ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 285, 2013

*Kamińska Agnieszka**

THE APPLICATION OF THE TOOLS OF SPATIAL STATISTICS TO EVALUATION REGIONAL DIFFERENTIATION OF POLISH AGRICULTURE

Abstract. The article presents the application of the spatial autocorrelation analysis in evaluation of regional differentiation of agriculture in Poland. The study based on the selected data for the sixteen provinces from the year 2010. In order to estimate the level of agriculture WAP methods were applied. On the basis of the synthetic measure, developed during the study, a ranking of regions was constructed. Additionally, the analyses were broadened by the use of spatial autocorrelation statistics which enabled to consider the existing spatial relations.

Key words: spatial autocorrelation, Moran's global statistic, Moran's local statistic, regional variability

I. INTRODUCTION

Membership in structures of European Union and use of union funds have capability on the level of Polish agriculture. Analyses of levels of agriculture development can be important instrument in creation of effective regional rural policy. These operations aim at liquidation of regional differences.

Situation of polish agriculture, especially regional disparity was analyzed by many researchers [Borkowski and Szczęsny (2002), Muszyńska (2009), Zegar (2003)].

In order to estimate regional differentiation methods of multivariate statistical analysis usually were applied [Młodak (2006)].

In the paper, on the basis of the synthetic variable, developed during the study, the regions were classified and grouped into clusters, according to its level of development. Additionally, the analyses were broadened by the use of spatial autocorrelation statistics which enabled to consider the existing spatial relations.

* Ph.D., Departament of Applied Mathematics and Computer Science, University of Life Sciences in Lublin.

II. MATERIAL AND METHODS

The study used selected data of the Central Statistical Office (Regional Data Bank) for the sixteen provinces from the year 2010. Both essential and statistic reasons decided about diagnostic features selection. Moreover mutually strong correlated features were eliminated to dispose of duplicate information. Coefficients of variation were also included in statistical analysis- quasi-constant variables were rejected.

Finally chosen diagnostic variables were:
X1- own revenue of voivodships budgets in PLN per capita;
X2- lifestock of cattle in heads per 100 ha of agricultural land;
X3- lifestock of pigs in heads per 100 ha of agricultural land;
X4- harvests of basic cereals per 1 ha of agricultural land [t];
X5- procurement of potatoes in t per 1 ha of agricultural land;
X6- procurement of sugar beets in t per 1 ha of agricultural land;
X7- procurement value of vegetables in t per 1 ha of agricultural land;
X8- procurement of fruits in t per 1 ha of agricultural land;
X9- unemployment rate in rural areas;
X10- proportion of agricultural lands in total area;
X11- average farm area in ha.

All of them were considered stimulants, except for X9 (destimulant).

In order to normalize the features the standarization was used. In order to estimate the level of agriculture synthetic measure was used, based on Hellwig method [(Hellwig(1968)]. He proposed a concept of the taxonomic measure of development understood as the arrangement of units investigated depending on their distance from the standard establishing the development pattern. In order to determine the degree of similarity between object and the standard point. The development measure is determined as follows:

$$z_i = 1 - \frac{d_i}{d_0}, \tag{1}$$

where: d_i- euclidean distance between the object i and the standard point,

$$d_0 = \bar{d} + 2s_d, \ \bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i, \ s_d = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i - \bar{d})^2}, \ 0 \le z_i \le 1. \tag{2}$$

The more developed the object, the higher the value of this indicator.

On the basis of synthetic measure, mean and standard deviation administrative units were divided into four typological classes representing different level of the research issue:

– I class: $z_i \geq \bar{z} + s_z$

– II class: $\bar{z} \leq z_i < \bar{z} + s_z$

– III class: $\bar{z} - s_z \leq z_i < \bar{z}$

– IV class: $z_i < \bar{z} - s_z$

where: $\bar{z}$ - mean, s_z - standard deviation

Spatial relationships were evaluated on the basis of global and local Moran's *I* coefficients [Anselin(1995)].

