A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 253, 2011

Elżbieta Szulc[*], *Iwona Müller-Frączek*[**], *Michał Bernard Pietrzak*[***]

MODELOWANIE ZALEŻNOŚCI MIĘDZY EKONOMICZNYMI PROCESAMI PRZESTRZENNYMI A POZIOM AGREGACJI DANYCH

Streszczenie. Statystyczno-ekonometryczne analizy zjawisk przestrzennych są przeprowadzane przy wykorzystaniu danych o różnym poziomie agregacji przestrzennej. Ma to określone konsekwencje dla możliwości odkrywania różnych własności i struktur procesów przestrzennych. Te zaś stanowią podstawę do odpowiedniej specyfikacji przestrzennych modeli zależności między procesami.

W artykule zostało przedyskutowane w jaki sposób agregacja przestrzenna wpływa na zmianę własności i struktur procesów przestrzennych, w szczególności zaś pokazano różnice w charakterze i sile wpływu agregacji przestrzennej na poszczególne składniki procesów.

Główną częścią artykułu jest omówienie wpływu agregacji danych na modelowanie zależności między procesami przestrzennymi. Szczególna uwaga została poświęcona ekonometrycznemu modelowaniu zjawisk na różnych poziomach agregacji danych w warunkach występowania autokorelacji przestrzennej. Ważnym efektem przeprowadzonej dyskusji jest wskazanie na problem „naiwnej" interpretacji modeli przestrzennych.

Rozważania przeprowadzone zostały na podstawie danych generowanych oraz zilustrowane przykładem empirycznym.

1. WPROWADZENIE

Procesy przestrzenne można zdefiniować jako losowe funkcje $Y(p)$ nielosowego argumentu $p = [p_1, p_2] \in R^2$. W tym sensie, stanowią one bezpośrednie odpowiedniki dynamicznych procesów stochastycznych. W niniejszym artykule rozważa się procesy $Y(p_i)$, obserwowane na płaszczyźnie, w przestrzennych lokalizacjach $p_i = [p_{1i}, p_{2i}]$, gdzie $i = 1, 2, ..., N$ oznacza numer jednostki przestrzennej.

W analizach procesów przestrzennych wykorzystuje się dane o różnym poziomie agregacji. Stanowią one informacje odnoszące się do jednostek przestrzennych różnych rozmiarów i kształtów (np.: gminy, powiaty, województwa). Dane przestrzenne tworzą określone konfiguracje, które mają zasadnicze znaczenie dla własności procesów przez nie opisywanych, a zatem również dla modeli tych procesów. Zmiana poziomu agregacji danych zmienia owe konfiguracje.

Prezentowane opracowanie nawiązuje 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). W opracowaniu, oprócz rozważań na temat możliwych zmian

[*] Prof., Katedra Ekonometrii i Statystyki, Uniwersytet Mikołaja Kopernika w Toruniu.
[**] Dr, Katedra Ekonometrii i Statystyki, Uniwersytet Mikołaja Kopernika w Toruniu.
[***] Dr, Katedra Ekonometrii i Statystyki, Uniwersytet Mikołaja Kopernika w Toruniu.

charakterystyk procesów przestrzennych ocenianych na podstawie danych przed agregacją oraz po agregacji, w szczególności poruszono następujące kwestie:
1) Postawiono pytanie, jakie są zmiany parametrów regresji w modelach przestrzennych szacowanych na podstawie danych przed agregacją i danych zagregowanych.
2) Sformułowano hipotezę, że znalezienie parametru mierzącego rzeczywisty wpływ procesu $X(p)$ na proces $Y(p)$ wymaga odkrycia własności i struktur badanych procesów.
3) 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.

2. PRZEDMIOT I ZAKRES BADANIA

Przeprowadzone badanie skoncentrowano na dwóch aspektach analizy procesów przestrzennych. Po pierwsze, przedmiotem zainteresowania były własności procesów przestrzennych przed i po agregacji. Po drugie, było to ekonometryczne modelowanie zależności między procesami przestrzennymi na różnych (w tym wypadku dwóch) poziomach agregacji.

W celu sprawdzenia, czy i jak pod wpływem agregacji zmieniają się charakterystyki procesów przestrzennych o różnych strukturach wewnętrznych rozważono następujące procesy:
- biały szum przestrzenny

$$\varepsilon(p) \sim \mathrm{N}[0, 1], \tag{1}$$

gdzie: $\mathrm{cov}[\varepsilon(p_i), \varepsilon(p_j)] = 0, \ \forall i \neq j$;

- proces autoregresyjny rzędu pierwszego

$$Y(p) = \rho \mathbf{W} Y(p) + \beta + \varepsilon(p), \tag{2}$$

gdzie: $\varepsilon(p)$ – j. w.,
$\mathbf{W}$ – macierz sąsiedztwa rzędu 1;
- proces z przestrzennym trendem deterministycznym stopnia s

$$Y(p) = \sum_{k=0} \sum_{m=0} \beta_{km} p_1^k p_2^m + \varepsilon(p), \tag{3}$$

gdzie: $k + m \leq s$; $\varepsilon(p)$ – j. w.;
- proces z przestrzennym trendem deterministycznym stopnia s oraz autoregresją rzędu 1.

$$Y(p) = \sum_{k=0} \sum_{m=0} \beta_{km} p_1^k p_2^m + \rho \mathbf{W} Y(p) + \varepsilon(p), \tag{4}$$

Analizę zależności między procesami przestrzennymi przeprowadzono przy założeniu analogicznych struktur wewnętrznych poszczególnych procesów. Dla uproszczenia uwagę ograniczono jedynie do dwóch procesów: $Y(p)$ (proces objaśniany), $X(p)$ (proces objaśniający). Wykorzystano dwie koncepcje modeli zależności – klasyczny model prze-

strzenny postaci: $Y(p)=\gamma X(p)+ \varepsilon(p)$, oraz tzw. model zgodny będący przestrzennym odpowiednikiem dynamicznego modelu zgodnego według koncepcji Zielińskiego[1].

Budowa zgodnego modelu przestrzennego przebiegała w następujących etapach:

Etap I.

Konstrukcja modeli podstawowych dla poszczególnych procesów, w tym:

(a) modeli z trendem przestrzennym stopnia, odpowiednio s_x, s_y:

$$Y(p)=\sum_{k=0}\sum_{m=0}\beta_{km}p_1^k p_2^m+\eta_y(p), \tag{5}$$

gdzie: $k+m\le s_y$.

$$X(p)=\sum_{k=0}\sum_{m=0}\alpha_{km}p_1^k p_2^m+\eta_x(p), \tag{6}$$

gdzie: $k+m\le s_x$;

(b) modeli autoregresyjnych rzędu 1:

$$\eta_y(p)=\rho_y\mathbf{W}\eta_y(p)+\varepsilon_y(p), \tag{7}$$

$$\eta_x(p)=\rho_x\mathbf{W}\eta_x(p)+\varepsilon_x(p), \tag{8}$$

Etap II.

