Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663
www.czasopisma.uni.lodz.pl/foe/
3(335) 2018

DOI: http://dx.doi.org/10.18778/0208-6018.335.05

Ewa Katarzyna Pośpiech
University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, posp@ue.katowice.pl

Adrianna Mastalerz-Kodzis
University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, adrianna.mastalerz-kodzis@ue.katowice.pl

Application of Spatial Regression in Employment Characteristics Modelling

Abstract: The article analyses the employment characteristics. The employment rate was studied in selected regions of Europe, and subsequently, for selected variables: total population employed, women employed and men employed, classic econometric models were constructed and the necessity of including the spatial factor in the process of modelling was verified. The demographic variables and GDP per capita were chosen as explaining variables of the model. It was analysed whether including a spatial approach in the models would improve their quality. Two basic spatial models were taken into consideration: the spatial error model and the spatial lag model, the former of which turned out to be the right tool for the analyses.

Keywords: spatial modelling, spatial error model, spatial lag model, employment

JEL: C33, C51, C52

1. Introduction

Spatial methods are becoming more and more popular amongst analysts dealing with regional studies. These methods are used increasingly to research issues, the development of which depends on geographical location and spatial connections. The importance of including the spatial dependences in one's research is indicated by Tobler's law, according to which neighbouring locations show greater similarity concerning the studied feature than locations further apart from each other (Tobler, 1970).

The issues featuring the importance of spatial dependences are: demographic processes, phenomena taking place on the real estate market or employment market, economic development and other (Overmars, de Koning, Veldkamp, 2003; Pietrzykowski, 2011; Pośpiech, 2015; 2016; Pośpiech, Mastalerz-Kodzis, 2015; 2016; Sikora, Woźniak, 2007; Zeug-Żebro, 2014). The spatial factor is becoming an increasingly important element improving the quality of the description of the researched phenomena and supporting their explanation.

An essential determinant of each regional policy is level of employment. Its rate and the factors forming it provide important information for local authorities. Monitoring of this phenomenon and appropriate administration of it improves, among others, the process of carrying out the right kind of socio-economic policy of a given region and enables the introduction of remedies aiming at prevention of undesirable phenomena.

The aim of this article is to identify the spatial factor in the description of employment level, and in the case of determining the importance of this factor – estimation of an appropriate spatial model, which would improve the quality of the description of the phenomenon. The article consists of two parts: theoretical and empirical. The theoretical part explains the idea of spatial models as well as the stages of their identification, whereas the empirical part illustrates the most important results and conclusions resulting from the analyses.

2. Selected spatial models and stages of their identification

In spatial modelling we use appropriate methods of estimation. In these methods we include matrices of neighbourhood and, depending on which spatial model is used, the matrices influence the explained variables in different ways. We can enumerate the following basic spatial model groups: spatial lag models, spatial error models, spatial cross-regressive models and mixed variants. The models used in the study are the spatial lag and spatial error models, therefore these are presented below.

In the spatial lag model (SLM) the component is the so called spatially lagged explained variable Wy (the model is based on spatial dependence). This model can be written as follows:

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

where:

ρ – spatial autocorrelation coefficient;

W – the matrix of spatial weights $[w_{ij}]$, $i = 1, \ldots, n$, $j = 1, \ldots, n$, of the following elements:

$$w_{ij} = \begin{cases} 1, & \text{when objects } i \text{ and } j \text{ share a border} \\ 0, & \text{when objects } i \text{ and } j \text{ do not share a border}\ ; \\ 0, & \text{for diagonal elements} \end{cases} \tag{2}$$

β – the vector of model coefficients;

X – the matrix of the explaining variables.

In the spatial error model (SEM) the component is the spatially lagged error **W**ξ; this model assumes a spatial autocorrelation of the model residuals. In the matrix form it is written as follows:

$$\mathbf{y} = \mathbf{X}\beta + \xi,\ \xi = \lambda\mathbf{W}\xi + \varepsilon,\ \varepsilon \sim N(0, \sigma^2\mathbf{I}), \tag{3}$$

where:

λ – spatial autocorrelation coefficient; other symbols as in the above.

