QUANTITATIVE METHODS IN ECONOMICS
Vol. XV, No. 2, 2014, pp. 188 – 197

APPLICATION OF SPATIAL TECHNIQUES FOR PANEL DATA ANALYSIS OF AGRICULTURAL REAL ESTATE MARKET IN THE YEARS 2004 – 2012

Robert Pietrzykowski
KERiMSG, Division of Quantitative Methods
Warsaw University of Life Sciences – SGGW
e-mail: robert_pietrzykowski@sggw.pl

Abstract: In the economic studies models based on panel data are increasingly used. The standard panel models are composed of a cross-section character of the data in the time, but do not include the interaction associated with the location of objects in the geographic space. Spatial panel models are based on the information contained cross-section data in the time with regard to space. The paper proposes a different approach to the changes in time on the basis of spatial matrix weight. The aim of this study was to show the possibility to apply spatial weights matrix with particular consideration of time. Data for the analysis came from the database of the CSO and the ARMA in period 2004-2012. In addition to working methods of spatial statistics classical taxonomic methods were also used to obtain a distance matrix.

Keywords: weight matrix, spatial analysis, cluster analysis, price of agricultural land

INTRODUCTION

In the economic studies models based on panel data are increasingly used. It is probably connected with the general accessibility of computer programs. The panel data have to contain the time and the object in the cross-section dimension. The property can be defined as the observation of phenomena located in space. The standard panel models are composed of a cross-section character of the data in the time, but do not include the interaction associated with the location of objects in the geographic space. Spatial panel models are based on the information containing cross-section data in the time with regard to space. Description of spatial panel models can be found in Elhorst [Elhorst 2003]. The paper proposes a different approach to the changes in time on the basis of spatial matrix weight. The aim of this study was to show the possibility to apply spatial weights matrix with particular consideration of time. The main aim of the paper was the concluding of time in the weight matrix. In this paper the specific objectives assumed: The first task was to define a links and interactions between neighbours in the time. The second task was to classify regions because of the price of agricultural land based on the modified matrix of weights. The study used Moran scatter plot to classify spatial objects. In addition to working methods of spatial statistics classical taxonomic methods were also used to obtain a distance matrix.

METHODOLOGY AND DATA

Data

Data for the analysis came from the database of the CSO and the ARMA in period 2004-2012. Province was adopted as a unit of spatial. It was due to the availability of data. The observed feature was the price of agricultural land of varying quality.

The matrix of weights - weights based on distances

Undoubtedly fundamental importance for spatial analysis is to define the weight matrix [Suchecka 2014] that reflects the interactions of studied objects. The simplest structure is the weight matrix in which for neighbours we recognize objects which have a common border as follow:

$$\boldsymbol{W} = \begin{cases} \mathrm{w_{ij}} = 1, & \text{the i} - \text{th object is a neighbor of j} - \text{th object} \\ \mathrm{w_{ij}} = 0, & \text{the i} - \text{th object is not a neighbor of j} - \text{th object} \\ \mathrm{w_{ij}} = 0, & when\ \ i = j \end{cases} \tag{1}$$

where:

W – matrix of weight,

w_{ij} – element of matrix of weight.

Because this approach to the problem may not always correspond to reality, therefore it seems to be natural to select another criterion which may be the selection of the neighbours due to the distance d km according to the formula 2. Take the example of two provinces in Poland: Mazowieckie and Wielkopolskie. However one cannot exclude the relation between these regions.

$$\boldsymbol{W} = \begin{cases} \mathrm{w_{ij}} = 1, & the\ i - th\ object\ is\ away\ an\ object\ j - th \leq\ d\ km \\ \mathrm{w_{ij}} = 0, & the\ i - th\ object\ is\ away\ an\ object\ j - th >\ d\ km \\ \mathrm{w_{ij}} = 0, & when\ \ i = j \end{cases} \tag{2}$$

where symbols are as in formula 1.