Konstrukcja modelu zależności na poziomie białych szumów przestrzennych $\varepsilon_y(p)$, $\varepsilon_x(p)$, tj.:

$$\varepsilon_y(p)=\gamma\varepsilon_x(p)+\varepsilon(p) \tag{9}$$

Etap III.

Konstrukcja modelu zależności na poziomie procesów rzeczywistych $Y(p)$ oraz $X(p)$, przy uwzględnieniu struktur tych procesów, wynikających z (5) – (8), tj.:

(a) uwzględnienie w modelu (9) struktur opisanych przez (7) i (8):

$$(\mathbf{I}-\rho_y\mathbf{W})\eta_y(p)=\gamma(\mathbf{I}-\rho_x\mathbf{W})\eta_x(p)+\varepsilon(p); \tag{10}$$

(b) uwzględnienie w (10) równań (5) i (6):

$$(\mathbf{I}-\rho_y\mathbf{W})\left(Y(p)-\sum_{k=0}\sum_{m=0}\beta_{km}p_1^k p_2^m\right)=\gamma(\mathbf{I}-\rho_x\mathbf{W})\left(X(p)-\sum_{k=0}\sum_{m=0}\alpha_{km}p_1^k p_2^m\right)+\varepsilon(p)\ ; \tag{11}$$

(c) dokonanie odpowiednich przekształceń, uporządkowanie i pogrupowanie wyrazów, co prowadzi do modelu:

$$Y(p)=\rho_y\mathbf{W}Y(p)+\gamma X(p)+\gamma^*\mathbf{W}X(p)+\sum_{k=0}\sum_{m=0}\beta^*{}_{km}p_1^k p_2^m+\varepsilon(p), \tag{12}$$

gdzie: $\gamma^*=-\gamma\rho_x$, $k+m\le s$, $s=\max\{s_y,s_x\}$; β^* parametry wynikające z agregacji parametrów modeli trendów procesów $Y(p)$ oraz $X(p)$, ważonych parametrami autoregresji tych procesów,

[1] Zob. Z. Zieliński, [1990], *Liniowe modele ekonometryczne jako narzędzie opisu i analizy przyczynowych zależności zjawisk ekonomicznych*, Wydawnictwo UMK, Toruń i Kufel T., [2002], *Postulat zgodności w dynamicznych modelach ekonometrycznych*, Wydawnictwo UMK, Toruń.

Model (12) jest zgodny w tym sensie, że uzgadnia struktury rozważanych procesów przestrzennych i objaśnia proces endogeniczny w takim stopniu, że proces resztowy jest białym szumem. Jednak pewnym mankamentem otrzymanego modelu jest to, iż inaczej niż w zgodnych modelach dynamicznych, proces resztowy nie jest niezależny od procesu objaśniającego. Aby podkreślić ten fakt, przestrzenne modele zgodne takie jak (12) należałoby nazywać raczej modelami „quasi" zgodnymi[2].

3. DANE

W przeprowadzonym badaniu wykorzystano dwa rodzaje danych. W głównej części badania wykorzystano dane generowane w przestrzennym układzie gmin i powiatów. Natomiast w celu empirycznej ilustracji rozważanych kwestii przeanalizowano przykład dotyczący bezrobocia oraz nakładów inwestycyjnych w układzie powiatów oraz podregionów.

Dane generowane jako realizacje pojedynczych procesów uzyskano w następujący sposób. Zgodnie z określonymi w punkcie 2. modelami, przyjmując odpowiednie parametry, wygenerowano dane w układzie gmin (niższy poziom agregacji), które następnie zagregowano do poziomu powiatów (wyższy poziom agregacji). Procedura generowania danych przebiegała, w zależności od założonej struktury procesu przestrzennego, w kolejnych krokach:

Krok I.
Wygenerowano dane czysto losowe dla procesu $\varepsilon(p) \sim \mathrm{N}[0, 1]$.

Krok II.
Wykorzystano dane z kroku I. do wygenerowania danych o strukturze autoregresyjnej według formuł:

$$\varepsilon(p) = (\mathbf{I} - \rho\mathbf{W})\eta(p), \tag{13}$$

$$\eta(p) = (\mathbf{I} - \rho\mathbf{W})^{-1}\varepsilon(p). \tag{14}$$

Krok III.
Nałożono na strukturę autoregresyjną trend przestrzenny stopnia s, tj.:

$$Y(p) = \eta(p) + \sum_{k=0}\sum_{m=0}\beta_{km}p_1^k p_2^m . \tag{15}$$

gdzie: $k + m \le s$,
otrzymując ostatecznie realizacje procesu, który można opisać równaniem:

$$Y(p) = \rho\mathbf{W}Y(p) + \sum_{k=0}\sum_{m=0}\beta_{km}p_1^k p_2^m + \varepsilon(p), \tag{16}$$

gdzie: $k + m \le s$.

Eksperyment dotyczący zależności między procesami przestrzennymi oparto na następującej procedurze.

[2] Por. E. Szulc, [2007], *Ekonometryczna analiza wielowymiarowych procesów gospodarczych*, Wydawnictwo UMK, Toruń, podrozdziały: 4.1.3, 4.2.2.

1. Wygenerowano zależne procesy białoszumowe $\varepsilon_y(p)$ oraz $\varepsilon_x(p)$ z poziomem zależności opisanym przez model $\varepsilon_y(p)=\gamma\varepsilon_x(p)+\varepsilon(p)$, przy czym $\gamma=0{,}7$.

2. Wygenerowano procesy $X(p)$ o postaci $X(p)=\rho_x\mathbf{W}X(p)+\sum_{k=0}\sum_{m=0}\alpha_{km}p_1^k p_2^m+\varepsilon_x(p)$ $k+m\le s_x$, wykorzystując kolejno następujące formuły: $\varepsilon_x(p)=(\mathbf{I}-\rho_x\mathbf{W})\eta_x(p)$, $\eta_x(p)=(\mathbf{I}-\rho_x\mathbf{W})^{-1}\varepsilon_x(p)$ oraz $X(p)=\eta_x(p)+\sum_{k=0}\sum_{m=0}\alpha_{km}p_1^k p_2^m$, $k+m\le s_x$, dla określonych wartości parametrów ρ_x, α_{km}, a także określonych stopni trendu s_x.

