Institute of Economic Research Working Papers

No. 14/2014

The application of local indicators for categorical data (LICD) in the spatial analysis of economic development

Michał Bernard Pietrzak

Justyna Wilk

Tomasz Kossowski

Roger Bivand

Toruń, Poland 2014

Michał Bernard Pietrzak[1]
Michal.Pietrzak@umk.pl
Nicolaus Copernicus University in Toruń, Department of Econometrics and Statistics, ul. Gagarina 13a, 87-100 Toruń

Justyna Wilk
justyna.wilk@ue.wroc.pl
Wrocław University of Economics, Department of Econometrics and Computer Science, ul. Nowowiejska 3, 58-500 Jelenia Góra

Tomasz Kossowski
tkoss@amu.edu.pl
Adam Mickiewicz University in Poznań, Department of Spatial Econometrics, ul. Dzięgielowa 27, 61-680 Poznań

Roger Bivand
Roger.Bivand@nhh.no
Norwegian School of Economics (NHH) in Bergen, Department of Economics, Helleveien 30, 5045 Bergen

The application of local indicators for categorical data (LICD) in the spatial analysis of economic development

JEL Classification: *C21, C51, J64, R11*

Keywords: *join-count test, spatial dependence, local indicators of spatial association (LISA), explorative spatial data analysis (ESDA), economic development*

Abstract: The objective of this paper is to identify classes of regions presenting different economic situations and apply a join-count test to examine spatial dependences between these classes. The test examines spatial autocorrelation on the basis of qualitative data. The global join-count test indicates general

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

interactions occurring between regions, while the local join-count test examines a tendency to form the spatial clusters (e.g. metropolitan areas).

The study covers the situations of 66 Polish NUTS 3 regions in 2011. Regions were divided into two classes presenting relatively low and high levels of economic development. Taxonomic methods of multivariate data analysis were applied in the research.

The global test proved spatial clustering of economically poor regions but was statistically insignificant as regards well-developed regions. Thus the join-count local join-count test was additionally applied. The test indicated the occurrence of five spatial clusters of NUTS 3 regions. Three of them include economically well-developed regions, while two of them present poor economic situations. Furthermore three spatial outliers (local growth centres), which deteriorate the economic situation of eastern Poland, were also recognized.

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 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 aim of the paper is to identify classes of regions presenting diversified economic situations and apply a join-count test to examine spatial dependences as regards these classes. The study covers the situations of 66 Polish NUTS 3 regions in 2011. Regions were divided into two groups presenting relatively low and relatively high levels of economic development. Groups were distinguished using taxonomic methods of multivariate data analysis.

The paper was divided into two main sections. The first section discusses statistical tests of spatial autocorrelation, presents their classification as regards a frame of reference and also data type, and also explains tests for qualitative data in detail. The second section covers an empirical study; presents distinguished classes of regions and discusses the results of global and local join-count tests.

Methodology of the research

The function of spatial autocorrelation is most often applied in the identification of spatial dependence with reference to socio-economic phenomena. Statistical tests, examining the statistical significance of spatial autocorrelation, are commonly included among the tools of explorative spatial data analysis (ESDA). Anselin distinguished global and local tests of spatial autocorrelation (Anselin 1999).

Global tests examine total spatial autocorrelation between regions, while local testes refer to the situations of individual regions; identify spatial clusters and also outlier regions. The results of the studies can support planning of the regional development policy and spatial management.

The most frequently applied global statistical test of spatial autocorrelation is Moran's I test (Cliff and Ord 1973, 1981, Anselin 1988, Florax and Nijkamp 2003), while Geary's C and Getis-Ord's G tests were also proposed (Cliff and Ord 1981). All these statistics are also available as local indicators of spatial association (LISA). They examine quantitative data set, e.g. the values of per capita Gross Domestic Product presented by regions (Anselin 1995).

In the field of economic research, regions are usually classified as regards their social and economic situations into levels, groups etc. to determine regional diversification. Multivariate data analysis methods, such as cluster analysis, factor analysis etc., are frequently used for these purposes. Classes can be equivalent (e.g. the economic profiles of regions) or ranked (e.g. the good, moderate or poor situations of regional labour markets).