From above considerations we can see that the matrix **W** has a very characteristic structure, composed of zeros and ones, and the diagonal is zero. Taking into account the distance in kilometers to determine the neighbours opens the possibility of using other measures of the position of objects in space. The Doreian and Conley proposition is to include social [Doreian 1980] and economic distance [Conley 1999] which are based on reciprocal trade relations, capital movements and migration between the two spatial units. However another approach is to determine the position of objects in space through the designation of measures by the formula:

$$d_{ik} = \left[\sum_{j=1}^{m} \left|x_{ij} - x_{kj}\right|^{p}\right]^{1/p} \tag{3}$$

where:

p – number determining the metrics type,

m – number of variables,

x_{ij}, x_{kj} – determine the accomplishment of j-feature in i-th object and k-th object

After determining the metrics type (p = 1 Manhatan distance, p = 2 Euclidean distance) we can combine the objects in the distance matrix. In the article was assumed p = 2. Note that the distance matrix as well as weight matrix is diagonal zero. In distance matrix instead of zeros and ones specifying the neighbours there are set distance measures for implementation of specific variables. It is therefore necessary to transform distance into the elements of matrix weight. The simplest formula would be written as:

$$w_{ij} = d_{ij} \tag{4}$$

It should be considered that with increasing distance the effect of a neighbor decreases, so the correct formula will be:

$$w_{ij} = \frac{1}{d_{ij}} \tag{5}$$

Other possible transform functions:

$$w_{ij} = e^{-\alpha d_{ij}} \tag{6}$$

$$w_{ij} = \left(d_{ij}\right)^{-\alpha} \tag{7}$$

where α is any positive.

The matrix weights were based on the distance matrix and can be used to analyze the interplay between the two neighbours. However in order to use it in the analysis it is subjected to rows or columns normalization. Due to the fact that the weight matrix should be a symmetric matrix we need a transformation based on sample moments approximation in order to correct estimation [Anselin 2001] according to the formula:

$$\boldsymbol{W} = \frac{(\boldsymbol{W}+\boldsymbol{W}^T)}{2} \tag{8}$$

The prepared matrix **W** can be used for further spatial analysis [Pietrzykowski 2011].

The matrix of weights - to include the time

In the example considered, we observe price per hectare of agricultural land in the period from 1999 to 2012. Observations are on various provinces. Standard weights matrix consists of zeros and ones. When we take into consideration time in the analysis, we typically have been using panel models. The work proposed the effect of time through proper preparation of weights matrix **W**. On the basis of the available data distance matrix was created according to the formula 3. In this study variable prices of agricultural land were observed each year. Distance matrix contained information about provinces (objects) and variables which were the prices of agricultural land in period 2004-2012. Distance matrix defined by:

$$T = \begin{vmatrix} t_{11} & t_{12} & \cdots & t_{1k} \\ t_{21} & t_{22} & \cdots & t_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ t_{i1} & t_{i2} & \cdots & t_{ik} \end{vmatrix} \tag{9}$$

where t_{ij} are distances for the characteristics including the effect of time at the k-th and i-th object.

So the resulting distance matrix **T** was transformed according to equations 5 and 8 giving weight matrix **W**, which was used in further analysis. On the basis of such a weight matrix **W** produced and using the Moran scatter plot [Anselin 1996, Bivand i inn. 2013] provinces were classified. Moran's scatter plot was used to present spatial relationships autocorrelation. Autocorrelation was calculated according to the formula:

$$R_{(M)} = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\boldsymbol{z}'\boldsymbol{W}\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}} \tag{10}$$

where: **W** – matrix weight, **z** – vector of elements $z_i = x_i - \mu$,

For comparison the classical method of classification using a flexible clustering technique SAHN was applied according to the formula:

$$d_{(ij);k} = \propto_1 d_{i,k} + \propto_2 d_{j,k} + \beta d_{ij} + \vartheta(d_{i,k} - d_{j,k}) \tag{11}$$

which assumed:

