

*Edyta Łaszkiwicz**

SKUTKI BŁĘDNEJ SPECYFIKACJI EFEKTÓW
PRZESTRZENNYCH W BAYESOWSKIM MODELU
AUTOREGRESJI PRZESTRZENNEJ.
WYNIKI SYMULACJI MONTE CARLO**

Z a r y s t r e ś c i. Celem artykułu jest identyfikacja skutków pominięcia dodatkowych charakterystyk procesu przestrzennego w modelu SAR. W badaniu skupiono się na dwóch procesach przestrzennych – cechujących się występowaniem skorelowanych przestrzennie i niezależnych efektów losowych. Wyniki symulacji Monte Carlo wykazały m.in., iż pominięcie efektów losowych w modelu SAR skutkuje przeszacowaniem parametru interakcji przestrzennych oraz wariancji składnika losowego.

S ł o w a k l u c z o w e: model przestrzennej autoregresji, efekty przestrzenne, MCMC, symulacja Monte Carlo.

K l a s y f i k a c j a J E L: C11, C21, C63.

WSTĘP

Jednym z podstawowych modeli ekonometrii przestrzennej jest przestrzenny model autoregresji (SAR), będący rozszerzeniem tradycyjnego modelu regresji liniowej o opóźnienie przestrzenne zmiennej objaśnianej (por. Szulc, 2007; Suchecki, 2010). Mimo iż pozwala on analizować procesy odznaczające się autoregresją przestrzenną, nie pozwala ustosunkować się do

* Adres do korespondencji: Edyta Łaszkiwicz, Uniwersytet Łódzki, Katedra Ekonometrii Przestrzennej, ul. Rewolucji 1905 r. 37, 90–214 Łódź, e-mail: elaszkiewicz@uni.lodz.pl.

** Praca współfinansowana z grantu Narodowego Centrum Nauki numer UMO-2013/08/T/HS4/00512.

ewentualnego występowania przestrzennej heterogeniczności procesu, jak również nie zezwala na występowanie bardziej złożonych, wielowymiarowych struktur zależności. Mając na względzie to, że zjawiska ekonomiczne i społeczne mogą cechować się skomplikowaną strukturą relacji w przestrzeni, jak również mając na uwadze popularność modelu SAR, istotne wydaje się zidentyfikowanie efektów niewłaściwej specyfikacji ww. modelu. W szczególności ważne, z punktu widzenia interpretacji efektów przestrzennych, będą skutki błędnego scharakteryzowania procesu przestrzennego, w postaci błędu estymatora parametru interakcji przestrzennych.

Pośród prac empirycznych podejmujących ww. zagadnienia wskazać można przede wszystkim te, które poświęcono skutkom pominięcia dodatkowej niejednorodności zjawiska w modelach wielopoziomowych¹ (por. Moerbeek, 2004; Van den Noortgate i in., 2005) oraz te, wskazujące konsekwencje pominięcia przestrzennej autoregresji procesu w tradycyjnych modelach regresji liniowej (por. Anselin, 1988; Anselin i Griffith, 1988). W przypadku modeli wielopoziomowych wiadomo, że zignorowanie dodatkowego „poziomu” (inaczej heterogeniczności) danych przekłada się na przeszacowanie wariancji składnika losowego oraz wariancji efektów losowych. Z uwagi na brak interakcji przestrzennych w tradycyjnych modelach wielopoziomowych, potencjalny wpływ pominięcia heterogeniczności procesu na oszacowania parametrów interakcji przestrzennych nie jest jednak znany.

Wśród prac poświęconych ekonometrycznym modelom przestrzennym wskazać można m.in. takie, które odnoszą się do problemu wpływu pominiętej zmiennej na wybór modelu przestrzennego (por. Lacombe i LeSage, 2013), czy też takie, które podejmują kwestie związane z wielkością błędu modelu przestrzennego wynikającego z niewłaściwej specyfikacji zależności przestrzennych w postaci macierzy wag (por. Liu i in., 2014). Na tym tle problematyka zawarta w niniejszej pracy wydaje się tematem rzadko podejmowanym w dyskusji.

Badanie ma na celu określenie skutków błędnej specyfikacji modelu SAR. W praktyce rzeczywisty charakter procesu generującego dane nie jest znany, a badacz dokonuje wyboru modelu w oparciu o kryteria prognostyczne lub stosując miary dobroci dopasowania modelu do danych. Mimo tego, jeśli wśród porównywanych modeli nie znajdzie się taki, który uwzględnia heterogeniczność procesu, istnieje szansa wyboru modelu SAR. Dotyczy to w szczególności tych sytuacji, kiedy mamy do czynienia z występowaniem zarówno autokorelacji, jak i heterogeniczności przestrzennej.

¹ Definicję modelu wielopoziomowego znaleźć można m.in. u Goldsteina (2011).

Błądną specyfikację modelu zdefiniowano w niniejszej pracy jako sytuację, w której proces przestrzenny o złożonej strukturze efektów przestrzennych jest traktowany przez badacza jako proces cechujący się wyłącznie autoregresją przestrzenną. W efekcie estymacja parametrów procesu odbywa się przy wykorzystaniu modelu SAR, który nie uwzględnia dodatkowych charakterystyk procesu.