In terms of qualitative data the measurement of spatial dependence, in the global perspective, is possible following the join-count test application (Cliff and Ord 1973, 1981). A local variant of the measure represents a family of local indicators for categorical data (LICD) (Boots 2003).

The values of the global test are determined jointly for all regions and the statistical properties of the test are well known (Cliff and Ord 1973, 1981). One of the most important issues while using a join-count test is to select the type of an adjacency matrix. This significantly affects the analysis results. A contiguity matrix is the most frequently used, while the other approach can be, for example, based on applying k-nearest neighbours method.

Let assume that “white” (W) means relatively poor economic situation, while “black” (B) – relatively good economic situation of a region. In case of two-colour chart, the idea of join-count statistics consists in counting the white-white (WW), white-black (WB) and black-black (BB) types of neighbourhoods (Cliff and Ord 1973, 1981):

$$BB = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} x_i x_j \qquad (1)$$

$$WW = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} (1 - x_i)(1 - x_j) \qquad (2)$$

$$BW = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} (x_i - x_j)^2 \qquad (3)$$

where: x_i, x_j – take the value of 1 for a region belonging to the black class (B) and the value of 0 for a region belonging to the white class (W),

w_{ij} – an element of an adjacency matrix.

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 indicates the random distribution of a variable.

These three statistics can be also used for testing the local dependences (in relation to each single unit). However using local tests is more difficult than using global tests. The first issue is that an neigbourhood matrix is determined separately for each region, using, for example, the contiguity matrix or k-nearest neighbours matrix. The second problem is that the statistical properties of the local test are unknown. Thus a significance of spatial dependences can not be statistically validated.

Classes of regions presenting different levels of economic development

Comparative studies examining the economic situation of regions and its territorial diversification frequently use taxonomic methods proposed in the field of multivariate data analysis. The first group of these methods tends to distinguish internally homogenous and externally separable classes of units.

This is the domain of cluster analysis, but the other approaches are also applied, e.g. multidimensional scaling etc. (Hair et al. 2006, Everitt and Dunn 2001). These methods are useful, among others, in the situation when the purpose of the study is to identify regional clusters featuring, for example, similar job market structure, similar economic profile, etc.

The second group covers methods used to arrange the units in accordance with a superior criterion. These methods determine the positions of units in comparison to the other units. International literature most frequently indicates factor analysis in the field (Hair et al. 2006, Everitt and Dunn 2001).

The last group represents one of the basic tools applied in the measurement of economic development levels of regions. This approach will be applied in the following study. The purpose of the research is to examine regional diversification of economic development in Poland in 2011. The study covers the situation of 66 Polish NUTS 3 regions, located in 16 NUTS 2 regions (Figure 1).

The economic development refers to a production level, economic growth, entrepreneurship, as well as the willingness to invest, and also the situation at regional job markets. Table 1 presents the set of diagnostic variables.

The set of variables met the following criteria: comparability, clear definition of the research problem, measurability and usefulness in the description of phenomena for NUTS 3 regions, relatively high statistical variation and low statistical correlation. The majority of selected variables indicate stimulants of economic development. Only the unemployment level adversely affects economic development of regions.

Table 1. The set of diagnostic variables

No	Name of variable	Unit
1	*Per capita* Gross Domestic Product	PLN
2	National economy entities included in the REGON register per 10,000 inhabitants	Entity
3	*Per capita* investment outlays in enterprises	PLN
4	Average monthly gross salaries and wages	PLN
5	Registered unemployment rate	%

Source: elaborated by the authors.

