

Spatial patterns of European innovation: a spatial econometric approach with technological proximity weights

Andrea Furková¹

¹ University of Economics in Bratislava, Department of Operations Research and Econometrics,
Dolnozemska cesta 1, 852 35 Bratislava, Slovakia, e-mail: andrea.furkova@euba.sk

Abstract. *The paper deals with a spatial econometric analysis of 220 European regions. The analysis follows the Mixed Geographically Weighted Regression – Spatial Autoregressive approach. Patent applications were a proxy for innovation output. Instead of traditionally applied geographical proximity, technological similarity was considered. The results supported the assumption of spatial differentiation of model parameters and indicated that regional innovation activities do not have only a local character in almost half of the regions. The region's technological similarity appears to be a significant factor stimulating innovation for regions that are not among the top innovators.*

Keywords: European Innovation, Spatial Econometrics, Technological similarity

JEL Codes: O31, R12

1. Introduction

The issue of innovation has been playing an essential role in the theoretical and empirical literature for a long time in connection with many economic phenomena. We can observe an increasing number of research studies focused on innovation's spatial component and innovative actors' geographical proximity. The results point to spatial innovation spillover effects, perceived as factors explaining innovation.

Although innovation is global and national policies create a platform for innovation, the main actors in innovation activities are regions or metropolitan areas where companies, research institutions, and government can interface more directly. The innovation clusters or innovation “hot spots” create suitable conditions for the creation of innovations, knowledge sharing, and their rapid dissemination. The research focused on innovation processes in European regions points to an uneven distribution of innovation processes and a tendency to cluster innovation activities (see, e.g., Charlot et al., 2015; Autor, 2019; Khan, 2012; Kumar, 2008; Moreno et al., 2005a or Moreno et al., 2005b). Regarding proximity among innovation actors, a study by Boschma (2005) critiqued the role of geographical proximity in learning and innovation processes. In his study, Boschma assumes that other forms of proximity can replace geographical proximity and that other forms of proximity dimensions, such as cognitive, organizational, social, or institutional proximity, may be more appropriate.

The paper's main aim is to explore an alternative measure of proximity between regions concerning the European innovation problem. From the traditionally considered geographical proximity (defined as the spatial distance between actors - European regions), we have moved to technological similarity based on the two indicators. These indicators are Small and medium-sized enterprises (SMEs) introducing product or process innovations, and Small and medium-sized enterprises (SMEs) introducing marketing or organizational innovations. The primary assumption of this analysis is that European regional innovations are unevenly distributed, probably spatially correlated, given the proximity of the regions in terms of technological similarity. We intend to simultaneously account for spatial effects such as spatial

autocorrelation and spatial heterogeneity. We see such an approach as beneficial to the innovation literature because although previous research on regional innovations has emphasized spatial autocorrelation and heterogeneity, these issues have been addressed separately.

Spatial econometrics offers limited tools suitable for simultaneously handling both spatial effects. Several notable exceptions can be found in the theoretical literature, e.g., Charlott et al. (2015) or Guastella and van Oort. (2015). These studies propose a geo-additive modeling framework to capture spatial autocorrelation and spatial heterogeneity simultaneously. Another study, the study by Cho et al. (Cho et al. 2010), suggests a model that combines Geographically Weighted Regression (GWR) and Spatial Error Model (SEM), called GWR-SEM. The latest advancement in this area is evident from the work of Geniaux and Martinetti (Geniaux and Martinetti, 2018) where a new class of data-generating processes (DGPs) called Mixed Geographically Weighted Regression – Spatial Autoregressive (MGWR-SAR) models was introduced. A significant advance of these models is that the regression and spatial autocorrelation parameters may vary in space, and the mixed case is also allowed. We chose this approach as the primary tool of our analysis.

The structure of the remaining parts of the paper is as follows: part 2 contains the data description, the specifications of the spatial weight matrices, and the main theoretical backgrounds of the study. In part 3, the empirical results are presented and interpreted. The main concluding remarks can be found in section 4. The paper ends with references.

2. Methodology

The following sections will present the data, the study area, the spatial weighting matrices, and the main methodological issues.

2.1. Study Area, Data and Spatial Weighting Schemes

The Regional Innovation Scoreboard (RIS) 2019 (Hollanders et al., 2019) and data from the Eurostat regional statistical database (Eurostat, 2021) were used to perform a spatial econometric analysis of the European innovation problem. The analysis includes the regions of EU Member States, Norway, Serbia, and Switzerland. Since the policy at the European Union (EU) level focuses mainly on the regional level, our database of 220 European regions is defined at the NUTS 1 or NUTS 2 level (NUTS - Nomenclature of Territorial Units for Statistical Purposes). The vast majority of regions are included at the NUTS 2 level. However, countries for which NUTS 1 and NUTS 2 levels coincide with the country's territory are included at the NUTS 1 level (country levels). In this paper, Patent Cooperation Agreement (PCT) applications represent the proxy output of regional European innovation, i.e., a modeled variable. Fig.1 presents Jenks's natural break map and a box plot for PCT applications to provide an overview of the actual spatial distribution of the modeled variable. A natural breaks map uses a nonlinear algorithm to group observations such that the within-group homogeneity is maximized (see Fisher, 1958 and Jenks, 1977). Using natural breaks maps seems more appropriate than, for instance, a quartile map because the natural breaks criterion is better at grouping the extreme observations. In contrast to the quantile maps, the number of observations in each category can be highly unequal. It is clear from Fig.1 that the regions of Greece, Spain, Portugal, southern Italy, and most regions in Eastern Europe have low levels of patents. On the contrary, patent production is concentrated mainly in the regions of Germany, Finland, Sweden, Norway, Switzerland, the Netherlands, and Denmark.