W pracy zostały rozpatrzone dwa procesy przestrzenne, odznaczające się: 1) występowaniem przestrzennej heterogeniczności (w postaci niezależnych efektów losowych); 2) dodatkową strukturą zależności przestrzennych i heterogenicznością (w postaci skorelowanych przestrzennie efektów losowych). W celu określenia efektów niewłaściwej specyfikacji modelu SAR posłużono się wynikami symulacji Monte Carlo.

1. BAYESOWSKA ESTYMACJA MODELU AUTOREGRESJI PRZESTRZENNEJ

W modelu SAR zakłada się istnienie powiązań między wartościami procesu objaśnianego (w określonej przestrzeni), które uwzględniane są jako tzw. opóźnienie przestrzenne zmiennej zależnej. Model ma następującą postać ogólną (Anselin, 1988):

$$\mathbf{Y} = \rho \mathbf{WY} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \sigma_{\varepsilon}^2 \mathbf{I}_N), \quad (1)$$

gdzie:

\mathbf{Y} – wektor $N \times 1$ objaśnianego procesu przestrzennego,

\mathbf{X} – macierz $N \times K$ procesów objaśniających,

\mathbf{W} – macierz $N \times N$ wag przestrzennych,

$\boldsymbol{\beta}$ – wektor $K \times 1$ parametrów strukturalnych,

ρ – parametr interakcji przestrzennych,

$\boldsymbol{\varepsilon}$ – wektor $N \times 1$ składnika losowego,

N – łączna liczba obserwacji w próbie.

Dla równania 1 estymator KMNK jest niezgodny (Lee, 2002). Pożądane własności estymatora można uzyskać, stosując metodę największej wiarygodności zmiennych instrumentalnych (por. Anselin, 1988), uogólnioną metodę momentów, czy też dwustopniową metodę najmniejszych kwadratów (por. Lee, 2007). Alternatywą dla powyższych jest zastosowanie podejścia bayesowskiego, rekomendowanego w przypadku estymacji bardziej zaawansowanych modeli przestrzennych (por. m.in. LeSage, 1997). Jedną z niewątpliwych zalet ww. podejścia jest możliwość uzyskania pełnego rozkładu *a posteriori* parametrów, co eliminuje konieczność asymptotycznych aproksymacji (wrażliwych np. w sytuacji dysponowania małymi próbami). Mając

na względzie powyższe, w niniejszym badaniu dokonano estymacji bayesowskiej modelu SAR.

Punktem wyjścia estymacji bayesowskiej (por. Osiewalski, 2001) jest przyjęcie założenia, że rozkłady *a posteriori* parametrów $\mathbf{E} = \{\lambda, \rho, \sigma_\varepsilon^2, \sigma_\mu^2, \boldsymbol{\beta}\}$ odzwierciedlają zarówno wiedzę *a priori* badacza $P(\mathbf{E})$, jak i informacje „z danych” $Dane = \{\mathbf{Y}, \mathbf{X}, \mathbf{W}\}$ wyrażone funkcją wiarygodności $P(Dane|\mathbf{E})$, co zapisać można następująco:

$$P(\mathbf{E} | Dane) \propto P(Dane | \mathbf{E}) \times P(\mathbf{E}). \quad (2)$$

Z uwagi na brak oczekiwań co do wartości parametrów zastosowano słabo informacyjne rozkłady *a priori* parametrów $P(\mathbf{E})$ następującej postaci:

$$\begin{aligned} P(\boldsymbol{\beta}) &\sim N(\mathbf{M}_0, \mathbf{T}_0), & P(\rho) &\sim U(1/v_{\min}, 1), \\ P(\sigma_\varepsilon^2) &\sim IG(c_0, d_0), \end{aligned} \quad (3)$$

gdzie:

v_{\min} – najmniejsza wartość własna macierzy \mathbf{W} ,

c_0, d_0 – parametr kształtu i skali odwrotnego rozkładu gamma,

\mathbf{M}_0 – wektor $K \times 1$ średnich wartości przestrzennych procesów objaśniających,

\mathbf{T}_0 – macierz $K \times K$ wariancji-kowariancji przestrzennych procesów objaśniających.

W celu wyznaczenia rozkładów *a posteriori* parametrów skorzystano z funkcji wiarygodności, która dla modelu SAR ma postać (LeSage i in., 2009, s. 47):

$$L(\mathbf{Y} | \mathbf{E}) = (2\sigma_\varepsilon^2)^{-N/2} |\mathbf{I}_N - \rho\mathbf{W}| \exp\left(-\frac{\mathbf{e}'\mathbf{e}}{2\sigma_\varepsilon^2}\right), \quad (4)$$

$$\mathbf{e} = \mathbf{Y} - \rho\mathbf{W}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}.$$

