DYNAMIC ECONOMETRIC MODELS
Vol. 11 – Nicolaus Copernicus University – Toruń – 2011

Elżbieta Szulc
Nicolaus Copernicus University in Toruń

Identification of the Structures of Spatial and Spatio-Temporal Processes and a Problem of Data Aggregation

A b s t r a c t. The paper concerns the measurement of the dependence between economic spatial and also spatio-temporal processes at various levels of data aggregation. The considerations refer to the investigations confirming efficiency of the so-called quasi-congruent spatial model as a tool of the measurement of the dependence between economic processes. The main problem of the paper is to discuss such a description of the spatial and spatio-temporal connections, which would be adequate to express the autodependence of the investigated processes. The application of the economic distance, which characterizes the similarity of the regions on the ground of the values of the analyzed processes, is proposed.

K e y w o r d s: quasi-congruent spatio-temporal model, autodependence, connectivity matrix, economic distance.

Introduction

Spatial models are estimated for the given spatial units, such as: countries, regions, provinces (voivodeships), census tracts and so on. Data, usually spatially correlated, collected for these units create the established configurations which determine quantification of the spatial autodependence. Configurations of the aggregated data simplify the connections of the smaller spatial units and then they may lead to the simplified description of some properties of the economic spatial processes and the dependence between them[1]. This is also the case of the spatio-temporal processes.

In the paper attention is paid to the spatial aspect of the data aggregation. The question of the time aggregation is not taken into account. Since the paper concerns the investigation of the spatial processes in time, some dynamic elements were introduced to the analysis.

[1] See e.g. Arbia (1988) for an early discussion of some of the issues.

The presented considerations refer to the previous investigations carried out on the basis of generated data (Szulc et al., 2011) and empirical data for one year (Szulc, 2011). At present the time range of the analysis has been extended to six years.

In Szulc et al. (2011) there was formulated and verified the thesis that finding out the value of the parameter which measures the actual influence of one spatial process on another spatial process needs to identify the properties and componential structures of the processes and to take them explicitly into account in the spatial regression model.

Formulation of the hypothesis mentioned above means the recommendation of construction of the quasi-congruent spatial model[2] in which, thanks to taking into account the components describing the internal structure of the processes, the filtration of the processes takes place and the estimation of the dependence is more precise, as not disturbed by the so-called indirect influences. Finally, on condition of the precise filtration, the regression coefficients in the quasi-congruent spatial models estimated at the individual level should be the same (identical) as the ones estimated at the aggregated level. The hypothesis and its consequences may be extended to the case of the analysis of the spatio-temporal processes.

In the quoted work there was shown, that the quasi-congruent spatial model was an effective tool of the measurement of the dependence between spatial processes (actual value of the regression coefficient), if the model description of the structures of the processes was correct and sufficient. The differences between the values of the regression coefficient obtained for the quasi-congruent models before and after the aggregation of the data were dependent on the structures of the separate processes. For the processes only with the trend structure, regardless of the degrees of the trend and of the values of their parameters, the regression coefficients estimated at the various levels of data aggregation did not differ from one another significantly and oscillated around the actual value. On the contrary, in terms of the spatial autocorrelation, the value of the regression coefficient after the aggregation was overestimated, when $\rho_y > \rho_x$, and underestimated, when $\rho_y < \rho_x$[3].

The regression coefficients in the spatial quasi-congruent models for the primary data and for the aggregated data with the spatial autocorrelation will differ in value first of all because the connectivity matrix in the aggregated model does not reflect the connections of the data before the aggregation. Thus, the autodependence existing at the individual level cannot be precisely described at the aggregated level. Moreover, the quantification of the spatial

[2] The procedure of constructing such a model was described in the quoted work. It was also explained what the quasi-congruency of the spatial model meant.

[3] ρ_y and ρ_x denote respectively the autoregression coefficients of the explaining and explanatory process.

dependence by means of the standard matrixes of the connections, within the scheme of the structure of the spatial connections existing at the given level of data aggregation as well, may be inadequate. This is the situation, when the strength of the connection between the spatial units depends not only on the physical distance and the fact of being contiguous, but also on e.g. the economic similarity of the units.

Looking for the more precise measurement of the spatial autodependence with the help of other than the standard matrixes of connections, in Szulc (2011) there was investigated the question of how the application of the certain matrix of the economic distance would influence the estimation of the dependence between the spatial processes. In the quoted investigation the difference between the parameters of the models for the data before aggregation and for the aggregated data slightly decreased, but it should not be treated as regularity. All the more it should not be expected that the difference will disappear because the proposed matrix is still based on the common border criterion.

The present investigation is the continuation of the works quoted above. Analogical analyses are carried out for the successive years in the period 2004–2009. This approach allows to verify whether the observed regularities for 2007 (see Szulc, 2011) are the same for other years. Moreover, the spatio-temporal models are considered. In the investigation both the standard connectivity matrixes and the matrixes of the economic distance are used. The economic distance matrixes in the spatio-temporal models have the dynamic features in the sense that for each of the years another connectivity matrix is applied.

1. Quantification of the Spatial Connections and the Measurement of the Spatial Autodependence

The basis of measurement of the spatial (auto)dependence is establishing the connections among spatial units. The spatial neighbours can be defined in a number of ways.