3. Wygenerowano procesy $Y(p)$ o postaci $Y(p)=\rho_y\mathbf{W}Y(p)+\sum_{k=0}\sum_{m=0}\beta_{km}p_1^k p_2^m+\varepsilon_y(p)$, $k+m\le s_y$, wykorzystując formuły analogiczne j. w., tj.: $\varepsilon_y(p)=(\mathbf{I}-\rho_y\mathbf{W})\eta_y(p)$, $\eta_y(p)=(\mathbf{I}-\rho_y\mathbf{W})^{-1}\varepsilon_y(p)$ oraz $Y(p)=\eta_y(p)+\sum_{k=0}\sum_{m=0}\beta_{km}p_1^k p_2^m$, $k+m\le s_y$, dla określonych wartości parametrów ρ_y, β_{km}, a także określonych stopni trendu s_y.

4. Wyznaczono współczynniki regresji $Y(p)$ względem $X(p)$ według klasycznych modeli regresji przestrzennej.

5. Oceniono regresję $Y(p)$ względem $X(p)$ na podstawie quasi - zgodnych modeli przestrzennych.

Parametry $\rho_x, \rho_y, \alpha_{km}, \beta_{km}$, a także stopnie trendu s_x, s_y zostały ustalone w taki sposób, aby rozważyć procesy $X(p)$, $Y(p)$ o różnych strukturach, zbudować dla nich modele klasyczne abstrahując od tych struktur oraz modele zgodne, uwzględniające owe struktury, a następnie dokonać odpowiednich porównań, w szczególności w kontekście zdolności odkrywania „czystego" związku pomiędzy badanymi procesami, przed i po agregacji.

Rozważono następujące sytuacje:

1) Procesy $X(p)$, $Y(p)$ są białymi szumami.

2) Procesy $X(p)$, $Y(p)$ są procesami z trendem stopnia 1. o jednakowych parametrach $(\alpha_{00}=\beta_{00}, \alpha_{10}=\beta_{10}, \alpha_{01}=\beta_{01})$.

3) Procesy $X(p)$, $Y(p)$ są procesami z trendem stopnia 1. o różnych parametrach $(\alpha_{00}\neq\beta_{00}, \alpha_{10}\neq\beta_{10}, \alpha_{01}\neq\beta_{01})$.

4) Procesy $X(p)$, $Y(p)$ są procesami z trendem stopnia 2. o jednakowych parametrach $(\alpha_{00}=\beta_{00}, \alpha_{10}=\beta_{10}, \alpha_{01}=\beta_{01}, \alpha_{11}=\beta_{11}, \alpha_{20}=\beta_{20}, \alpha_{02}=\beta_{02})$.

5) Procesy $X(p)$, $Y(p)$ są procesami z trendem stopnia 2. o różnych parametrach $(\alpha_{00}\neq\beta_{00}, \alpha_{10}\neq\beta_{10}, \alpha_{01}\neq\beta_{01}, \alpha_{11}\neq\beta_{11}, \alpha_{20}\neq\beta_{20}, \alpha_{02}\neq\beta_{02})$.

6) Proces $X(p)$, jest procesem z trendem stopnia 1. natomiast proces $Y(p)$ wykazuje trend stopnia 2.

7) Procesy $X(p)$, $Y(p)$ są procesami autoregresyjnymi rzędu 1. o jednakowym poziomie autoskorelowania $(\rho_x=\rho_y)$.

8) Procesy $X(p)$, $Y(p)$ są procesami autoregresyjnymi rzędu 1. o różnym poziomie autoskorelowania $(\rho_x \neq \rho_y)$.

9) Procesy $X(p)$, $Y(p)$ są procesami wykazującymi trendy stopnia 1. oraz autozależności rzędu 1., przy czym: $\alpha_{00} = \beta_{00}, \alpha_{10} = \beta_{10}, \alpha_{01} = \beta_{01}, \rho_x = \rho_y$.

10) Procesy $X(p)$, $Y(p)$ są procesami wykazującymi trendy stopnia 1. oraz autozależności rzędu 1., przy czym: $\alpha_{00} \neq \beta_{00}, \alpha_{10} \neq \beta_{10}, \alpha_{01} \neq \beta_{01}, \rho_x = \rho_y$.

11) Procesy $X(p)$, $Y(p)$ są procesami wykazującymi trendy stopnia 1. oraz autozależności rzędu 1., przy czym: $\alpha_{00} = \beta_{00}, \alpha_{10} = \beta_{10}, \alpha_{01} = \beta_{01}, \rho_x \neq \rho_y$.

12) Procesy $X(p)$, $Y(p)$ są procesami wykazującymi trendy stopnia 1. oraz autozależności rzędu 1., przy czym: $\alpha_{00} \neq \beta_{00}, \alpha_{10} \neq \beta_{10}, \alpha_{01} \neq \beta_{01}, \rho_x \neq \rho_y$.

13) Procesy $X(p)$, $Y(p)$ są procesami z trendem stopnia 2. o jednakowych parametrach $(\alpha_{00} = \beta_{00}, \alpha_{10} = \beta_{10}, \alpha_{01} = \beta_{01}, \alpha_{11} = \beta_{11}, \alpha_{20} = \beta_{20}, \alpha_{02} = \beta_{02})$ oraz autozależnościach rzędu 1., przy czym $\rho_x = \rho_y$.

14) Procesy $X(p)$, $Y(p)$ są procesami z trendem stopnia 2. o różnych parametrach $(\alpha_{00} \neq \beta_{00}, \alpha_{10} \neq \beta_{10}, \alpha_{01} \neq \beta_{01}, \alpha_{11} \neq \beta_{11}, \alpha_{20} \neq \beta_{20}, \alpha_{02} \neq \beta_{02})$ oraz autozależnościach rzędu 1., przy czym $\rho_x = \rho_y$.

15) Procesy $X(p)$, $Y(p)$ są procesami z trendem stopnia 2. o jednakowych parametrach $(\alpha_{00} = \beta_{00}, \alpha_{10} = \beta_{10}, \alpha_{01} = \beta_{01}, \alpha_{11} = \beta_{11}, \alpha_{20} = \beta_{20}, \alpha_{02} = \beta_{02})$ oraz autozależnościach rzędu 1., przy czym $\rho_x \neq \rho_y$.

16) Procesy $X(p)$, $Y(p)$ są procesami z trendem stopnia 2. o różnych parametrach $(\alpha_{00} \neq \beta_{00}, \alpha_{10} \neq \beta_{10}, \alpha_{01} \neq \beta_{01}, \alpha_{11} \neq \beta_{11}, \alpha_{20} \neq \beta_{20}, \alpha_{02} \neq \beta_{02})$ oraz auto-zależnościach rzędu 1., przy czym $\rho_x \neq \rho_y$.

17) Proces $X(p)$ jest procesem z trendem stopnia 1. natomiast proces $Y(p)$ wykazuje trend stopnia 2. Oba procesy wykazują autozależności rzędu 1., przy czym $\rho_x = \rho_y$.

