego (różne współczynniki korelacji, regresji)².

W Szulc i inni (2011) sformułowano i poddano weryfikacji hipotezę, że znalezienie parametru mierzącego rzeczywisty wpływ procesu przestrzennego na inny proces przestrzenny wymaga odkrycia własności i składnikowych struktur poszczególnych procesów. Badając różne, przykładowe struktury procesów wygenerowanych na poziomie danych z określonych podstawowych jednostek przestrzennych, stwierdzono, że mogą one zmieniać się przy przechodzeniu na poziom danych zagregowanych. Jedynie ich odkrycie i uwzględnienie *explicite* w modelu pozwoli prawidłowo ocenić związek między badanymi procesami.

Zmiany, o których mowa, dotyczą w szczególności przestrzennej struktury autoregresyjnej procesu oraz wpływu autokorelacji przestrzennej na jego strukturę trendową przy przechodzeniu na wyższy poziom agregacji danych. Stwierdzono, iż struktura autoregresyjna, w tym siła autozależności, zależy od poziomu agregacji danych. Przejściu na wyższy poziom agregacji towarzyszy spadek współczynnika autoregresji, a także obniżenie jego istotności. Obecność autokorelacji przestrzennej w badanym procesie powoduje zmiany parametrów trendu w modelach szacowanych na podstawie danych zagregowanych w porównaniu z parametrami otrzymanymi na podstawie danych pierwotnych. Parametry otrzymane na poziomie zagregowanym są wyraźnie wyższe. Jednocześnie średnie błędy ocen tych parametrów wzrastają.

Przyjęcie wspomnianej wyżej hipotezy oznacza zalecenie budowy przestrzennego modelu quasi-zgodnego³, w którym dzięki uwzględnieniu *explicite* składników opisujących wewnętrzną strukturę procesów dokonuje się filtracji tych procesów, w wyniku czego pomiar zależności jest dokładniejszy, niezakłócony tzw. wpływami pośrednimi. W konsekwencji, pod warunkiem dokładnej filtracji, współczynniki regresji quasi-zgodnych modeli przestrzennych szacowane na poziomie danych indywidualnych powinny być takie same jak te szacowane na poziomie danych zagregowanych.

¹ Procesy przestrzenne zazwyczaj wykazują przestrzenną autokorelację. Koncepcja autokorelacji przestrzennej oznacza, że obserwacje zjawiska w jednym regionie są systematycznie zależne od obserwacji w regionach sąsiadujących, natomiast zależność maleje dla regionów bardziej odległych.

² Wczesne dyskusje na ten temat, patrz: Arbia (1989) oraz literatura tamże.

³ Procedura konstrukcji takiego modelu została opisana w cytowanym opracowaniu. Wyjaśniono także, dlaczego można mówić jedynie o tzw. quasi-zgodności modelu przestrzennego.

W Szulc i inni (2011) pokazano, iż konstrukcja przestrzennego modelu quasi-zgodnego na podstawie wygenerowanych danych⁴ we wszystkich wypadkach zapewniła otrzymanie prawdziwej wartości współczynnika regresji. Wartości współczynnika regresji otrzymane w modelach quasi-zgodnych po agregacji danych zależały od struktury poszczególnych procesów. Dla procesów o strukturze jedynie trendowej, bez względu na stopnie trendu i ich parametry, wartość współczynnika regresji oscylowała wokół prawdziwej wartości. Natomiast w warunkach obecności autokorelacji w danych wartość współczynnika była po agregacji zawyżona, gdy $\rho_y > \rho_x$, a zaniżona, gdy $\rho_y < \rho_x$ ⁵.

Przestrzenne modele quasi-zgodne dla danych pierwotnych i zagregowanych w warunkach autokorelacji przestrzennej będą się różnić pod względem oceny rzeczywistej zależności między procesami. Wynika to przede wszystkim z faktu, iż macierz powiązań przestrzennych (zwykle jest to macierz konstruowana na podstawie kryterium wspólnej granicy) w modelu zagregowanym nie odzwierciedla powiązań danych przed agregacją. W ten sposób autozależności istniejące na poziomie podstawowym (przed agregacją) nie mogą zostać precyzyjnie opisane na poziomie zagregowanym.