Wiedząc, że funkcję wiarygodności utożsamiamy z $P(Dane|\mathbf{E})$, rozkłady *a posteriori* poszczególnych parametrów wyznaczyć można następująco:

$$P(\boldsymbol{\beta} | \mathbf{Y}, \rho, \sigma_\varepsilon^2) \sim N(\mathbf{M}_\beta, \mathbf{T}_\beta), \quad (5)$$

$$\text{gdzie: } \mathbf{M}_\beta = \mathbf{T}_\beta \left[\frac{1}{\sigma_\varepsilon^2} \mathbf{X}'(\mathbf{Y} - \rho\mathbf{W}\mathbf{Y}) + \mathbf{T}_0^{-1}\mathbf{M}_0 \right], \quad \mathbf{T}_\beta = \left[\frac{1}{\sigma_\varepsilon^2} \mathbf{X}'\mathbf{X} + \mathbf{T}_0^{-1} \right]^{-1}.$$

$$P(\sigma_\varepsilon^2 | \mathbf{Y}, \boldsymbol{\beta}, \rho) \sim IG(c_e, d_e), \quad (6)$$

gdzie: $c_e = N/2 + c_0$, $d_e = 0,5 \times \mathbf{e}'\mathbf{e} + d_0$.

$$P(\rho | \mathbf{Y}, \boldsymbol{\beta}, \sigma_\varepsilon^2) \propto | \mathbf{I}_N - \rho \mathbf{W} | \sigma_\varepsilon^{-(N+1)} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} (\mathbf{e}'\mathbf{e}) \right\}. \quad (7)$$

Obliczenie charakterystyk rozkładów *a posteriori*, wyrażonych równaniami 5–7, umożliwiło wykorzystanie metod Monte Carlo typu łańcuchów Markowa (MCMC). Dla parametrów $\boldsymbol{\beta}$ i σ_ε^2 zastosowano losowanie Gibbsa, będące jedną z częściej wykorzystywanych metod numerycznych stosowanych w badaniach empirycznych. Polega ono na wielokrotnym generowaniu w sposób sekwencyjny liczb pseudolosowych z przedstawionych rozkładów *a posteriori*.

W przeciwieństwie do dwóch pierwszych parametrów, rozkład *a posteriori* parametru ρ nie odzwierciedla żadnego ze znanych rozkładów. W związku z tym, w celu wyznaczenia kolejnych wartości z rozkładu *a posteriori* parametru ρ , zastosowano metodę odwrotnej dystrybuanty (generowanie liczb losowych z odwracaniem dystrybuanty). W każdej iteracji wyznaczono skumulowany rozkład ρ , dokonując numerycznej integracji rozkładu gęstości (por. m.in. Pace i in., 1997, LeSage i in., 2009), a następnie losowano wartość ρ .

Zgodnie z metodą MCMC w każdej kolejnej iteracji wykorzystywano wartości parametrów z poprzedniej iteracji, uzyskując tzw. łańcuch Markowa. Stosowany łańcuch Markowa był zbieżny do rozkładu stacjonarnego przy łącznej liczbie iteracji ustalonej na 10 000, przyjmując 5000 jako liczbę cykli spalonych².

2. ZAŁOŻENIA EKSPERYMENTU SYMULACYJNEGO

Symulację przeprowadzono oddzielnie dla dwóch procesów przestrzennych, cechujących się autokorelacją przestrzenną i/lub przestrzenną heterogenicznością. Rozpatrzono następujące przypadki:

$$\mathbf{Y} = (\mathbf{I}_N - \rho \mathbf{W})^{-1} (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} + \boldsymbol{\Lambda}(\mathbf{I}_J - \lambda \mathbf{M})^{-1} \boldsymbol{\mu}), \quad (8)$$

$$\mathbf{Y} = (\mathbf{I}_N - \rho \mathbf{W})^{-1} (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} + \boldsymbol{\Lambda}\boldsymbol{\mu}), \quad (9)$$

$$\boldsymbol{\mu} \sim N(\mathbf{0}, \sigma_\mu^2 \mathbf{I}_J), \quad (10)$$

gdzie:

- \mathbf{M} – standaryzowana wierszami macierz $J \times J$ wag przestrzennych,
- ρ, λ – parametry interakcji przestrzennych,
- $\boldsymbol{\mu}$ – wektor $J \times 1$ efektów losowych,

² Badanie zbieżności łańcuchów Markowa przeprowadzono w programie R Cran, korzystając z pakietu *coda* (Plummer i in., 2005).

J – łączna liczba efektów losowych w próbie.

Przez Δ oznaczono macierz blokowo-diagonalną o wymiarach $N \times J$ następującej postaci:

$$\Delta = \begin{bmatrix} \mathbf{1}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_J \end{bmatrix}, \quad (11)$$

gdzie:

$\mathbf{0}$ – wektor $n_j \times 1$ zer,

$\mathbf{1}_j$ – wektor $n_j \times 1$ jedynek,

n_j – liczba jednostek przestrzennych w grupie j .

Jak wskazuje Szulc (2007, s. 47), procesy przestrzenne stanowią funkcje $Y(s)$ argumentów $s = [s_1, s_2] \in R^2$. Można zatem przyjąć, że procesy wyrażone równaniami 8–10 są procesami $Y(s_i)$, tj. obserwowanymi na płaszczyźnie, w lokalizacjach przestrzennych $s_i = [s_{1i}, s_{2i}]$, gdzie $i = 1, 2, \dots, N$ jest numerem jednostki przestrzennej.