18) Proces $X(p)$ jest procesem z trendem stopnia 1. natomiast proces $Y(p)$ wykazuje trend stopnia 2. Oba procesy wykazują autozależności rzędu 1., przy czym $\rho_x \neq \rho_y$.

Wykorzystano następujące modele generujące struktury procesu $X(p)$:

M 1. $X(p) = \varepsilon_x(p)$

M 2. $X(p) = 5 + 0{,}8\mathbf{W}X(p) + \varepsilon_x(p)$

M 3. $X(p) = 5 + 0{,}5\mathbf{W}X(p) + \varepsilon_x(p)$

M 4. $X(p) = 5 + 0{,}9p_1 + 2p_2 + \varepsilon_x(p)$

M 5. $X(p) = 2 + 0{,}5p_1 + 1{,}5p_2 + \varepsilon_x(p)$

M 6. $X(p) = 5 + 0{,}9p_1 + 2p_2 + 0{,}8\mathbf{W}X(p) + \varepsilon_x(p)$

M 7. $X(p) = 2 + 0{,}5p_1 + 1{,}5p_2 + 0{,}8\mathbf{W}X(p) + \varepsilon_x(p)$

M 8. $X(p)=5+0{,}9p_1+2p_2+0{,}5\mathbf{W}X(p)+\varepsilon_x(p)$

M 9. $X(p)=2+0{,}5p_1+1{,}5p_2+0{,}5\mathbf{W}X(p)+\varepsilon_x(p)$

M 10. $X(p)=5+6{,}51p_1+6{,}23p_2+0{,}12p_1p_2-0{,}61p_1^2-0{,}76p_2^2+\varepsilon_x(p)$

M 11. $X(p)=10+3{,}5p_1+7{,}2p_2+0{,}06p_1p_2-0{,}3p_1^2+0{,}4p_2^2+\varepsilon_x(p)$

M 12. $X(p)=5+6{,}51p_1+6{,}23p_2+0{,}12p_1p_2-0{,}61p_1^2-0{,}76p_2^2+0{,}8\mathbf{W}X(p)+\varepsilon_x(p)$

M 13. $X(p)=10+3{,}5p_1+7{,}2p_2+0{,}06p_1p_2-0{,}3p_1^2+0{,}4p_2^2+0{,}8\mathbf{W}X(p)+\varepsilon_x(p)$

M 14. $X(p)=5+6{,}51p_1+6{,}23p_2+0{,}12p_1p_2-0{,}61p_1^2-0{,}76p_2^2+0{,}5\mathbf{W}X(p)+\varepsilon_x(p)$

M 15. $X(p)=10+3{,}5p_1+7{,}2p_2+0{,}06p_1p_2-0{,}3p_1^2+0{,}4p_2^2+0{,}5\mathbf{W}X(p)+\varepsilon_x(p)$

Z kolei, wykorzystano następujące modele generujące struktury procesu $Y(p)$:

M 1. $Y(p)=\varepsilon_y(p)$

M 2. $Y(p)=5+0{,}8\mathbf{W}Y(p)+\varepsilon_y(p)$

M 3. $Y(p)=5+0{,}5\mathbf{W}Y(p)+\varepsilon_y(p)$

M 4. $Y(p)=5+0{,}9p_1+2p_2+\varepsilon_y(p)$

M 5. $Y(p)=5+0{,}9p_1+2p_2+0{,}8\mathbf{W}Y(p)+\varepsilon_y(p)$

M 6. $Y(p)=5+0{,}9p_1+2p_2+0{,}5\mathbf{W}Y(p)+\varepsilon_y(p)$

M 7. $Y(p)=5+6{,}51p_1+6{,}23p_2+0{,}12p_1p_2-0{,}61p_1^2-0{,}76p_2^2+\varepsilon_y(p)$

M 8. $Y(p)=10+6{,}51p_1+6{,}23p_2+0{,}12p_1p_2-0{,}61p_1^2-0{,}76p_2^2+\varepsilon_y(p)$

M 9. $Y(p)=5+6{,}51p_1+6{,}23p_2+0{,}12p_1p_2-0{,}61p_1^2-0{,}76p_2^2+0{,}8\mathbf{W}Y(p)+\varepsilon_y(p)$

M 10. $Y(p)=5+6{,}51p_1+6{,}23p_2+0{,}12p_1p_2-0{,}61p_1^2-0{,}76p_2^2+0{,}5\mathbf{W}Y(p)+\varepsilon_y(p)$

M 11. $Y(p)=10+6{,}51p_1+6{,}23p_2+0{,}12p_1p_2-0{,}61p_1^2-0{,}76p_2^2+0{,}8\mathbf{W}Y(p)+\varepsilon_y(p)$

M 12. $Y(p)=10+6{,}51p_1+6{,}23p_2+0{,}12p_1p_2-0{,}61p_1^2-0{,}76p_2^2+0{,}5\mathbf{W}Y(p)+\varepsilon_y(p)$

Wykorzystując otrzymane dane oszacowano i zweryfikowano klasyczne modele przestrzenne $Y(p)$ względem $X(p)$ oraz modele quasi-zgodne. Analogiczne modele oszacowano i zweryfikowano na podstawie danych zagregowanych.

4. UZYSKANE WYNIKI

4.1. Agregacja danych a własności procesów przestrzennych

Wyniki oszacowań parametrów rozważanych modeli, otrzymane dla danych przed agregacją oraz danych zagregowanych przedstawiają tabele 1 – 7.

W tym wątku rozważań sformułowano następujące wnioski:

1. Agregacja nie zmienia niezależnej struktury białego szumu przestrzennego. Średnia procesu pozostaje na tym samym poziomie. Zmniejsza się jedynie zmienność procesu.
2. Agregacja nie narusza także struktury trendowej procesu. Parametry trendu pozostają średnio biorąc takie same przed, jak i po agregacji. Zmienność procesu zmniejsza się.
3. Struktura autoregresyjna zależy od poziomu agregacji danych. Przejściu na wyższy poziom agregacji towarzyszy spadek wartości współczynnika autoregresji, a także obniżenie jego istotności. Zmienność procesu autoregresyjnego zmniejsza się.
4. Obecność autokorelacji przestrzennej powoduje zmiany parametrów trendu w modelach szacowanych na podstawie danych zagregowanych, w stosunku do tych otrzymanych na podstawie danych pierwotnych (przed agregacją). Parametry te są wyraźnie wyższe. Nadal obserwuje się prawidłowość, polegającą na zmniejszaniu się pod wpływem agregacji zmienności procesu oraz na wzroście ogólnego dopasowania modelu.