The identification and the specification of the spatial model takes part in stages (Anselin, 2006; Anselin, Bera, 1998; Arbia, 2006; Cliff, Ord, 1981; Kopczewska, 2011; Suchecki, 2010). The first one is the estimation of linear model using the ordinary least squares method (OLS). The next one is the analysis of the model in consideration to the presence of spatial autocorrelation of the model residuals; in this case the Moran's *I* test for residuals is used, which is:

$$I = \frac{n}{S_0} \cdot \frac{\mathbf{u}^{\mathrm{T}}\mathbf{W}\mathbf{u}}{\mathbf{u}^{\mathrm{T}}\mathbf{u}}, \tag{4}$$

where:

n – the number of included regions;

u – *n*-dimensional column vector of the model residuals;

S_0 – the sum-total of weights matrix elements, $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$; other symbols as in the above.

This test verifies hypotheses concerning the absence (H_0) or the presence (H_1) of spatial autocorrelation of residuals. However, the presence of the autocorrelation of residuals can, for example, mean that the nonlinear dependence is estimated with the use of linear model or that important variables have been omitted in the model. In this situation, a graphic illustration of residuals[1] and an analysis of their distribution with the use of join-count test can help – the non-random residuals distribution will signify the existence of other factors influencing the forming of the explained variable.

The next stage of the procedure is to carry out *LM* (Lagrange multiplier) tests – ordinary and robust ones, in order to make a specification of a spatial model which would describe a phenomenon in a better way. The tests are based on OLS residuals. In the ordinary version, the *LM* test statistic for the spatial error model LM_{ERROR} with asymptotic distribution χ^2 (1) is expressed in the formula:

$$LM_{ERROR} = \frac{1}{T_1}\left(\frac{\mathbf{u}^T \mathbf{W}\mathbf{u}}{\hat{\sigma}^2}\right)^2, \tag{5}$$

where:

$\hat{\sigma}$ – estimated standard error;

T_1 – parameter illustrated in the formula $T_1 = \mathrm{tr}[(W^T + W)W]$; other symbols as in the above.

Statistic (5) enables to verify hypothesis concerning the statistical significance of λ coefficient (H_0: $\lambda = 0$, H_1: $\lambda \neq 0$). Test statistic for spatial lag model LM_{LAG} also has asymptotic distribution χ^2 (1) and takes the following form:

$$LM_{LAG} = \frac{1}{T_2}\left(\frac{\mathbf{u}^T \mathbf{W}\mathbf{y}}{\hat{\sigma}^2}\right)^2, \tag{6}$$

where:

$$T_2 = T_1 + \frac{(\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}})^T \mathbf{M}(\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}})}{\hat{\sigma}^2}; \tag{7}$$

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T; \tag{8}$$

$\hat{\beta}$ – the estimated vector of coefficients; other symbols as in the above.

[1] The analysis of positive and negative residuals allows to identify the areas, in which the model underestimates the values (positive residuals) or overestimates them (negative residuals).

The statistic given in formula (6) enables the verification of the hypothesis concerning the statistical significance of ρ coefficient (H_0: $\rho = 0$, H_1: $\rho \neq 0$). A higher (statistically significant) value of the *LM* statistic indicates the spatial model (if $LM_{LAG} > LM_{ERROR}$, than the spatial lag model is assigned, otherwise – the spatial error model). If unanimous information is not received on the base of ordinary tests, analogical conclusions are drawn on the base of robust ones.

The last stage in the process of the identification and specification of the spatial model is to carry out additional diagnostic tests: likelihood ratio test (*LR*) and Wald's test. The following interdependences indicate a proper model specification – for the spatial error model: $\text{Wald}(\lambda) \geq LR_{ERROR} \geq LM_{ERROR}$, for the spatial lag model: $\text{Wald}(\rho) \geq LR_{LAG} \geq LM_{LAG}$.

3. Empirical analysis

The subject of the study were the regions of chosen European countries, specified on the base of EU classification, NUTS 2 system. Six countries were taken into consideration: Austria, Czech Republic, Germany, Hungary, Poland and Slovakia. The period of the study were years 2011–2014. The following variables were considered:
1) Employment rates at the age 15–64 (%), total – ERT,
2) Employment rates at the age 15–64 (%), males – ERM,
3) Employment rates at the age 15–64 (%), females – ERK,
4) Employment at the age 15–64 (thousands), total – ET,
5) Employment at the age 15–64 (thousands), males – EM,
6) Employment at the age 15–64 (thousands), females – EF,
7) Regional gross domestic product (PPS per inhabitant) – GDP,
8) Population density (persons per square kilometer) – PD,
9) Population, age 15 to 64 (thousands), total – PT,
10) Population, age 15 to 64 (thousands), males – PM,
11) Population, age 15 to 64 (thousands), females – PF.