W poszukiwaniu dokładniejszego opisu autozależności przestrzennych za pomocą innych niż standardowe macierzy powiązań, w niniejszym opracowaniu zbadano wpływ wykorzystania macierzy specjalnie zdefiniowanej odległości ekonomicznej na pomiar zależności między procesami przestrzennymi.

1. KWANTYFIKACJA POWIĄZAŃ PRZESTRZENNYCH A POMIAR PRZESTRZENNYCH AUTOZALEŻNOŚCI

Podstawą pomiaru autozależności przestrzennych jest określenie powiązań między jednostkami przestrzennymi (regionami). Przestrzenni sąsiedzi mogą być określani na wiele sposobów.

Kwantyfikacji powiązań przestrzennych dokonuje się za pomocą odpowiedniej macierzy wag w_{ij} . Macierz taką oznacza się przez \mathbf{W} . Przy założeniu, że rozważa się N regionów, macierze \mathbf{W} mają wymiary $N \times N$. Każdy wiersz macierzy \mathbf{W} zawiera elementy niezerowe w kolumnach, które odpowiadają obiektom

⁴ Dane generowane uzyskano jako realizacje zależnych procesów przestrzennych o zadanych strukturach wewnętrznych, w układzie gmin (w cytowanym opracowaniu opisano odpowiednie procedury generowania danych), które następnie zagregowano do poziomu powiatów.

⁵ ρ_y, ρ_x oznaczają odpowiednio współczynniki autoregresji procesu objaśnianego i procesu objaśniającego.

sąsiedzkim, określonym według przyjętego kryterium. Ponadto przyjmuje się, że elementy diagonalne macierzy \mathbf{W} są zerami, co oznacza, że dany obiekt nie jest swoim sąsiadem. Punktem wyjścia w ustalaniu powiązań przestrzennych jest binarna macierz sąsiedztwa \mathbf{S} o elementach:

$$s_{ij} = \begin{cases} 1, & \text{gdy } j \in N(i) \\ 0, & \text{gdy } j \notin N(i) \end{cases} \quad (1)$$

gdzie $N(i)$ oznacza zbiór sąsiadów przestrzennej jednostki i .

Sąsiadów wyznacza się zazwyczaj według kryterium wspólnej granicy. Następnie normuje się wiersze w \mathbf{S} w taki sposób, aby sumy elementów w każdym wierszu były równe jedności (tzw. standaryzacja wierszami do 1). Zatem jeśli

$$\mathbf{D} = \text{diag}\{d_i\}, \text{ gdzie } d_i = \frac{1}{\sum_{j=1}^N s_{ij}}, i = 1, 2, \dots, N, \text{ to } \mathbf{W} = \mathbf{D}\mathbf{S} \text{ i } \sum_{j=1}^N w_{ij} = 1.$$

Wagi w_{ij} ustalane w ten sposób oznaczają, że każdy j -ty sąsiad jednostki i jest traktowany tak samo, a siła jej powiązań z sąsiadami jest tym większa, im mniej ma ona sąsiadów.

Inaczej jest, gdy wagi w_{ij} są funkcjami pewnych własności miejsc siatki, np. długości wspólnej granicy, odległości między centrami regionów lub też innych miar podobieństwa regionów, np. tzw. odległości ekonomicznej⁶ między nimi. Można wskazać różne typy wag w_{ij} w zależności od określonych kryteriów (np. Haining, 2005, s. 83–84).

Oczywiście, tak jak wyżej, dla wszystkich jednostek i oraz j $w_{ij} \geq 0$, gdy $i \neq j$, oraz $w_{ij} = 0$, gdy $i = j$. Takie wagi tworzą uogólnioną macierz sąsiedztwa, której standaryzacja wierszami do jedności odbywa się w drodze przekształcenia jej elementów według wzoru $w_{ij}^* = \frac{w_{ij}}{\sum_{j=1}^N w_{ij}}$.

W niniejszym artykule powiązania między regionami zostaną określone przy zastosowaniu dwóch podejść. Pierwsze z nich to podejście tradycyjne, w którym wykorzystuje się standardową macierz sąsiedztwa z równymi wagami⁷, drugie

⁶ Stwierdzenie, iż w pomiarze zależności przestrzennych należy uwzględnić odległość ekonomiczną między regionami, ma zarówno teoretyczne, jak i praktyczne uzasadnienie.