Tab. 1. Szacunki parametrów modelu białego szumu

Model: $\varepsilon(p) \sim N(0, \sigma)$ **Parametry przyjęte do symulacji:** $\mu = 0$; $\sigma = 1$								
	Przed agregacją				**Po agregacji**			
	μ	$S(\mu)$	I	p-value	μ	$S(\mu)$	I	p-value
Średnie	-0,0005	1,0017	-0,0014	0,5264	0,0001	0,5411	-0,0052	0,5228
Odchylenia	0,0192	0,0155	0,0113	0,2839	0,0246	0,0320	0,0325	0,2787

Źródło: opracowanie własne.

Tab. 2. Szacunki parametrów modelu autoregresyjnego (autoregresja dodatnia)

Model: $Y(p) = \rho WY(p) + \beta + \varepsilon(p)$; $\varepsilon(p) \sim N(0, \sigma)$ **Parametry przyjęte do symulacji:** $\rho = 0{,}8$; $\beta = 5$; $\sigma = 0{,}2$								
Przed agregacją								
	ρ	$S(\rho)$	β	$S(\beta)$	σ	R^2	I	p-value
Średnie	0,7535	0,0166	6,1623	0,0040	0,1995	0,5490	0,0081	0,2695
Odchylenia	0,0116	0,0003	0,2881	0,0001	0,0025	0,0124	0,0065	0,1700
Po agregacji								
		$S(\rho)$	β	$S(\beta)$	σ	R^2	I	p-value
Średnie	0,5144	0,0636	12,1381	0,0089	0,1738	0,8222	0,0173	0,2861
Odchylenia	0,0579	0,0027	1,4523	0,0003	0,0052	0,0382	0,0096	0,0935

Źródło: opracowanie własne.

Tab. 3. Szacunki parametrów modelu autoregresyjnego (autoregresja ujemna)

Model: $Y(p) = \rho \mathbf{W} Y(p) + \beta + \varepsilon(p)$; $\varepsilon(p) \sim N(0, \sigma)$ **Parametry przyjęte do symulacji:** $\rho = -0{,}8$; $\beta = 5$; $\sigma = 0{,}2$								
Przed agregacją								
	ρ	$S(\rho)$	β	$S(\beta)$	σ	R^2	*I*	***p*-value**
Średnie	-0,8423	0,0315	5,1173	0,0040	0,1982	0,7086	-0,0060	0,6647
Odchylenia	0,0302	0,0001	0,0830	0,0001	0,0029	0,0208	0,0044	0,1215
Po agregacji								
	ρ	$S(\rho)$	β	$S(\beta)$	σ	R^2	*I*	***p*-value**
Średnie	-0,3977	0,0960	3,8837	0,0059	0,1143	0,9265	-0,0089	0,5719
Odchylenia	0,1188	0,0032	0,3304	0,0002	0,0048	0,0386	0,0067	0,0761

Źródło: opracowanie własne.

Tab. 4. Szacunki parametrów modelu z trendem stopnia 1

Model: $Y(p) = \beta_{00} + \beta_{10} p_1 + \beta_{01} p_2 + \varepsilon(p)$; $\varepsilon(p) \sim N(0, \sigma)$ **Parametry przyjęte do symulacji:** $\beta_{00} = 5$; $\beta_{10} = 0{,}9$; $\beta_{01} = 2$; $\sigma = 1{,}7$					
Przed agregacją					
	β_{00}	$S(\beta_{00})$	β_{10}	$S(\beta_{10})$	β_{01}
Średnie	5,0598	0,1726	0,8965	0,0217	1,9899
Odchylenia	0,1450	0,0024	0,0179	0,0003	0,0225
	$S(\beta_{01})$	σ	R^2	***I*-Morana**	***p*-value**
Średnie	0,0231	1,6999	0,7645	-0,0017	0,5165
Odchylenia	0,0003	0,0234	0,0076	0,0121	0,2988
Po agregacji					
	β_{00}	$S(\beta_{00})$	β_{10}	$S(\beta_{10})$	β_{01}
Średnie	5,0820	0,2356	0,8949	0,0304	1,9902
Odchylenia	0,2114	0,0090	0,0197	0,0012	0,0285
	$S(\beta_{01})$	σ	R^2	***I*-Morana**	***p*-value**
Średnie	0,0315	0,9314	0,9191	-0,0021	0,5010
Odchylenia	0,0012	0,0354	0,0067	0,0263	0,2549

Źródło: opracowanie własne.

Tab. 5. Szacunki parametrów modelu z trendem stopnia 1. oraz autoregresją

Model: $Y(p) = \rho \mathbf{W} Y(p) + \beta_{00} + \beta_{10} p_1 + \beta_{01} p_2 + \varepsilon(p)$; $\varepsilon(p) \sim N(0, \sigma)$						
Parametry przyjęte do symulacji						
$\rho = 0{,}8$; $\beta_{00} = 5$; $\beta_{10} = 0{,}9$; $\beta_{01} = 0{,}9$; $\sigma = 1{,}7$						
Przed agregacją						
	ρ	$S(\rho)$	β_{00}	$S(\beta_{00})$	β_{10}	$S(\beta_{10})$
Średnie	0,7626	0,0161	5,9090	0,1712	0,7883	0,0215
Odchylenia	0,0172	0,0006	0,4001	0,0019	0,0660	0,0003
	β_{01}	$S(\beta_{01})$	σ	R^2	***I*-Morana**	***p*-value**
Średnie	1,7594	0,0229	1,6859	0,7186	0,0061	0,3082
Odchylenia	0,1384	0,0003	0,0186	0,0279	0,0041	0,1144
Po agregacji						
	ρ	$S(\rho)$	β_{00}	$S(\beta_{00})$	β_{10}	$S(\beta_{10})$
Średnie	0,4929	0,0579	12,5236	0,3855	1,7133	0,0498
Odchylenia	0,0926	0,0046	2,4073	0,0118	0,2951	0,0015
	β_{01}	$S(\beta_{01})$	σ	R^2	***I*-Morana**	***p*-value**
Średnie	3,7925	0,0516	1,5236	0,9346	0,0486	0,0970
Odchylenia	0,7010	0,0016	0,0467	0,0208	0,0209	0,0967

Źródło: opracowanie własne.

