The identification of spatial dependence in the analysis of regional economic development – join-count test application

Michał B. Pietrzak[1], Justyna Wilk[2], Tomasz Kossowski[3], Roger Bivand[4]

Abstract

The content of the article is focused on the problem of spatial dependence identification, referring to the analysis of regional economic development. Economic development, as an economic category, represents a very extensive concept. Within its framework the following problems should be considered: the influence of institutions performing social services, environmental conditions of residence and the overall complex of general economic phenomena. The correct measurement of economic development requires the selection of adequate diagnostic variables as well as the construction of the taxonomic measure of development. The implementation of taxonomic measure of development allows for the division of regions into homogenous classes, however, reduces the possibility of statistical tests usage for quantitative data, which decide about the occurrence of spatial dependence.

The objective of the article is to identify spatial dependence of regions covered by classes characterized by the defined economic development level having applied the joint-count test. This test allows detecting the spatial autocorrelation for quantitative data. The test application allows detecting whether the regions, within the framework of each specified class, are grouped based on the existing spatial dependence and create clusters featuring similar economic development level.

***Keywords**: explorative spatial data analysis (ESDA), spatial dependence, join-count test, economic development, taxonomic measure of development.*

JEL Classification: C21, C51, J64, R11

1. Introduction

The problem of spatial dependence is more and more frequently discussed within the framework of spatial economic research. This particular concept is of vital importance since it indicates the occurrence of certain phenomena intensity depending on their spatial location. In case of the majority of socio-economic phenomena the existence of positive spatial dependence is their natural property. This observation was presented in the form of Tobler's First Law of Geography according to which the higher the level of interaction between

[1] Corresponding author: Nicolaus Copernicus University, Department of Econometrics and Statistics, Gagarina 13A, 87-100 Torun, Poland, e-mail: Pietrzak@umk.pl.

[2] Wrocław University of Economics, The Department of Econometrics and Computer Science, Nowowiejska 3, 58-500 Jelenia Góra, Poland, e-mail: Justyna.Wilk@ue.wroc.pl.

[3] A. Mickiewicz University in Poznań, The Department of Spatial Econometrics, ul. Dzięgielowa 27, 61-680 Poznań, Poland, e-mail: tkoss@amu.edu.pl.

[4] Norwegian School of Economics (NHH), Department of Economics, Helleveien 30, NO-5045 Bergen, Norway, e-mail: Roger.Bivand@nhh.no.

regions the closer they are spatially located (Tobler, 1970). Failure to include the existing spatial dependence in economic research can lead to cognitive errors (Paelinck and Nijkamp, 1975, Paelinck and Klaassen, 1979, Anselin, 1988, Haining, 2003, Arbia, 2006, LeSage and Pace, 2009).

The function of spatial autocorrelation is most often applied in the identification of spatial dependence with reference to socio-economic phenomena. Statistical tests, allowing statistical significance testing of spatial autocorrelation, are commonly included among the tools of explorative spatial data analysis (ESDA) (Anselin, 1999). The discussed tests can be divided into these used for the purposes of spatial autocorrelation significance identification in both, global and local sense. Spatial autocorrelation, as the global measure of spatial dependence, applies jointly to all the analyzed regions. Spatial autocorrelation, understood in the local sense, refers to the analyzed individual regions. In case of the globally approached spatial autocorrelation Moran's I test represents the generally applied tool (Cliff and Ord, 1973, 1981, Anselin, 1988). Moran's I statistics allows for the visualization of spatial regional properties, by means of Moran scatter plot, relevant to their coverage by the specified classes (Florax and Nijkamp, 2003). Geary's C test, based on statistical data, is less frequently applied (Cliff and Ord, 1981). The set of tests applied for the purposes of spatial correlation significance, in the local perspective, is referred to in literature references as local indicator of spatial association (LISA) (Anselin, 1995). Currently LISA covers Moran's I local statistics, Geary's C local statistics and Getis-Ord G local statistics (Anselin, 1995). Based on the results obtained within LISA framework it is possible to perform spatial clusters identification and distinguish the outlier regions, which are significantly different from the identified neighbours in line with the adopted neighbourhood matrix. The specification of global spatial autocorrelation and the identification of individual outlier regions and local spatial clusters allows for better understanding of the spatial nature relevant to the occurring phenomena. It can also turn out helpful in the regional development policy planning and with reference to spatial management.

