



Ewa Pośpiech

Uniwersytet Ekonomiczny w Katowicach
Wydział Zarządzania
Katedra Matematyki
posp@ue.katowice.pl

MODELOWANIE PRZESTRZENNE CHARAKTERYSTYK RYNKU PRACY

Streszczenie: W artykule dokonano oceny potrzeby zastosowania liniowego modelu przestrzennego dla wybranych charakterystyk rynku pracy: pracujących ogółem, pracujących kobiet oraz pracujących mężczyzn. Jako zmienne objaśniające modelu uwzględniono zmienne demograficzne oraz zmienną z kategorii Wynagrodzenia. Badano, czy włączenie do uzyskanych modeli czynnika przestrzennego poprawi ich jakość. W rozważaniach wzięto pod uwagę dwa podstawowe modele przestrzenne: model błędu przestrzennego oraz model opóźnienia przestrzennego; w wyniku przeprowadzonej analizy wskazano i oszacowano model przestrzenny odpowiedni dla badanych zmiennych.

Słowa kluczowe: modelowanie przestrzenne, autokorelacja przestrzenna reszt, charakterystyki rynku pracy.

Wprowadzenie

Wykorzystywanie metod przestrzennych w analizach społeczno-ekonomicznych spotyka się z coraz większym zainteresowaniem wśród badaczy oraz praktyków zajmujących się na co dzień analizami regionalnymi. Metody te znajdują zastosowanie w badaniach zjawisk, na których kształtowanie ma wpływ lokalizacja oraz interakcje przestrzenne. Należą do nich m.in. procesy demograficzne, zjawiska zachodzące na rynku pracy, poziom płac, rozwój gospodarczy¹; czynnik przestrzenny odgrywa coraz większą rolę w ich opisie i wyjaśnianiu.

Jedną z najczęściej uwzględnianych charakterystyk rynku pracy jest stopa bezrobocia, jednak równie istotnymi charakterystykami są poziomy zatrudnienia

¹ Inne zastosowania można znaleźć m.in. w [Pietrzykowski, 2011].

(ogółem oraz w podziale według różnych cech). Identyfikacja czynników, które kształtują poziomy rozpatrywanych cech jest ważnym elementem polityki regionalnej pozwalającym na monitorowanie zjawiska i zarządzanie nim.

Celem artykułu jest ocena potrzeby modelowania przestrzennego oraz identyfikacja i estymacja odpowiedniego modelu przestrzennego, którego zmiennymi endogenicznymi są wybrane zmienne charakteryzujące rynek pracy. W pierwszej części artykułu przedstawiono teoretyczny opis narzędzi uwzględnianych w analizie przestrzennej, natomiast w drugiej – zaprezentowano wyniki przeprowadzonych badań i uzyskane wnioski.

1. Modelowanie przestrzenne

Uwzględnienie w modelowaniu czynnika przestrzennego (co oznacza wzięcie pod uwagę powiązań przestrzennych w ramach danego obszaru, ale także w ramach obszarów sąsiednich) może pozytywnie wpływać na tłumaczenie zmienności rozważanych cech oraz poprawić jakość modelu ekonometrycznego (u podstaw tych rozważań jest prawo, które podał W. Tobler, głoszące, iż lokalizacje sąsiednie wykazują większe podobieństwo pod względem badanej cechy niż lokalizacje dalej od siebie położone [Tobler, 1970]).

1.1. Modele przestrzenne

Konieczność uwzględnienia czynnika przestrzennego w modelu ekonometrycznym skutkuje uwzględnieniem przestrzennych metod estymacji. Wyróżnia się następujące podstawowe grupy modeli przestrzennych: modele opóźnienia przestrzennego, modele błędu przestrzennego, modele regresji krzyżowej oraz warianty mieszane. W rozważaniach brano pod uwagę dwa pierwsze modele i te dwa są poniżej przedstawione. W wymienionych modelach macierz sąsiedztwa ma różny wpływ na zmienną objaśnianą.

- Model opóźnienia przestrzennego – jego składową jest tzw. przestrzennie opóźniona zmienna endogeniczna $\mathbf{W}y$ (model autoregresyjny, podstawą modelu jest zależność przestrzenna). Model ten w notacji macierzowej przyjmuje postać:

$$\mathbf{y} = \rho \mathbf{W}y + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (1)$$

gdzie:

ρ – współczynnik autokorelacji przestrzennej,

\mathbf{W} – macierz wag przestrzennych,

$\boldsymbol{\beta}$ – wektor współczynników modelu,

\mathbf{X} – macierz zmiennych egzogenicznych.