Na potrzeby wyznaczenia elementów w_{ij} macierzy wag przestrzennych \mathbf{W} oraz \mathbf{M} wygenerowano regularną siatkę punktów (grid) z lokalizacjami przestrzennymi $N=900$ jednostek³. Dodatkowo wydzielono $J=100$ jednakowych grup, zawierających po $n_j=9$ jednostek przestrzennych. Każdą grupę j traktować można jako wyższy stopień agregacji jednostek i . Dla tak wyznaczonych lokalizacji obliczono odległości d_{ij} między każdą parą punktów i oraz j , stosując metrykę euklidesową. W przypadku odległości między J grupami w kalkulacji odległości wykorzystano informacje o położeniu centroidu grupy. Dla regularnej siatki punktów lokalizację centroidu grupy j można obliczyć jako $s_{(29+j \times 3) \times q(j/10-0,1)}$, gdzie q jest częścią całkowitą dzielenia.

W poszczególnych macierzach wag przestrzennych zastosowano następujące funkcje ważenia odległości:

– dla macierzy \mathbf{W} :

$$\forall i, j \in \{1, \dots, N\} \wedge i \neq j : w_{ij} = \exp(-d_{ij}^2), \quad (12)$$

– dla macierzy \mathbf{M} :

$$\begin{aligned} \forall i, j, l \in \{1, \dots, J\} : w_{ij} = 1 &\Leftrightarrow \forall j \in KNN(i, J), \\ \forall l \in J - KNN(i, J), d_{ij} &\leq d_{il}, \end{aligned} \quad (13)$$

³ W pracy zrezygnowano z wprowadzenia warunkowania macierzy \mathbf{W} i \mathbf{M} , jak również z analizy ogólniejszych procesów, np. nieskorelowanych przestrzennie.

gdzie: $KNN(i, J)$ jest zbiorem $k=1$ obserwacji pochodzących ze zbioru J grup, będących najbliższym sąsiadem dla i .

Tabela 1. Wartości docelowe parametrów

Warianty dla procesu 1	Warianty dla procesu 2	ρ	λ	σ_{μ}^2
1	1	0,20	0,20	0,20
2		0,20	0,40	0,20
3		0,20	0,60	0,20
4		0,20	0,80	0,20
	2	0,20		0,40
	3	0,20		0,60
	4	0,20		0,80
	5	0,40		0,20
5	6	0,40	0,20	0,40
6		0,40	0,40	0,40
7		0,40	0,60	0,40
8		0,40	0,80	0,40
	7	0,40		0,60
	8	0,40		0,80
	9	0,60		0,20
	10	0,60		0,40
9	11	0,60	0,20	0,60
10		0,60	0,40	0,60
11		0,60	0,60	0,60
12		0,60	0,80	0,60
	12	0,60		0,80
	13	0,80		0,20
	14	0,80		0,40
	15	0,80		0,60
13	16	0,80	0,20	0,80
14		0,80	0,40	0,80
15		0,80	0,60	0,80
16		0,80	0,80	0,80

Źródło: opracowanie własne.

Symulację przeprowadzono dla 16 kombinacji wartości parametrów ρ , λ i σ_{μ}^2 (por. tabela 1). Dla procesu wyrażonego równaniem 8 zastosowano kombinację wartości parametrów ρ , λ oraz σ_{μ}^2 w celu określenia skali błędów ocen parametrów wynikających z różnej siły interakcji przestrzennych na dwóch poziomach agregacji przestrzennej oraz skali przestrzennej heterogeniczności. W przypadku procesu wyrażonego równaniem 9 w poszczególnych wariantach różnicowano wartości parametrów ρ oraz σ_{μ}^2 , podobnie jak w równaniu 8 kontrolując siłę oddziaływań przestrzennych i skalę niejednorodności, przy czym założono brak dodatkowych interakcji na wyższym poziomie agregacji przestrzennej.

Dla wszystkich procesów przyjęto stałą wartość wariancji składnika losowego $\sigma_e^2=0,2$ oraz wartości parametrów strukturalnych $\beta=[0,3, 0,7]$. Liczbę objaśniających procesów przestrzennych ustalono na $K=2$. Wygenerowane zostały niezależnie od siebie, przy czym wartości obu pochodziły z rozkładu normalnego o zerowej wartości średniej i wariancji równej jeden.

Dla każdej kombinacji wartości docelowych parametrów, w poszczególnych procesach, ustalono $R=50$ replikacji. Niewielka, w porównaniu z tradycyjnie przyjmowaną w symulacjach Monte Carlo, liczba replikacji wynikała z długiego czasu estymacji modelu SAR metodami bayesowskimi⁴. Łączy liczbą replikacji dla każdego procesu wynosiła 800.

W badaniu wykorzystano dwie najpowszechniej stosowane miary, pozwalające na ocenę wielkości błędów ocen parametrów i błędów standardowych tychże ocen:

- względny błąd estymatora \hat{E} parametru E :

$$B(\hat{E}) = \frac{\bar{\hat{E}} - E}{E} \cdot 100\%, \quad (14)$$

gdzie: $\bar{\hat{E}}$ – średnia arytmetyczna obliczona z R replikacji na podstawie wartości oczekiwanej rozkładu *a posteriori* parametru E (por. Domański i in., 2000)⁵.

Analogicznie jak Hoogland i in. (1998, s. 329–367) przyjęto maksymalną dopuszczalną wartość błędu 5%.

