

SPATIAL MODELING OF INVESTMENT ACTIVITY OF ENTERPRISES IN SERVICE SECTOR

Katarzyna ZEUG-ŻEBRO

University of Economics in Katowice; katarzyna.zeug-zebro@ue.katowice.pl, ORCID: 0000-0001-7786-3478

Purpose: Due to the visible disproportions, the problem of innovation is increasingly often perceived regionally. These inequalities result from the concentration of knowledge, resources and the amount of expenditure on innovation in a few regions. The aim of the paper is to study the spatial dependence between Polish voivodships in terms of expenditure on innovative activities in enterprises in the service sector.

Design/methodology/approach: For the selected variable conditioning the innovative activity of the enterprise, a classical econometric model will be built and the necessity to include the spatial factor in the modeling process will be verified. For this purpose, two spatial models will be considered: the spatial error model and spatial lag model.

Findings: During the study, the hypothesis on the legitimacy of introducing spatial relationships to the econometric model describing the shaping of the amount of expenditure on innovative activities in enterprises in the service sector in Polish voivodeships was verified. The hypothesis has been verified positively – there are spatial relationships between the examined objects.

Research limitations/implications: The need to take into account the spatial factor in the econometric model results in taking into account spatial estimation methods. The research used selected spatial models. Due to the limitations resulting from the availability of source data, the analysis was conducted only for voivodships in selected years. The analysis should be further deepened, e.g., by even more precise identification of the models and taking into account other neighborhood matrices - only the first-order neighborhood matrix was included in the study.

Practical implications: Modeling the phenomenon of innovation.

Social implications: An essential condition influencing the innovative activity of enterprises is their environment. It is the regional factors that largely influence innovation and faster development of enterprises.

Originality/value: Introducing spatial relationships to the econometric model of outlays on innovative activities of enterprises in the service sector.

Keywords: spatial modeling, regional innovation, factors shaping innovation.

Category of the papers: research paper.

1. Introduction

The analysis and evaluation of innovative activities of enterprises is complicated and raises many doubts. The innovation process is the result of a number of factors both inside the company and outside. Due to the importance of this topic, many works have been written on the problem of innovation. Most of the time it is international literature. Local analyzes are rare due to the lack of proper data. The set of relevant indicators is available only for voivodships (and not really to the full extent).

The research on the problem of innovation shows increasingly the importance of regions in the innovation process. The processes of learning and knowledge transfer are related to the location in the geographical space. It is at the local level that enterprises have the ability to create knowledge through interactions with other enterprises in the process of mutual learning (Maskell, and Malmberg, 1999). Taking into account the regional nature of innovative activities of enterprises, the aim of the paper is to check whether such relationships exist between the regions studied, i.e., to study the significance of the space factor in modeling outlays on innovative activities.

The analysis of the innovation process in enterprises from the service sector was carried out on the basis of data aggregated to the voivodship level. They were obtained from the Local Data Bank of the Central Statistical Office. Due to the limitations resulting from the availability of source data, the analysis was performed only for selected years (2006, 2010, 2014, and 2017). All calculations were performed with R Cran and Microsoft Excel.

2. Innovative service enterprises in Poland

The concept of innovation was introduced into the economic literature in 1911 by J.A. Schumpeter. According to him, innovation is a new combination of means of production and capital in the following cases: introducing a new product, introducing a new production method, opening a new market, acquiring a new source of raw materials or semifinished products, carrying out a new organization of industry (Schumpeter, 1960). On the other hand, according to Haffer's definition, innovations are all changes which, in given spatial and temporal conditions, are perceived as carriers of novelty, concerning products of material and nonmaterial culture as much (Zastempowski, 2010).

Currently, economic growth is largely determined by the development of service enterprises, including those that implement innovation the fastest. This is evidenced by the growing share of services in generating GDP, growth in investments and expenditure on innovation, domestic and foreign trade, and employing a large number of employees (Czubała, 2015).

In 2017, service companies in Poland spent the most funds in terms of expenditure on innovative activities on research and development activities – 43.4%. On the other hand, the lowest outlays in these enterprises were incurred to introduce new or improved products or business processes. The main source of financing these outlays was own funds, which in 2017 in service enterprises accounted for 85.7% of all outlays incurred for this purpose (GUS, 2020).