For the synthetic indicator *z* global Moran's *I* coefficient was calculated according to formula:

$$I = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2} \tag{3}$$

where:

n - number of observations

z_i - value of variable for *i*-th location

z_j - value of variable for *j*-th location

$\bar{z}$ - average value of variable

w_{ij} – weight between locations *i* and *j* based on contiguity where the definition of neighbor was based on sharing a common boundary. Contiguity spatial relationships were assigned a value of 1 to neighboring locations and 0 to all other ones. Spatial weights are standardized by row. Each weight is divided by its row sum.

Significant value of *I* greater than 0 proves positive autocorrelation. It means that objects have similar values of variables (high values of the variable in location *i* tend to be clustered with high values of the same variable in locations that are neighbors of *i*, and vice versa). Value of *I* less than 0 proves negative autocorrelation. Negative autocorrelation means, that exist big differences between values for neighboring objects (high values in a variable in location *i* tend to be co-located with lower values in the neighboring locations).Value of *I* equal to 0 or similar to 0 means random spatial distribution.

Local Moran's I was calculated for each observation unit according to formula:

$$I_i = \frac{(z_i - \bar{z})\sum_{j=1}^{n} w_{ij}(z_j - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2 / n} \tag{4}$$

For each location, values of I_i allow for the computation of its similarity with its neighbours and also to test its significance. Five scenarios may emerge [Janc (2006)]:

- Locations with high values with similar neighbors (known as "hot spots")
- Locations with low values with similar neighbors (known as "cold spots")
- Locations with high values with low-value neighbors (potential outlier)
- Locations with low values with high-value neighbors (potential outlier)
- Locations with no significant local autocorrelation.

Significance test for global and local Moran's I statistics were presented in Cliff and Ord (1981). Testing of significance of autocorrelation was based on empirical value of the Z statistic which follows normal distribution: $Z_I = \frac{I - E(I)}{\sqrt{Var(I)}} \sim N(0,1)$.

These specific configurations can be identified from a Moran scatterplot. This graph depicts a standardized variable in the x-axis versus the spatial lag of that standardized variable. The spatial lag is a summary of the effects of the neighboring spatial units. In essence, Moran scatterplot presents the relation of the variable in the location i with respect the values of that variable in the neighboring locations. By construction the slope of the line in the scatter plot is equivalent to the Moran's I coefficient. The four quadrants in the scatterplot box thus represent different types of association between the values at a given location and its spatial lag. he upper right and lower left quadrants represent positive spatial association, in the sense that a location is surrounded by similar valued locations. For the upper right this is association between high values , while for the lower left quadrant this is association between low values. The upper left and lower right quadrants correspond to negative association, low values are surrounded by high values (upper left) and high values are surrounded by low values (lower right). The relative densities of these quadrants indicate which of these patterns of negative spatial association (in the traditional sense) dominate [Anselin (1993)].

III. RESULTS

Table 1 contains the values of the synthetic measure *z* representing the level of agriculture development.

Table 1. Values of synthetic measure in Poland voivodships in the year 2010

Voivodships	z_i	Class
Wielkopolskie	0,525	I
Kujawsko-Pomorskie	0,372	
Opolskie	0,345	II
Łódzkie	0,324	
Pomorskie	0,312	
Warmińsko-Mazurskie	0,277	
Mazowieckie	0,268	
Dolnośląskie	0,261	
Podlaskie	0,252	
Śląskie	0,226	III
Małopolskie	0,175	
Zachodniopomorskie	0,172	
Świętokrzyskie	0,168	
Lubelskie	0,124	
Lubuskie	0,057	IV
Podkarpackie	0,017	
mean	0,242	

Source: own work.

There was a clastering trend in Poland's provincial level development (represented by indicator *z*). Voivodships with the high values of synthetic measure *z* were along the line from Pomorskie to Opolskie. Low values of *z* was observed in the north-western and south-eastern corners of Poland [Figure 1].