Figure 1. The 16 Polish NUTS 2 regions (dark bold line) and 66 Polish NUTS 3 regions (grey line)

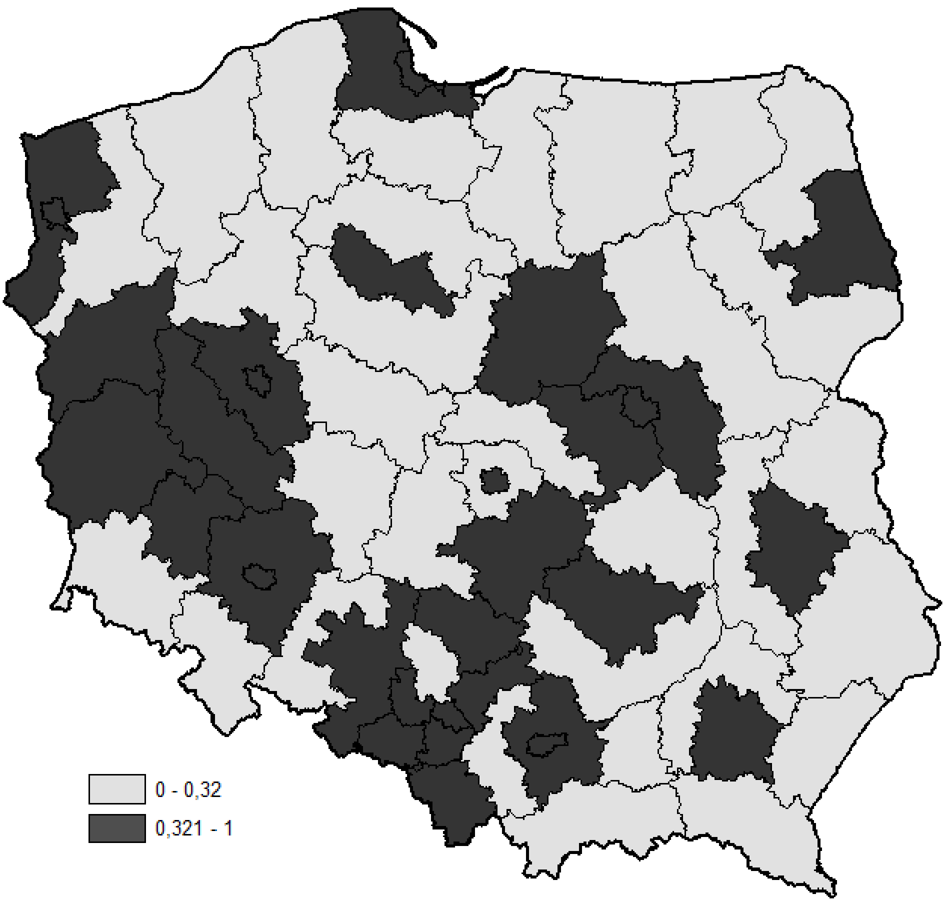
Source: elaborated by the authors.

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 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 was normalized using unitization with zero minimum. After normalization the variables take values in the range [0.0, 1.0]. Then the values of variables with negative impact (registered unemployment rate) were translated into variables exerting positive impact by subtracting the value of 1.

Figure 2. Two classes of NUTS 3 Polish regions presenting different levels of economic development



Source: elaborated by the authors on the basis of data provided by Local Data Bank of the Central Statistical Office of Poland (BDL GUS).

Following the above, Euclidean distances between each region (ith region) and the positive ideal pattern (PIP) and also between each region (ith region) and the negative ideal pattern (NIP), were calculated. Then the values of the synthetic measure for each region (SMi) were calculated (Hwang and Yoon 1981):

$$SM_i = \frac{NIP_i}{NIP_i + PIP_i} \qquad (4)$$

The synthetic measure takes its values in the range [0.0, 1.0], where 1 is determined for a region presenting 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 Mazowieckie NUTS 2 region); the capital city of Poland.

The second position was taken by the city of Poznań (0.703) located in the Wielkopolskie NUTS 2 region. The rest of regions took values in the range [0.654, 0.182]. The lowest value was recorded for Ełcki region, located in the Warmińsko-Mazurskie NUTS 2 region.

The interval of synthetic measure values was divided into two classes representing, respectively, the low level of economic development (white colour) and the relatively high level (black colour) using the median value. Figure 2 presents the analysis results.

There are visible clusters characterized by the low level of economic development, e.g. in the northern Poland (apart Trójmiejski NUTS 3 region and its surrounding Gdański NUTS 3 region). Figure 2 also illustrates clusters featuring a relatively high level of economic development, in western Poland, southern Poland and also central Poland. We can also notice outlier well-developed regions, e.g. in eastern Poland.