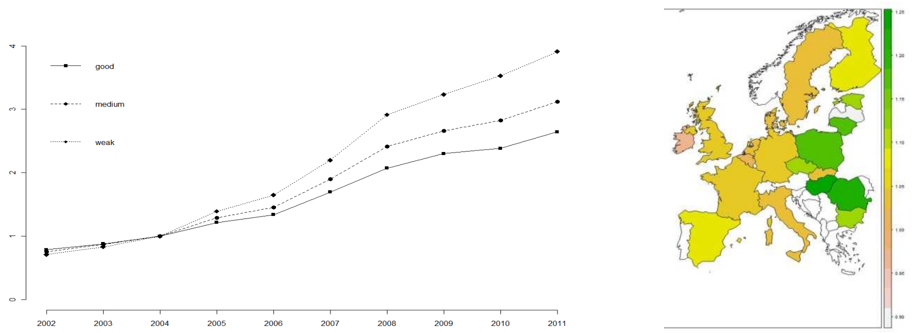
$$\propto_1 + \propto_2 + \beta = 1;\ \propto_1 = \propto_2;\ \beta \le 1;\ \vartheta = 0$$

This resulted in the division of objects similar to the techniques of furthest neighbor ($where\ \beta = 1$).

RESULTS

Changes in the prices of agricultural land are particularly noticeable in Poland after accession to the EU in 2004. Prior to 2004 agricultural land prices remained at a low level. Price increase from year to year was relatively small, and the largest changes were observed for a good price, then the medium and poor. Since 2004, this situation has been changing. The largest price changes can be observed for the poor land prices (Figure 1) then these changes relate to the medium land prices and good quality. Regarding the situation in Europe, it can be seen that the average variation in the period regard Poland which due to these changes is at the forefront of the countries that joined the EU (Figure 1). A more detailed description of the problem can be found in Pietrzykowski [Pietrzykowski 2012]. In later work hierarchical cluster analysis method was used. The provinces (NUT2) were divided using agricultural land prices in the period from 2004 to 2012 according to the formula 11. Figure 2 shows the distribution obtained by using Euclidean distances and SAHN techniques. In addition dendrogram also contains a visualization of the spatial distribution of prices land on the map of Poland. Analyses were carried out for the land of good, medium and poor. The results have been presented only for price of agricultural land of poor quality due to their specific use [Pietrzykowski 2012].

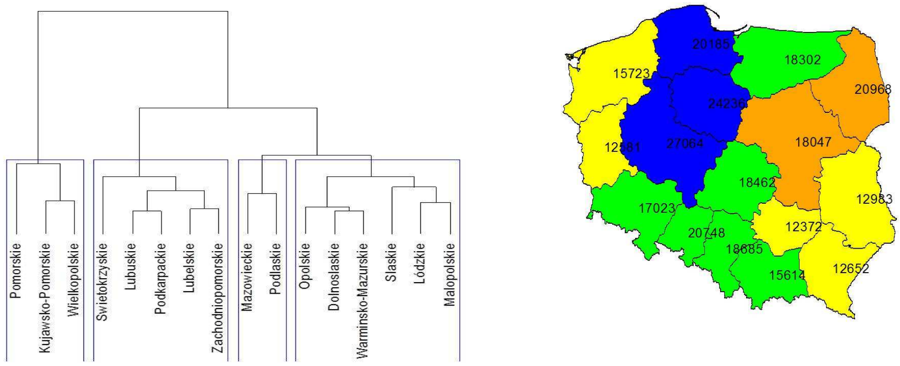
Figure 1. Dynamics of changes in agricultural land prices in the years 2002 – 2011 and the average change in prices in Europe in the period from 2005 – 2011



Source: own calculations [Pietrzykowski 2012]

Figure 2 shows the division of the four clusters. This proposal follows the methodology of work. The Moran scatter plot is a graphic representation that enables a description of the schema of spatial relation. This scatter plot has the standardized value of the variable on the horizontal axis and standardized lagged value on the vertical axis. The Moran scatter plot has four quadrants, where each quadrant corresponds to a specific spatial affiliation that can exist between a region and its neighbours [Anselin 1996, Bivand i inn. 2013]. Because as mentioned earlier it was planned to compare classification of the resulting methods of cluster and spatial analysis divided into four groups.