- wskaźnik pokrycia przedziału z 95% wiarygodnością:

$$C(CI(\hat{E})) = \frac{\sum CI(\hat{E}_r)}{R} \cdot 100\%, \quad (15)$$

$$CI(\hat{E}_r) = \begin{cases} 1 & \text{gdy } E \in CI \\ 0 & \text{gdy } E \notin CI \end{cases}$$

gdzie: CI – 95% przedział wartości parametru E , którego dolną granicę obliczono jako kwantyl rzędu 0,025 rozkładu *a posteriori* parametru, a górną jako kwantyl rzędu 0,975. W celu stwierdzenia, czy wartości względnego błędu estymatora różnią się w zależności od wartości docelowej parametru, przeprowadzono analizę ANOVA.

⁴ W efekcie skala dokładności oszacowań względnych błędów estymatora (pow. równanie 14) jest niższa od tej, jaką uzyskano by, stosując większą liczbę replikacji.

⁵ Rozkłady brzegowe *a posteriori* wszystkich rozważanych parametrów były symetryczne, jednomodalne. W takim przypadku 95% przedziały wartości parametrów pokrywają się z przedziałami o najwyższych wartościach funkcji gęstości *a posteriori* (HPD).

3. WYNIKI SYMULACJI MONTE CARLO

Wyniki symulacji zaprezentowano w podrozdziałach odpowiadających rozważanym procesom przestrzennym. W przypadku estymacji modelu SAR dla procesu opisanego równaniem 9 uzyskane wyniki interpretować można w kontekście efektu pominięcia przestrzennej heterogeniczności (podrozdział 3.1). Wyniki estymacji procesu opisanego równaniem (8) pozwalają określić wpływ pominięcia przestrzennej heterogeniczności oraz obecności dodatkowych interakcji przestrzennych dla wyższego stopnia agregacji (podrozdział 3.2).

3.1. EFEKT POMINIĘCIA PRZESTRZENNEJ HETEROGENICZNOŚCI

Oceny parametrów modelu SAR dla procesu z heterogenicznością przestrzenną (wyrażoną losowymi efektami) wskazują, że dla ocen parametrów stojących przy zmiennych objaśniających względny błąd estymatora nie przekracza 5%. Oznacza to, że zarówno estymator parametru β_1 , jak i β_2 jest nieobciążony (por. tabela 2). Pominięta heterogeniczność przestrzenna nie wpłynęła również negatywnie na 95% bayesowski przedział ufności obu parametrów, o czym świadczą wysokie wartości wskaźnika pokrycia.

Tabela 2. Wpływ pominiętej przestrzennej heterogeniczności na oszacowania parametrów i 95% obszaru ufności w modelu SAR

Parametr	Względny błąd estymatora	Wskaźnik pokrycia 95% przedziału
β_1	1%	95%
β_2	2%	80%
ρ	36%	0%
σ_ε^2	23%	0%

Obliczono na podstawie $R=800$.

Źródło: obliczenia własne w R Cran.

Sytuacja ulega zmianie w przypadku oceny parametru interakcji przestrzennych ρ oraz wariancji składnika losowego σ_ε^2 . Względny błąd estymatora dla obu przekracza znacząco wartość 5%, potwierdzając znaczące przeszacowanie. Wyznaczone dla obu parametrów 95% przedziały wiarygodności nie zawierają rzeczywistej wartości parametrów, co znajduje odzwierciedlenie w zerowych wartościach wskaźników pokrycia.

Uzyskane wyniki dają jedynie ogólną informację dotyczącą skutków pominięcia heterogeniczności przestrzennej w modelu SAR. W celu określenia związku między skalą niejednorodności przestrzennej (wartość σ_μ^2) a wartościami błędów estymatora wykonano analizę ANOVA (tabela 3).

Wynika z niej, że wartość parametru ρ różnicuje wielkość błędu estymatora parametru β_1 , β_2 oraz ρ , przy czym dla β_1 i β_2 nie stwierdzono obciążenia. W przypadku ρ zauważono, że im większa wartość parametru, tym mniejszy względny błąd estymatora. Dla wartości $\rho=0,20$ względny błąd wyniósł aż 53%, dla $\rho=0,40$ było to 43%. Wartości $\rho=0,60$ oraz $\rho=0,80$ odznaczały się zaś najmniejszym względnym błędem estymatora, który wyniósł odpowiednio 32% i 18%. Wynik ten sugeruje, że dla procesów silnie skorelowanych przestrzennie negatywny efekt pominięcia heterogeniczności w modelu SAR, wyrażony wielkością błędu estymatora, będzie nieco mniejszy, chociaż uzyskana ocena parametru ρ wciąż będzie przeszacowana.

Tabela 3. Wpływ wartości ρ i σ_μ^2 na oceny parametrów modelu SAR – wyniki analizy ANOVA

Parametr	ρ	σ_μ^2
β_1	0,00	0,48
β_2	0,00	0,20
ρ	0,00	0,00
σ_ϵ^2	0,36	0,00

Obliczono na podstawie $R=200$ (wariant parametru). W tabeli podano wartości p-value.

Źródło: obliczenia własne w R Cran.

