

## The use of random fields in the Modifiable Areal Unit Problem

Michał Bernard Pietrzak<sup>1</sup>, Bartosz Ziemkiewicz<sup>2</sup>

### Abstract

The focus of the research will be on the modifiable areal unit problem (MAUP) within which two aspects will be considered: the scale problem and the aggregation problem. In the article we consider the use of random fields theory for the needs of the “Scale Problem” issue. The Scale Problem is defined as a volatility of the results of analysis as a result of a change in the aggregation scale. In the case of the scale problem empirical studies should be conducted with application of simulations. Within the simulation analysis the realisations of random fields referred to irregular regions will be generated. First, the internal structure of spatial processes will be analysed. Next, we consider the theoretical foundations for random fields relative to irregular regions. The accepted properties of random fields will be based on the characteristics established for economic phenomena. The outcome of the task will be the development of a procedure for generating the vector of random fields with specified properties. Procedure for generating random fields will be used to simulations within the scale problem too. The research is funded by National Science Centre, Poland under the research project no. 2015/17/B/HS4/01004.

**Keywords:** *spatial econometrics, Scale Problem, random fields, Modifiable Areal Unit Problem, simulations*

**JEL Classification:** C10, C15, C21

### 1 Introduction

The article deals with the problem of Modifiable Areal Unit Problem (MAUP) which is found to be significant in the area of economic spatial analysis (see: Arbia, 1989). The MAUP issue concerns the possibility of obtaining different results due to changes in the level of aggregation (see: Pietrzak, 2014). The main purpose of this work is to consider the Scale Problem, which is one of the aspects of the issue of MAUP. Analysis of the Scale Problem will be conducted based on the example of the assessment of socio-economic development of which one aspect can be expressed by means of number of entities of the national economy per capita. Examining the number of entities of the national economy makes only one aspect of the complex phenomenon of socio-economic development. Various aspects of socio-economic development both at the regional and national levels have been examined in many works (see: Ciburiene, 2016; Łyszczarz, 2016; Balcerzak, 2016a, 2016b; Jantón-Drozdowska

---

<sup>1</sup> Corresponding author: Nicolaus Copernicus University, Department of Econometrics and Statistics, ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: pietrzak@umk.pl.

<sup>2</sup> Nicolaus Copernicus University, Department of Probability Theory and Stochastic Analysis, ul. Chopina 12/18, 87-100 Toruń, Poland, e-mail: bartek@mat.umk.pl.

and Majewska, 2016; Małkowska and Głuszak, 2016; Pietrzak and Balcerzak, 2016; Żelazny and Pietrucha, 2017).

Research conducted in this work enabled us to consider the scale problem based on the example of spatial development of the number of business entities. Empirical analysis of the properties of this phenomenon allowed the spatial trend in its internal structure to be identified. Then, a simulation analysis was performed with the assumption of identified empirical properties. It turned out that the simulation analysis performed at various levels of aggregation led to the obtainment of similar parameter estimates of the spatial trend and a different correlation structure for simulated processes. This allowed the evaluation of selected elements of the internal structure of spatial processes.

## **2 Analysis of the internal structure of selected spatial processes**

In the performed analysis emphasis was laid on distinguishing processes expressed in the absolute quantities from those expressed in relative quantities. This results from the fact that spatial studies ought to be predominantly based on the analysis of processes expressed in relative quantities referred to certain values characterizing the selected region (area, number of residents). This ensures the comparability of data and the correctness of the results obtained. In case of spatial economic analysis of business entities, final conclusions should be based on a process expressed in relative quantities, *i.e.*, on the number of business entities referred to the number of residents. This process combines two processes expressed in the absolute quantities, the number of business entities and the number of residents. The processes adopted in the study are:  $X_1$  - population in 2016,  $X_2$  - number of entities of the national economy in 2016 and  $Y$  - number of entities of the national economy per capita in 2016. In addition, it is assumed that data used in the context of spatial economic analyses may be treated as a realization of a two-dimensional random field  $X(u_1, u_2)$ , where  $u_1, u_2$  denote the coordinates on the plane (see: Arbia, 1989). The two-dimensional random field defined in such a way will be further referred to in the work as a spatial process.