Tab. 6. Szacunki parametrów modelu z trendem stopnia 2

Model: $Y(p) = \beta_{00} + \beta_{10} p_1 + \beta_{01} p_2 + \beta_{11} p_1 p_2 + \beta_{20} p_1^2 + \beta_{02} p_2^2 + \varepsilon(p)$; $\varepsilon(p) \sim N(0, \sigma)$								
Parametry przyjęte do symulacji								
$\beta_{00} = 10$; $\beta_{10} = 6{,}51$; $\beta_{01} = 6{,}23$; $\beta_{11} = 0{,}12$; $\beta_{20} = -0{,}61$; $\beta_{02} = -0{,}76$; $\sigma = 1{,}7$								
Przed agregacją								
	β_{00}	$S(\beta_{00})$	β_{10}	$S(\beta_{10})$	β_{01}	$S(\beta_{01})$	β_{11}	$S(\beta_{11})$
Średnie	10,0011	0,7895	6,5237	0,1789	6,2301	0,1970	0,1235	0,0168
Odchylenia	0,6533	0,0126	0,1665	0,0029	0,1826	0,0031	0,0106	0,0003
Średnie	β_{20}	$S(\beta_{20})$	β_{02}	$S(\beta_{02})$	σ	R^2	***I***	***p*-value**
	-0,613	0,0133	-0,763	0,0160	1,7013	0,6879	-0,0018	0,5321
Odchylenia	0,0128	0,0002	0,0181	0,0003	0,0271	0,0084	0,0122	0,2940
Po agregacji								
	β_{00}	$S(\beta_{00})$	β_{10}	$S(\beta_{10})$	β_{01}	$S(\beta_{01})$	β_{11}	$S(\beta_{11})$
Średnie	10,1259	1,1106	6,5065	0,2535	6,1859	0,2749	0,1266	0,0230
Odchylenia	0,9207	0,0719	0,1980	0,0164	0,2558	0,0178	0,0134	0,0015
Średnie	β_{20}	$S(\beta_{20})$	β_{02}	$S(\beta_{02})$	σ	R^2	***I***	***p*-value**
	-0,612	0,0190	-0,760	0,0225	0,9297	0,08859	-0,0116	0,5956
Odchylenia	0,0178	0,0012	0,0242	0,0015	0,0602	0,0138	0,0328	0,2666

Źródło: opracowanie własne.

Tab. 7. Szacunki parametrów modelu z trendem stopnia 2. oraz autoregresją

Model: $Y(p) = \rho \mathbf{W} Y(p) + \beta_{00} + \beta_{10} p_1 + \beta_{01} p_2 + + \beta_{11} p_1 p_2 + \beta_{20} p_1^2 + \beta_{02} p_2^2 + \varepsilon(p)$; $\varepsilon(p) \sim \mathrm{N}(0, \sigma)$								
Parametry przyjęte do symulacji								
$\rho = 0{,}8$; $\beta_{00} = 10$; $\beta_{10} = 6{,}51$; $\beta_{01} = 6{,}23$; $\beta_{11} = 0{,}12$; $\beta_{20} = -0{,}61$; $\beta_{02} = -0{,}76$; $\sigma = 1{,}7$								
Przed agregacją								
	ρ	$S(\rho)$	β_{00}	$S(\beta_{00})$	β_{10}	$S(\beta_{10})$	β_{01}	$S(\beta_{01})$
Średnie	0,7490	0,0163	12,099	0,7894	6,1350	0,1788	5,8503	0,1970
Odchylenia	0,0124	0,0006	1,3568	0,0094	0,2626	0,0021	0,2604	0,0023
	β_{11}	$S(\beta_{11})$	β_{20}	$S(\beta_{20})$	β_{02}	$S(\beta_{02})$	σ	R^2
Średnie	0,1135	0,0169	-0,576	0,0133	-0,713	0,0160	1,7010	0,6579
Odchylenia	0,0221	0,0002	0,0271	0,0002	0,0316	0,0002	0,0203	0,0198
Po agregacji								
	ρ	$S(\rho)$	β_{00}	$S(\beta_{00})$	β_{10}	$S(\beta_{10})$	β_{01}	$S(\beta_{01})$
Średnie	0,4012	0,0527	27,309	1,8175	14,992	0,4148	15,234	0,4498
Odchylenia	0,0716	0,0032	5,1802	0,0272	1,5551	0,0062	1,5484	0,0067
	β_{11}	$S(\beta_{11})$	β_{20}	$S(\beta_{20})$	β_{02}	$S(\beta_{02})$	σ	R^2
Średnie	0,2440	0,0377	-1,394	0,0311	-0,717	0,0368	1,5214	0,9340
Odchylenia	0,0646	0,0006	0,1534	0,0005	0,1956	0,0005	0,0227	0,0149

Źródło: opracowanie własne.

5. AGREGACJA DANYCH A ZALEŻNOŚCI MIĘDZY PROCESAMI PRZESTRZENNYMI

Analiza zależności między procesami o różnych strukturach wewnętrznych na podstawie danych wygenerowanych na dwóch różnych poziomach agregacji pozwoliła na zaobserwowanie pewnych prawidłowości. Na najważniejsze z nich wskazano poniżej.

Klasyczne modele regresji $Y(p)$ względem $X(p)$ poprawnie identyfikują czyste zależności pomiędzy badanymi procesami jedynie w sytuacji, gdy oba procesy są białymi szumami lub procesami autoregresyjnymi o jednakowym rzędzie i poziomie autozależności. Nie ma przy tym znaczenia, czy analizy dokonuje się na poziomie danych przed agregacją, czy też po agregacji. Dla białych szumów klasyczny model przestrzenny jest oczywiście tożsamy z modelem zgodnym.

Obecność trendów w badanych procesach zakłóca pomiar faktycznej zależności między $Y(p)$ i $X(p)$ za pomocą modelu klasycznego. Współczynnik regresji w takim modelu mierzy nie tylko bezpośredni, czysty wpływ procesu objaśniającego na proces objaśniany, ale także zależność między trendami rozważanych procesów. Gdy trendy są jednakowe, wartość współczynnika wzrasta, gdy zaś są one różne wartość ta zazwyczaj maleje. Nie zaobserwowano istotnych różnic pomiędzy współczynnikami regresji dla danych przed agregacją i po agregacji w modelach dla procesów posiadających jednakowe trendy. Natomiast w modelach dla procesów posiadających różne trendy różnice takie pojawiały się. Prawidłowości tych nie zmienia pojawienie się dodatkowo jednakowych autozależności w rozważanych procesach ($\rho_x = \rho_y$).

Różny poziom autozależności w poszczególnych procesach powoduje, że współczynniki regresji w modelach klasycznych dla danych przed agregacją i po agregacji różnią się. Zauważono, iż w sytuacji gdy współczynnik autoregresji modelu podstawowego procesu $Y(p)$ jest większy niż współczynnik autoregresji w modelu procesu $X(p)$, tj.: $\rho_y > \rho_x$, współczynnik regresji wzrasta przy przejściu z niższego do wyższego poziomu agregacji danych. Z kolei, współczynnik regresji maleje przy agregacji danych, gdy pochodzą one z procesów o różnym poziomie autoskorelowania, przy czym $\rho_y < \rho_x$.