- Model błędu przestrzennego – jego składową jest opóźniony przestrzennie błąd $\mathbf{W}\xi$; model ten zakłada autokorelację przestrzenną reszt modelu. W notacji macierzowej przedstawia się następująco:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi} \quad (2)$$

$$\boldsymbol{\xi} = \lambda\mathbf{W}\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \quad (3)$$

gdzie:

λ – współczynnik autokorelacji przestrzennej,

pozostałe oznaczenia jw.

1.2. Etapy wyboru modelu przestrzennego

Istnieją różne sposoby mające na celu wybór i specyfikację modelu najlepiej opisującego dane zagadnienie. W przypadku doboru modelu przestrzennego także istnieje kilka możliwości specyfikacji takiego modelu. Jedną z nich jest prosta procedura wyboru modelu regresji przestrzennej, która przebiega według schematu [Anselin, 2006; Suhecki (red.), 2010; Kopczewska, 2011]:

1. Estymacja modelu liniowego KMNK.
2. Diagnostyka modelu pod względem występowania autokorelacji przestrzennej reszt modelu – na tym etapie wykorzystywany jest test I Morana dla reszt, który przyjmuje postać:

$$I = \frac{n}{S_0} \cdot \frac{\mathbf{u}^T \mathbf{W} \mathbf{u}}{\mathbf{u}^T \mathbf{u}}, \quad (4)$$

gdzie:

n – liczba rozważanych regionów,

\mathbf{u} – n -wymiarowy kolumnowy wektor reszt modelu wyznaczonego MNK,

S_0 – suma elementów macierzy wag, $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$,

\mathbf{W} – macierz wag $[w_{ij}]$, $i = 1, \dots, n, j = 1, \dots, n$, o elementach postaci

$$w_{ij} = \begin{cases} 1, & \text{gdy jest obiekty } i \text{ oraz } j \text{ mają wspólną granicę} \\ 0, & \text{gdy obiekty } i \text{ oraz } j \text{ nie mają wspólnej granicy.} \\ 0, & \text{dla elementów diagonalnych} \end{cases} \quad (5)$$

Za pomocą tego testu weryfikowane są następujące hipotezy:

H_0 : brak autokorelacji przestrzennej reszt.

H_1 : występowanie autokorelacji przestrzennej reszt.

Występowanie autokorelacji reszt nie musi od razu oznaczać konieczności stosowania modeli przestrzennych – przyczynami takiego stanu rzeczy może być np. estymowanie zależności nieliniowej modelem liniowym. Można wówczas przeanalizować reszty modelu szacowanego na logarytmach zmiennych [Cliff i Ord, 1970] – gdy autokorelacja nadal występuje, może to oznaczać pominięcie istotnych zmiennych w modelu. Przydatnym narzędziem jest wówczas graficzne przedstawienie reszt² oraz analiza ich rozkładu za pomocą testu join-count – brak losowości rozmieszczenia reszt oznaczać będzie istnienie innych czynników wpływających na kształtowanie się zmiennej objaśnianej.

- Przeprowadzenie testów diagnostycznych LM (mnożnika Lagrange'a) w celu wyspecyfikowania modelu przestrzennego lepiej opisującego zjawisko – testy te są oparte na resztach MNK. Wyróżnia się testy zwykłe oraz odporne. W wersji zwykłej statystyka testowa LM dla modelu błędu przestrzennego LM_{ERROR} o rozkładzie asymptotycznym $\chi^2(1)$ przyjmuje postać:

$$LM_{ERROR} = \frac{1}{T_1} \left(\frac{\mathbf{u}^T \mathbf{W} \mathbf{u}}{\hat{\sigma}^2} \right)^2, \quad (6)$$

gdzie:

$\hat{\sigma}$ – oszacowany błąd standardowy,

T_1 – parametr wyrażony wzorem $T_1 = \text{tr}[(\mathbf{W}^T + \mathbf{W})\mathbf{W}]$,

pozostałe oznaczenia jw.