Zauważono również, że skala heterogeniczności przestrzennej, rozumiana jako wartość wariancji efektów losowych (σ_μ^2), w sposób istotny statystycznie różnicuje względny błąd estymatora parametru ρ oraz σ_ϵ^2 . W szczególności w przypadku wariancji składnika losowego wzrost heterogeniczności przestrzennej powoduje wzrost wartości względnego błędu estymatora. Dla $\sigma_\mu^2=0,20$ względny błąd estymatora wariancji składnika losowego wyniósł zaledwie 11%, dla $\sigma_\mu^2=0,40$ wartość błędu wyniosła już 19%, a przy $\sigma_\mu^2=0,60$ względny błąd estymatora $\sigma_\epsilon^2=27\%$. Dla procesu o najsilniejszej skali przestrzennej heterogeniczności (0,80) względny błąd estymatora wariancji składnika losowego w modelu SAR wyniósł 33%.

3.2. EFEKT POMINIĘCIA SKORELOWANYCH PRZESTRZENNIE EFEKTÓW LOSOWYCH

W dalszej kolejności wzięto pod uwagę proces przestrzenny, w przypadku którego wprowadzono zarówno heterogeniczność przestrzenną, jak i dodatkowe interakcje przestrzenne w postaci skorelowanych przestrzennie efektów losowych. Uzyskane wyniki (tabela 4) wskazują, że podobnie jak w przypadku pominięcia heterogeniczności, estymator parametru β_1 i β_2 pozostaje nieobciążony. Względny błąd estymatora w obu przypadkach jest nieznaczny, co wskazuje na niedoszacowanie o 1 i 2%. Analogicznie,

wskaźniki pokrycia 95% przedziału wiarygodności są wysokie (93% i 81%), co potwierdza brak negatywnego wpływu pominiętych efektów przestrzennych na oszacowania błędów ocen parametrów.

Tabela 4. Wpływ pominięcia skorelowanych przestrzennie efektów losowych na oszacowania parametrów i 95% przedziału wiarygodności w modelu SAR

Parametr	Względny błąd estymatora	Wskaźnik pokrycia 95% przedziału
β_1	-1%	93%
β_2	-2%	81%
ρ	38%	0%
σ^2	28%	0%

Obliczono na podstawie $R=800$.

Źródło: obliczenia własne w R Cran.

Sytuacja ta ulega zmianie dla ocen parametru interakcji przestrzennych ρ oraz wariancji składnika losowego. Względny błąd estymatora parametru ρ w sytuacji, gdy w modelu SAR nie uwzględniono skorelowanych przestrzennie efektów losowych, wyniósł 38%. W porównaniu do modelu, w którym pominięto nieskorelowane przestrzennie efekty losowe, jest to wynik porównywalny. Sugeruje to, że uwzględnienie efektów losowych ma dominujące znaczenie dla uzyskania prawidłowej oceny parametru ρ , a dodatkowe interakcje przestrzenne efektów losowych jedynie nieznacznie zwiększają przeszacowanie wartości parametru.

Inaczej jest w przypadku oszacowań wariancji składnika losowego. Podobnie jak w modelu SAR z pominiętymi efektami losowymi, wielkość względnego błędu estymatora przekracza 5%, potwierdzając przeszacowanie wariancji składnika losowego w sytuacji pominięcia skorelowanych przestrzennie efektów losowych. Jednocześnie wartość błędu jest o 5 p.p. wyższa od tej, którą odnotowano w przypadku pominięcia nieskorelowanych przestrzennie efektów losowych. Fakt ten świadczyć może o tym, że pominięcie skorelowanych efektów losowych znajduje odzwierciedlenie w podwyższeniu względnego błędu estymatora wariancji składnika losowego. Co więcej, można się spodziewać, że uzyskane na podstawie modelu SAR reszty będą skorelowane przestrzennie, przy czym struktura zależności będzie wyrażona w postaci interakcji grupowych, analogicznych do tych, które zawarto w macierzy \mathbf{M} .

W celu określenia, czy poziom skorelowania efektów losowych (wyrażony parametrem interakcji przestrzennych λ) oraz skala przestrzennej heterogeniczności (wyrażona wariancją efektów losowych σ_μ^2) różnicują w spo-

sób istotny statystycznie wartości względnych błędów estymatora parametrów modelu SAR, przeprowadzono analizę ANOVA (tabela 5).

Tabela 5. Wpływ wartości λ i σ_μ^2 na oceny parametrów modelu SAR – wyniki analizy ANOVA

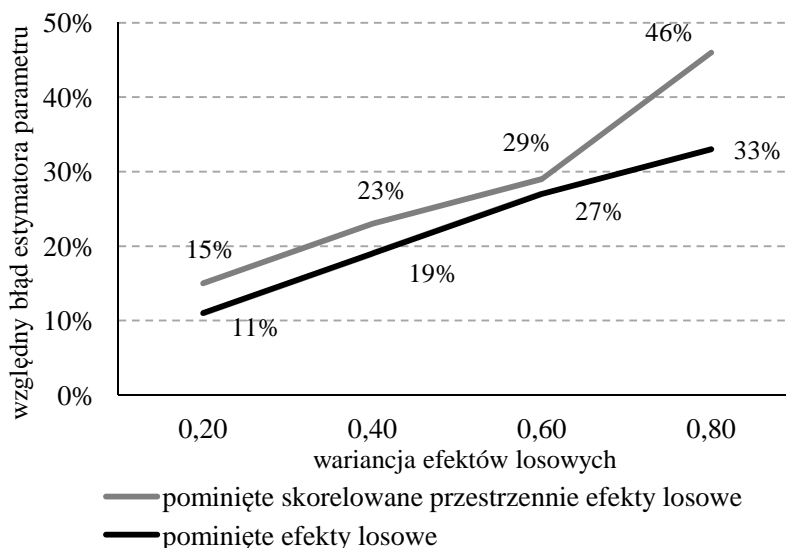
Parametr	λ	σ_μ^2
β_1	0,52	0,28
β_2	1,00	0,00
ρ	0,00	0,00
σ_ε^2	0,00	0,00