Tab. 8. Średnie wartości współczynników regresji w modelach klasycznych przed i po agregacji

$Y(p)$ / $X(p)$	M 1. M 7.	M 2. M 8.	M 3. M 9.	M 4. M 10.	M 5. M 11.	M 6. M 12.
M 1.	0,7001					
M 1.	0,7012					
M 2.		0,7030	0,5450			
M 2.		0,6923	0,4740			
M 3.		0,8473				
M 3.		0,9775				
M 4.		0,0874		0,9602		
M 4.		0,0041		0,9881		
M 5.				1,2150		
M 5.				1,3144		
M 6.					0,9795	0,5369
					0,9893	0,5342
M 6.					0,0932	0,0564
					0,0192	0,0085
M 7.					1,1301	0,7109
M 7.					1,3370	0,7151
M 8.					1,7417	
					1,8223	
M 8.					0,1720	
					0,0148	
M 9.					2,1674	
M 9.					2,3919	
M 10.	0,9742					
M 10.	0,9929					
M 11.	-0,0140					
M 11.	-0,0271					
M 12.			0,9732	0,5387		
M 12.			0,9867	0,5382		
M 13.			-0,0149	-0,0077		
M 13.			-0,0268	-0,0147		
M 14.			1,6723			
M 14.			1,7861			
M 15.			-0,0233			
M 15.			-0,0464			

[*] *Współczynniki uzyskane po agregacji danych zaznaczono ciemniejszym tłem.*
Źródło: opracowanie własne.

Gdy struktury procesów $Y(p)$ oraz $X(p)$ są bardziej rozbudowane, tj. posiadają zarówno składniki autoregresyjne, jak i trendowe, związane z agregacją zmiany współczynników regresji w modelach klasycznych są wypadkową zmian autozależności oraz zmian podobieństwa trendów w poszczególnych procesach. Na przykład, gdy $\rho_y > \rho_x$ oraz $\alpha_{00} = \beta_{00}$, $\alpha_{10} = \beta_{10}$, $\alpha_{01} = \beta_{01}$, (rozważane procesy wykazują jednakowe zmiany trendowe), to współczynnik regresji był zawyżony, a agregacja powodowała dodatkowy jego wzrost. Z kolei, gdy np. $\rho_y < \rho_x$ oraz proces $Y(p)$ wykazywał trend stopnia 2, natomiast proces $X(p)$ – trend stopnia 1, to współczynnik regresji był zaniżony, a agregacja powodowała dodatkowy spadek tego współczynnika.

Tab. 9. Średnie wartości współczynników regresji w modelach zgodnych przed i po agregacji

$Y(p)$ / $X(p)$	M 1. M 7.	M 2. M 8.	M 3. M 9.	M 4. M 10.	M 5. M 11.	M 6. M 12.
M 1.	0,7001					
M 1.	0,7012					
M 2.		0,6978	0,7001			
M 2.		0,6947	0,5213			
M 3.		0,7026				
M 3.		0,8946				
M 4.		0,6894		0,6962		
M 4.		0,7379		0,6306		
M 5.				0,7006		
M 5.				0,6612		
M 6.					0,7003	0,7022
					0,7176	0,5127
M 6.					0,7012	0,7048
					0,6853	0,5154
M 7.					0,6991	0,6976
M 7.					0,7169	0,5294
M 8.					0,6924	
					0,8996	
M 8.					0,6974	
					0,8976	
M 9.					0,6938	
M 9.					0,9117	
M 10.	0,6925					
M 10.	0,7058					
M 11.	0,6950					
M 11.	0,5840					
M 12.			0,7064	0,6956		
M 12.			0,7183	0,5181		
M 13.			0,6922	0,6982		
M 13.			0,5797	0,4165		
M 14.			0,6980			
M 14.			0,8943			
M 15.			0,6977			
M 15.			0,8296			

Źródło: opracowanie własne.

Klasyczne modele przestrzenne budowane dla procesów o zadanych strukturach wewnętrznych nie mają wartości poznawczej z jeszcze jednego ważnego powodu, mianowicie w resztach tych modeli pojawia się autokorelacja.

Quasi-zgodne modele przestrzenne okazały się znacznie bardziej efektywne w odkrywaniu faktycznych zależności między rozważanymi procesami. Konstrukcja modelu quasi-zgodnego na podstawie wygenerowanych danych we wszystkich wypadkach zapewniła otrzymanie prawdziwej wartości współczynnika γ. Wartości współczynnika γ otrzymane w modelach quasi-zgodnych po agregacji danych zależały od struktury procesów $Y(p)$ oraz $X(p)$. Dla procesów o strukturze jedynie trendowej, w zasadzie bez względu na stopnie trendu i ich parametry, wartość współczynnika γ oscylowała wokół prawdziwej wartości 0,7. W warunkach obecności autokorelacji w danych wartość współczynnika γ była po agregacji zawyżona, gdy $\rho_y > \rho_x$, a zaniżona, gdy $\rho_y < \rho_x$.

Podsumowując ocenę quasi-zgodnego modelu przestrzennego jako narzędzia odkrywania prawdziwych, czystych zależności między procesami, należy stwierdzić, że spełnia on takie zadanie, o ile modelowy opis struktur poszczególnych procesów jest prawidłowy. Modele quasi-zgodne dla danych pierwotnych i zagregowanych w warunkach autokorelacji przestrzennej będą się różnić przede wszystkim dlatego, że macierz powiązań przestrzennych (macierz **W**) w modelu zagregowanym nie odzwierciedla powiązań danych przed agregacją. Zatem wygenerowane na poziomie podstawowym autozależności nie mogą zostać precyzyjnie opisane (wyodrębnione w modelu) na poziomie zagregowanym.

6. PRZYKŁAD EMPIRYCZNY

Zbadano zależność między stopą bezrobocia a nakładami inwestycyjnymi (w zł) zarejestrowanymi w Polsce w 2007 r. w układzie powiatów i podregionów. Dane pochodzą ze strony internetowej GUS (www.gov.stat.pl). Wykorzystano koncepcję modelu zgodnego. Zbadano struktury trendowo-autoregresyjne poszczególnych procesów na obu poziomach agregacji danych konstruując odpowiednie modele podstawowe. Następnie zbudowano modele quasi-zgodne opisujące zależności między badanymi procesami i porównano otrzymane wyniki.

Tabele 10 – 11 przedstawiają wybrane wyniki badania bezrobocia i inwestycji na poziomie powiatów, natomiast tabele 12 – 13 dotyczą podregionów.