The data used in the analyses comes from the Eurostat base, whereas the figures and calculations were prepared with the use of MS Excel and R CRAN.

3.1. Employment rate – distribution in regional terms

In the introductory stage of the study an analysis of the employment rate in the years 2011 and 2014 was conducted, for the total of the population and separately, according to gender. The considerations focused on a group of people in their productive age (15–64). Percentage values, divided into four groups, according to the three averages rule, are illustrated in Figures 1 and 2. Average values, minimum and maximum, defined for 2011 and 2014, as well as for the specified categories jointly, are included in Table 1 (marked are: group with the lowest values of average employment rate as group I, the following as group II, the next as group III, and the last one, with the highest levels of employment rate, as group IV).

Table 1. Values of selected characteristics

Characteristic	Value
Minimum	44.3
Mean 1	60.03
Mean 2	68.23
Mean 3	74.33
Maximum	83.7

Source: own study

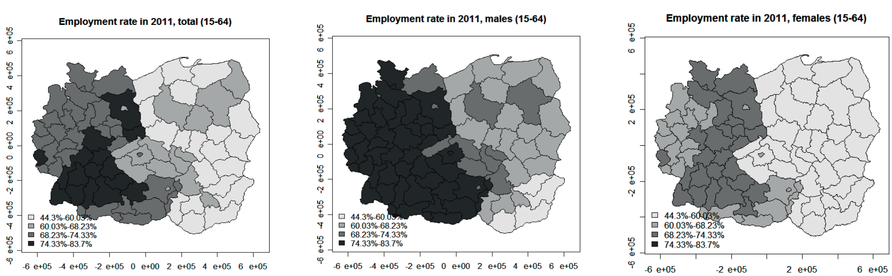

Figure 1. Employment rate grouped into regions and by gender, year 2011

Source: own study

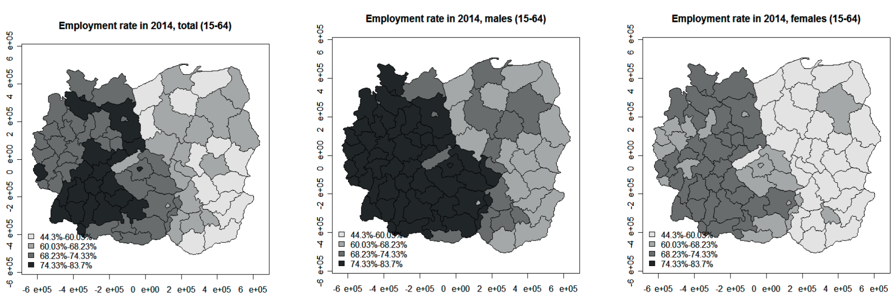

Figure 2. Employment rate grouped into regions and by gender, year 2014

Source: own study

The division of the area into western (south-western) and eastern part is visible in the figures obtained. The regions with the highest employment rate are the regions of Germany, Austria and Czech Republic. The regions of Poland, Slovakia and Hungary feature lower values of the index level. An important difference between the values of female and male employment rates is noticeable (the difference is about 10% in favour of men). Generally, throughout the considered four years, the employment rate increased. Assuming fixed interval limits, we can compare changes which took place in the level of employment in 2014 in comparison to 2011, as well as within the three considered categories. In 2014 in relation to 2011 no region in each of the considered categories dropped to a lower group. The structure of employment level in the 'total' category changed only a little in the western block of the studied region, whereas slightly bigger changes were observed in north-eastern and central regions (employment rate increased significantly, especially in most of the regions of Czech Republic and Poland). A similar situation took place in men's group – the greatest increase in the percentage of employment was observed in Czech Republic, and in the western parts of Slovakia and Hungary, as well as in central regions of Poland. A positive change in the percentage of employment is also noticeable in the women's group, although, in this group the average employment rate did not exceed 74.33% (in none of the regions did the employment rate find itself in a group with the highest levels of this characteristic). In this category, the group with higher values of employment rate (group III) was joined by several regions of Germany and Austria, group II, in turn, was joined by five central regions of Czech Republic and one region in Poland (mazowieckie) as well as in Hungary (Kozep-Magyarorszag) – capital regions. Therefore, a slow process of rebalancing of average employment rate levels in the considered regions (within each of the categories) is taking place, but geographical dependences are clearly visible. This might suggest the importance of the spatial factor in the considered phenomenon. We can, therefore, assume that the level of the considered phenomenon is dependent on the location.