The scale problem should be examined only for a composition of territorial units forming Quasi Composition of Regions. Therefore, the next step in the study of the Scale Problem should consist in determining Quasi Composition of Regions for the examined number of business entities in Poland. The concept of Quasi Composition of Regions was introduced by Pietrzak (2014a). This concept refers to a composition of regions which is formed of individual compositions of territorial units for subsequent levels of aggregation. Individual compositions of territorial units need to be selected in such a way that research conducted

based on them should provide valid conclusions. In this study, Quasi Composition of Regions will be limited to two individual regions from NUTS 4 and NUTS 3 territorial units.

Since economic analysis will be performed based on selected spatial processes, it is very important to examine their internal structure (see: Pietrzak 2014). Therefore, after determining Quasi Composition of Regions, another step consisted in studying the internal structure of selected spatial processes  $X_1$ ,  $X_2$ ,  $Y$ . Studying the internal structure of spatial processes means providing a correct description of their properties. The following components of the internal structure can be distinguished: the component related to unsystematic heterogeneity, the component related to systematic heterogeneity, and the component of the structure where the spatial process is homogeneous (see: Pietrzak 2014a).

The analysis focused on the study of systematic heterogeneity in the form of parameters estimation of the model of spatial trend<sup>3</sup>. For each process  $X_1$ ,  $X_2$ ,  $Y$  we estimated a linear spatial trend model determined by the following equation

$$\mathbf{Y} = \alpha_0 + \alpha_1 \mathbf{U}_1 + \alpha_2 \mathbf{U}_2 + \boldsymbol{\varepsilon} \quad (1)$$

where  $\mathbf{Y}$  is the vector of spatial process,  $\mathbf{U}_1$ ,  $\mathbf{U}_2$  are vectors of geographic coordinates,  $\boldsymbol{\varepsilon}$  is the vector of spatial noise,  $\alpha_0$ ,  $\alpha_1$ ,  $\alpha_2$  are parameters.

It is expected that a spatial linear trend in the process of the number of business entities per capita will occur and the phenomenon is expected to become more intense in the east-west directions. The cause of the occurrence of a linear spatial trend specified in such a way results from a higher level of socio-economic development of western parts of Poland (see: Pietrzak et al., 2013; Hadaś-Dyduch, 2016; Kondratiuk-Nierodzińska; 2016; Czaplak, 2016; Murawska, 2016). Therefore, for the selected processes we estimated the spatial model parameters of a linear trend following formula 1. The results obtained are shown in Table 1. The identification of the spatial line trend was made only for the  $Y$  process - the number of entities of the national economy per capita. The parameter with the variable referring to longitude coordinates proved to be statistically significant. The negative parameter estimate indicates a spatial increase in the number business entities per capita in Poland, as long as we move to the west.

After the estimation of linear spatial trend models, we estimated the value of Pearson's correlation coefficients between the processes, where for the  $Y$  process a further spatial line

---

<sup>3</sup> We deliberately omitted here the internal structure in the form of spatial autocorrelation (spatial homogeneity) to focus in the article solely on the analysis of heterogeneity in the form of a systematic spatial trend. The problems related to spatial autocorrelation were raised in the works of Pietrzak et al., (2014), Pietrzak (2016).

trend was identified. The performed analysis of the results led to the conclusion that there are strong positive correlations among all processes. At the NUTS 4 aggregation level the highest level of correlation dependence occurs between the processes X1 and X2 and reaches level of 0.971. For the pairs of the processes (Y, X1) and (Y, X2) the dependence reaches similar levels of 0.481 and 0.421 respectively. The correlation coefficients change during the transition to a higher level of aggregation. At the NUTS 3 aggregation level the correlative dependence between the processes of X1 and X2 decreased to the level of 0.889. In turn, the correlative dependence between the processes of Y and X2 remained practically unchanged and reaches level of 0.433, and between the processes Y and X1 it rose to 0.741.