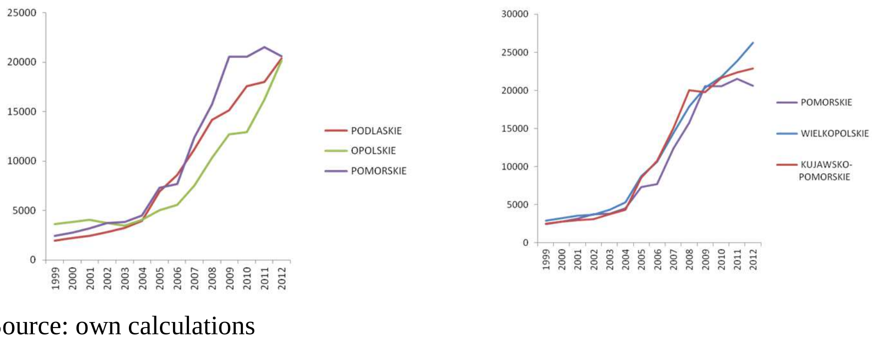
Figure 2. Dendrogram for the provinces due to prices of agricultural land of poor quality in the period from 2004 to 2012 and visualization of four clusters on the spatial map of Poland



Source: own calculations

Map of Poland gives the average prices of agricultural land for the poor quality of the region during the period of time. As you can see the division into four clusters is quite coherent. The first group includes three provinces: Pomorskie, Kujawsko-Pomorskie and Wielkopolskie. In these provinces the average price of land is the highest. One may wonder why Pomorskie (20185 PLN) was also included in this cluster because the average price of land is similar to that in other clusters (Podlaskie 20748 PLN, Opolskie 20968 PLN). Figure 3 is an explanation of this situation. The provinces which are in first cluster had similar prices of land in the time period 1999 - 2012 (Figure 3 right). The prices of agricultural land in provinces Opolskie and Podlaskie differ from price of land in province Pomorskie (see Figure 3, left) in the years 1999 - 2012.

Figure 3. Prices of agricultural land of poor quality for selected regions in the years 1999 – 2012



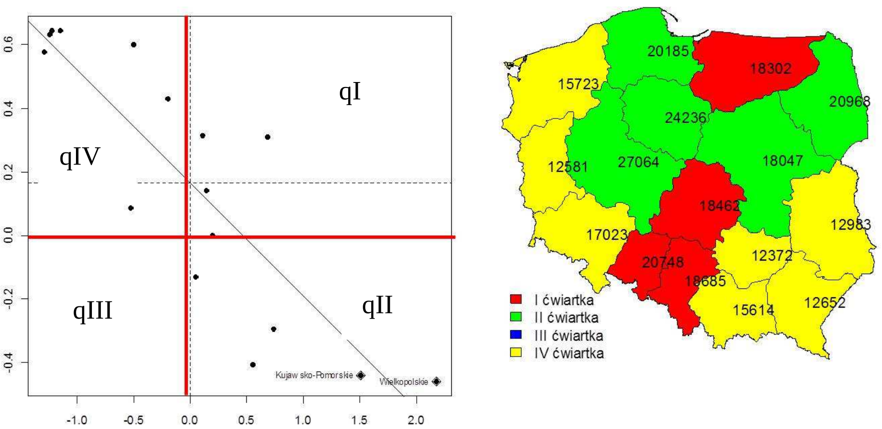
Source: own calculations

The second cluster includes: Świętokrzyskie, Lubuskie, Zachodniopomorskie, Podkarpackie and Lubelskie. The third cluster are two provinces: Mazowieckie and

Podlaskie. The fourth cluster includes six provinces: Opolskie, Dolnośląskie, Warmińsko-Mazurskie, Śląskie, Łódzkie and Małopolskie. Cluster analysis does not account for possible interactions between neighbours, which are the province here.