In terms of qualitative data the measurement of spatial dependence, in the global perspective, is possible following the joint-count test application (vide: Cliff and Ord, 1973, 1981). This test allows for testing dependence for an unlimited number of quality variable variants. There is also a joint-count test developed version allowing for spatial dependence measurement in local perspective. The local joint-count test version is referred to in literature as local indicators for categorical data (LICD) (Boots, 2003). Similarly to quantity variables, local indicators for categorical data allow for the identification of outliers or the regions covered by spatial clusters.

Attention should be paid to the fact that spatial economic research applies, quite frequently, qualitative data in the form of regional division into classes characterized by the analyzed phenomenon similar level. This division is most often performed based on quantitative synthetic variable (taxonomic measure of development). The taxonomic measure of development (TMD) constitutes the resultant of quantitative variables referring to diverse aspects of the studied phenomenon. Therefore, the subject matter discussed in the article is focused on the problem of spatial dependence testing for qualitative variables. Within the framework of the analyzed problem joint-count test will be applied for the purposes of spatial dependence identification in terms of regional economic development level. The analysis will be conducted for (NUTS 3) sub-regions in Poland in 2011. The test will be used to identify the occurrence of spatial dependence for two variants of a variable, i.e. low and relatively high economic development level. The discussed variants will be determined based on the taxonomic measure of development specified for the sub-regions. The obtained test results will allow the identification of spatial dependence for the selected test variants in the global perspective.

2. The measurement of economic development level

Taxonomic methods are applied, in particular, in spatial studies of economic development level and its changes. Cluster analysis algorithms represent the first group of these methods. They allow distinguishing internally homogenous and externally separable classes of objects. They are useful, among others, in the situation when the purpose of the study is to identify regional clusters featuring similar job market structure, similar economic profile, etc. The second group covers methods aimed at the arrangement of objects (regions) in accordance with the primary criterion which is not subject to direct measurement. They allow determining the distance of objects from a given, most frequently predetermined pattern of development. They represent one of the basic tools applied in economic development level measurement. The obtained results constitute the basis for distinguishing regional clusters characterized by a diversified economic development level.

The purpose of the research is to specify the level of economic development for 66 Polish (NUTS 3) sub-regions in 2011. The economic development level measurement is, however, a complex task and requires many economic aspects to be considered. The set of variables referring to production level, economic growth, entrepreneurship, as well as the willingness to invest, the financial condition of enterprises and the situation at regional job market, were included in the study. The set of variables covered GDP per capita (X1), national economy entities included in the Official Company Register (REGON) per 10 000 population (X2), investment outlays in enterprises per capita (X3), average monthly gross salary (X4) and the registered unemployment rate (X5), for which the relevant data covering 2011 were available. The variables were selected in the way which allowed meeting the following criteria: comparability, clear definition of the problem, measurability and usefulness in the description of phenomena at NUTS-3 regional level. The majority of variables (apart from unemployment rate) represent stimulants.

In the next step, the arrangement of objects, based on variables values, was conducted. The TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method served this particular purpose (Hwang and Yoon, 1981). The regions were compared by referring their economic situation to an ideal positive pattern and also the negative ideal pattern. In the presented study, the positive ideal pattern takes the form of an artificial object which represents the best values of variables, i.e. the highest (real reached) values of variables having positive impact (stimulants) and the lowest (real reached) values of variables having negative impact (destimulants). The negative ideal pattern is calculated inversely.

Next, the data were normalized using unitization with zero minimum ([x–minimum]/range, where x means variable implementation). After normalization the variables take values in the range [0, 1]. Then the values of variables with negative impact (registered unemployment rate) were translated into variables exerting positive impact by (x–1).

Following the above, Euclidean distances between each region and the positive ideal pattern (PIP) and also between each region and the negative ideal pattern (NIP), were calculated. Then the values of the synthetic measure of economic development level were calculated ([distance to NIP / (distance to NIP + distance to PIP)]).