Analiza bezrobocia i inwestycji na poziomie powiatów wykazała przestrzenne trendy stopnia 1. oraz przestrzenne autozależności rzędu 1. w obu procesach (patrz tabela 10). Z ustaleń tych wynika postać pełnego modelu quasi-zgodnego. Model ten zawiera nieistotne zmienne, dlatego dokonano ich eliminacji metodą selekcji *a posteriori*, uzyskując model zredukowany (patrz tabela 11). Ostateczny model nie zawiera trendu oraz przesuniętych przestrzennie inwestycji. Zatem na poziom bezrobocia w danym powiecie wpływają nakłady inwestycyjne ponoszone w tymże powiecie (każde wydatkowane sto złotych powoduje spadek stopy bezrobocia średnio o 0,09 punktu procentowego) oraz poziom bezrobocia w powiatach sąsiadujących (zmiana stopy bezrobocia w danym powiecie o około 0,89 punktu procentowego jest związana z jednoprocentową zmianą stopy bezrobocia w powiatach sąsiadujących).

Tab. 10. Charakterystyki modeli z trendem liniowym i autoregresją dla powiatów.

Model podstawowy dla bezrobocia: $Y(p) = \beta_{00} + \beta_{10}p_1 + \beta_{01}p_2 + \rho_y \mathbf{W}Y(p) + \varepsilon_y(p)$				
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
ρ_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: $X(p) = \alpha_{00} + \alpha_{10}p_1 + \alpha_{01}p_2 + \rho_x \mathbf{W}X(p) + \varepsilon_x(p)$				
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
ρ_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: opracowanie własne.

Tab. 11. Charakterystyki modeli quasi-zgodnych dla powiatów

Model pełny: $Y(p) = \beta^*_{00} + \beta^*_{10}p_1 + \beta^*_{01}p_2 + \rho\mathbf{W}Y(p) + \gamma X(p) + \gamma^*\mathbf{W}X(p) + \varepsilon(p)$				
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
ρ = 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: $Y(p) = \beta^*_{00} + \rho\mathbf{W}Y(p) + \gamma X(p) + \varepsilon(p)$				
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
ρ = 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: opracowanie własne.

Bezrobocie i inwestycje na poziomie podregionów zostały opisane za pomocą przestrzennych modeli autoregresyjnych bez trendów (patrz tabela 12). Obecność nieistotnych składników w quasi-zgodnym modelu pełnym spowodowała, również w tym wypadku, konieczność jego redukcji (patrz tabela 13). Redukcji uległy nieistotne przesunięte przestrzennie inwestycje. Struktura modelu zredukowanego dla podregionów jest analogiczna jak dla powiatów.

Tab. 12. Charakterystyki modeli autoregresyjnych dla podregionów

Model podstawowy dla bezrobocia: $Y(p) = \beta_{00} + \rho_y \mathbf{W}Y(p) + \varepsilon_y(p)$				
Parametry	**Szacunki parametrów**	**Błędy standardowe**	**Statystyki Z**	**Pr(>|Z|)**
β_{00}	5,6762	1,5715	3,6120	0,0003
ρ_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: $X(p) = \alpha_{00} + \rho_x \mathbf{W}X(p) + \varepsilon_x(p)$				
Parametry	**Szacunki parametrów**	**Błędy standardowe**	**Statystyki Z**	**Pr(>|Z|)**
α_{00}	1889,74	476,03	3,9698	0,0000
ρ_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: opracowanie własne.

Tab. 13. Charakterystyki modeli quasi-zgodnych dla podregionów

Model pełny: $Y(p) = \beta_{00}^{*} + \rho \mathbf{W}Y(p) + \gamma X(p) + \gamma^{*} \mathbf{W}X(p) + \varepsilon(p)$				
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
ρ = 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: $Y(p) = \beta_{00}^{*} + \rho \mathbf{W}Y(p) + \gamma X(p) + \varepsilon(p)$				
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
ρ = 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: opracowanie własne.

Współczynnik mierzący wpływ inwestycji na stopę bezrobocia w podregionie różni się od analogicznego współczynnika ocenianego na poziomie powiatów. Wynik ten jest zbieżny z rezultatami przeprowadzonych badań, uzyskanymi w oparciu o dane generowane. Głównym powodem zaobserwowanej różnicy jest autokorelacja przestrzenna bezrobocia, a także inwestycji. Tak jak należało się spodziewać, wartość współczynnika γ dla podregionów jest większa niż dla powiatów $\rho_y > \rho_x$. Z kolei, współczynnik autoregresji ρ mierzący związki między stopami bezrobocia na poziomie podregionów jest wyraźnie mniejszy niż analogiczny parametr obliczony dla powiatów.

7. PODSUMOWANIE

Na podstawie przeprowadzonych badań można sformułować następujące wnioski ogólne:

1. Własności i struktury procesów przestrzennych identyfikowane na podstawie danych pierwotnych i danych zagregowanych mogą różnić się. W szczególności zmienia się siła i struktury autozależności.
2. Zmiany własności i struktur procesów przestrzennych, zachodzące pod wpływem agregacji danych mają wpływ na zależności między procesami.
3. Quasi-zgodne modele przestrzenne nie muszą zapewnić odkrycia faktycznej zależności między badanymi procesami. Należy doprecyzować opis autozależności za pomocą odpowiednich macierzy powiązań przestrzennych.
4. Nie można bezpośrednio przenosić wyników badań uzyskanych z danych zagregowanych na zależności procesów i zjawisk zachodzące na poziomie mniejszych jednostek przestrzennych. Nie można zatem ulec pokusie naiwnej interpretacji, przenoszącej wyniki tzw. makro-szacunków na mikro-relacje.

LITERATURA

Arbia G., [1988], *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*, Kluwer Academic Press, Dordrecht.

Kufel T., [2002], *Postulat zgodności w dynamicznych modelach ekonometrycznych*, Wydawnictwo UMK, Toruń.

Szulc E., [2007], *Ekonometryczna analiza wielowymiarowych procesów gospodarczych*, Wydawnictwo UMK, Toruń.

Zieliński Z., [1990], *Liniowe modele ekonometryczne jako narzędzie opisu i analizy przyczynowych zależności zjawisk ekonomicznych*, Wydawnictwo UMK, Toruń.

MODELING OF DEPENDENCE BETWEEN SPATIAL ECONOMIC PROCESSES AND THE LEVEL OF DATA AGGREGATION

Statistic and econometric analyses of spatial phenomena use the data of different levels of spatial aggregation. It has particular consequences for the possibilities of discovering different properties and structures of the spatial processes. These properties are the basis for the proper specification of the spatial models of the dependence between the processes.

In the paper it is discussed how the spatial aggregation affects the change of the properties and the structures of the spatial processes. In particular the differences as regards the character and the strength of the spatial aggregation influence on the separate components of the processes are shown.

The main part of the paper is the discussion on the data aggregation influence on the modelling of the dependence between the spatial processes. A special attention is paid to the econometric modelling of the phenomena observed at different levels of the data aggregation with the presence of the spatial autocorrelation. The important result of the discussion is to point at the problem of the "naive" spatial models interpretation.

The considerations are based on the generated data and they are illustrated with an empirical example.