⁷ W tym zakresie będzie to odwołanie się do wcześniejszych badań (Szulc i inni, 2011), w celu porównania uzyskanych tamże wyników z wynikami otrzymanymi po zastosowaniu pewnej zmodyfikowanej macierzy sąsiedztwa.

zaś polega na uwzględnieniu w macierzy powiązań odległości ekonomicznej, której istotą jest ustalenie podobieństwa regionów na podstawie wartości analizowanego procesu ekonomicznego i ewentualnie procesów ekonomicznych determinujących jego zmienność.

W ogólnej postaci odległość ekonomiczna między regionami i oraz j wyrażona jest następująco (Pietrzak, 2010, s. 75):

$$d_{ij} = \begin{cases} |y_i - y_j|^{\delta_1} + \frac{1}{k_1 + 1} \sum_{z=0}^{k_1} |x_{1i,t-z} - x_{1j,t-z}|^{\delta_2} + \frac{1}{k_2 + 1} \sum_{z=0}^{k_2} |x_{2i,t-z} - x_{2j,t-z}|^{\delta_3} + \\ + \dots + \frac{1}{k_n + 1} \sum_{z=0}^{k_n} |x_{ni,t-z} - x_{nj,t-z}|^{\delta_{n+1}}, & \text{dla } i \neq j \\ 0, & \text{dla } i = j, \end{cases} \quad (2)$$

gdzie: y_i, y_j – wartości badanego ekonomicznego procesu przestrzennego w przestrzennej lokalizacji, odpowiednio i oraz j ; x_{li}, x_{lj} ($l = 1, 2, \dots, n$) – wartości procesów objaśniających, determinujących zmienność procesu objaśnianego w przestrzennych lokalizacjach jw.; k_1, k_2, \dots, k_n – stałe oznaczające wielkość opóźnień w czasie dla rozważanych procesów; $\delta_1, \delta_2, \dots, \delta_{n+1}$ – stałe normalizujące.

Przy tym podejściu elementy macierzy sąsiedztwa równać się będą

$$w_{ij} = \begin{cases} \frac{1}{d_{ij}}, & \text{dla } i \neq j \\ 0, & \text{dla } i = j \end{cases} \quad (3)$$

Ostatecznie, w wyniku standaryzacji wierszami do jedności uzyskuje się macierz sąsiedztwa opartą na odległości ekonomicznej.

Nie oczekuje się, iż efekt przestrzennej agregacji danych w postaci zmiany parametru regresji w modelu przestrzennym zniknie, ponieważ zagregowana macierz sąsiedztwa nadal odnosi się do innej niż przed agregacją konfiguracji danych. Jednak dokładniejszy, bo oparty na kryterium ekonomicznym pomiar autozależności przestrzennych powinien skutkować bardziej precyzyjną filtracją badanych procesów (objaśnianego i objaśniającego), a w konsekwencji prowadzić do prawidłowej oceny zależności regresyjnej. Inaczej mówiąc, im dokładniejszy opis autozależności przestrzennych, tym skuteczniejsza filtracja procesów i pomiar zależności między procesami odbywa się na poziomie danych coraz bardziej oczyszczonych z tych autozależności. Czyste zależności między procesami byłyby takie same bez względu na poziom agregacji danych.

2. PRZYKŁAD EMPIRYCZNY

Zbadano zależność między stopą bezrobocia a nakładami inwestycyjnymi przedsiębiorstw (w zł) w Polsce w 2007 r. w układzie powiatów i podregionów (odpowiednio 379 i 66 jednostek przestrzennych). Dane pochodzą ze strony internetowej GUS (www.gov.stat.pl). Wykorzystano koncepcję przestrzennego modelu quasi-zgodnego. Zbadano struktury trendowo-autoregresyjne poszczególnych procesów na obu poziomach agregacji danych, konstruując odpowiednie modele podstawowe⁸. Do opisu struktury trendowej wykorzystano modele trendu przestrzennego, natomiast struktura autoregresyjna została określona w dwóch wersjach, przy wykorzystaniu (1) standardowej macierzy z równymi wagami – wariant I oraz (2) macierzy odległości ekonomicznej (wagi zróżnicowane) – wariant II. Następnie zbudowano modele quasi-zgodne opisujące zależności między badanymi procesami i porównano otrzymane wyniki.