The synthetic measure takes its values in the range [0, 1], where 1 is determined for a region representing the most favourable variables values, while 0 is presented by a region noting the most unfavourable variables values. The highest value (0.998) was recorded for the city of Warsaw; the capital city of Poland. The second position was taken by the city of Poznań (0.703). The rest of sub-regions' values were classified in the range [0.654, 0.182], while the lowest value was recorded for Ełcki sub-region, located in Warmińsko-Mazurskie region (NUTS 2).

The interval of synthetic measure values was divided into two classes, using the value of median, representing the low level of economic development (class I) and the relatively high level (class II). Figure 1 presents the conducted analysis results with visible regional clusters characterized by the low level of economic development. Figure 1 also illustrates regional clusters featuring a relatively high level of economic development (class I), as well as individual sub-regions from class I representing outliers at the background of low development level regions (class II). The application of joint-count test should allow the verification of presented conclusions based on the visual analysis of the spatial diversification of economic development.

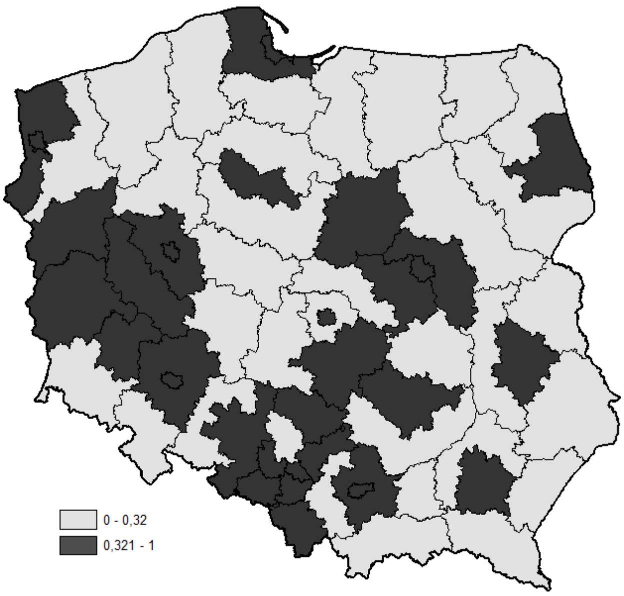

Fig. 1. The division of sub-regions in Poland in terms of economic development level.

3. The identification of spatial dependence in the analysis of economic development

The idea underlying the joint-count statistics development results from the problem related to the analysis of spatial distribution of qualitative variables. In the simplest case it is adopted that the quality variable takes only two values. These values can be presented on a map by using two colours: white and black. The spatial distribution of quality variable can be of random nature, or can present the tendency for spatial clustering. Regional clustering is illustrated on a map in the form of colours grouping. Owing to the quality-specific nature of data the assessment of this autocorrelation level is not possible based on the variable values. In case of positive spatial autocorrelation occurrence the neighbourhood of units marked by the same colour should be the dominating one over the neighbourhood of units having different colours. Otherwise a negative correlation can be adopted. If “one-colour” neighbourhoods are not distinctively dominant over the “two-colour” ones it can indicate the random distribution of a variable on a map. In case of two-colour maps the idea of joint-count statistics consist in counting the white-white (WW), white-black (WB) and black-black (BB) types of neighbourhoods. The distributions of joint-count test statistics, as well as the generalization of joint-count statistics into quality variables adopting m different values, are discussed in research studies by Cliff and Ord (1973, 1981). In case of multivalent qualitative variables both, statistical forms and their distribution moments are complicated in a significant way. In such situation the frequently applied approach is the division of colours into two groups or testing one of them against the remaining ones.

The joint-count test was conducted based on the specified division of sub-regions into two classes in terms of economic development (figure 1). The test verified two model hypotheses, the spatial clustering of class I sub-regions more often than in case of spatial randomized distribution and spatial clustering of class II sub-regions. The obtained results are presented in table 1. Following the conducted joint-count tests the occurrence of positive spatial dependence in case of sub-regions presenting low level of socio-economic development (class I) and the insignificant spatial dependence for sub-regions characterized by a relatively high development level (class II) was observed. It means that sub-regions featuring low development level present the tendency for spatial clustering which is manifested in the form of spatial clusters occurrence.