Global join-count test of spatial dependence between classes

The application of a join-count test will verify conclusions made in previous chapter which were based on the visual analysis of the regional diversification of economic situation in Poland. Table 2 presents the analysis results.

The test proved the occurrence of positive spatial dependence in case of regions presenting low level of socio-economic development (white) and the insignificant spatial dependence for regions characterized by a relatively high development level (black). Therefore regions featuring low development level present the tendency for spatial clustering.

Table 2. Global join-count test results

Type of tested relation	Statistics	Expected value	Variance	Z-value
WW	10.4591	8.1230	0.6968	2.798
BB	8.8863	8.1230	0.6968	0.914
BW	13.6545	16.7538	2.0044	–2.189

Source: elaborated by the authors on the basis of spdep package (Bivand et al. 2014) of R-CRAN.

Spatial clustering of regions presenting low development level indicates that these regions present slow, however, ongoing withdrawal of resources such as enterprises, human capital, etc. 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 black class (well developed) regions.

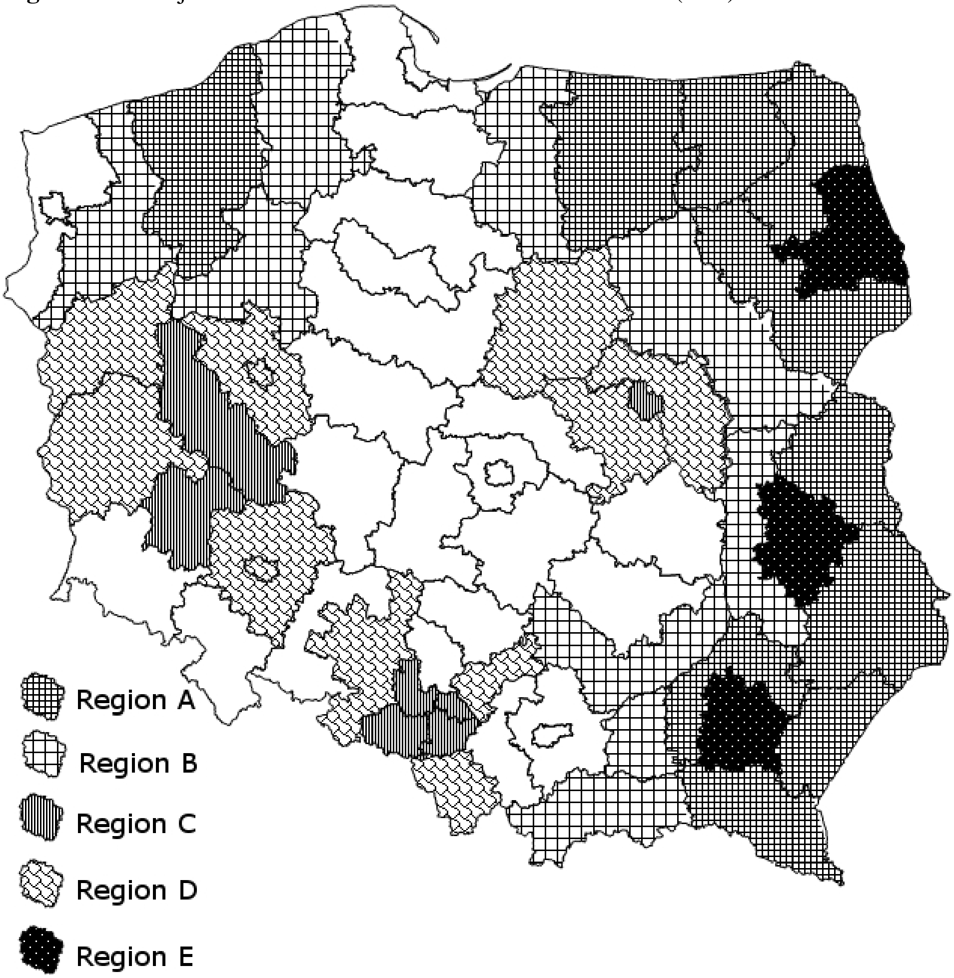
The results of join-count test for black-to-black relation are contradictory to the visual assessment of regions' spatial distribution. Note that the applied join-count test examined only global tendency. The results were then averaged for all regions from black class. The presence of outlier regions can significantly influence such test results. Furthermore some of analyzed NUTS 3 regions are also classified as LAU 2 units and function as independent territorial units (e.g. cities).