WARIANT I⁹

Tabele 1–2 przedstawiają wybrane wyniki badania bezrobocia i inwestycji na poziomie powiatów, a tabele 3–4 dotyczą badania tych samych procesów na poziomie podregionów.

Analiza bezrobocia i inwestycji na poziomie powiatów wykazała przestrzenne trendy stopnia 1. oraz przestrzenne autozależności rzędu 1. w obu procesach (tabela 1). Składniki te uwzględniono w konstruowanym dalej quasi-zgodnym modelu przestrzennym. Model ten zawierał nieistotne zmienne, dlatego dokonano ich eliminacji metodą selekcji a posteriori, uzyskując model zredukowany (tabela 2). Model zredukowany nie zawiera trendu oraz przesuniętych przestrzennie inwestycji. Oznacza to, że bezrobocie w danym powiecie nie zależy od inwestycji w powiatach sąsiadujących.

Na poziom bezrobocia w danym powiecie wpływają nakłady inwestycyjne ponoszone w tymże powiecie (każde wydatkowane tysiąc złotych powoduje spadek stopy bezrobocia średnio o 0,9 punktu procentowego) oraz poziom

⁸ Model podstawy oznacza tu model opisujący składnikową strukturę procesu przestrzennego. Rozważaniom towarzyszy założenie, że w ekonomicznych procesach przestrzennych potencjalnie występują przynajmniej dwa składniki: trend przestrzenny oraz powiązania autoregresyjne.

⁹ Szulc i inni, 2011.

bezrobocia w powiatach sąsiadujących (zmiana stopy bezrobocia w danym powiecie o około 0,69 punktu procentowego jest związana z jednoprocetową zmianą stopy bezrobocia w powiatach sąsiadujących).

Tabela 1. Wyniki estymacji i weryfikacji modeli z trendem i autoregresją dla powiatów

Model podstawowy dla bezrobocia: $Y(\mathbf{p}) = \beta_{00} + \beta_{10}p_1 + \beta_{01}p_2 + \rho_y \mathbf{W}Y(\mathbf{p}) + \varepsilon_y(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
β_{00}	1,8861	1,1880	1,5876	0,1124
β_{10}	0,1150	0,1496	0,7686	0,4421
β_{01}	0,3752	0,1625	2,3085	0,0210
$\rho_y = 0,6888$ Test LR: 156; p-value: 0,0000				
Statystyka Walda: 233,59; p-value: 0,0000 AIC: 2277,2 (AIC dla lm: 2431,2)				
Autokorelacja reszt Test LM: 0,0001; p-value: 0,9916				
Model podstawowy dla inwestycji: $X(\mathbf{p}) = \alpha_{00} + \alpha_{10}p_1 + \alpha_{01}p_2 + \rho_x \mathbf{W}X(\mathbf{p}) + \varepsilon_x(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
α_{00}	3403,828	591,209	5,7574	0,0000
α_{10}	-205,305	68,036	-3,0176	0,0025
α_{01}	-118,810	68,716	-1,7290	0,0838
$\rho_x = 0,2026$ Test LR: 6,7374; p-value: 0,0094				
Statystyka Walda: 8,0894; p-value: 0,0045 AIC: 6852,8 (AIC dla lm: 6857,6)				
Autokorelacja reszt Test LM: 1,0482; p-value: 0,3059				

Źródło: obliczenia własne wykonane w programie R-CRAN.