Process (Aggregation level)							
<b>Process Y - Aggregation level NUTS 4</b>				<b>Process Y - Aggregation level NUTS 3</b>			
Parameter	Estimate	p-value	R <sup>2</sup>	Parameter	Estimate	p-value	R <sup>2</sup>
$\alpha_1$	-0.661	~0.000	0.186	$\alpha_1$	-0.748	0.001	0.245
$\alpha_2$	0.143	0.072	-	$\alpha_2$	0.187	0.318	-
<b>Process X1 - Aggregation level NUTS 4</b>				<b>Process X1 - Aggregation level NUTS 3</b>			
Parameter	Estimate	p-value	R <sup>2</sup>	Parameter	Estimate	p-value	R <sup>2</sup>
$\alpha_1$	-9.113	0.8116	0.002	$\alpha_1$	112.248	0.473	0.031
$\alpha_2$	-64.551	0.103	-	$\alpha_2$	-166.083	0.308	-
<b>Process X2 - Aggregation level NUTS 4</b>				<b>Process X2 - Aggregation level NUTS 3</b>			
Parameter	Estimate	p-value	R <sup>2</sup>	Parameter	Estimate	p-value	R <sup>2</sup>
$\alpha_1$	-530.752	0.421	0.007	$\alpha_1$	-1606.984	0.604	0.005
$\alpha_2$	-298.757	0.662	-	$\alpha_2$	109.743	0.972	-

**Table 1.** The results of the estimation of model parameters of a linear spatial trend.

Based on the empirical analysis of the properties of processes we can conclude that the linear trend did not change during the transition to a higher level of aggregation. Also in the case of correlations dependence, there were no significant changes during the transition from the NUTS4 level to the NUTS3 level. The next phase of the study should consist in performing simulation analysis and checking whether similar results will be obtained. The simulated processes should have similar properties to those set for the empirical processes. Simulation analysis should allow us to find out whether between the simulated processes at

the aggregate NUTS 4 level and NUTS 3 level there are any differences in a form of a spatial linear trend as well as in the structure of correlations dependence.

Therefore, for the purposes of simulation analysis we assumed for spatial processes an adequate correlation structure, for the pairs of the processes  $(Y, X1) - 0.39$ ,  $(Y, X2) - 0.71$   $(X1, X2) - 0.91$ . We also assumed adequate parameters of the spatial trend in the case of process Y, where  $\alpha_0=10$ ,  $\alpha_1=-0.4$ ,  $\alpha_2=0.1$ . In addition, it was assumed that between the processes there is a relation defined by the following equation

$$Y = \frac{X2}{X1}. \quad (2)$$

The relationship defined by the formula (2) causes a problem in the simulation of the random fields vector with given properties (see Arbia, 1989), hence the assumptions in correlation structure and parameters of the spatial trend are different from the properties of empirical spatial processes. Accordingly, the simulation of the random field vector was performed in the following steps. In the first step, a simulation was performed of two random fields Y and X1 with the assumed correlation dependence at the level of 0.39. Next the spatial trend values were added to the received results of the process Y. Then, based on the spatial processes Y and X1, the values of the process X2 were determined according to the formula 2. In the last step a correction of the correlation structure was made.

As a result of the simulation, we obtained simulated values of the random fields vector, where every individual component is a spatial process related to the composition of NUTS 4 spatial units. Thus we obtained a starting set of spatial processes at the aggregation NUTS 4 level with specified properties. This was followed by aggregation of simulated spatial processes so that the aggregated processes could relate to the composition of NUTS 3 territorial units. Aggregation of spatial processes was carried out by an appropriate summation of the spatial processes X1 and X2. We calculated the sum of the values of the process from respective regions at the NUTS 4 level that make up a selected region at the NUTS 3 level. Then for the aggregated processes X1 and X2 we determined their quotient and thus we received the values of the spatial process Y referred to the NUTS 3 composition. Therefore, aggregation was performed only for the processes X1 and X2. However, the values of the process Y at the NUTS 3 level were obtained based on the formula 2.

As a result of the simulation, we received a thousand simulated values of the random fields vector at the aggregate NUTS 4 level. In turn, the aggregation performed allowed to obtain a thousand of simulated values of the random fields vector at the aggregate NUTS 3 level. The possessed simulated values of variables at the two levels of aggregation, i.e.,

NUTS 4 and NUTS 3, allowed us to estimate model parameters of the linear spatial trend, and then to determine the value of the correlation dependence between the processes devoid of spatial trend. In this way two resultant sets were obtained - a set of estimations of correlation coefficients and a set of parameter estimations of the spatial model. Based on the two sets, the mean values and standard deviations were determined (see: Table 2 and Table 3).