3.2. Estimation of spatial linear models

In the next stage of the study, with the use of the ordinary least squares method (OLS), we estimated linear econometric models. Twelve models were constructed. The explained variables were the variables determining employment – ET, EM and EF, whereas the explaining variables were GDP, PD and accordingly PT, PM, PF. The results of model estimation are illustrated in Table 2.

Table 2. The results of linear estimations of econometric models

Year	Explained variable	Explanatory variables					Intercept	R^2
		GDP	PD	PT	PM	PF		
2011	ET	0.0083	–0.0414	0.666	×	×	–188.0	0.981
	EM	0.0004	–0.0242	×	0.726	×	–90.8	0.988
	EF	0.0043	–0.0170	×	×	0.606	–97.8	0.968
2012	ET	0.0080	–0.0377	0.669	×	×	–185.0	0.981
	EM	0.0038	–0.0217	×	0.729	×	–89.4	0.987
	EF	0.0049	–0.0159	×	×	0.610	–96.5	0.970
2013	ET	0.0078	–0.0368	0.673	×	×	–180.0	0.981
	EM	0.0036	–0.0212	×	0.729	×	–83.5	0.987
	EF	0.0042	–0.0154	×	×	0.616	–97.3	0.970
2014	ET	0.0070	–0.0380	0.687	×	×	–171.0	0.986
	EM	0.0032	–0.0219	×	0.742	×	–78.3	0.991
	EF	0.0038	–0.0156	×	×	0.631	–92.8	0.976

Source: own study

All coefficients of the estimated models are statistically significant, which means that the studied value is considerably influenced by the given variables. What is more, the determination coefficients R^2 are on a high level, they do not fall lower than the value of 0.97 – the adjustment of the models is therefore very good. In order to state whether the spatial factor is important in the description of the studied variables, we have to study the spatial autocorrelation of the residuals. The results are shown in Table 3.

Table 3. Moran's *I* statistics for models residuals

Year	ET model		EM model		EF model	
	I	p-value	I	p-value	I	p-value
2011	0.27	1.1×10^{-5}	0.18	0.0014	0.34	5.8×10^{-8}
2012	0.25	3.3×10^{-5}	0.18	0.0016	0.32	3.1×10^{-7}
2013	0.27	1.2×10^{-5}	0.19	0.0010	0.33	9.2×10^{-8}
2014	0.26	2.3×10^{-5}	0.20	0.0001	0.32	1.5×10^{-7}

Source: own study

Moran's I statistic for all estimated models is positive and statistically significant. Such a situation indicates the presence of spatial autocorrelation, that is, the absence of randomness in the distribution of residuals. For models describing accordingly: the number of the employed in total (ET), the number of employed men (EM) and the number of employed women (EF), the values of Moran's statistic vary, however, within a given model, throughout the period of time from 2011 to 2014, these statistics are similar to each other. Figure 3 illustrates the distribution of residuals of selected models divided into positive and negative residuals – because of similar values of statistic I in each of the years, one example for each of the models has been presented.