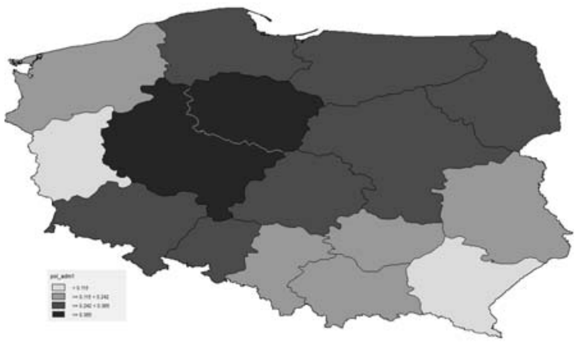

Figure 1. Values of indicator z for voivodships
Source: own work.

The value of global Moran's I was equal to 0,26 and proved significant positive spatial autocorrelation (p-value=0,023). Table 2 contains the values of local Moran's I_i. Significant positive values of I_i were obtained by Kujawsko-Pomorskie, Łódzkie, Opolskie and Podkarpackie. It means that locations were surrounded by similar neighbors.

Table 2. Values of local Moran's I in voivodships in the year 2010

Voivodship	I_i	p-value
Wielkopolskie	0,341	0,103
Kujawsko-Pomorskie	0,812	0,005
Opolskie	0,577	0,044
Łódzkie	0,379	0,018
Pomorskie	0,414	0,038
Mazowieckie	0,019	0,360
Podlaskie	–0,010	0,416
Warmińsko-Mazurskie	0,144	0,156
Dolnośląskie	0,078	0,16
Śląskie	–0,006	0,433
Zachodniopomorskie	–0,243	0,208
Małopolskie	0,452	0,06
Lubelskie	0,487	0,094
Świętokrzyskie	0,234	0,085
Lubuskie	–0,93	0,12
Podkarpackie	1,24	0,037

Source: own work.

The provinces with significant values for I_i (using significance level of 0.05) are depicted in Figure 2. The light gray shade on the map indicates a spatial claster of voivodships with high values of z and the dark gray shade corresponds with claster of low z values.

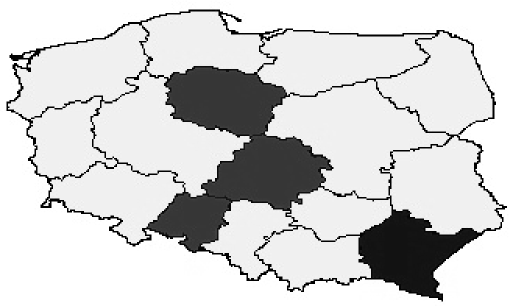

Figure 2. Map illustrating significance of Local Moran's I_i

Source: own work.

Figure 3 is the Moran scatterplot for this data, with a linear smoother superimposed. Almost all of the associations fall in the lower left and upper right quadrants indicating presence of clasters. All points in the upper right quadrant (high-high association) corresponding to claster of high values and its similar neighbors (Kujawsko-Pomorskie, Opolskie, Łódzkie, Pomorskie, Warmińsko-Mazurskie, Wielkopolskie, Mazowieckie i Dolnośląskie). Points in the lower left quadrant (low-low association) corresponding to Podkarpackie voivodship and all its neighbors (Lubelskie, Świętokrzyskie i Małopolskie).

While the overall tendency portrayed in the scatterplot is one of positive association, one voivodship show the opposite: low value surrounded by high values for Lubuskie. Lubuskie was the potential outlier, but the value of I_i for that region wasn't significant.