Smaller urban centres, exerting limited impact, establish individual growth regions the boundaries of which are closed within a given sub-region. This situation is true in case of the following regions: the Podlaskie NUTS 2 region (the Białostocki NUTS 3 region), the Lubelskie NUTS 2 region (the Lubelski NUTS 3 region), the Podkarpackie NUTS 2 region (the Rzeszowski NUTS 3 region) and the Kujawsko-Pomorskie NUTS 2 region (the Bydgosko-Toruński NUTS 3 region).

Good economic situations of these NUTS 3 regions result exclusively from the influence of urban centres in the above-mentioned regions, i.e. in the Lubelskie NUTS 2 region from the influence of the city of Lublin, in the Podkarpackie NUTS 2 region – the city of Rzeszów, and in the Kujawsko-Pomorskie NUTS 2 region – the cities of Bydgoszcz and Toruń. This also exerts the impact on the economic situation of the remaining

NUTS 3 regions, covered by these NUTS 2 regions, which presents low level of economic development.

Figure 3. Local join-count test results – five economic classes (A-E)



Source: elaborated by the authors.

In the situation when urban centres are very strong areas of economic development, their impact extends outside and may cover the neighbouring NUTS 3 regions. The first type of economic centres is made up of medium impact centres of growth which is most often limited to one NUTS 3 region.

Among the first type centres the following are listed: the Tri-city NUTS 3 region (the Pomorskie NUTS 2 region), the city of Szczecin (the

Zachodniopomorskie NUTS 2 region) and the city of Cracow (the Małopolskie NUTS 2 region), where the influence of these centres spreads over the neighbouring, individual NUTS 3 regions.

The second type of economic centres covers strong impact centres where the impact is strong enough to create, around the growth centre, sub-regional spatial clusters presenting a relatively high level of economic development. This refers to three spatial clusters. The first cluster covers the city of Warsaw in the Mazowieckie NUTS 2 region.

The second one refers to the cities of Poznań and Wrocław, and covers the NUTS 3 regions of the Wielkopolskie, Dolnośląskie and Lubuskie NUTS 2 regions, as well as the cluster referring to strong urban centres of the Śląskie NUTS 2 region, such as the citied of Katowice, Tychy and Gliwice, which can also cover NUTS 3 regions of the Opolskie, Łódzkie and Małopolskie NUTS 2 regions.

Local join-count test of spatial dependence between classes

Specifying the value of local indicators for categorical data seems to be the natural supplementation of the results obtained based on join-count test which points to spatial dependence of global nature. It can become the tool for spatial clusters identification especially in the situation when the join-count test indicates statistical insignificance of spatial dependence.

In the first step the contiguity matrices were determined for each region. Furthermore there were selected 25% of regions presenting the highest values of BB and WW statistics (Equations 1-2). It was assumed that for these regions there is the highest probability of occurring local positive spatial dependences. Figure 3 shows the analysis results.

The A-class includes the NUTS 3 regions noted the highest values of WW statistics. Furthermore the economically poor NUTS 3 regions neighbouring to the A-class regions are defined as the B-class. The regions of A-class and B-class form two spatial clusters presenting poor economic situation.

The largest cluster ("eastern bloc") includes NUTS 3 regions of eastern, south eastern and also north eastern Poland. This cluster comprises almost 25% of the country's total area and approximately 20.0% of the total population. It represents an area with the highest share of agricultural sector in relation to the rest parts of Poland. The Podkarpackie, Lubelskie, Podlaskie and Świętokrzyskie NUTS 2 regions belong also to the poorest regions within the European Union. Their per capita GDP is much below the national average, while the unemployment rate is much above the national average.