Tabela 2. Wyniki estymacji i weryfikacji modeli quasi-zgodnych dla powiatów

Model pełny: $Y(\mathbf{p}) = \beta_{00}^* + \beta_{10}^* p_1 + \beta_{01}^* p_2 + \rho \mathbf{W}Y(\mathbf{p}) + \gamma X(\mathbf{p}) + \gamma^* \mathbf{W}Y(\mathbf{p}) + \varepsilon(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
β_{00}^*	6,7884	1,5823	4,2901	0,0000
β_{10}^*	-0,1736	0,1538	-1,1288	0,2590
β_{01}^*	0,2807	0,1517	1,8508	0,0642
γ	-0,0009	0,0001	-8,1468	0,0000
γ^*	-0,0002	0,0002	-0,8003	0,4235
$\rho = 0,6456$ Test LR: 133,05; p-value: 0,0000				
Statystyka Walda: 181,2; p-value: 0,0000 AIC: 2216 (AIC dla lm: 2347)				
Autokorelacja reszt Test LM: 0,0744; p-value: 0,7851				
Model zredukowany: $Y(\mathbf{p}) = \beta_{00}^* + \rho \mathbf{W}Y(\mathbf{p}) + \gamma X(\mathbf{p}) + \varepsilon(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
β_{00}^*	6,1623	0,6950	8,8672	0,0000
γ	-0,0009	0,0001	-8,3536	0,0000
$\rho = 0,6851$ Test LR: 182,3; p-value: 0,0000				
Statystyka Walda: 263,92; p-value: 0,0000 AIC: 2215,8 (AIC dla lm: 2396,1)				
Autokorelacja reszt Test LM: 0,7907; p-value: 0,3739				

Źródło: obliczenia własne wykonane w programie R-CRAN.

Bezrobocie i inwestycje na poziomie podregionów zostały opisane za pomocą przestrzennych modeli autoregresyjnych bez trendów (tabela 3). Struktura składnikowa quasi-zgodnego modelu bezrobocia względem inwestycji została uproszczona w wyniku redukcji nieistotnego składnika w modelu pełnym (tabela 4). Redukcji uległy przesunięte przestrzennie inwestycje. Zatem struktura modelu zredukowanego dla podregionów jest analogiczna jak dla powiatów.

Współczynnik mierzący wpływ inwestycji na stopę bezrobocia w podregionie różni się od analogicznego współczynnika ocenianego na poziomie powiatów. Głównym powodem zaobserwowanej różnicy jest autokorelacja przestrzena bezrobocia, a także inwestycji. Wartość współczynnika γ dla podregionów jest większa niż dla powiatów. Wynik ten jest zbieżny ze wskazanymi we wprowadzeniu rezultatami wcześniejszych analiz, uzyskanymi na podstawie danych generowanych. (Gdy $\rho_y > \rho_x$, to współczynnik regresji obliczany na podstawie danych zagregowanych jest zawyżony). Z kolei współczynnik autoregresji ρ mierzący związki między stopami bezrobocia na obszarach sąsiadujących na poziomie podregionów jest wyraźnie mniejszy niż analogiczny parametr obliczony dla powiatów.

Tabela 3. Wyniki estymacji i weryfikacji modeli autoregresyjnych dla podregionów

Model podstawowy dla bezrobocia: $Y(\mathbf{p}) = \beta_{00} + \rho_y \mathbf{W}Y(\mathbf{p}) + \varepsilon_y(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
β_{00}	5,6762	1,5715	3,6120	0,0003
$\rho_y = 0,5145$ Test LR: 10,799; p-value: 0,0010				
Statystyka Walda: 16,848; p-value: 0,0000 AIC: 392,07 (AIC dla lm: 400,86)				
Autokorelacja reszt Test LM: 4,002; p-value: 0,045				
Model podstawowy dla inwestycji: $X(\mathbf{p}) = \alpha_{00} + \rho_x \mathbf{W}X(\mathbf{p}) + \varepsilon_x(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
α_{00}	1889,74	476,03	3,9698	0,0000
$\rho_x = 0,3532$ Test LR: 3,8826; p-value: 0,0488				
Statystyka Walda: 5,8179; p-value: 0,0159 AIC: 1176,9 (AIC dla lm: 1178,7)				
Autokorelacja reszt Test LM: 1,3882; p-value: 0,2387				

Źródło: obliczenia własne wykonane w programie R-CRAN.