All calculations and graphical visualization of the results (maps and box plots) were carried out in GeoDa, GeoDaSpace, and RStudio.

The geographical differences in the distribution of innovation resulting from Fig.1 and many empirical studies point to an uneven distribution of innovation processes across European regions. Although the role of geographical proximity in innovation processes is often emphasized, this paper aims to examine another form of proximity among European regions. In contrast to geographical proximity (geographical distance between innovation actors - European regions), we consider technological similarities based on two

indicators. The first indicator is Small and medium-sized enterprises (SMEs) introducing product or process innovations, and the second indicator Small and medium-sized enterprises (SMEs) introducing marketing or organizational innovation is considered. Both indicators are defined as a percentage of SMEs. The neighborhood of the regions will be defined based on the difference in the number of SMEs. Fig.2 shows Jenks's natural break maps for the two indicators. As in the case of patent applications, we see that the highest concentration of innovators is evident in the regions of countries such as Germany, Finland, Norway, Switzerland, the Netherlands, and Denmark, and the number of innovators is lower for the regions of Spain and most regions of Eastern Europe. On the other hand, Greek or Portuguese regions have low patent performance and are characterized by many SMEs.

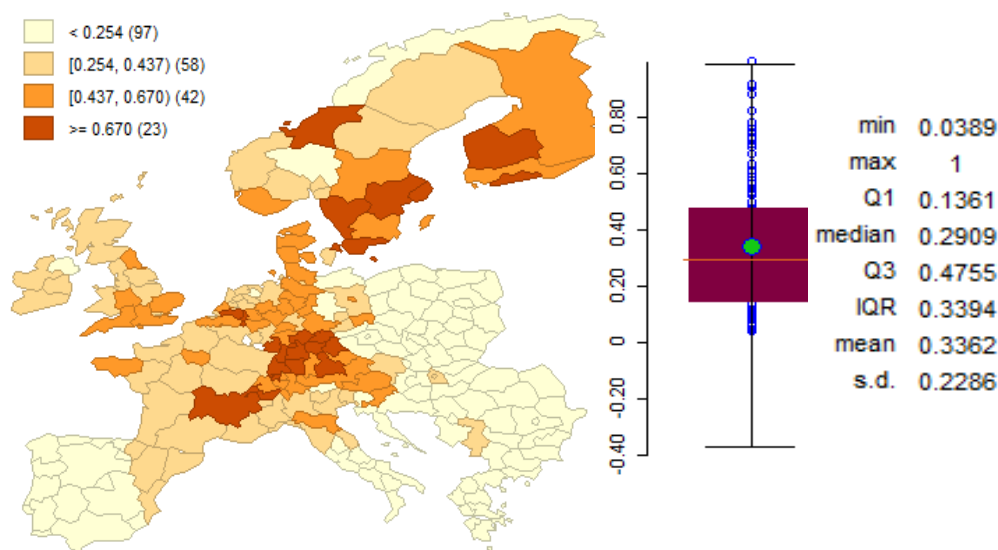
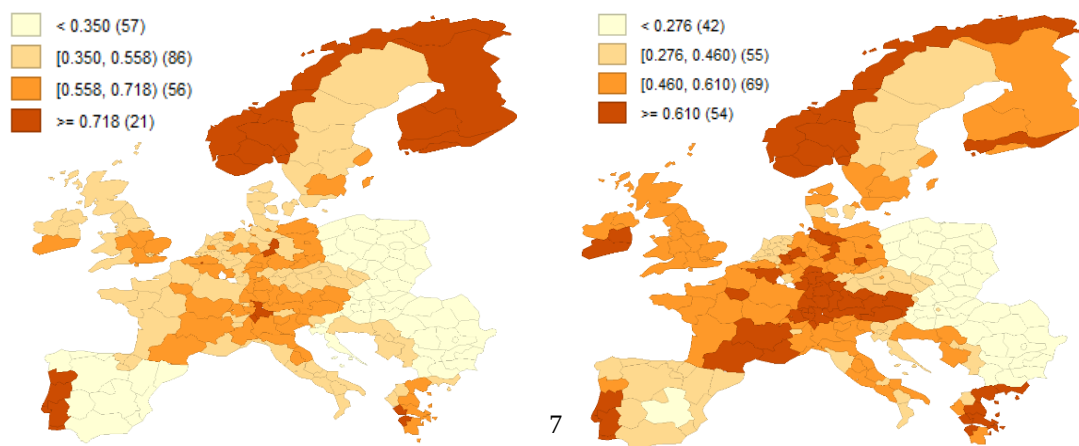


Fig. 1: Jenks's natural break map and box plot for Patent applications
 Source: RIS 2019 data (Hollanders et al., 2019); author's calculations
 Notes: Q1 – First quartile; Q3 – Third quartile; s.d. – Standard deviation



(a)

(b)

Fig. 2: Jenks's natural break maps for (a) SMEs - product or process innovations (b) SMEs - marketing or organizational innovations

Source: RIS 2019 data (Hollanders et al., 2019); author's calculations

The central premise of this analysis is that European regional innovations are unevenly distributed, probably spatially correlated, considering the proximity of regions in terms of technological similarity. The Moran's I test was performed as a starting point for spatial autocorrelation detection. The Moran's I value for PCT applications (technological similarity is considered) was 0.311 and statistically significant (pseudo p -value 0.001). The resulting potential spatial spillovers and uneven distribution of innovation have led us to consider spatial autocorrelation and spatial heterogeneity. The new MGWR-SAR models (presented in the next section) allow simultaneous consideration of both spatial effects (spatial autocorrelation and heterogeneity).