Taking into account the territorial division (Fig. 1), the highest percentage of innovative enterprises in the service sector was recorded in the following voivodships: Lublin (28.73%), Swietokrzyskie (26.24%), Maslovian (25.19%) in 2006; Masovian (15.56%), Subcarpathian (13.75%), Opole (13.67%) in 2010; in Lublin (17.53%), Opole (15.73%), Masovian (15.09%) in 2014; Masovian (14.17%), Pomeranian (13.21%) and Lesser Poland (11.77%) in 2017. However, the lowest percentage can be observed in the following voivodships: Lubusz (12.61%), Kuyavian-Pomeranian (13.53%), Lodz (14.87%) in 2006; Podlachian (8.33%), Warmian-Masurian (8.74%), Kuyavian-Pomeranian (9.36%) in 2010; Warmian-Masurian (4.38%), Lubusz (5.48%), Greater Poland (7.68%) in 2014; Opole (2.51%), Warmian-Masurian (3.89%) and Swietokrzyskie (5.15%). Additionally, it is worth noting a significant decrease in the number of innovative enterprises in 2017 (compared to 2006).

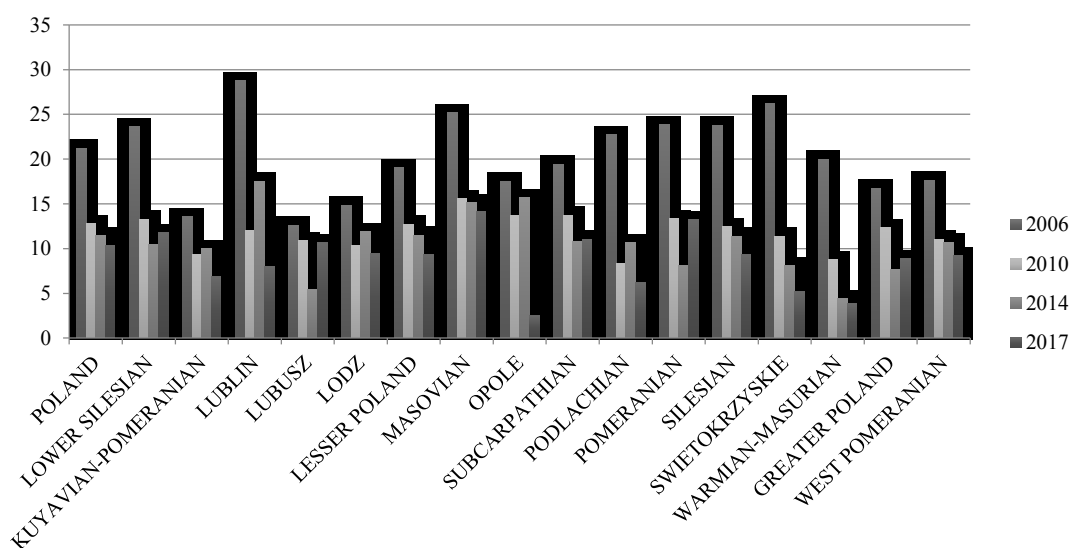


Figure 1. The percentage of innovative enterprises in the service sector in 2006, 2010, 2014 and 2017.

The presentation of the results confirms the regional differentiation in the level of innovation among enterprises. The distance between the highest and the lowest percentage of innovative service enterprises in voivodships was 16.11 percentage points in 2006, 7.23 percentage points in 2010, 13.15 percentage points in 2014, and 11.6 percentage points in 2017.