Za pomocą powyższej statystyki weryfikowana jest hipoteza o istotności współczynnika λ ($H_0: \lambda = 0$, $H_1: \lambda \neq 0$).

Statystyka testowa dla modelu opóźnienia przestrzennego LM_{LAG} także ma asymptotyczny rozkład normalny $\chi^2(1)$. Wyraża się wzorem:

² Analiza reszt w podziale na dodatnie i ujemne umożliwia zidentyfikowanie obszarów, na których model nie doszacowuje wartości badanego zagadnienia (reszty dodatnie) lub przeszacowuje (reszty ujemne).

$$LM_{LAG} = \frac{1}{T_2} \left(\frac{\mathbf{u}^T \mathbf{W} \mathbf{y}}{\hat{\sigma}^2} \right)^2 \quad (7)$$

gdzie:

$$T_2 = T_1 + \frac{(\mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}})^T \mathbf{M} (\mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}})}{\hat{\sigma}^2} \quad (8)$$

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \quad (9)$$

oraz

$\hat{\boldsymbol{\beta}}$ – oszacowany wektor współczynników,

pozostałe oznaczenia jw.

Statystyka dana wzorem (7) umożliwia weryfikację hipotezy o istotności współczynnika ρ ($H_0: \rho = 0$, $H_1: \rho \neq 0$).

O wyborze modelu przestrzennego (według strategii hybrydowej) decyduje wyższa (istotna) wartość statystyki LM (jeśli $LM_{LAG} > LM_{ERROR}$, to wyznacza się model opóźnienia przestrzennego, w przeciwnym wypadku – model błędu przestrzennego). Jeśli nie uzyskuje się jednoznacznej informacji opierając się na zwykłych testach, analogiczne wnioski wyciągane są na podstawie testów odpornych³.

4. Przeprowadzenie dodatkowych testów diagnostycznych: ilorazu wiarygodności LR i testu Walda oraz porównanie wartości statystyk tych testów z wartościami statystyki LM . O prawidłowej specyfikacji modelu świadczą następujące zależności:
 - dla modelu błędu przestrzennego: $Wald(\lambda) \geq LR_{ERROR} \geq LM_{ERROR}$,
 - dla modelu opóźnienia przestrzennego: $Wald(\rho) \geq LR_{LAG} \geq LM_{LAG}$.

2. Analiza empiryczna

2.1. Uwarunkowania badań

Badaniem objęto powiaty Polski; w rozważaniach uwzględnione zostały lata 2005 oraz 2013. Jako zmienne objaśniane przyjęto następujące zmienne charakteryzujące rynek pracy:

- Liczba osób pracujących ogółem (PO),

³ Statystyki testowe dla testów odpornych można znaleźć np. w [Suchecki (red.), 2010].

- Liczba pracujących kobiet (PK),
 - Liczba pracujących mężczyzn (PM),
- natomiast jako zmienne objaśniające wybrano zmienną opisującą kategorię Wynagrodzenia oraz zmienne z kategorii Ludność:
- Przeciętne miesięczne wynagrodzenia brutto (PW),
 - Wskaźnik obciążenia demograficznego (OD),
 - Gęstość zaludnienia na 1 km² (G),
 - Ludność w wieku przedprodukcyjnym (Lpr),
 - Ludność w wieku produkcyjnym (L),
 - Ludność w wieku poprodukcyjnym (Lpo).

Dane wykorzystane w analizach pochodzą z Banku Danych Lokalnych GUS, a zaprezentowane rysunki oraz obliczenia wykonano w programach MS Excel oraz R CRAN.

2.2. Szacowanie liniowych modeli przestrzennych

Pierwszy etap analiz obejmował estymację liniowych modeli ekonometrycznych, które zbudowane zostały dla wyszczególnionych lat 2005 oraz 2013. W tabelach 1 i 2 zamieszczono wyniki szacowania modeli ekonometrycznych (nieistotny parametr oznaczono symbolem „*”) wraz ze współczynnikiem dopasowania R^2 .

Tabela 1. Wyniki estymacji modeli ekonometrycznych (2005 r.)