In all subsequent spatial econometric models, PCT applications at the European Patent Office (PCT applications in the EPO per billion regional Gross Domestic Product (GDP)) are used as a modeled variable - innovation output. The following variables are considered as innovation inputs: top 10% most-cited publications (as a percentage of total scientific publications), Research & Development (R&D) expenditure in the business sector (as a percentage of regional GDP), human resources in science and technology (as a percentage of the active population) and population density (persons per square kilometer).

All data are normalized using the min-max procedure, while RIS 2019 data are already reported as normalized. We also assume that lagging European regions (with a GDP per capita below 75% of the EU-27 average GDP) are weaker in patenting activities. The dummy variable is considered to control additional regional characteristics. Dummy variable for lagging regions ($DUMMY_{LAGG}$) is defined as: $DUMMY_{LAGG} = 1$, if region belongs to the category of lagging regions, others=0.

The estimation of spatial econometric models is based on the specification of the spatial weighting matrix $\mathbf{W} = (w_{ij} : i, j = 1, \dots, N)$ which summarizes the spatial relationships between N regions and is a key to the specification of the spatially lagged dependent variable. We consider technological similarity/proximity weights ($\mathbf{W}^{\text{TECH.PROX}}$) – the weights are a negative power distance function of distance (Euclidean distance metric was considered) of the following form:

$$w_{ij} = d_{ij}^{(-\alpha)} \quad (1)$$

where

$$\alpha = 1;$$

d_{ij} is the distance between observations i and j (based on the SMEs - product or process innovations).

In local regressions, i.e., MGWR-SAR estimation, a bandwidth parameter and a kernel function type must be provided. Spatial kernel weights ($\mathbf{W}_i^{\text{KERNEL-TECH.PROX}}$) - the adaptive Gaussian-shaped kernel function based on 50 nearest neighbours is used. The adaptive Gaussian-shaped kernel function is given as:

$$w_{ij} = \exp\left(-1/2\left(d_{ij}/h_{i(k)}\right)^2\right) \quad (2)$$

where

w_{ij} is the j -th element of the diagonal of the matrix $\mathbf{W}_i^{\text{KERNEL-TECH.PROX}}$;

d_{ij} is the distance (Euclidean distance metric was considered) between observations i and j (based on the SMEs - product or process innovations and SMEs - marketing or organisational innovations);

h is the bandwidth parameter.

The bandwidth parameter h is a function of the k nearest neighbours (50 nearest neighbours) and the same number of observations - locations are included in each local estimate.

To illustrate the spatial weights of the neighbourhood, we present Fig.3, which depicts the Swedish region Sydsverige (dark marked region) and its neighbouring regions. This region is one of the leaders in innovation activities (represented by patents); it is characterized by a diversified business environment, with several established clusters and several strong universities, which are also very beneficial in innovation. Fig.3(b) shows its neighbouring regions based on technological similarity/proximity weights, and for comparison, Fig.3(a) shows a neighbourhood based on the very well-known queen contiguity spatial weights. This figure clearly illustrates the differences in neighbouring regions. Higher variability in the number of neighbours occurs when we consider the matrix $\mathbf{W}^{\text{TECH.PROX}}$.

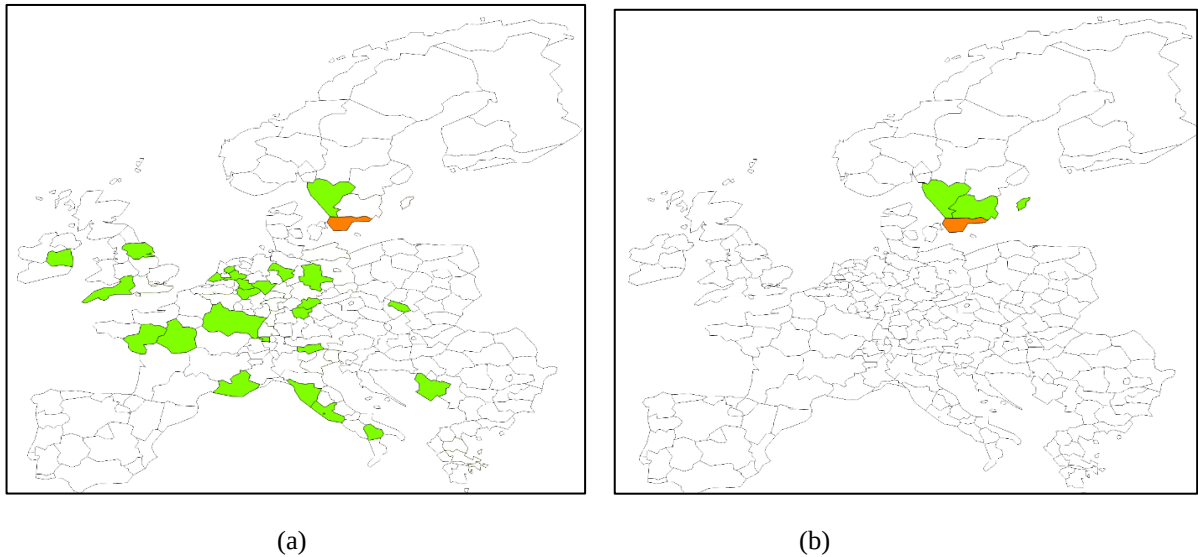


Fig.3: Swedish region Sydsverige (dark marked region) and its neighbouring regions - queen contiguity weights (a) and inverse distance technological proximity weights (b)