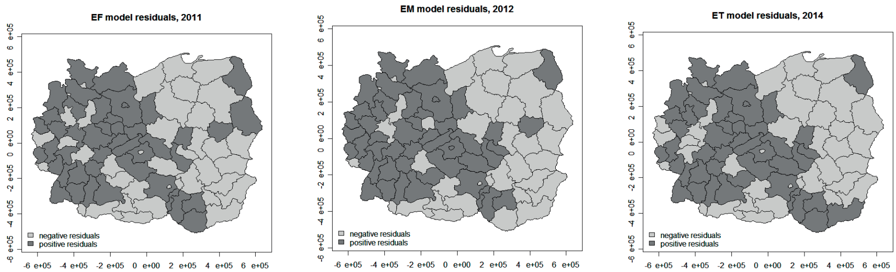

Figure 3. Residuals of the models: EF in 2011, EM in 2012, ET in 2014

Source: own study

The analysis performed using the join-count test showed, in most cases, non-randomness of positive and negative residuals. Only for EF model, in the years 2011, 2012 and 2014, negative residuals showed randomness. The achieved results encourage us to use spatial modelling in order to improve the quality of the description of considered characteristics. *LM* tests, which allow to indicate the appropriate spatial model, were conducted. From the considered models, for each of the cases, the most recommended one turned out to be the spatial error model. Presented below are example values of information criteria, allowing for the comparison of models estimated by the OLS method with spatial models (SEM, SLM). Taken into consideration were Akaike criterion (AIC) and Bayesian criterion (BIC) – low values of these criteria indicate a better model, as well as logLik criterion, high values of which recommend the model. The results achieved for year 2014 are included in Table 4.

Table 4. The values of information criteria for models (2014)

Information criteria	ET model			EM model			EF model		
	OLS	SEM	SLM	OLS	SEM	SLM	OLS	SEM	SLM
AIC	936.93	916.30	938.21	794.03	781.94	795.98	853.18	822.93	853.02
BIC	949.09	930.89	952.79	806.18	796.53	810.56	865.33	837.51	867.60
logLik	–463.47	–452.15	–463.11	–392.01	–384.97	–391.99	–421.59	–405.46	–420.51

Source: own study

The values of the information criteria confirm the usage of the spatial error model in modelling the number of employed (in 2014) with the use of selected variables. For the remaining years the same regularity took place, therefore, the suggested spatial model was used for all of the models. Results of the estimation are illustrated in Table 5.

Table 5. Results of spatial error models estimation

Year	Explained variable	Explanatory variables					Intercept	λ
		GDP	PD	PT	PM	PF		
2011	ET	0.0037	–0.0345	0.673	×	×	–89.0	0.74
	EM	0.0019	–0.0193	×	0.732	×	–47.3	0.71
	EF	0.0019	–0.0152	×	×	0.614	–44.6	0.75
2012	ET	0.0036	–0.0317	0.677	×	×	–88.0	0.73
	EM	0.0018	–0.0174	×	0.735	×	–46.6	0.69
	EF	0.0018	–0.0144	×	×	0.619	–44.1	0.75
2013	ET	0.0032	–0.0301	0.683	×	×	–81.2	0.74
	EM	0.0017	–0.0170	×	0.739	×	–41.8	0.69
	EF	0.0017	–0.0134	×	×	0.627	–42.2	0.76
2014	ET	0.0032	–0.0325	0.697	×	×	–88.2	0.72
	EM	0.0018	–0.0194	×	0.751	×	–48.9	0.66
	EF	0.0015	–0.0133	×	×	0.643	–41.0	0.77

Source: own study

Model coefficients as well as autocorrelation coefficients turned out to be statistically significant in each case. In order to confirm the appropriate model specification additional statistics values of selected diagnostic tests were determined. These are included in Table 6. As it has been mentioned in section 2, a model is appropriately specified, if the following interdependence (in the case of the spatial error model): Wald(λ) $\geq LR_{ERROR} \geq LM_{ERROR}$, is realised. The statistics values presented in Table 6 meet these conditions, confirming at the same time the validity of the model choice.

Table 6. Statistics values of selected diagnostic tests

Year	Model	Statistics values		
		Wald (λ)	LR_{ERROR}	LM_{ERROR}
2011	ET	79.73	24.70	13.74
	EM	64.22	16.00	6.36
	EF	89.57	31.71	22.18
2012	ET	74.39	22.31	12.04
	EM	57.12	14.64	6.21
	EF	87.64	29.35	19.48
2013	ET	80.02	24.94	13.70
	EM	55.56	15.40	6.89
	EF	97.85	32.83	21.46
2014	ET	72.84	22.63	12.66
	EM	46.30	14.09	7.51
	EF	99.66	32.25	20.68

Source: own study

4. Conclusion

Selected regions of Europe were analysed in relation to employment: first of all we analysed the employment rate and then, we constructed econometric models, which were verified in relation to spatial features. The explained variables of the estimated models were the total number of people employed as well as grouped according to gender. For explaining variables GDP per inhabitant, population density and the number of people in their productive age (in total and grouped according to gender) were chosen. All of the obtained models featured spatial autocorrelation, which encouraged the construction of spatial models. Analyses showed that, in the considered case, spatial error models worked best and appropriate diagnostic tests confirmed the proper specification of the models.