Zmienne	Wartości współczynników		
	Model (PO)	Model (PK)	Model (PM)
Wyraz wolny	18 100	11 270	5 078,56*
PW	6,48	1,73	3,84
OD	-1,47	-7,04	-615,28
G	-	-	-
Lpr	-1,20	-0,65	-0,79
L	-	-	0,14
Lpo	2,80	1,44	1,02
R^2	0,961	0,9652	0,9536

Tabela 2. Wyniki estymacji modeli ekonometrycznych (2013 r.)

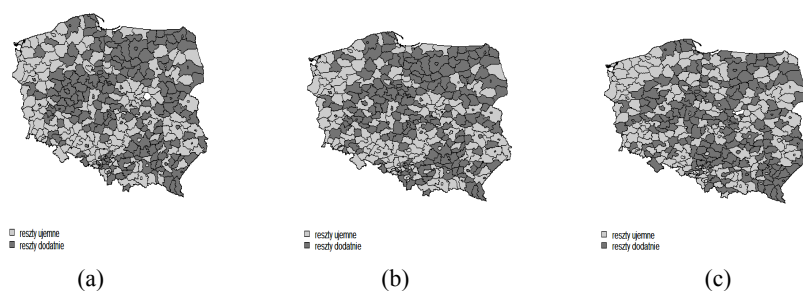
Zmienne	Wartości współczynników		
	Model (PO)	Model (PK)	Model (PM)
Wyraz wolny	14 754,98	12 410	2 343,99*
PW	6,34	2,23	4,11
OD	-1 533,16	-841,7	-691,46
G	4,03	2,09	1,94
Lpr	1,89	1,16	0,73
L	-1,02	-0,65	-0,38
Lpo	3,69	2,16	1,53
R^2	0,9641	0,9670	0,9569

W roku 2005 dla wszystkich modeli nieistotna była zmienna G, natomiast zmienna L okazała się istotna jedynie w modelu PM. Współczynniki determinacji wszystkich modeli kształtowały się na bardzo wysokim poziomie, dopasowanie modeli jest zatem bardzo dobre. W kolejnym kroku analiz zbadano przestrzenną autokorelację błędów. Wyniki przedstawia tabela 3.

Tabela 3. Statystyki I Morana dla reszt modelu

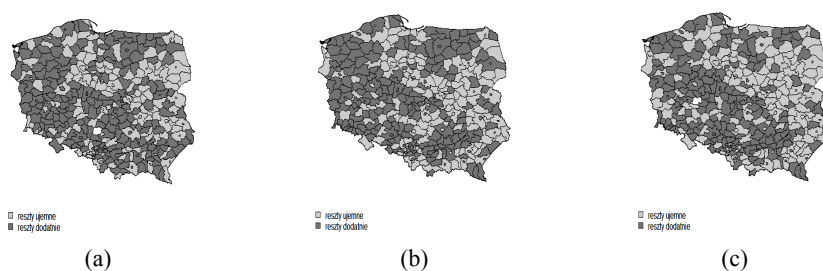
Rok	Model (PO)		Model (PK)		Model (PM)	
	I	p -value	I	p -value	I	p -value
2005	-7,9E-04	0,46	3,7E-03	0,41	0,01	0,34
2013	-0,03	0,75	-0,42	0,86	5,8E-06	0,45

Dla wszystkich oszacowanych modeli statystyka I Morana jest statystycznie nieistotna, co sugeruje brak autokorelacji przestrzennej, czyli losowe rozłożenie reszt (rozmięszczenie nie zależy od położenia). Na rysunkach 1 i 2 przedstawiono rozkład reszt modeli w podziale na reszty dodatnie i ujemne.



Rys. 1. Reszty modeli PO – (a), PK – (b), PM – (c) w podziale na dodatnie i ujemne (2005 r.)

Źródło: Opracowanie przy użyciu programu R CRAN.



Rys. 2. Reszty modeli PO – (a), PK – (b), PM – (c) w podziale na dodatnie i ujemne (2013 r.)

Źródło: Opracowanie przy użyciu programu R CRAN.

Za pomocą testu join-count, przeanalizowano pod względem losowości reszty dodatnie oraz ujemne uzyskując następujące wyniki (tabela 4).