Source: author's calculations

2.2. Econometric models

In the empirical part of this paper, we will focus on the estimation of the following three econometric models:

1. Non-spatial Ordinary Least Squares (OLS) regression:

$$y_i = \beta_0 + \sum_{k=1}^K \beta_k x_{ki} + \varepsilon_i, \quad i = 1, 2, \dots, N \quad (3)$$

where

y_i is the dependent variable at location i (N is the number of regions);

x_{ki} is the value of k -th explanatory variable at location i (K is the number of explanatory variables);

β_k is the regression coefficient of k -th explanatory variable;

β_{0i} is the intercept parameter;

ε_i is the random error at location i (independent and identically distributed).

The OLS estimation method will be used to estimate model (3).

2. Spatial Autoregressive model (SAR) - $\mathbf{W}^{\text{TECH.PROX}}$ spatial weights:

$$y_i = \rho \sum_{j=1}^N w_{ij} y_j + \beta_0 + \sum_{k=1}^K \beta_k x_{ki} + \varepsilon_i, \quad i = 1, 2, \dots, N \quad (4)$$

where

ρ is a spatial autoregressive parameter;

w_{ij} is an element of spatial weighting matrix $\mathbf{W}^{\text{TECH.PROX}}$.

The remaining model (4) terms were defined with the previous model (3). The issue of endogeneity problem appears in spatial autoregressive models such as model (4). In such models, the spatially lagged dependent variable $\mathbf{W}y$ is present. Therefore, the OLS method is not a suitable estimation method. The parameters of the model (4) can be estimated through Instrumental Variable approach, Maximum Likelihood, or generalized moment method. However, the modification of spatial aspects must be done. Estimating the SAR model in this paper is performed using the spatial two-stage least square (S2SLS) technique with $\mathbf{X}$ (matrix of regressors) and $\mathbf{W}\mathbf{X}$ (matrix of spatially lagged of regressors) as a set of instruments. Estimation details can be found, e.g., Anselin and Rey (2014) or Geniaux and Martinetti (2018).

3. Mixed Geographically Weighted Regression - Spatial Autoregressive (MGWR-SAR) - spatial kernel weights $\mathbf{W}_i^{\text{KERNEL-TECH.PROX}}$ and $\mathbf{W}^{\text{TECH.PROX}}$ spatial weights:

$$y_i = \rho_i \sum_{j=1}^N w_{ij} y_j + \beta_{0i} + \sum_{k=1}^K \beta_{ki} x_{ki} + \varepsilon_i, \quad i = 1, 2, \dots, N$$

(5)

where y_i is the dependent variable at location i ;

x_{ki} is the value of k -th explanatory variable at location i ;

β_{ki} is the local regression coefficient of k -th explanatory variable at location N ;

β_{0i} is the intercept parameter at location i ;

ρ_i is a spatial autoregressive parameter at location i ;

w_{ij} is an element of spatial weighting matrix $\mathbf{W}^{\text{TECH.PROX}}$;

ε_i is the random error at location i (independent and identically distributed).

Several MGWR-SAR model specifications can be derived from equation (5). This group of models is estimated by S2SLS method (see Geniaux and Martinetti, 2018).

The estimation of all econometric models in this paper follows the concept of an extended Regional Knowledge Production Function, which accounts for the role of territorial characteristics and spatial processes (for more details, see, e.g., Moreno et al., 2025a).

3. Results and Discussions

Tabl.1 contains the main estimation results. If we consider the results of the spatial dependence tests (see Tab.1), which were carried out based on OLS estimation (based on the formula (3)) and technological similarity/proximity weights matrix, we find that the problem of spatial autocorrelation is present. Therefore, discussing the OLS estimation in more detail is not necessary. We consider the OLS estimation and the SAR spatial model estimation (based on the formula (4)) to be the starting step of the econometric analysis. The estimation results of the SAR spatial model indicate the adequacy of the inclusion of all considered innovation inputs and point to the importance of spatial innovation spillovers. The models also contain dummy variables to capture the diversity of the innovation process in lagging regions. The statistical significance of both dummy variables (in absolute and interaction forms) indicated the expected differences. The estimates and statistical significance of the spatial autoregressive parameter (see the spatial lag parameter in Tab.1) indicate that the effect on innovation output using technological similarity as a measure of regional proximity is not negligible. Also, all indirect (results not shown due to lack of space, available on

request from the author) effects resulting from SAR estimation indicate significant spatial spillovers among regions.

In addition to the problem of spatial autocorrelation, in an area as large as the European space, it can also be expected that the strength of influence of individual innovation determinants may differ in individual regions. Therefore, the problem of spatial heterogeneity should also be considered. For this purpose, the simultaneous consideration of these two spatial effects within one model, we used a new approach to MGWR-SAR models outlined in Geniaux and Martinetti (2018) and presented by formula (5). We assume a model with all spatially variable parameters, including the spatial autoregressive parameter. In the MGWR-SAR model estimation process, an adaptive Gaussian kernel based on the 50 nearest neighbours defined by formula (2) was used.