The second cluster was formed in north western Poland. In both clusters, the A-class regions establish the cores of clusters, while the B-class regions determine the spatial borders of these clusters.

The C-class covers the NUTS 3 regions presented the highest values of BB statistics. The economically well developed NUTS 3 regions neighbouring with the C-class are belong to D-class. Both classes of regions (C and D) form three spatial clusters of well developed NUTS 3 regions.

Economically the strongest cluster is located in central Poland whose the core is the city of Warsaw. The second cluster covers the most industrialized area of southern Poland. The biggest cluster includes NUTS 3 regions located in the middle of western part of Poland.

Furthermore the Białostocki, Lubelski and Rzeszowski NUTS 3 regions were classified into the E-class due to presenting high values of BW statistics. All of them are located in the eastern Poland, within economically poor spatial cluster. They can be defined as outlier regions due to performing negative spatial dependences with reference to their neighbouring regions.

In other words, outliers establish local growth centres due to presenting relatively good economic situations, while being unable to form economic clusters. They negatively affect economic situations of surrounding regions. The outliers contribute to drain neighbouring regions due to occurring one-way flows of well qualified human resources, being a place of concentration of investment outlays etc. This leads to deteriorating of economic situations of the other "eastern bloc" regions.

Conclusions

The paper made an attempt to apply a join-count test in the analysis of spatial dependences between classes of regions presenting different economic situations. Two approaches were included in the study. The first one examined the global spatial interactions, while the second one concerned particular situations of each region.

The global join-count test resulted in indicating statistically significant spatial dependence exclusively for the NUTS 3 regions featuring low level of economic development (white). In case of NUTS 3 regions presenting a relatively high development level, the test pointed to statistically insignificance of spatial dependence. The test result could have been influenced by the occurrence of individual NUTS 3 regions (like a city) and mainly constituting the regional growth centres. The results pointed out the strengths and weaknesses of the global test.

Using local join-count test facilitated more extensive analysis of the studied problem. It proved that the NUTS 3 regions featuring a relatively high development level constitute the main reason for the spatial diversification of NUTS 3 regions characterized by the low economic development level and the occurrence of spatial dependence.

The results of the local join-count test indicated the occurrence of five economic clusters and also three outlier NUTS 3 regions. Three clusters present relatively good economic situation. The first one includes the NUTS 3 regions of south western Poland. The second one covers regions located in central Poland. The third one includes NUTS 3 regions located in southern part of Poland.

Two clusters of NUTS 3 regions present relatively poor economic situation. The first one is located in the "Polish eastern bloc", while the second one covers north western Poland. Furthermore the Rzeszowski, Lubelski and Białostocki NUTS 3 regions, located within "eastern bloc", are outlier regions (local growth centres).

Although the statistical properties of the local join-count test are unknown and the statistical validation of its results cannot be conducted, the results of the study are very close to real economic situation in Poland. The selection of the highest values of statistics of local BB and local WW tests supports revealing spatial clusters, i.e. core regions and their economically related neighbours. Thus the application of local indicators of spatial association (LISA) can indicate convergence processes, regions exceptionally exposed to poverty, processes of forming metropolitan areas etc. This preliminary study can be a starting point for further research.

References

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

Anselin L. (1995), *Local indicators of spatial association – LISA*, "Geographical Analysis", No. 27.

Anselin L. (1999), *The future of spatial analysis on the social sciences*, "Geographical Information Sciences", No. 5.

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

Boots B. (2003), *Developing local measures of spatial association for categorical data*, "Journal of Geographical Systems", No. 5, http://dx.doi.org/10.1007/s10109-003-0110-3.

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

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

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

Everitt B.S., Dunn G. (2001), *Applied Multivariate Data Analysis*, Arnold, London.

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 University Press, Cambridge.

Hair J.F., Black W.C., Babin B.J, Anderson R.E., Tatham R.L. (2006), *Multivariate Data Analysis*, Pearson Prentice Hall, New Jersey.

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

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

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

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

Tobler W. (1970), *A computer movie simulating urban growth in the Detroit region,* "Economic Geography", No. 46(2).