Tabela 4. Wyniki analizy losowości reszt dodatnich oraz ujemnych

Model	Rozłożenie reszt dodatnich			Rozłożenie reszt ujemnych		
	Wartość statystyki	p-value	wniosek	Wartość statystyki	p-value	wniosek
PO (2005)	0,60	0,27	losowe	3,24	5,9E-04	nielosowe
PK (2005)	-0,12	0,55	losowe	3,09	1,0E-03	nielosowe
PM (2005)	0,47	0,32	losowe	1,89	2,9E-02	nielosowe
PO (2013)	0,73	0,23	losowe	4,88	5,3E-07	nielosowe
PK (2013)	0,22	0,41	losowe	4,71	1,3E-06	nielosowe
PM (2013)	1,02	0,15	losowe	4,45	4,3E-06	nielosowe

Reszty dodatnie wszystkich modeli wykazują losowe rozmieszczenie, natomiast reszty ujemne nie są rozmieszczone losowo – wykazują przestrzenną autokorelację. Dla każdego modelu przeprowadzono testy *LM*, których zadaniem jest zbadanie istnienia zależności przestrzennych w modelach liniowych. Wyniki tej analizy (wskazanie, który z modeli przestrzennych byłby bardziej odpowiedni) zawarto w tabeli 5.

Tabela 5. Rekomendowany model przestrzenny

Model	Typ modelu przestrzennego
PO (2005)	–
PK (2005)	–
PM (2005)	–
PO (2013)	model opóźnienia przestrzennego
PK (2013)	model opóźnienia przestrzennego
PM (2013)	model opóźnienia przestrzennego (przy poziomie p-value = 0,0752)

Powyższe wyniki sugerują, że w przypadku modeli dla roku 2005 szacowanie modeli przestrzennych nie jest wskazane, natomiast dla roku 2013 najlepiej jest zbudować modele opóźnienia przestrzennego. Dodatkowo, w tabeli 6 przedstawiono wartości kryteriów informacyjnych: Akaike (AIC), bayesowskiego (BIC) oraz logLik, które pozwalają porównać model oszacowany MNK z modelami przestrzennymi (MBP – model błędu przestrzennego, MOP – model opóźnienia przestrzennego). Najlepszy model to ten, dla którego kryteria AIC oraz BIC przyjmują najniższe wartości, natomiast kryterium logLik przyjmuje wartości najwyższe.

Tabela 6. Wartości kryteriów informacyjnych dla modeli z roku 2013

Model		Wartości kryteriów		
		AIC	BIC	logLik
PO	MNK	7960,68	7992,12	-3972,34
	MBP	7962,22	7997,59	-3972,11
	MOP	7956,39	7991,76	-3969,19
PK	MNK	7444,61	7476,05	-3714,30
	MBP	7445,59	7480,96	-3713,80
	MOP	7438,73	7474,10	-3710,37
PM	MNK	7470,47	7501,91	-3727,24
	MBP	7472,47	7507,84	-3727,24
	MOP	7468,61	7503,98	-3725,31

Wartości kryteriów informacyjnych modeli zbudowanych dla zmiennych Pracujący ogółem (PO) oraz Pracujące kobiety (PK) jednoznacznie wskazują na model opóźnienia przestrzennego. W przypadku modelu dla zmiennej Pracujący mężczyźni (PM) kryteria AIC oraz logLik wskazują model opóźnienia przestrzennego, natomiast kryterium BIC wskazuje model ekonometryczny szacowany metodą MNK (w modelu tym nieistotny statystycznie był wyraz wolny, co może wpływać na niespójność ocen). Niemniej jednak, dla każdego modelu z roku 2013 zbudowano sugerowany model przestrzenny; wyniki estymacji zamieszczono w tabeli 7 (gwiazdką oznaczono parametr nieistotny statystycznie).

Tabela 7. Wyniki estymacji modeli opóźnienia przestrzennego (2013 r.)

Zmienne	Wartości współczynników		
	Model (PO)	Model (PK)	Model (PM)
Wyraz wolny	13 299,56	11 575	1 765,71*
PW	6,77	2,46	4,29
OD	-1 500,71	-823,24	-678,42
G	4,19	2,18	2,01
Lpr	2,08	1,28	0,80
L	-1,06	-0,67	-0,39
Lpo	3,68	2,16	1,52
ρ	-0,0478	-0,0512	-0,0409

Dla wszystkich trzech modeli współczynniki ρ są statystycznie istotne, a także prawie wszystkie współczynniki modelu są bardziej istotne niż w modelu szacowanym metodą MNK (wyjątkiem jest wyraz wolny). O właściwej specyfikacji modeli świadczą także wartości statystyk wyznaczone dla poszczególnych modeli (tabela 8).