Now we will focus on the main estimation results of the MGWR-SAR model (see Tab.1). The estimates are also presented in box plots (see Fig.4). Even the basic descriptive statistics (see Fig.4) draw a non-negligible variability of the estimated parameters.

Table 1: Estimation results generated by OLS, SAR, and MGWR-SAR model

Variable/innovation input	OLS	SAR	MGWR-SAR Percentage of significant cases at 95 %
Intercept	-0.1691***	-0.1831***	53.18%
Most-cited publications	0.2440***	0.1948***	47.73%
R&D expenditure in the business sector	0.5899***	0.5806***	100.00%
Human resources in science and technology	0.2898***	0.2777***	80.00%
Density	-0.3498***	-0.3386***	89.09%
DUM(Lagging regions)	0.1265***	0.1196***	71.82%
DUM(Lagging regions)* Human resources in science and technology	-0.5002***	0.4690***	93.18%
Spatial Lag	—	0.1352**	47.73%
R^2	0.7472	0.7582	—
Spatial pseudo R^2	—	0.7525	—
AIC	-314.465	—	-363.5662
Moran's I (Patent applications)	0.3107***	—	—
Moran's I (OLS residuals)	3.3453***	—	—
Lagrange Multiplier (SAR)	8.9227***	—	—
Robust Lagrange Multiplier (SAR)	4.4134**	—	—
Lagrange Multiplier (SEM)	4.8214**	—	—

Robust Lagrange Multiplier (SEM)	0.3121	–	–
Anselin–Kelejian test	–	0.321	–
Number of data points	220	220	Kernels type: Gauss Adapt, bandwidth=50

Notes: Parameters are statistically significant at **p-value<0.05; ***p-value< 0.01, LM – Lagrange Multiplier
Source: author's calculations

The boxes in Fig.4 represent the interquartile range, and they do not overlap with one another. So, there are significant discrepancies in the estimated parameters in magnitude. All boxplots (box width) are presented concerning the percentage of significant cases. According to our results, R&D expenditure in the business sector seems to be the most significant innovation input; this factor was statistically significant in every region (see Tab.1). Moreover, its impact on innovation output reached the highest values (for instance, mean value was 0.5706). Another highly significant factor seems to be input Human resources in science and technology, which achieve high efficiency (mean value 0.3161) and a high proportion of statistically significant regions (80%).

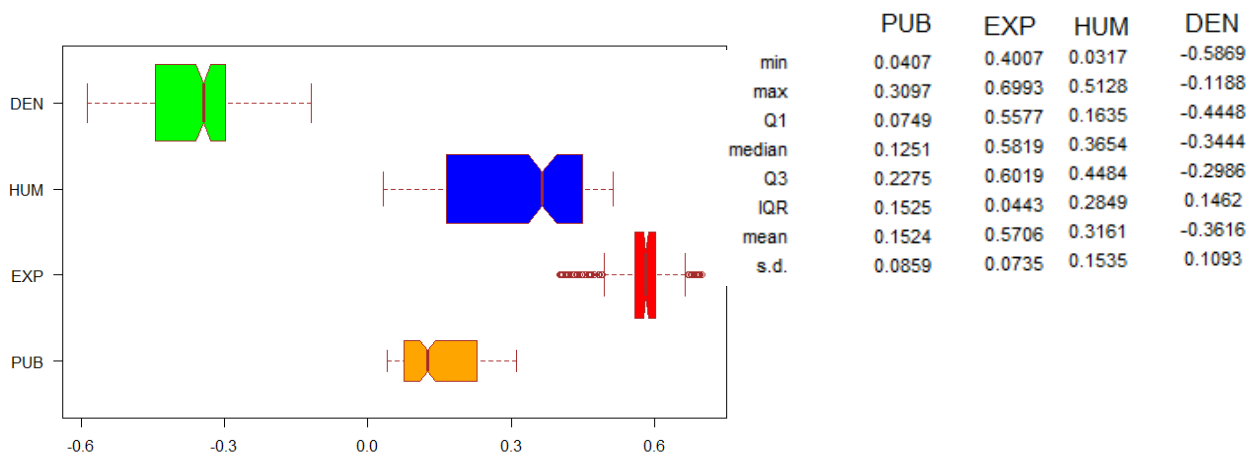


Fig. 4: Multiple boxplots for comparison – estimated parameters

Notes: PUB – Most-cited publications; EXP – R&D expenditure in the business sector; HUM – Human resources in science and technology; DEN – Density; Q1 – First quartile; Q3 – Third quartile; IQR – Interquartile range; s.d. – Standard deviation

Source: author's calculations

Both dummy variables were also included in the MGWR-SAR estimation as local variables, which is possible because the MGWR-SAR function allows dropping variables with low local variance in the local regression, which allows dummies to be considered without problems. MGWR-SAR also allows local outliers to be dropped in the local regression.

Finally, we constructed cluster maps based on the Machine Learning clustering algorithm to assess the parameters' spatial heterogeneity. Our clustering method was the density-based spatial clustering of applications with noise (DBSCAN). We decided on this because the DBSCAN does not require specifying the number of clusters to use. Only a function to calculate the distance between values need to be specified. Euclidean distance metric was applied. Another advantage compared to other algorithms (e.g., K-means clustering) is the different way the clusters are formed. In, e.g., K-means, clusters depend on the mean value of cluster elements, so each data point influences the forming of clusters. Even a slight change in data points can affect the clustering result (for details, see, e.g., Anselin, 2020).