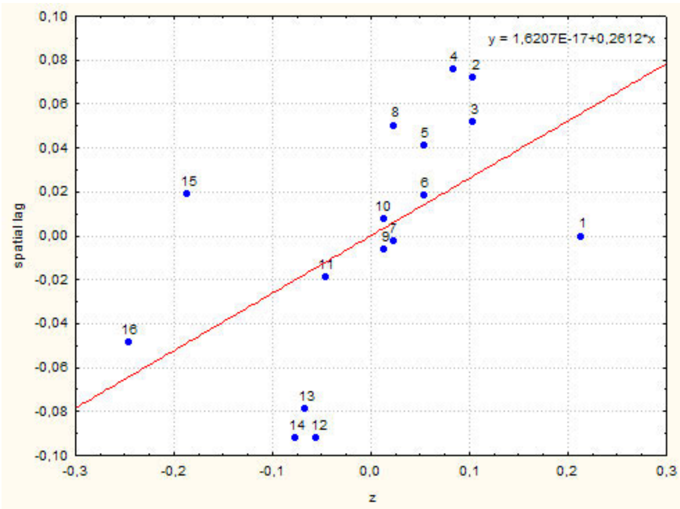

Figure 3. Moran scatterplot for indicator z

Source: own work.

IV. CONCLUSIONS

The performed analyses showed regional differentiation of agriculture in Poland.

Voivodships with the high level of development were along the line from Pomorskie to Opolskie. Low level of agriculture was observed in the north-western and south-eastern corners of Poland. The best agricultural standing was found for Wielkopolskie voivodship and relatively the worst for Podkarpackie and Lubuskie. That results are partly consistent with those reported in previous studies.

Additionally, the comparative analysis was broadened by the use of the tools of spatial autocorrelation statistics which enabled to consider the existing spatial relations.

Presence of positive spatial autocorrelation for agriculture was proved. It means clastering trend in Poland's provincial level development. Claster of high level of agriculture was formed by Pomorskie, Opolskie, Łódzkie and Warmińsko-Mazurkie.

The claster of low values was formed by Podkarpackie.

Generalizing, spatial autocorrelation statistics are very useful tool in multivariate regional analyses, in particular Polish agriculture. They inform about the kind and the strength of spatial dependence, make possible a determination of associations among objects and establishing spatial structure better than using the traditional methods**.**

REFERENCES

Anselin L. (1993), The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association, Research Paper 9330.

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

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

Borkowski B., Szczęsny W. (2002), Metody taksonomiczne w badaniach przestrzennego zróżnicowania rolnictwa. Roczniki Nauk Rolniczych, seria G., t. 89, z.2, Warszawa.

Hellwig Z. (1968), Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. Przegląd statystyczny 15, 307–327.

Janc K. (2006), Zjawisko autokorelacji przestrzennej na przykładzie statystyki I Morana oraz lokalnych wskaźników, zależności przestrzennej. Idee i praktyczny uniwersalizm geografii, nr 33, IGiPZ PAN, Warszawa.

Młodak A. (2006), Analiza taksonomiczna w statystyce regionalnej. Difin, Warszawa

Muszyńska A. (2009), Regionalne zróżnicowanie rolnictwa w Polsce w 2007 roku. Roczniki Naukowe Seria, t.XI, z.4. Warszawa.

Zegar J. (red). (2003), Zróżnicowanie regionalne rolnictwa. GUS, Warszawa.

Bank Danych Regionalnych: Główny Urząd Statystyczny.

Kamińska Agnieszka

ZASTOSOWANIE METOD WAP DO OCENY POZIOMU PRZESTRZENNEGO ZRÓŻNICOWANIA ROZWOJU ROLNICTWA W POLSCE

Przedmiotem badań była analiza regionalnego zróżnicowania rolnictwa w Polsce. Ocenę poziomu rolnictwa i jego dyspersji przestrzennej opracowano na podstawie danych statystycznych gromadzonych przez Główny Urząd Statystyczny, wykorzystując narzędzia Wielowymiarowej Analizy Porównawczej (WAP). Na podstawie skonstruowanej zmiennej syntetycznej utworzono ranking województw. Dodatkowo wykorzystano narzędzia statystyki przestrzennej w celu identyfikacji przestrzennych zależności.