On the basis of the conducted analyses, we can state that when studying the employment issues, it is justified to include spatial modelling. It allows for an improvement in the description quality of the considered phenomenon and it can also support construction of a more effective prediction tool. In socio-economic issues it is important therefore to valuate a phenomenon in relation to spatial properties and in the case of their identification – to apply spatial methods.

References

Anselin L. (2006), *Spatial Analysis with GeoDa. 4. Spatial Regression*, University of Illinois, Urbana Champaign.

Anselin L., Bera A. (1998), *Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics*, [in:] A. Ullah, D.E.A. Giles (eds.), *Handbook of Applied Economic Statistics*, Springer-Verlag, Berlin.

Arbia G. (2006), *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*, Springer-Verlag, Berlin.

Cliff A.D., Ord J.K. (1981), *Spatial Process: Models and Applications*, Pion, London.

Eurostat, http://ec.europa.eu/eurostat/web/regions/data/database [accessed: 24.10.2016].

Kopczewska K. (2011), *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R Cran*, CeDeWu, Warszawa.

Overmars K.P., Koning G.H.J. de, Veldkamp A. (2003), *Spatial autocorrelation in multi-scale land use models*, "Ecological Modelling", no. 164, pp. 257–270, http://dx.doi.org/10.1016/S0304-3800(03)00070-X.

Pietrzykowski R. (2011), *Wykorzystanie metod statystycznej analizy przestrzennej w badaniach ekonomicznych*, "Roczniki Ekonomiczne Kujawsko-Pomorskiej Szkoły Wyższej w Bydgoszczy", vol. 4, pp. 97–112.

Pośpiech E. (2015), *Analiza przestrzenna bezrobocia w Polsce*, "Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach", vol. 227, pp. 59–74.

Pośpiech E. (2016), *Modelowanie przestrzenne charakterystyk rynku pracy*, "Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach", vol. 265, pp. 69–79.

Pośpiech E., Mastalerz-Kodzis A. (2015), *Autokorelacja przestrzenna wybranych charakterystyk społeczno-ekonomicznych*, "Metody Ilościowe w Badaniach Ekonomicznych", vol. 16, no. 4, pp. 85–94.

Pośpiech E., Mastalerz-Kodzis A. (2016), *Spatial and Temporal Analysis of Labour Market Characteristics*, "Folia Oeconomica Stetinensia" [in print].

Sikora J., Woźniak A. (2007), *Autokorelacja przestrzenna wskaźników infrastruktury wodno-ściekowej woj. małopolskiego*, "Infrastruktura i Ekologia Terenów Wiejskich", vol. 4, no. 2, pp. 315–329.

Suchecki B. (ed.) (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", vol. 46, no. 2, pp. 234–240.

Zeug-Żebro K. (2014), *Analiza przestrzenna procesu starzenia się polskiego społeczeństwa*, "Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania", vol. 36, no. 2, pp. 441–456.

Zastosowanie regresji przestrzennej do modelowania charakterystyk zatrudnienia

Streszczenie: W artykule analizowano zagadnienie poziomu zatrudnienia. Zbadano stopę zatrudnienia w wybranych regionach Europy, a następnie dla wybranych zmiennych – ludność pracująca ogółem, pracujące kobiety oraz pracujący mężczyźni – zbudowano klasyczne modele ekonometryczne i zweryfikowano konieczność uwzględnienia w modelowaniu badanego zjawiska czynnika przestrzennego. Jako zmienne objaśniające modelu wybrano zmienne demograficzne oraz PKB na mieszkańca. Badano, czy uwzględnienie w konstrukcji modeli podejścia przestrzennego poprawi ich jakość. W rozważaniach wzięto pod uwagę dwa podstawowe modele przestrzenne – model błędu przestrzennego oraz model opóźnienia przestrzennego, spośród których ten pierwszy okazał się dobrym narzędziem analiz.

Słowa kluczowe: modelowanie przestrzenne, model błędu przestrzennego, model opóźnienia przestrzennego, zatrudnienie

JEL: C33, C51, C52

Received: 2016-12-28; verified: 2017-11-23. Accepted: 2018-02-21