Tabela 4. Wyniki estymacji i weryfikacji modeli quasi-zgodnych dla podregionów

Model pełny: $Y(\mathbf{p}) = \beta_{00}^* + \rho \mathbf{W}Y(\mathbf{p}) + \gamma X(\mathbf{p}) + \gamma^* \mathbf{W}Y(\mathbf{p}) + \varepsilon(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
β_{00}^*	14,8476	3,1771	4,6733	0,0000
γ	-0,0014	0,0003	-5,2884	0,0000
γ^*	-0,0008	0,0007	-1,1823	0,2371
$\rho = 0,2635$ Test LR: 2,6541; p-value: 0,1033				
Statystyka Walda: 2,957; p-value: 0,0855 AIC: 367,91 (AIC dla lm: 368,56)				
Autokorelacja reszt Test LM: 2,0835; p-value: 0,1489				
Model zredukowany: $Y(\mathbf{p}) = \beta_{00}^* + \rho \mathbf{W}Y(\mathbf{p}) + \gamma X(\mathbf{p}) + \varepsilon(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
β_{00}^*	12,1359	2,0108	6,0353	0,0000
γ	-0,0015	0,0003	-5,8125	0,0000
$\rho = 0,3369$ Test LR: 5,1063; p-value: 0,0024				
Statystyka Walda: 6,2839; p-value: 0,0122 AIC: 367,43 (AIC dla lm: 370,54)				
Autokorelacja reszt Test LM: 0,5196; p-value: 0,47099				

Źródło: obliczenia własne wykonane w programie R-CRAN.

WARIANT II

Analogicznie jak w wariancie I, tabele 5–6 oraz 7–8 przedstawiają wyniki badania bezrobocia i inwestycji odpowiednio na poziomie powiatów oraz podregionów.

Zarówno bezrobocie, jak i inwestycje analizowane na poziomie powiatów wykazują przestrzenne trendy stopnia 1. oraz przestrzenne autozależności rzędu 1. Identyfikacji autozależności dokonano przy wykorzystaniu macierzy powiązań ze zróżnicowanymi wagami, uwzględniającymi podobieństwo powiatów określone na podstawie wartości bezrobocia oraz inwestycji w powiązanych powiatach¹⁰. Zatem odległość ekonomiczna została określona następująco:

¹⁰ W ten sam sposób określono podobieństwo podregionów.

$$d_{ij} = \begin{cases} |y_i - y_j|^{\delta_1} + |x_i - x_j|^{\delta_2}, & \text{dla } i \neq j \\ 0, & \text{dla } i = j, \end{cases} \quad (4)$$

gdzie $\delta_1 = \delta_2 = 0,5$.

Tabela 5. Wyniki estymacji i weryfikacji modeli z trendem i autoregresją dla powiatów

Model podstawowy dla bezrobocia: $Y(\mathbf{p}) = \beta_{00} + \beta_{10}p_1 + \beta_{01}p_2 + \rho_y \mathbf{W}^*Y(\mathbf{p}) + \varepsilon_y(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
β_{00}	0,9597	1,0408	0,9221	0,3565
β_{10}	0,1453	0,1342	1,0825	0,2790
β_{01}	0,2675	0,1429	1,8716	0,0613
$r_y = 0,7707$ Test LR: 226,25; p-value: 0,0000				
Statystyka Walda: 442,78; p-value: 0,0000 AIC: 2207 (AIC dla lm: 2431,2)				
Autokorelacja reszt Test LM: 7,5665; p-value: 0,00595				
Model podstawowy dla inwestycji: $X(\mathbf{p}) = \alpha_{00} + \alpha_{10}p_1 + \alpha_{01}p_2 + \rho_x \mathbf{W}^*X(\mathbf{p}) + \varepsilon_x(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
α_{00}	3018,171	576,401	5,2362	0,0000
α_{10}	-174,162	66,466	-2,6203	0,0088
α_{01}	-114,086	67,543	-1,6891	0,0912
$\rho_x = 0,3277$ Test LR: 15,207; p-value: 0,0000				
Statystyka Walda: 25,201; p-value: 0,0000 AIC: 6844,3 (AIC dla lm: 6857,6)				
Autokorelacja reszt Test LM: 25,0456; p-value: 0,0000				

Źródło: obliczenia własne wykonane w programie R-CRAN.

W tym wariancie analizy struktura badanych procesów nie różniła się od struktury ustalonej w wariancie I. Również analogiczne okazały się struktury składnikowe pełnego oraz zredukowanego quasi-zgodnego modelu przestrzennego bezrobocia względem inwestycji. Różnice natomiast dotyczą wartości parametrów konstruowanych modeli.