In the following part of the work on the basis of the obtained distance matrix weight matrix **W** was made. The Moran scatter plot plotted according to the resulting the matrix **W** (Figure 4). The points of Moran scatter plot have been divided into four groups based on four quadrants of this graph. The division due to zero values was made (thick lines in Figure 4, left). As a result of the analysis in the first quarter, four provinces are obtained: Warmińsko-Mazurskie, Łódzkie, Śląskie and Opolskie. In the second quarter: Podlaskie, Mazowieckie, Kujawsko-Pomorskie, Wielkopolskie and Pomorskie. In the third quarter there are no provinces and in the fourth there are seven provinces: Zachodniopomorskie, Lubuskie, Dolnośląskie, Małopolskie, Świętokrzyskie, Podkarpackie and Lubelskie. The Moran scatter plot noted Kujawsko-Pomorskie and Wielkopolskie as points or outliers in the provinces of poor agricultural where land prices obtained the highest value.

Figure 4. The Moran scatter plot for weight matrix based on the distance matrix and their spatial visualization on map of Poland



Source: own calculations

Regarding Moran coefficient it was found to be significant (p-value: 1,181 E-12) was negative -0.3561. It can be concluded that the price of agricultural land of poor quality are not randomly distributed among regions. The matrix weight used instruments that allow the indirect influence of the state on the price of agricultural land. This is due to its construction which takes into account changes variable over time. You can also see that the distribution using the Moran scatter plot allows the identification of regions where there are low prices of agricultural land (fourth quarter) compared to the other provinces (figure 4, right).

Besides it seems that such large provinces like Mazowieckie and Wielkopolskie should interact with each other exactly the price of the land. The division obtained by the spatial analysis confirms this assumption. Table 1 shows a comparison of divisions using the two methods (differences in bold).

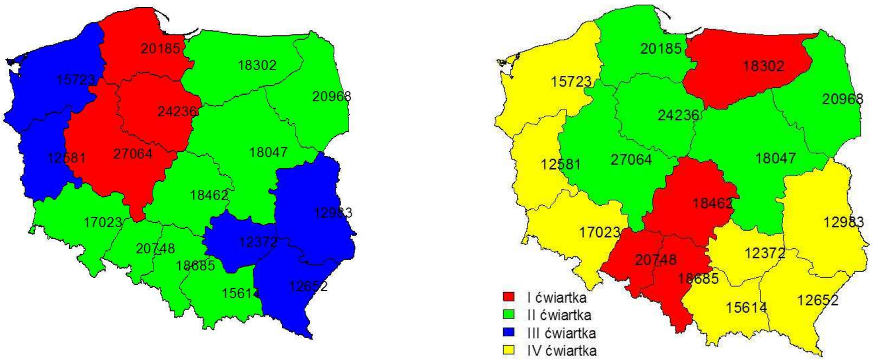
Table 1. Comparison of division the use of cluster analysis and spatial analysis

Cluster (quarter)	Cluster analysis (Figure 2)	Spatial analysis (Figure 4)
First (I)	warmińsko-mazurskie łódzkie, śląskie, opolskie, **dolnośląskie, małopolskie**	warmińsko-mazurskie, łódzkie, śląskie, opolskie
Second (II)	podlaskie, mazowieckie	podlaskie, mazowieckie, kujawsko-pomorskie, wielkopolskie, pomorskie
Third (III)	**kujawsko-pomorskie, wielkopolskie, pomorskie**	
Fourth (IV)	zachodniopomorskie, lubuskie, świętokrzyskie, podkarpackie, lubelskie	zachodniopomorskie, lubuskie, **dolnośląskie**, **małopolskie**, świętokrzyskie, podkarpackie, lubelskie

Source: own calculations

Although at the beginning of the substantive reasons division into four clusters was established, the result of the spatial analysis were obtained for three groups of provinces. Therefore, further analysis examined how it appeared to split into three groups using cluster analysis. Figure 5 compares the resulting division by two methods.