Obliczono na podstawie $R=200$ (wariant parametru). W tabeli podano wartości p-value.

Źródło: obliczenia własne w R Cran.

Uzyskane wyniki wskazują, że zarówno wartość λ , jak i σ_μ^2 różnicują skalę przeszacowania ocen parametrów ρ i σ_ε^2 . Względny błąd estymatora parametru ρ wahał się od 34% dla $\lambda=0,20$ do 43% dla $\lambda=0,80$, wykazując stopniowy wzrost wraz ze wzrostem wartości parametru λ . Analogiczną sytuację odnotowano w przypadku względnego błędu estymatora wariancji składnika losowego, uzyskując kolejno: dla $\lambda=0,20$ i $\lambda=0,40$ wartość błędu $\sigma_\varepsilon^2=24\%$, dla $\lambda=0,60$ błąd $\sigma_\varepsilon^2=26\%$, a dla $\lambda=0,80$ błąd $\sigma_\varepsilon^2=38\%$. Oznacza to, że im bardziej efekty losowe są skorelowane przestrzennie, tym większe jest przeszacowanie parametru interakcji przestrzennych i wariancji składnika losowego w modelu SAR.

Podobne wnioski można sformułować, analizując wpływ wartości σ_μ^2 na oceny ww. parametrów. W szczególności wzrost skali heterogeniczności przestrzennej znajduje swoje odzwierciedlenie w coraz większym przeszacowaniu wariancji składnika losowego (rysunek 1). Dla każdej z czterech wartości wariancji efektów losowych wartość względnego błędu estymatora parametru σ_ε^2 była wyższa dla procesu cechującego się skorelowaniem efektów losowych, niż w przypadku efektów losowych nieskorelowanych przestrzennie. Warto zauważyć, że dla procesu o najwyższej skali heterogeniczności przestrzennej względny błąd estymatora różni się aż o 13 p. p. w zależności od tego, czy efekty losowe są skorelowane przestrzennie, czy też są niezależne.



Rysunek 1. Względny błąd estymatora parametru σ_ϵ^2 w zależności od wartości σ_μ^2
 Źródło: obliczenia własne.

Uzyskane wyniki wskazują, że pominięcie skorelowanych przestrzennie lub niezależnych od siebie efektów losowych wpływa negatywnie wyłącznie na oceny parametrów interakcji przestrzennych oraz wariancji składnika losowego w modelu SAR. Wielkość przeszacowania parametru interakcji przestrzennych ρ jest porównywalna dla obu przypadków i wynika ze skali pominiętej heterogeniczności przestrzennej (σ_μ^2). W przypadku oszacowań wariancji składnika losowego zauważyć można, że wielkość przeszacowania jest wyższa w sytuacji, gdy pominięte efekty losowe są skorelowane przestrzennie.

4. WNIOSKI

Uzyskane wyniki wskazują, że pominięcie heterogeniczności przestrzennej (w postaci efektów losowych), jak również pominięcie heterogeniczności i dodatkowych zależności przestrzennych (rozumianych jako obecność skorelowanych przestrzennie efektów losowych), wpływa negatywnie na błąd estymatora parametrów interakcji przestrzennych, który każdorazowo jest przeszacowany w modelu SAR. Przeszacowaniu, wskutek pominięcia dodatkowych charakterystyk procesu przestrzennego, ulega również wariancja składnika losowego. Co ważne, nieuwzględnienie efektów losowych (skore-

lowanych przestrzennie bądź niezależnych) nie wpływa na oszacowania parametrów zmiennych objaśniających. Dla tych ostatnich względny błąd estymatora jest wciąż niewielki (poniżej 5%), a wskaźnik pokrycia 95% przedziału wiarygodności wysoki.

Różnice między wartością względną błędów estymatora parametru interakcji przestrzennych w modelu SAR w przypadku pominięcia niezależnych i skorelowanych przestrzennie efektów losowych są nieznaczące (2 p.p.). Oznacza to, że o ile sam fakt pominięcia efektów losowych powoduje przeszacowanie parametru interakcji przestrzennych, o tyle to, czy efekty losowe pozostają niezależne, czy też cechują się określoną strukturą relacji, ma znaczenie drugorzędne w przypadku oszacowań parametru interakcji przestrzennych.