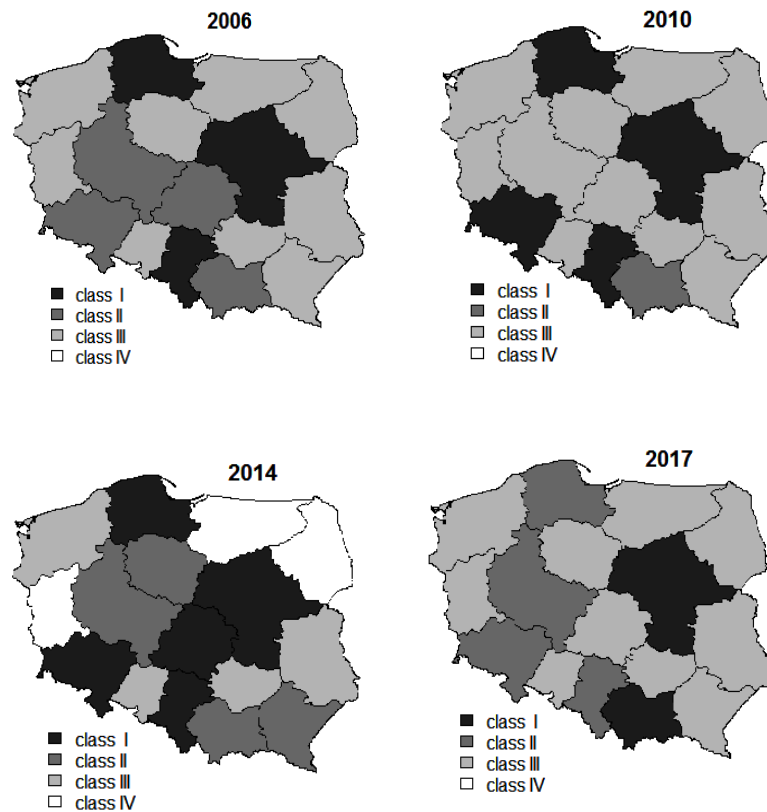


Figure 2. Innovative activity expenditure in enterprises according to the sources of financing innovative activities in 2006, 2010, 2014 and 2017.

This regional differentiation in the level of innovation among enterprises is also confirmed by the analysis of outlays on innovative activities in enterprises according to the sources of financing innovative activities. Figure 2 shows the division into classes depending on the intensity of the studied phenomenon. The first class includes the voivodeships with the highest level of expenditure on innovative activities (Masovian, Silesian and Pomeranian – 2006; Masovian, Silesian, Lower Silesian and Pomeranian - 2010; Masovian, Silesian, Lower Silesian, Pomeranian and Lodz -2014; Masovian and Lesser Poland - 2017). However, in the fourth grade, the voivodeships with the lowest level can be observed (none in 2006, 2010 and 2017; Lubusz, Podlachian and Warmian Masurian in 2014). In 2006, 2010 and 2017, most voivodships belonged to the class with a low level of expenditure on innovative activities.

3. Spatial modeling

Observation of social, economic, and natural phenomena almost always allows us to state that each of them is conditioned by the action of other phenomena. Estimating the relationships that occur between the phenomena that are affected and the factors shaping these phenomena

is performed using regression analysis. In regression modeling using the CLS, it is assumed that the modeled phenomena are independent of their location, so there are no interactions between the examined objects. This assumption does not always apply to the analysis of social and economic phenomena in spatial terms. Therefore, when analyzing and modeling localized data, one should take into account the spatial interactions that may relate to both the dependent variable and the random component (Sucheckki, 2010).

The construction of the spatial model is aimed at improving the quality of the econometric model. Inclusion of spatial relationships within a given area as well as within neighboring areas can have a positive influence on the translation of the variability of the features under consideration. The following basic groups of spatial models are distinguished: spatial lag models, spatial error models, cross-regression models and mixed variants. In paper the first two models were considered.

The spatial lag model includes the spatially delayed endogenous variable $\mathbf{W}y$, ie, it is an autoregressive model (the basis of the model is spatial dependence). The general form of this model is described by formula:

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

where:

α – spatial autocorrelation coefficient,

\mathbf{W} – spatial weight matrix,

β – vector of model coefficients,

\mathbf{X} – matrix of exogenous variables,

ε – model error.

The model test if $\alpha = 0$ is the significance of the dependent variable which is spatially delayed. Spatial delay $\mathbf{W}y$ is interpreted as the level of the dependent variable y in neighboring regions. If this is significant, then the level of y in the i -th region can be explained by the level of the phenomena in the neighborhood and other factors represented by the remaining explanatory variables.

The spatial error model contains spatially delayed errors. This model assumes the spatial autocorrelation of the rest of the model. The general form of the model is given by the formula:

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

where:

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

μ – spatial autocorrelation coefficient, other signs as above.

$\mathbf{W}\varepsilon$ is a spatially delayed error, which should be interpreted as the average error from neighboring locations, and φ is an independent error of the model. In the model, we test whether $\mu = 0$ ie lack of spatial autocorrelation.

There are different ways to select and specify the model that best describes the issue. One of them is a simple procedure for choosing a spatial regression model, which follows the algorithm (Anselin, 2006; Sucheccki, 2010; Kopczewska, 2011):

1. Estimation of the linear model by the classical least squares method (CLS).
2. Study of spatial autocorrelation of model residues (Anselin, and Rey, 1991) – at this stage the Moran test (Moran, 1950) is used for residues, which takes the form:

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

where:

n – number of regions,

\mathbf{u} – column vector of residual from the CLS model,

S_0 – sum of elements of the weights matrix, $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$,

\mathbf{W} – spatial weight matrix degree n , defining the structure of the neighborhood,

w_{ij} – element of the weights matrix \mathbf{W} .