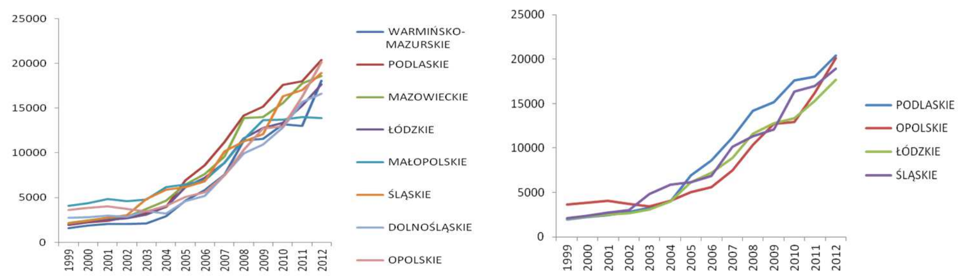
Figure 5. The division by the use of cluster analysis (left) and spatial analysis (right) on the map of Poland.



Source: own calculations

In the case of division into three groups using hierarchical methods we obtain the division into eight provinces (Warmińsko-Mazurskie, Podlaskie, Mazowieckie, Łódzkie, Dolnośląskie, Śląskie, Opolskie, Małopolskie), contrary to the spatial analysis. In this group there are provinces that clearly differ because of the price of land. A thorough analysis of agricultural land prices leads to the conclusion. Comparing the price of land in these provinces, note the differences for provinces Warmińsko-Mazurskie since 2011 and Małopolskie since 2009. It seems that more accurate grouping is achieved by spatial analysis. For example, let us consider the change in the price of land in the years 1999-2012 (Figure 6 right) which is located in the first quadrant (Figure 4).

Figure 6. Prices of agricultural land of poor quality for selected provinces in the years 1999 – 2012.



Source: own calculations

Because of this it seems that the division into clusters, which uses information about reciprocal interactions and changes in prices over time would allow for a better assessment of the market for agricultural land prices.

SUMMARY

The paper proposes a modification of the weight matrix **W** in order to obtain information on changes in the studied phenomenon in time. The use of the weight matrix takes into account the construction aspect of time and information about spatial interactions. The use of spatial analysis allowed us to obtain a division to clusters, which in a more complementary manner presented the phenomenon than the classical cluster analysis did. The paper also shows the possibility of the use of Moran scatter plot and spatial analysis to classify objects. In addition, the work was characterized by changes in the prices of agricultural land of poor quality in Poland in the years 1999-2012.

REFERENCES

Anselin, L. (1996), The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association, in M. Fischer, H. Scholten, and D. Unwin, eds., Spatial Analytical Perspectives on GIS, London: Taylor and Francis, s. 111–125

Anselin L. (2001): Spatial econometrics, Oxford, Basil Blackwell

Bivand R. S., Pebesma E., Gómez-Rubio V. (2013): Applied spatial data analysis with R, Springer, London

Conley T. G. (1999): GMM estimation with cross selection dependence, Journal of Econometrics, 92(1), s. 1 – 45

Doreian P. (1980): Linear models with spatial distributed data. Spatial disturbances or spatial effects, Sociological Methods and Research, 9, 29 – 60

Elhorst J.P. (2003): Specification and estimation of spatial panel data models, International Science Review, nr 26(3), s. 224-268

Pietrzykowski R. (2011): Koncepcja i zastosowanie modyfikacji macierzy wag w przestrzennych badaniach ekonomicznych, Metody ilościowe w badaniach ekonomicznych. Vol. 12, nr 2 / red. Bolesław Borkowski s. 270-278

Pietrzykowski R. (2013): Agricultural land market in Poland and the European Union, Economic Science for Rural Development. Proceedings of the International Scientific Conference 2013, nr 32, s. 76-82

Suchecka J. (2014): Statystyka przestrzenna. Metody analiz struktur przestrzennych, Wydawnictwo C.H. Beck Warszawa