Inaczej jest w przypadku oszacowań wariancji składnika losowego. W przypadku procesu cechującego się niezależnymi efektami losowymi względny błąd estymatora wariancji składnika losowego jest mniejszy, niż gdy efekty losowe pozostają ze sobą skorelowane przestrzennie (por. rysunek 1). Co więcej, w przypadku silnego zróżnicowania efektów losowych (wyższa wartość σ_{μ}^2) fakt skorelowania tychże efektów powoduje znaczący wzrost względnego błędu estymatora wariancji składnika losowego, co nie jest zauważalne w przypadku nieskorelowanych efektów losowych.

Dla obu procesów przestrzennych zauważono, iż skala przeszacowania parametru interakcji przestrzennych, jak również wariancji składnika losowego wynika ze skali niejednorodności przestrzennej zjawiska oraz siły interakcji przestrzennych. Im bardziej niejednorodny proces (wyższa wartość σ_{μ}^2), tym wyższy względny błąd estymatora wariancji składnika losowego. Analogicznie, im bardziej efekty losowe są ze sobą przestrzennie skorelowane, tym wyższy względny błąd estymatora parametru interakcji przestrzennych i wariancji składnika losowego w modelu SAR.

PODSUMOWANIE

Niniejsze badanie miało na celu wskazanie skutków błędnej specyfikacji modelu SAR, rozumianej jako pominięcie faktu występowania dodatkowych charakterystyk procesu przestrzennego. Skoncentrowano się na dwóch procesach – odznaczających się występowaniem niezależnych i skorelowanych przestrzennie efektów losowych. W celu określenia skutków pominięcia wspomnianych efektów, w modelu SAR, skorzystano z symulacji Monte Carlo.

Wyniki symulacji wskazują na wrażliwość oszacowań parametru interakcji przestrzennych oraz wariancji składnika losowego – w modelu SAR – na pominięcie dodatkowych charakterystyk procesu przestrzennego. Oba są

przeszacowane w przypadku pominięcia niezależnych bądź skorelowanych przestrzennie efektów losowych. W przypadku parametru interakcji przestrzennych względny błąd estymatora jest porównywalny, gdy pominięte efekty losowe są skorelowane przestrzennie bądź niezależne. Skala przeszacowania wariancji składnika losowego różni się jednak dla obu procesów i jest wyższa w sytuacji, gdy pominięte efekty losowe są skorelowane przestrzennie.

LITERATURA

- Anselin L. (1988), *Spatial Econometrics: Methods and Models*, Vol. 4. Springer.
- Anselin L., Griffith D. A. (1988), *Do spatial effects really matter in regression analysis?*, „Papers in Regional Science”, 65(1), 11–34.
- Domański Cz., Pruska K. (2000), *Nieklasyczne metody statystyczne*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Goldstein H. (2011), *Multilevel Statistical Models*, John Wiley & Sons.
- Hoogland J., Boomsma A. (1998) *Robustness studies in covariance structure modeling: An overview and a meta-analysis*, „Sociological Methods and Research”, 26(3), 329–367.
- Lee L. F. (2002), *Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models*, „Econometric Theory”, 18(02), 252–277.
- Lee L. F. (2007), *GMM and 2SLS estimation of mixed regressive, spatial autoregressive models*, „Journal of Econometrics”, 137(2), 489–514.
- LeSage J. (1997), *Bayesian estimation of spatial autoregressive models*, „International Regional Science Review”, 20(1–2), 113–129.
- LeSage J., Pace R. K. (2009), *Introduction to Spatial Econometrics*, CRC Press.
- Liu A., Folmer H., Oud J. H. (2014), *Estimation of autoregressive models with two types of weak spatial dependence by means of the W-based and the latent variables approach: evidence from Monte Carlo simulations*, „Environment and Planning”, A, 46(1), 186–202.
- Moerbeek M. (2004), *The consequence of ignoring a level of nesting in multilevel analysis*, „Multivariate Behavioral Research”, 39.1, 129–149.
- Osiewalski J. (2001), *Ekonometria bayesowska w zastosowaniach*, Wyd. AE w Krakowie, Kraków.
- Pace R. K., Barry R. (1997), *Quick computation of spatial autoregressive estimators*, “Geographical analysis”, 29.3, 232–247.
- Plummer M., Best N., Cowles K., Vines K. (2005), *CODA: output analysis and diagnostics for MCMC*, R package version 0.9–2.
- Suchecky B. (red.), (2010), *Ekonometria przestrzenna*, Wydawnictwo C.H.Beck, Warszawa.
- Szulc E. (2007), *Ekonometryczna analiza wielowymiarowych procesów gospodarczych*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.
- Van den Noortgate W., Opdenakker M. C., Onghena P. (2005), *The effects of ignoring a level in multilevel analysis*, „School Effectiveness and School Improvement”, 16(3), 281–303.

MISSPECIFICATION OF SPATIAL EFFECTS IN THE BAYESIAN
SPATIAL AUTOREGRESSIVE MODEL. THE RESULTS FROM
THE MONTE CARLO SIMULATION

A b s t r a c t: The aim of this paper is to analyze the impact of the SAR model misspecification. We concentrate on the effect of ignoring random effects (both independent and spatially correlated) in the SAR model. The results from the Monte Carlo simulation suggests the overestimation of the spatial parameter and error variance if the random effects are omitted in the SAR model.

K e y w o r d s: spatial autoregressive model, Bayesian MCMC, Monte Carlo simulation.