Test type	**Statistics**	**Expected value**	**Variance**	**Z-value**
1:1	10.4591	8.1230	0.6968	2.798
2:2	8.8863	8.1230	0.6968	0.914
1:2	13.6545	16.7538	2.0044	-2.189
Jtot	13.6545	16.7538	2.0044	-2.189

Table 1 Joint-count test results.

The preliminary, visual analysis of the discussed classes' spatial distribution indicated clustering of regions characterized by both, the relatively high economic development level and the regions showing low development level (vide: Figure 1). However, based on the conducted joint-count test the occurrence of positive dependence was recorded only in case of sub-regions featuring low development level (class I). In case of regions presenting a relatively high development level statistically insignificant spatial dependence were detected. The obtained results indicate the occurrence of spatial regional clusters characterized by the low level of economic development. Spatial clustering of regions presenting low development level indicates that in the area of these sub-regions slow, however, ongoing withdrawal of resources such as enterprises, human capital, etc., is taking place. It results in the advancing deterioration of the situation in the regions grouped in such spatial cluster. It also brings about the expansion of spatial cluster boundaries by more regions featuring low development level. The observed positive spatial dependence illustrates that this situation is difficult to change and, additionally, it will keep advancing by further decrease in the level of development comparing to class II sub-regions.

The obtained joint-count test results are contradictory to the visual assessment of sub-regions' spatial distribution where the grouping of class II regions is visible. In spite of the statistically insignificant spatial dependence for the sub-regions presenting a relatively high development level their spatial distribution should be analyzed once again. The applied joint-count test represents global statistics, which means that the obtained result is averaged for all sub-regions from class II. The presence of individual sub-regions can significantly influence such test results. Additionally, attention should be paid to the fact that the analyzed unit was the sub- region which can be referred to as an independent macro-region in terms of its economic situation. Smaller urban centres, exerting limited impact, establish individual growth macro-regions the boundaries of which are closed within a given sub-region. This situation is true in case of the following regions: Podlaskie (Białostocki sub-region), Lubelskie (Lubelski sub-region), Podkarpackie (Rzeszowski) and in Kujawsko-Pomorskie region (Bydgosko-Toruński sub-region). Such situation could be manifested in the value of joint-count test statistics and result in the conclusion presenting statistically insignificant spatial dependence. Positive economic situation of these sub-regions results exclusively from the influence of urban centres in the above-mentioned regions, i.e. in Lubelskie region from the influence of Lublin, in Podkarpackie – Rzeszów and in Kujawsko-Pomorskie – Bydgoszcz and Toruń. This also exerts impact on the economic situation of the remaining sub-regions, covered by these regions, which presents low level (class I). It has to be emphasized that the majority of sub-regions create spatial clusters featuring low development level, visible on figure 1.

In the situation when urban centres constitute very strong poles of economic growth, their impact extends outside sub-regional areas and may cover the neighbouring sub-regions. In case of Poland strong centres of growth can be divided into two types of centres. The first type is made up of medium impact centres of growth which is most often limited to one sub-region. The second type covers strong impact centres where the impact is strong enough to create, around the growth centre, sub-regional spatial cluster presenting a relatively high level of socio-economic growth.

Among the first type centres the following are listed: Tri-city (Pomorskie region), Szczecin (Zachodniopomorskie region) and Cracow (Małopolskie region), where the influence of these centres spreads over the neighbouring, individual sub-regions. The second case refers to three spatial clusters. The first cluster covers Warsaw in Mazowieckie region, the second one refers to Poznań and Wrocław and covers the sub-regions of Wielkopolskie, Dolnośląskie and Lubuskie regions, as well as the cluster referring to strong urban centres of Śląskie region, such as Katowice, Tychy and Gliwice, which can also cover sub-regions of Opolskie, Łódzkie and Małopolskie regions.

Therefore it seems founded to perform further quantitative analysis taking the form of local indicators for categorical data (LICD). This analysis would probably present the occurrence of individual sub-regions featuring a relatively high level of development surrounded by sub-regions characterized by low development level, where an urban centre, representing the growth pole, exerts a relatively limited impact. Simultaneously the statistically significance of spatial dependence occurrence in sub-regions of relatively high development level could be observed (the above-listed three spatial clusters).