Tabela 8. Wartości statystyk wybranych testów

Model	Wartości statystyk		
	Wald (ρ)	LR_{LAG}	LM_{LAG}
PO (2013)	7,65	6,29	5,07
PK (2013)	9,69	7,87	6,27
PM (2013)	4,62	3,86	3,17

Podsumowanie

Analizy pokazały, że mimo nieistotnej wartości statystyki I Morana, zasadnym jest w pewnych przypadkach stosowanie modelowania przestrzennego. Zbudowanych zostało sześć modeli za pomocą KMNK. Modele, w których zmienną endogeniczną były odpowiednio: liczba osób pracujących, liczba pracujących kobiet oraz liczba pracujących mężczyzn, nie cechowały się autokorelacją przestrzenną, ale po podziale reszt modelu na dodatnie i ujemne, dla tych drugich autokorelacja była istotna. Sytuacja ta skłoniła do zbadania testami LM możliwości estymowania modeli przestrzennych. Zarekomendowane zostały modele opóźnienia przestrzennego dla modeli z roku 2013. Dobra specyfikacja modeli przestrzennych została potwierdzona wartościami odpowiednich statystyk.

Przeprowadzone badania pokazały, że modelowanie przestrzenne dla omawianych zagadnień jest uzasadnione, chociaż trzeba podkreślić, że analizę należałoby jeszcze pogłębić (np. o jeszcze dokładniejszą identyfikację modeli oraz uwzględnienie innych macierzy sąsiedztwa – w analizach uwzględniono tylko macierz sąsiedztwa pierwszego rzędu). Niemniej jednak kolejne badania ukazują istotność implementacji w analizach zjawisk społeczno-ekonomicznych metod przestrzennych.

Literatura

- Anselin L. (2006), *Spatial Analysis with GeoDa. 4. Spatial Regression*, University of Illinois, Urbana-Champaign.
- Cliff A. & Ord J.K. (1970), *Spatial Autocorrelation: A Review of Existing and New Measures with Applications*, "Economic Geography", Vol. 24.
- Kopczewska K. (2011), *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R Cran*, CeDeWu, Warszawa.
- Kopczewska K., Kopczewski T. i Wójcik P. (2013), *Metody ilościowe w R. Aplikacje ekonomiczne i finansowe*, CeDeWu, Warszawa.
- Mastalerz-Kodzis A. i Pośpiech E. (2015), *Autokorelacja przestrzenna wybranych charakterystyk społeczno-ekonomicznych*, „Metody Ilościowe w Badaniach Ekonomicznych”, T. XVI, nr 4.
- Mastalerz-Kodzis, A. i Pośpiech, E. (2015), *Spatial and Temporal Analysis of Labor Market Characteristics*, "Folia Oeconomica Stetinensia" (na etapie recenzji).
- Pietrzykowski R. (2011), *Wykorzystanie metod statystycznej analizy przestrzennej w badaniach ekonomicznych*, „Roczniki Ekonomiczne Kujawsko-Pomorskiej Szkoły Wyższej w Bydgoszczy”, nr 4.
- Pośpiech E. (2015), *Analiza przestrzenna bezrobocia w Polsce*, „Studia Ekonomiczne”, nr 227, s. 59-74.

Sucheckı B., red. (2010), *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, C.H. Beck, Warszawa.

Tobler W. (1970), *A Computer Model Simulating Urban Growth in Detroit Region*, „Economic Geography”, No. 46(2).

[www 1] <http://stat.gov.pl/bdl> (dostęp: 7.09.2015).

SPATIAL MODELLING OF LABOR MARKET CHARACTERISTICS

Summary: In many cases, when considered phenomena are settled on spatial localization, the spatial factor can be very important – it can help to explain the variation of chosen characteristics and can improve the quality of estimated models. The main purpose of the paper is to identify if it is necessary to use spatial modelling for selected labor market characteristics. In the analyses two spatial models were taken into consideration: the spatial error model and spatial lag model. The models were estimated, verified and the best one was indicated.

Keywords: spatial modelling, spatial autocorrelation of residuals, labor market characteristics.