Tabela 6. Wyniki estymacji i weryfikacji modeli quasi-zgodnych dla powiatów

Model pełny: $Y(\mathbf{p}) = \beta_{00}^* + \beta_{10}^* p_1 + \beta_{01}^* p_2 + \rho \mathbf{W}^* Y(\mathbf{p}) + \gamma X(\mathbf{p}) + \gamma^* \mathbf{W}^* Y(\mathbf{p}) + \varepsilon(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
β_{00}^*	5,6816	1,4559	3,9024	0,0001
β_{10}^*	-0,1369	0,1413	-0,9690	0,3326
β_{01}^*	0,1914	0,1334	1,4347	0,1514
γ	-0,0008	0,0001	-7,6444	0,0000
γ^*	-0,0003	0,0003	-1,0607	0,2888
$\rho = 0,7244$ Test LR: 118,48; p-value: 0,0000				
Statystyka Walda: 321,25; p-value: 0,0000 AIC: 2147 (AIC dla lm: 2333,5)				
Autokorelacja reszt Test LM: 5,2397; p-value: 0,0221				
Model zredukowany: $Y(\mathbf{p}) = \beta_{00}^* + \rho \mathbf{W}^* Y(\mathbf{p}) + \gamma X(\mathbf{p}) + \varepsilon(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
β_{00}^*	4,8470	0,5992	8,0890	0,0000
γ	-0,0008	0,0001	-8,3243	0,0000
$\rho = 0,7567$ Test LR: 253,3; p-value: 0,0000				
Statystyka Walda: 446,03; p-value: 0,0000 AIC: 2144,8 (AIC dla lm: 2396,1)				
Autokorelacja reszt Test LM: 1,669; p-value: 0,1964				

Źródło: obliczenia własne wykonane w programie R-CRAN.

Bezrobocie i inwestycje na poziomie podregionów zostały opisane za pomocą przestrzennych modeli autoregresyjnych, definiowanych – podobnie jak w przypadku analizy powiatów – w warunkach nowej macierzy powiązań \mathbf{W}^* . Inne określenie macierzy powiązań przestrzennych nie spowodowało zmiany struktury modeli podstawowych badanych procesów, zmieniły się jedynie wartości parametrów tych modeli (por. tabele 3 i 7).

Tabela 7. Wyniki estymacji i weryfikacji modeli autoregresyjnych dla podregionów

Model podstawowy dla bezrobocia: $Y(\mathbf{p}) = \beta_{00} + \rho_y \mathbf{W}^* Y(\mathbf{p}) + \varepsilon_y(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
β_{00}	4,0871	1,2556	3,2552	0,0011
$\rho_y = 0,6437$ Test LR: 19,848; p-value: 0,0000				
Statystyka Walda: 40,705; p-value: 0,0000 AIC: 383,02 (AIC dla lm: 400,86)				
Autokorelacja reszt Test LM: 9,5614; p-value: 0,0200				
Model podstawowy dla inwestycji: $X(\mathbf{p}) = \alpha_{00} + \rho_x \mathbf{W} X(\mathbf{p}) + \varepsilon_x(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
α_{00}	1833,62	452,55	4,0518	0,0001
$\rho_x = 0,3748$ Test LR: 4,9755; p-value: 0,0257				
Statystyka Walda: 7,45; p-value: 0,0063 AIC: 1175,8 (AIC dla lm: 1178,7)				
Autokorelacja reszt Test LM: 1,077; p-value: 0,2994				

Źródło: obliczenia własne wykonane w programie R-CRAN.

Tabela 8 prezentuje wyniki oszacowania i weryfikacji pełnego, a następnie zredukowanego modelu quasi-zgodnego na poziomie podregionów.

Współczynnik mierzący wpływ inwestycji na stopę bezrobocia w podregionie nadal różni się od analogicznego współczynnika ocenianego na poziomie powiatów. Jednak różnica reakcji stopy bezrobocia zmniejszyła się o 0,1 punktu procentowego na każde 1000 zł inwestycji. Odnotowano również mniejszą różnicę wartości współczynników autoregresji w końcowych modelach ocenianych na poziomie powiatów i podregionów.