<b>Correlation (aggregation level)</b>			
<b>Correlation between processes (aggregation level NUTS 4)</b>			
Statistics	(Y, X1)	(Y,X2)	(X1,X2)
Mean	0.393	0.713	0.912
Standard deviation	0.002	0.002	0.002
<b>Correlation between processes (aggregation level NUTS 3)</b>			
Statistics	(Y, X1)	(Y,X2)	(X1,X2)
Mean	0.138	0.337	0.972
Standard deviation	0.102	0.094	0.004

**Table 2.** The results of the estimation of the correlation structure on simulated data.

The results obtained allowed us to draw up the following conclusions. For all spatial processes there were no changes in the nature of the linear spatial trend due to the aggregation process. According to the assumption made, in case of the processes X1 and X2, a spatial trend was not present at the NUTS 4 level and after aggregation it also proved to be statistically insignificant at the NUTS 3 level. However, in the case of the process Y, the presence of a spatial trend at the aggregation NUTS 4 level was assumed. As a result of the estimation of the trend model parameters for the process Y at the aggregation NUTS 3 level, we obtained similar evaluation parameters to the aggregation at the NUTS 4 level. Based on the simulation performed, it can be concluded that the results of parameters estimation of a linear model of the spatial trend do not change depending on the choice of the aggregation level. Obviously, a condition must be fulfilled that the adopted compositions of territorial units belong to Quasi Composition of Regions. It should be further noted that the higher the aggregation level, the better fit of the spatial trend model to empirical data (see Table 3).

The simulation analysis conducted also allowed us to conclude that the correlation structure of spatial processes changed as a result of the aggregation process. There was decline in the value of the correlation dependence between the pairs of processes (Y, X1) and (Y, X2) and an increase in the level of the correlation dependence between the pairs of

processes (X1, X1) (see Table 2). In case of the analysis of empirical processes, the spatial correlation structure of these processes did not change. It is possible that this is due to the presence of other properties, including spatial autocorrelation. Therefore, taking into account spatial autocorrelation in simulations should be the subject of further study.

Statistics	$\alpha_0$	p-value	$\alpha_1$	p-value	$\alpha_2$	p-value	R <sup>2</sup>
Process Y (Aggregation level NUTS 4)							
Mean	10.470	~0.000	-0.424	~0.000	0.099	0.129	0.216
Standard deviation	0.370	~0.000	0.039	~0.000	0.047	0.201	0.027
Process Y (Aggregation level NUTS 3)							
Mean	10.528	~0.000	-0.413	~0.000	0.101	0.202	0.501
Standard deviation	0.441	~0.000	0.055	~0.000	0.053	0.263	0.095
Process X1 (Aggregation level NUTS 4)							
Mean	1116.218	~0.000	-25.210	0.071	2.910	0.506	0.017
Standard deviation	75.284	~0.000	9.522	0.154	11.286	0.313	0.013
Process X1 (Aggregation level NUTS 3)							
Mean	4787.751	0.003	-23.327	0.826	21.784	0.303	0.009
Standard deviation	4371.663	0.003	557.875	0.129	6.453	0.131	0.012
Process X2 (Aggregation level NUTS 4)							
Mean	11707.71	~0.000	-643.677	~0.000	116.828	0.402	0.086
Standard deviation	872.7975	~0.000	110.782	~0.000	119.756	0.299	0.025
Process X2 (Aggregation level NUTS 3)							
Mean	53283.251	0.001	-2480.641	0.204	239.923	0.231	0.039
Standard deviation	5040.059	0.001	643.462	0.121	696.878	0.126	0.022

**Table 3.** Estimation of the spatial linear trend model parameters based on simulated data.

## Conclusions

The subject of the article concerned the scale problem whose presence may lead to different results obtained from spatial economic analysis. The scale problem was analysed based on an empirical example of the formation of number of entities of the national economy per capita in Poland. For the needs of this research, we established Quasi Composition of Regions which consisted of two single areas of territorial units, the NUTS 4 and NUTS 3 systems. The performed analysis of the empirical properties of processes at both levels of aggregation allowed the identification of their internal structure. The existence of a spatial trend for the

number of entities of the national economy per capita was established, and after taking into account this fact the correlation structure was determined for selected spatial processes.

Then a simulation analysis was made where the simulated spatial processes displayed similar properties to the empirical processes examined. It turned out that the simulation analysis performed at various levels of aggregation led to the obtainment of similar parameter evaluations of the spatial trend. This means that the scale problem does not significantly influence the nature of the spatial trend when affected by an aggregation process. Simulation analysis also allowed the identification of changes in the correlation structure of the processes resulting from aggregation. The comparison of the results obtained for empirical processes and results gained from the simulation analysis indicate that the lack of changes in the correlation structure for empirical processes may result from the presence of spatial autocorrelation. Therefore, the obtained results indicate the need to broaden the scope of research into the scale problem by the inclusion of the issue of spatial autocorrelation.