Having analyzed the occurrence of statistically significant spatial dependence for sub-regions included in the class presenting low level of economic development, a hypothesis can be put forward that such situation results mainly from the sub-regions covered by the class featuring a relatively high development level. The underlying reason for that is the fact that the sub-regions constituting strong growth centres facilitate the economic situation improvement in the selected sub-regions only by establishing spatial clusters with them. In case of the remaining sub-regions it results in their valuable capital and human resources drainage which, on the other hand, is manifested by further deterioration of their development level. The discussed hypotheses confirmation would mean that the regional spatial clusters, characterized by the poor economic situation, are established not only as the result of poor situation in the relevant sub-regions, but mostly due to the influence of the closest growth centres to which the most valuable resources are moved from the sub-regions of low economic development. It results in an even worse situation in regions made up of spatial clusters characterized by low development level and stimulates the divergence phenomenon. Ultimately, the statistical significance of spatial dependence for class I regions was probably influenced by class II sub-regions.

Specifying the value of local indicators for categorical data seems to be the natural supplementation of the results obtained based on joint-count test which points to spatial dependence of global nature. It can become the tool for spatial clusters identification in the situation when the joint-count test indicates statistical insignificance of spatial dependence. Obviously, if statistically significant spatial dependence is recorded in the selected class of regions, the local indicators for categorical data, as a statistical tool, play the same role.

Conclusions

The article discusses the problem of testing spatial dependence for qualitative variables. The major objective of the study was to consider the joint-count test and its application in an empirical example referring to the analysis of economic development level in (NUTS 3) sub-regions in Poland in 2011. The division of regions into two classes was conducted using the taxonomic measure of development. The set of diagnostic variables, characterizing the chosen aspects of regional economic situation, was selected for the purposes of an indicator construction. Such division allowed for the joint-count test application and the identification of spatial dependence for the selected test variants. The obtained test results resulted in observing statistically significant spatial dependence exclusively for the sub-regions featuring low level of economic development (class I).

The analysis of dependence, by means of joint-count test, also pointed to its weaknesses owing to the global nature of the applied measure. In case of sub-regions presenting a relatively high development level (class II), the test pointed to statistically insignificance of spatial dependence. The test result could have been influenced by the occurrence of individual sub-regions covered by class II and mainly constituting the regional growth centres.

It has to be emphasized that the joint-count test discussed in the article can also be applied for any other economic phenomena, as well as any possible arrangement of territorial units. Beyond any doubt, in the future the article should be extended by the application of local indicators for categorical data (LICD). The obtained results would be used for the purposes of outlier regions identification, as well as local spatial clusters, in terms of the analyzed economic problem. In the discussed case of economic situation the application of local indicators for categorical data could facilitate more extensive analysis of the studied problem. It could also enhance the confirmation of the research hypothesis put forward in the article, i.e. that the sub-regions featuring a relatively high development level constitute the main reason for the spatial diversification of sub-regions characterized by the low economic development level and the occurrence of spatial dependence.

Acknowledgements

The project was co-financed by Nicolaus Copernicus University in Toruń within the UMK research grant no. 1481-E.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic.

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis*, 27, 93-115.

Anselin, L. (1999). The future of spatial analysis on the social sciences. *Geographical Information Sciences,* 5, 67-76.

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

BOOTS, B. (2003). Developing local measures of spatial association for categorical data. *Journal of Geographical Systems*, 5, 139-160. DOI: 10.1007/s10109-003-0110-3.

Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V. (2008). *Applied Spatial Data Analyses with R*. New York: Springer.

Cliff, A. D., & Ord, J. K. (1973). *Spatial Autocorrelation*. London: Pion.

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

Florax, R., & Nijkapm, P. (2003). *Misspecification in Linear Spatial Regression Models*. Tinbergen Institute Discussion Paper, TI 2003-081/3.

Haining, R. (2003). *Spatial Data Analysis: Theory and Practice.* Cambridge: Cambridge University Press.

Hwang, C. L., & Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications*. New York: Springer-Verlag.

LeSage, J., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Boca Raton: Chapman & Hall/CRC.

Paelinck, J. H. P., & Klaassen, L. H. (1979). *Spatial Econometrics*. Farnborough: Saxon House.

Paelinck, J. H. P., & Nijkamp, P. (1975). *Operational theories and method in regional analysis*. Farnborough: Saxon House.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(2), 234-240.