This test verifies the following hypotheses: H_0 : no spatial autocorrelation of residues, H_1 : occurrence of spatial autocorrelation of residues.

The occurrence of autocorrelation of residuals does not necessarily use the need for spatial models, and the residuals of the model (model estimated on the logarithms of variables) can then be analyzed (Cliff, and Ord, 1970).

3. Carry out LM tests (Lagrange Multiplier) to specify a spatial model that better describes the phenomenon (Anselin et al., 1996) – these tests are based on the CLS model residuals. Two basic LM tests are LMerror and LMlag. The LMerror test tests the spatial dependence of the error, and the LMlag tests the significance of the dependent variable which is spatially delayed. LMerror with asymptotic distribution $\chi^2(1)$ takes the form:

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

where:

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

$$T_1 = \text{tr}[(\mathbf{W}^T + \mathbf{W})\mathbf{W}] \quad (6)$$

other signs as above.

Using the above statistic, significance of the coefficient μ is verified.

LMlag with asymptotic distribution $\chi^2(1)$ takes the form:

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

where:

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

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

$\hat{\boldsymbol{\beta}}$ – estimated vector of the coefficients, other signs as above.

Using the above statistic, significance of the coefficient α is verified.

The choice of the spatial model (by hybrid strategy) is determined by the higher (significant) value of the LM statistics (if LMlag > LMerror, then the spatial lag model is determined, otherwise the spatial error model).

4. Carry out additional diagnostic tests: LR test and Wald test and compare the statistic values of these tests with the LM statistic values. Then for:
 - the spatial error model: $\text{Wald}(\mu) \geq \text{LRerror} \geq \text{LMerror}$,
 - the spatial lag model: $\text{Wald}(\alpha) \geq \text{LRlag} \geq \text{LMlag}$.

4. Empirical analysis – evaluation of models

The analysis of the innovation process in enterprises from the service sector was carried out on the basis of data aggregated to the voivodeship level. For this purpose, data obtained from the Central Statistical Office for 2006, 2010, 2014, and 2017 were used (www1). The data used is annual data. The study began with the selection of variables on the basis of which three regression models will be estimated. The dependent variable Y characterizing the innovation process in enterprises from the service sector is the amount of expenditure on innovative activities in enterprises from the service sector from all investment sources (PLN thousand). The set of explanatory variables related to innovation was selected according to the following principles (Cywiński, 2013): the use of advanced production technologies, information technologies, intangible investment, measurement of organizational changes and non-technological innovations in enterprises, research on society's attitudes towards science and technology. As a result, 14 indicators were selected:

- X1 – share of innovative enterprises in the service sector by type of introduced innovation – new, or improved products (%);
- X2 – share of innovative enterprises in the service sector by type of introduced innovation – new, or improved processes (%);
- X3 – expenditure on innovative activity of enterprises from the service sector for the purchase of software (PLN thousand);
- X4 – expenditure on innovative activities of enterprises in the service sector on personnel training directly related to the introduction of product or process innovations (PLN thousand);
- X5 – employed in R&D (FTE);
- X6 – expenditure on innovative activities in enterprises per 1 professionally active person (PLN);
- X7 – internal expenditure on R&D (PLN thousand);
- X8 – share of enterprises with computers (%);
- X9 – share of enterprises with Internet access (%);
- X10 – number of patents granted by the PPO – total;
- X11 – average monthly gross wages and salaries in the enterprise sector (PLN);
- X12 – graduate studies at the master's level;
- X13 – students of postgraduate studies and participants of doctoral studies per 1000 inhabitants (people);
- X14 – expenditure on innovative activities of enterprises from the service sector on marketing related to the introduction of new or significantly improved products (PLN thousand).

In the first stage of the study, three linear econometric models were estimated. The models examined the relationship between the increase in the years: 2006-2017 (Model I), 2010-2017 (Model II), 2014-2017 (Model III), the amount of expenditure on innovative activities in enterprises from the service sector from all investment sources, and the increase in explanatory variables.