Tabela 8. Wyniki estymacji i weryfikacji modeli quasi-zgodnych dla podregionów

Model pełny: $Y(\mathbf{p}) = \beta_{00}^* + \rho \mathbf{W}^* Y(\mathbf{p}) + \gamma X(\mathbf{p}) + \gamma^* \mathbf{W}^* X(\mathbf{p}) + \varepsilon(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
β_{00}^*	12,3117	2,7234	4,5207	0,0000
γ	-0,0012	0,0003	-4,9238	0,0000
γ^*	-0,0008	0,0006	-1,4082	0,1591
$\rho = 0,4291$ Test LR: 8,4641; p-value: 0,0036				
Statystyka Walda: 11,242; p-value: 0,0008 AIC: 359,33 (AIC dla lm: 365,8)				
Autokorelacja reszt Test LM: 7,4793; p-value: 0,0062				
Model zredukowany: $Y(\mathbf{p}) = \beta_{00}^* + \rho \mathbf{W}^* Y(\mathbf{p}) + \gamma X(\mathbf{p}) + \varepsilon(\mathbf{p})$				
Parametry	Szacunki parametrów	Błędy standardowe	Statystyki z	Pr (> z)
β_{00}^*	9,6631	1,7254	5,6005	0,0000
γ	-0,0013	0,0002	-5,6001	0,0000
$\rho = 0,4987$ Test LR: 12,965; p-value: 0,0003				
Statystyka Walda: 20,214; p-value: 0,0001 AIC: 359,57 (AIC dla lm: 370,54)				
Autokorelacja reszt Test LM: 0,0208; p-value: 0,8853				

Źródło: obliczenia własne wykonane w programie R-CRAN.

PODSUMOWANIE

Zmiany własności i struktur procesów przestrzennych zachodzące przy agregacji danych mają wpływ na określanie zależności między procesami. Quasi-zgodne modele przestrzenne nie muszą zapewnić odkrycia faktycznej zależności między badanymi procesami. Jednak dokładniejszy opis autozależności w tych procesach prowadzi do łągodzenia efektu zmiany parametrów regresji spowodowanego agregacją przestrzenną.

Modele konstruowane przy wykorzystaniu macierzy odległości ekonomicznej są lepiej dopasowane do danych niż modele, w których zastosowano macierz standardową. Można to stwierdzić, porównując wartości odpowiednich statystyk

testowych (tabele 2 i 6 oraz 4 i 8). Również wartości współczynnika „pseudo- R^2 ” są wyższe w wariancie II niż w wariancie I analizy (w zredukowanych modelach quasi-zgodnych dla powiatów wynoszą one odpowiednio 0,63 oraz 0,53; podobnie dla podregionów – 0,54 oraz 0,46).

Przy zastosowaniu macierzy odległości ekonomicznej różnica między współczynnikami regresji w modelach dla danych przed i po agregacji się zmniejszyła. Celem kolejnych badań będzie sprawdzenie, czy efekt ten zostanie potwierdzony dla innych lat.

LITERATURA

- Arbia G. (1988), *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*, Kluwer Academic Press, Dordrecht.
- Haining R. (2005), *Spatial Data Analysis. Theory and Practice*, Cambridge University Press, 3th ed., Cambridge.
- Pietrzak M. B. (2010), *Dwuetałowa procedura budowy przestrzennej macierzy wag z uwzględnieniem odległości ekonomicznej*, „Oeconomia Copernicana”, nr 1, 65–78.
- Szulc E., Müller-Frączek I., Pietrzak M. B. (2011), *Modelowanie zależności między ekonomicznymi procesami przestrzennymi a poziom agregacji danych*, [w:] Suchecka J. (red.), *Ekonometria przestrzenna i regionalne analizy ekonomiczne*, „Folia Oeconomica”, nr 253, 327–344.

ECONOMETRIC IDENTIFICATION OF THE STRUCTURES OF SPATIAL PROCESSES AND A PROBLEM OF DATA AGGREGATION

A b s t r a c t. The paper concerns the measurement of the dependence between economic spatial processes at various levels of data aggregation. The considerations refer to the investigations confirming efficiency of the so-called quasi-congruent spatial model as a tool of the measurement of the dependence between economic processes, on condition that the model description of the structures of the processes is correct and sufficient. The main problem of the paper is to discuss such a description of the spatial connections, which would be adequate to express the autodependence of the investigated processes. The application of the economic distance, which characterizes the similarity of the regions on the ground of the values of the analyzed processes, is proposed.

Key words: quasi-congruent spatial model, autodependence, connectivity matrix, economic distance.