The resulting cluster maps are shown in Fig.5 and Fig.6. Regarding local parameter estimates of Most cited publications and R&D expenditures in the business sector, we can see some similarities (see Fig.5). Both DBSCAN cluster maps contain one large cluster (with 157 and 142 observations, respectively) and some small ones. In addition, one region and seven regions are classified as outliers (noise). We see a different spatial pattern in the cluster maps for the Human resources in science and technology and Density. As we can see in the cluster maps in Fig.6, the clusters are now much smaller. In both cases, one large cluster in the previous case, covering mainly the Central and Western European regions, is being decomposed into several smaller sub-clusters. In the case of the innovation input Human resources in science and technology, up to 21 regions were identified as outliers. These results support the assumption of spatial heterogeneity, the spatial differentiation of parameters is very evident for human resources and population density, as we found many small clusters. On the other hand, the magnitudes of scientific publications and R&D expenditures seem more homogeneous across European regions.

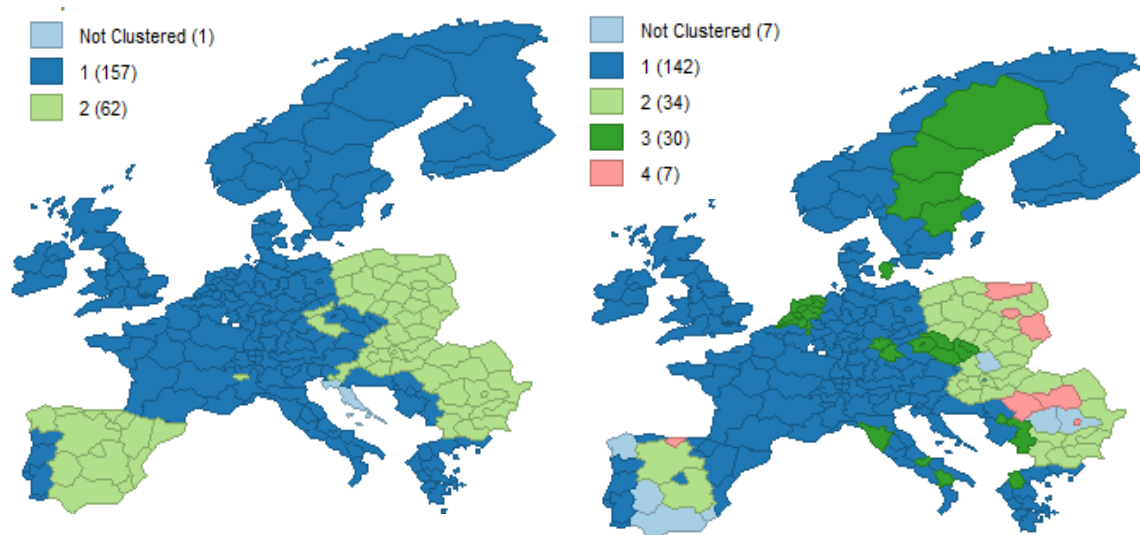


Fig. 5: DBSCAN cluster maps for Most-cited publications for R&D expenditure in the business sector
 Source: author's calculations

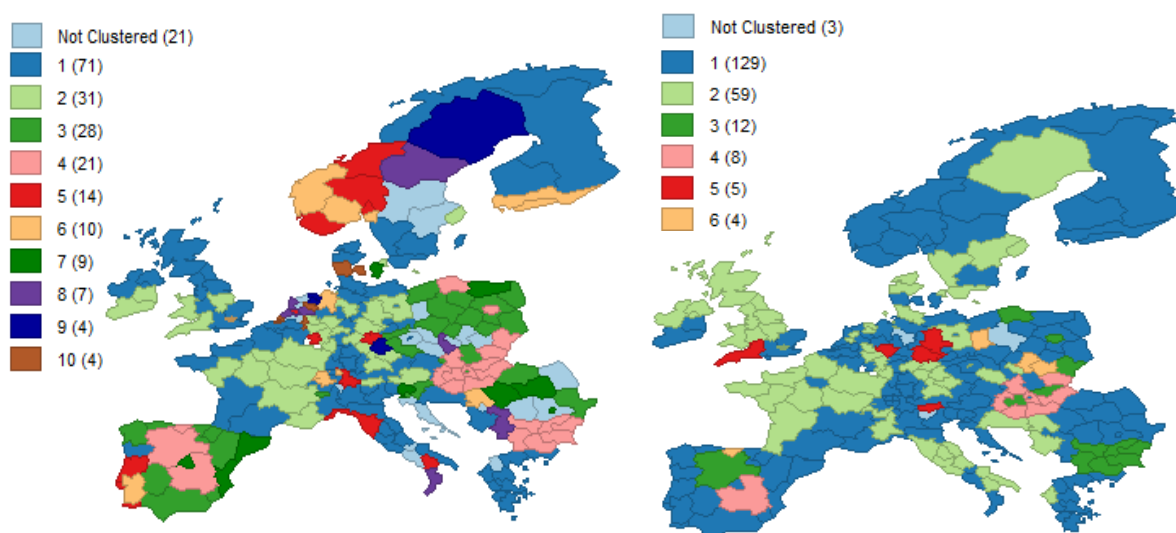


Fig. 6: DBSCAN cluster map for Human resources in science and technology and for Density
Source: author's calculations