Modeling of the global relationships between the dependent variable and the independent variables remaining after the elimination procedure was carried out using the R program. Assuming a linear relationship between the adopted variables, the multiple regression model was estimated using the Classical Least Squares (CLS) method. As not all parameters of the estimated model showed statistical significance, the posteriori elimination procedure was used, assuming in the next step the removal from the original model (model with all variables) of those variables that have the least significant impact on the dependent variable in a given step. After each elimination step, the model is subjected to statistical verification. The procedure ends when a model is obtained, all the parameters of which show statistical significance (Guzik, 2008). As a result of the applied stepwise elimination procedure, the selected explanatory variables remained in the final version of the model (Table 1).

Table 1.
Results of estimation of econometric models

Variables	Value of coefficients		
	Model I	Model II	Model III
Intercept	3.0253	1.0967	-1.4242
X1	1.5575	-	1.2677
X3	-	1.6462	0.7306
X4	0.6563	-	-
X5	-0.3839	-	-
X7	0.1041	-0.1888	1.2932
X8	-	-60.3745	25.4611
X12	4.1538	-	6.7403
X14	-	-	-0.9709
R ²	0.8082	0.7354	0.9296

The symbol "-" means that the parameter was statistically insignificant.

The estimated regression coefficients show that part of the variability in the expenditure on innovative activities of enterprises from the service sector from all investment sources was explained by: the share of innovative enterprises from the service sector by the types of innovations introduced – new or improved products (Model I and III), expenditures on innovative activities of enterprises from the service sector for the purchase of software (Model II and III), the expenditure on innovative activities of enterprises from the service sector on training of personnel related directly to the introduction of product or process innovations, employed in R&D (Model I), internal expenditure on R&D (Model I, II, III), the share of enterprises using computers (Model II and III), the number of graduates of studies at the master's level (Model I and III), the expenditure on innovative activities of enterprises from the service sector on marketing related to the introduction of new or is totally improved products (Model III). The determination coefficients of all models were relatively high, therefore the fit of the models is good.

In the next step of analysis, the spatial autocorrelation of errors was studied. Results are presented in Table 2.

Table 2.
Moran statistics for the residuals of the model

Model I		Model II		Model III	
<i>I</i>	<i>p</i> -value	<i>I</i>	<i>p</i> -value	<i>I</i>	<i>p</i> -value
-0.2336	0.8940	-0.1436	0.7024	0.0413	0.2639

Moran's statistic *I* is statistically insignificant for all estimated models, which suggests the lack of spatial autocorrelation, i.e., a random distribution of the residuals (the distribution does not depend on the location).

In the next step of the study, LM tests were performed for each model to investigate the existence of spatial relationships in linear models. The results of this analysis (indicating which spatial model would be more appropriate) are contained in Table 3.

Table 3.
Recommended spatial model

Model	LM tests				Type of spatial model
	LMerror	<i>p</i> -value	LMlag	<i>p</i> -value	
Model I	4.6314	0.0314	2.9375	0.0866	spatial error model
Model II	0.7796	0.3773	0.2338	0.3773	-
Model III	0.1975	0.6568	0.2330	0.6293	-

The above results suggest that in the case of Models II and III, estimating spatial models is not advisable, while for Model I it is best to build a spatial error model. In addition, Table 4 shows the values of the information criteria: Akaike (AIC), Bayesian (BIC), and logLik, to compare the model estimated by CLS with the spatial models (SEM – spatial error model, SLM – spatial lag model). The best model is the one for which the AIC and BIC criteria take the lowest values, while the logLik criterion assumes the highest values.

Table 4.
Values of the information criteria

Model		Values of criteria		
		AIC	BIC	logLik
I	SLM	44.0644	50.2451	-14.0322
	SEM	32.6687	38.8494	-8.3343
	CLS	42.1606	47.5687	-14.0803

The values of the information criteria for Model I clearly indicate the spatial error model. For this model, a suggested spatial model has been built; the results of the estimation are given in Table 5. The spatial factor μ was significant.

Table 5.
Results of estimation of spatial error models

Variables	Value of coefficients
	Model I
Intercept	4.2209
X1	3.1856
X4	0.6752
X5	-0.3684
X7	0.1035
X12	5.9436
μ	-1.5620

According to all criteria of information, spatial error models are definitely better than linear models.

The modeling carried out shows that in the analysis of the factors of the size of expenditure on innovative activities in enterprises from the service sector from all investment sources, it is worth taking into account the assumption of the presence of certain spatial relationships, significantly influencing the variability of the intensity of the Y variable. Omission of this relationship may lead to incorrect model specifications and wrong conclusions.