### **Acknowledgements**

The research is funded by National Science Centre, Poland under the research project no. 2015/17/B/HS4/01004.

### **References**

- Arbia, G. (1989), *Spatial Data Configuration in Statistical Analysis of Regional Economics and Related Problems*, Kluwer Academic Publisher, Dordrecht.
- Balcerzak, A. P. (2016a). Technological Potential of European Economy. Proposition of Measurement with Application of Multiple Criteria Decision Analysis. *Montenegrin Journal of Economics*, 12(3), 7-17. DOI: 10.14254/1800-5845.2016/12-3/1.
- Balcerzak, A. P. (2016b). Multiple-criteria Evaluation of Quality of Human Capital in the European Union Countries. *Economics & Sociology*, 9(2), 11-27.
- Ciburienė, J. (2016). The Evaluation of Economic Development Index of Poland and Lithuania in the Context of the European Union. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(3), 437-449.
- Czaplak, J. (2016). Analysis of the Effectiveness of Public Services Provision by Local Governments in the Years 2007–2013. *Oeconomia Copernicana*, 7(2), 223-243.
- Hadaś-Dyduch M. (2016). Econometric-wavelet prediction in the aspect of space-time. In: Papież, M., Śmiech, S. (eds.), *Proceedings of the 10th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*,.



- Conference Proceedings*, Cracow: Foundation of the Cracow University of Economics, 45-52.
- Jantoń-Drozdowska, E., & Majewska, M. (2016). Investment Attractiveness of Central and Eastern European Countries in the Light of New Locational Advantages Development. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(1), 97-119.
- Kondratiuk-Nierodzińska M. (2016). New knowledge generation capabilities and economic performance of polish regions. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(3), 451-471. DOI: <http://dx.doi.org/10.12775/EQUIL.2016.021>
- Łyszczarz, B. (2016). Public-private Mix and Performance of Health Care Systems in CEE and CIS Countries. *Oeconomia Copernicana*, 7(2), 169-185.
- Małkowska, A., & Głuszak, M. (2016). Pro-investment Local Policies in the Area of Real Estate Economics - Similarities and Differences in the Strategies Used by Communes. *Oeconomia Copernicana*, 7(2), 269-283.
- Murawska A. (2016). Differences in unemployment among persons in a special situation on the labour market on the example of polish provinces. *Oeconomia Copernicana*, 7(3), 375-388. DOI: <http://dx.doi.org/10.12775/OeC.2016.022>
- Pietrzak M. B, Wilk J., & Siekaniec M. (2013). The Impact of Metropolitan Areas on Internal Migrations in Poland. The Case of Southern Regions. In: Papież, M., Śmiech, S. (eds.), *Proceedings of the 7th Professor AleksanderZelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, Cracow: Foundation of the Cracow University of Economics, 124-132.
- Pietrzak M. B. (2014a). Redefining The Modifiable Areal Unit Problem Within Spatial Econometrics, The Case Of The Scale Problem. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 9(2), 111-132.
- Pietrzak M.B., Wilk J., Kossowski T., & Bivand R. (2014). The Identification of Spatial Dependence in the Analysis of Regional Economic Development - Join-count Test Application. In: Papież, M., Śmiech, S. (eds.), *Proceedings of the 8th Professor AleksanderZelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*. Cracow: Foundation of the Cracow University of Economics, 135-144.
- Pietrzak, M. B. (2016). The Problem of the Inclusion of Spatial Dependence Within the TOPSIS Method. *Montenegrin Journal of Economics*, 12(3), 69-86.
- Pietrzak, M. B., & Balcerzak, A. P. (2016). Assessment of Socio-Economic Sustainability in New European Union Members States in the years 2004-2012. In: Papież, M., Śmiech, S. (eds.), *The 10th Professor Aleksander Zelias International Conference on Modelling and*

*Forecasting of Socio-Economic Phenomena. Conference Proceedings, Cracow:*  
Foundation of the Cracow University of Economics, 120-129.

Żelazny R., & Pietrucha J. (2017). Measuring Innovation and Institution: the Creative Economy Index. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(1).