At the end of the empirical part, we will separately deal with the estimates of the spatial autoregressive parameter. The estimated values of this parameter varied relatively high among regions, as indicated by both the natural breaks maps and the larger number of DBSCAN clusters (see Fig.7(a) and Fig.7(c)). Furthermore, it can be noted that the spatial autoregressive parameter was statistically significant in almost half of all regions (47.73%). Higher intensities of interregional innovation links were found in areas where the values of the parameter estimates were statistically significant (see Fig.7(a) and Fig.7(b)). An interesting finding is that spatial interactions are not a significant factor affecting innovation in the regions with the highest patenting activities (see Fig.1). However, we must remind that the proximity was defined through the similarity of the regions regarding SMEs. The geographical proximity of the most innovative regions and the resulting benefits for driving innovation seems more important than the similarity of these regions in terms of SMEs. These conclusions regarding the geographical proximity of the top innovation actors are evident from many empirical analyses. Clusters of innovative companies and the geographical proximity of research centres or universities allow their cooperation, sharing, and faster dissemination of new knowledge and technologies. On the other hand, the similarity of the regions in terms of SMEs seems to be a significant factor stimulating innovation processes for regions that are not among the top innovators.

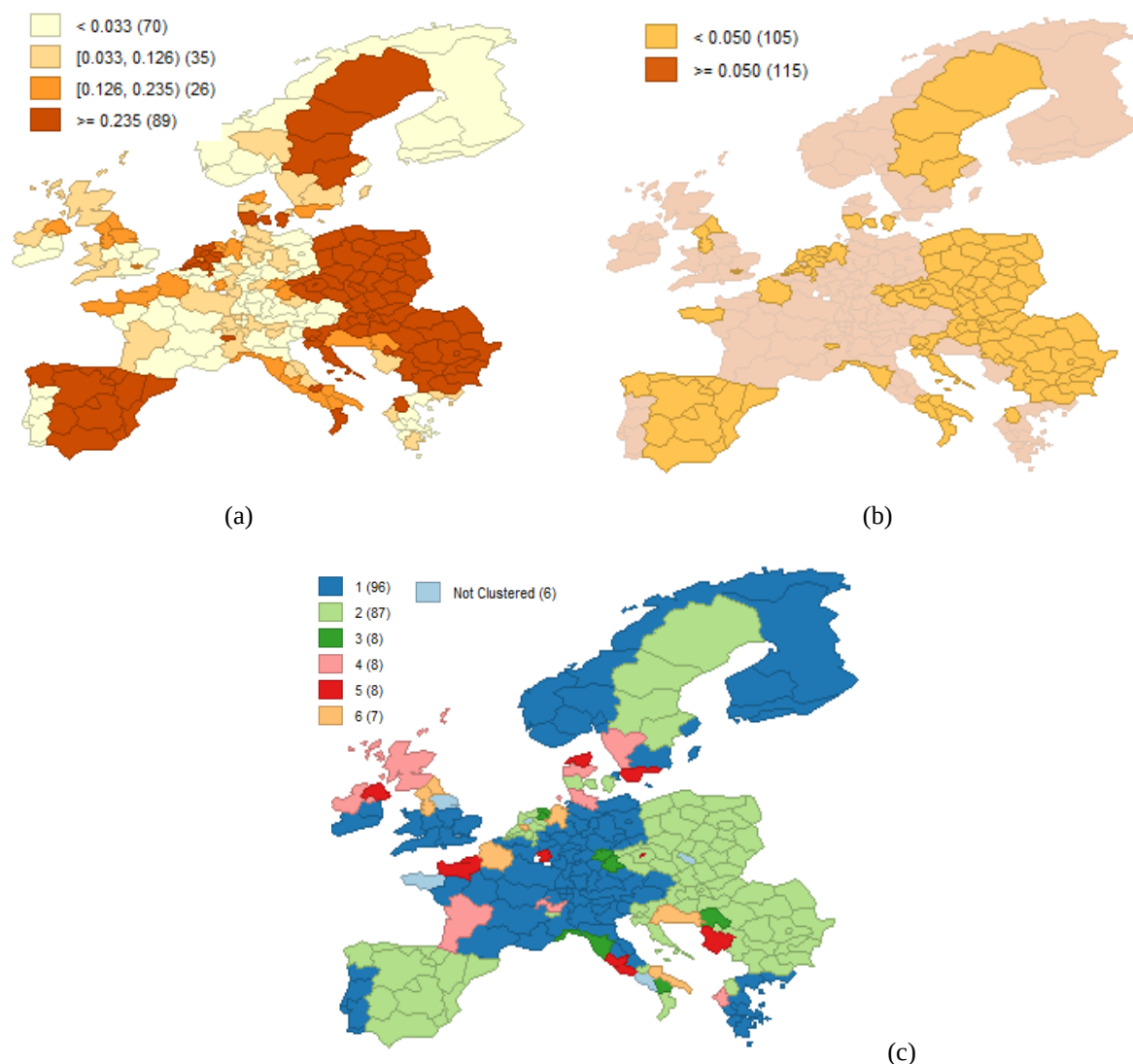


Fig. 7: Jenks's natural break map (a), statistical significance map (b), and DBSCAN cluster map (c) for spatial autoregressive parameter estimates
 Source: author's calculations

4. Conclusions and some policy implications

This paper focuses on a spatial econometric analysis of 220 European regions based on a new class of data-generating processes, the MGWR-SAR approach. A significant advantage of this class of models is the simultaneous consideration of spatial autocorrelation of the endogenous variable and spatial heterogeneity of the parameters. As for European innovation, several studies have investigated both spatial effects, and their conclusions consider both spatial aspects. However, the simultaneous consideration of these effects within a single model is scarce, and therefore we intended to contribute to the innovation literature by estimating just such a model. Another aspect that distinguishes the analysis of this paper from the majority of analyses in a similar area can be considered the use of an alternative measure of proximity between regions concerning the European innovation problem. We have moved from geographical proximity to technological similarity based on the two indicators (Small and medium-sized enterprises (SMEs) introducing product or process

innovations and Small and medium-sized (SMEs) introducing marketing or organizational innovations. Patent applications were considered a proxy of innovation output.