5. Conclusions

The analysis of the factors influencing the amount of expenditure on innovative activities of enterprises from the service sector from all investment sources, carried out in this study, shows that it depends to the greatest extent on: the share of innovative enterprises in the service sector by the type of introduced innovations (new, or improved products); the expenditure on innovative activities of enterprises in the service sector on personnel training directly related to the introduction of product or process innovations; number of employed in R&D; internal expenditure on R&D; number of graduates studies at the master's level. These findings are particularly important in the context of current trends related to the concentration of knowledge and resources, the transfer of R&D employees, and the development of new technologies resulting in an increase in the number of students in the technical faculty. Research has shown that the observed trends are also related to the location in geographical space. The voivodeships characterized by the highest level of expenditure on innovative activities in the analyzed years were: Masovian, Silesian, Pomeranian, Lower Silesian, Lodz and Lesser Poland. Most voivodships belong to the class with a low level of expenditure on innovative activities. The obtained results can be used in the spatial planning process to indicate the appropriate directions for the management of public funds for financing innovative activities.

During the study, the hypothesis about the legitimacy of introducing spatial relationships to the econometric linear model describing the shaping of the amount of investment outlays in industrial enterprises in Polish voivodships was verified. In the light of the obtained results, it was found that spatial effects occur for the data used in Model I (they do not occur in the other two models). This means that if the factors influencing the expenditure on innovative activities in enterprises from the service sector from all investment sources are analyzed, the estimation of regression model parameters using the least squares method changes the properties of the estimators, which may make the results inaccurate. According to the current state of knowledge, an adequate solution in such a situation is the use of spatial regression models. The spatial modeling carried out showed that the spatial error model (for Model 1) gives the highest precision and correctness of the results in the analysis of the amount of expenditure on innovative activities in enterprises from the service sector from all sources of investment. The good specification of the spatial model was confirmed by the values of the relevant statistics. The conducted research has shown that despite the insignificant value of the I Moran statistics, spatial modeling for the discussed issues is justified, although it should be emphasized that the analysis should be further deepened taking into account, for example, a different neighborhood matrix.

References

1. Anselin, L. (2006). *Spatial Analysis with GeoDa. 4. Spatial Regression*. University of Illinois, Urbana-Champaign.
2. Anselin, L., and Rey, S.J. (1991). Properties of tests for spatial dependence in linear-regression models. *Geographical Analysis, No. 23*, pp. 112-131.
3. Anselin, L., and Bera, A.K., and Florax, R., and Yoon, M.J. (1996). Simple diagnostic tests for spatial dependence. *Regional science and urban economics, 26*, pp. 77-104.
4. Cliff, A., and Ord, J.K. (1970). Spatial Autocorrelation: A Review of Existing and New Measures with Applications. *Economic Geography, Vol. 46*, pp. 269-292.
5. Cywiński, M. (2013). Próba identyfikacji spójnego systemu oceny działalności innowacyjnej przedsiębiorstw. *Zarządzanie i Finanse, Vol. 1, Nr 4*, pp. 19-39.
6. Czubała, A. (2015). Innowacje w sektorze usług w Polsce. *Zeszyty Naukowe Małopolskiej Wyższej Szkoły Ekonomicznej w Tarnowie, 1(26)*, pp. 35-45.
7. GUS (2020). *Działalność innowacyjna przedsiębiorstw w latach 2016-2018*. Warszawa.
8. Guzik, B. (2008). *Podstawy ekonometrii*. Poznań: Wydawnictwo Akademii Ekonomicznej.
9. Kopczyńska, K. (2011). *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R Cran*. Warszawa: CeDeWu.
10. Maskell, P., and Malmberg, A. (1999). Localized Learning and industrial Competitiveness. *Cambridge Journal of Economics, Vol. 23*, pp 167-185.
11. Moran, P.A.P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika, 37(1)*, pp. 17-23.
12. Schumpeter, J.A. (1960). *Teoria rozwoju gospodarczego*. Warszawa: PWN.
13. Suhecki, B. (ed.) (2010). *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*. Warszawa: C.H. Beck.
14. www1, <https://bdl.stat.gov.pl/BDL/dane>, 1.09.2020.
15. Zastempowski, M. (2010). *Uwarunkowania budowy potencjału innowacyjnego polskich małych i średnich przedsiębiorstw*. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.