The results supported the assumption of spatial heterogeneity; spatial differentiation of parameters was evident for innovation inputs, human resources, and population density, as we found many small clusters. It also appears that magnitudes of innovation input scientific publications and R&D expenditures are more homogeneous within the European regions. Considering the spatial spillovers, we found that regional innovation activities do not have only a local character in almost half of the regions. Their effects can be spread and might have impacts beyond the boundaries of the regions. An interesting finding was that spatial interactions were not a significant factor affecting innovation in the regions with the highest patenting activities. It should be remembered that proximity was defined through regional similarity in terms of SMEs. The geographical proximity of the most innovative regions and the resulting benefits for driving innovation seems more important than the similarity of these regions in terms of SMEs. These conclusions regarding the geographical proximity of the top innovation actors are evident from many empirical analyses (see, e.g., Autor, 2019). On the other hand, the similarity of the regions in terms of SMEs seems to be a significant factor stimulating innovation processes for regions that are not among the top innovators.

As our results provided individual regional estimates of parameters, innovation, and cluster policy instruments might be implemented more effectively, and a more place-based policy strategy can be pursued. For example, it is possible to predict whether the potential invested funds will only have a local benefit, or whether their benefit will spread to other regions as well. The results of such analyses help determine the targets of innovation and cluster policies. The interest does not have to be focused exclusively on the leading regions - the top innovators. However, policymakers can emphasize less developed regions or a specific group of actors, such as SMEs.

5. Acknowledgements

This work was supported by the Grant Agency of the Slovak Republic VEGA 1/0047/23 „The importance of spatial spillover effects in the context of the EU's greener and carbon-free Europe priority“.

6. References

- [1] Anselin, L. (2020). Density-Based Clustering Methods. [online]. [cit. 19.09.2022]. Available at: https://geodacenter.github.io/workbook/99_density/lab9b.html
- [2] Anselin, L., Rey, S. J. (2014). Modern Spatial Econometrics in Practice. GeoDa Press LLC, Chicago.
- [3] Basile, R., Durbán, M., Mínguez, R., Montero, J.M., Mur, J. (2014). Modeling regional economic dynamics: spatial dependence, spatial heterogeneity and nonlinearities. *Journal of Economic Dynamics and Control*, 48: 229–245.
- [4] Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, 39: 61–74.
- [5] Charlot, S., Crescenzi, R., Musoleli, A. (2015). Econometric modelling of the regional knowledge production function in Europe. *Journal of Economic Geography*, 15: 1227–1259.
- [6] Cho, S. H., Lambert, D., Roberts, R., Kim, S. (2010). Moderating urban sprawl: Is there a balance between shared open space and housing parcel size? *Journal of Economic Geography*, 10(5): 763–783.
- [7] Eurostat (2021): Eurostat regional statistical database. [online]. [cit. 01.06.2021]. Available at: <http://ec.europa.eu/eurostat/>
- [8] Fisher, W. D. (1958). On Grouping for Maximum Homogeneity. *Journal of the American Statistical Association*, 53: 789–98.
- [9] Autor (2019).
- [10] Geniaux, G, Martinetti, D. (2018). A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. *Regional Science and Urban Economics*, 72: 74–85.
- [11] Guastella, G., van Oort, F. (2015). Regional Heterogeneity and Interregional Research Spillovers in European Innovation: Modelling and Policy Implications. *Regional Studies*, 49(11): 1-16.

- [12] Hollanders, H., Es-Sadki, N., Mekelbach, I. (2019). Regional Innovation Scoreboard 2019. [online]. [cit. 01.03.2020]. Available at: <https://ec.europa.eu/growth/sites/growth/files/ris2019.pdf>
- [13] Jenks, G. F. (1977). Optimal Data Classification for Choropleth Maps. Occasional Paper no. 2. Department of Geography, University of Kansas.
- [14] Khan, B. Z. (2012). Of Time and Space: A Spatial Analysis of Technological Spillovers among Patents and Unpatented Innovations in the Nineteenth Century. [online]. [cit. 01.06.2017]. Available at: <http://www.nber.org/papers/w20732>
- [15] Kumar, I. (2008). Innovation Clusters: A Study of Patents and Citations ESRI International User Conference. [online]. [cit. 01.06.2017]. Available at: <https://www.pcrdpurdue.edu/files/media/Innovation-Clusters-A-Study-of-Patents-and-Citations.pdf>
- [16] Moreno, R., Paci, R., Usai, S. (2005a). Spatial Spillovers and Innovation Activity in European Regions. *Environment and Planning A: Economy and Space*, 37(10): 1793-1812.
- [17] Moreno, R., Paci, R., Usai, S. (2005b). Innovation clusters in the European regions. [online]. [cit. 10.03.2017]. Available at: <http://crenos.unica.it/crenos/sites/default/files/wp/05-12.pdf>

Manuscript received: 16.11.2022

Manuscript received in revised form: 28.11.2023

Manuscript accepted: 25.04.2024