The spatial connections are represented in the form of connectivity matrix **W**. Assuming that there are N regions (spatial units), the matrix has as many rows and columns as there are the regions, i.e. N by N matrix **W** is considered. Each row of the matrix contains non-zero elements in columns which correspond to the connected objects (the so-called neighbours), according to the received criterion. Furthermore, the given object cannot be connected to itself, i.e. it cannot be a neighbour of itself, so $w_{ij} = 0$ for all $i = j$. Thus, the diagonal elements of **W** are zeros.

Starting point in establishing the spatial connections is the binary matrix **S**, with elements:

$$s_{ij} = \begin{cases} 1, \text{ if } \ j \in N(i) \\ 0, \text{ if } \ j \notin N(i), \end{cases} \tag{1}$$

where $N(i)$ denotes the set of neighbours of spatial units i.

The neighbours usually are established according to the common border criterion. Then the rows in **S** are normalized, so that the row sums equal 1, as a result of dividing each entry on a row by the sum of the row values (the so-called row standardization to one). So, if $\mathbf{D} = \mathrm{diag}\{d_i\}$, where $d_i = \frac{1}{\sum_{j=1}^{N} s_{ij}}$, $i = 1, 2, \ldots, N$, then $\mathbf{W} = \mathbf{DS}$ and $\sum_{j=1}^{N} w_{ij} = 1$.

The weights w_{ij} which are established in this way signify that each j-th neighbour of the i-th spatial units is treated identically, and the greater the strength of its interactions with the neighbours is, the less neighbours it has.

Another is the case when weights w_{ij} are the functions of some properties of cells of the lattice, e.g. of the length of the common border, of the distance between the centres of the regions, or of other measures of similarity between the regions, e.g. of the so-called economic distance between them as well. Various types of weights w_{ij} may be pointed out according to the established criteria (see Haining, 2005, p. 83–84). As above, for all units i and j $w_{ij} \geq 0$, when $i \neq j$, and $w_{ij} = 0$, when $i = j$. Such weights create the generalized matrix of neighbourhood, which is row standardized to one by transforming its elements according to the formula $w_{ij}^{*} = \frac{w_{ij}}{\sum_{j=1}^{N} w_{ij}}$.

In the paper the connections between the regions will be defined by using two approaches. The first one is traditional. In it the standard matrix of neighbourhood with the equal weights is used. The second one consists in that in the matrix of connections the economic distance (the essence of which is to establish similarity of the regions on the basis of the value of the analyzed economic process and the economic processes which determine its changeability) is taken into consideration.

In the general form the economic distance between regions i and j is expressed as follows (see Pietrzak, 2010, p. 75):

$$d_{ij} = \begin{cases} |y_i - y_j|^{\delta_1} + \frac{1}{k_1+1}\sum_{z=0}^{k_1} |x_{1i,t-z} - x_{1j,t-z}|^{\delta_2} + \frac{1}{k_2+1}\sum_{z=0}^{k_2} |x_{2i,t-z} - x_{2j,t-z}|^{\delta_3} \\ + \ldots + \frac{1}{k_n+1}\sum_{z=0}^{k_n} |x_{ni,t-z} - x_{nj,t-z}|^{\delta_{n+1}}, \ \text{for } i \neq j \\ 0, \quad \text{for } i = j, \end{cases} \tag{2}$$

where: y_i, y_j – values of the investigated spatial economic process at spatial location i and j respectively; x_{li}, x_{lj} (l = 1, 2, …, n) – values of explanatory processes, which determine changeability of the explained process at spatial locations as above; k_1, k_2, …, k_n – constants denoting time lags for the considered processes; δ_1, δ_2, …, δ_{n+1} – normalizing constants.

In this approach, the elements w_{ij} of the matrix of neighbourhood will be as follows:

$$w_{ij} = \begin{cases} \dfrac{1}{d_{ij}}, \text{ if } i \neq j \\ 0, \text{ if } i = j \ . \end{cases} \tag{3}$$

Finally, as a result of row standardization to one the matrix of neighbourhood based on the economic distance is obtained.

The more exact, because based on the economic criterion, measurement of the spatial autodependence should effect more precise filtration of the investigated processes (explained and explanatory), and as a consequence, lead to the appropriate estimation of the regression dependence. In other words, the more precise description of the spatial autodependence the more efficient filtration of the processes and the measurement of the dependence between the processes takes place at the level of the data which are more and more cleaned from the autodependence. The pure dependence between the processes would be the same, apart from the level of the data aggregation.

In the approach proposed the effect of spatial aggregation of the data as the change of the regression parameter in the spatial model will be not eliminated, because the aggregated matrix of neighbourhood has still referred to another one than before aggregation of the data configuration.

2. Empirical Example

The dependence between the unemployment rate and investment outlays of enterprises (in PLN) in Poland in years 2004–2009 across poviats (districts) and sub-regions was investigated. The data are from the internet source: http://www.stat.gov.pl. The conception of the spatial and spatio-temporal quasi-congruent model was used. The trend-autoregressive structures of the individual processes at the both levels of the data aggregation were investigated by constructing the appropriate fundamental models[4]. For describing the trend structure the models of the spatial and spatio-temporal trend were used. The autoregressive structure was defined in two versions, with the use of (1) the

[4] The fundamental model denotes here the model describing the componential structure of the spatial and spatio-temporal process. The considerations are accompanied by the foundation, that in economic processes at least two components are potentially present: the trend and autoregressive connections.

standard matrix with the equal weights – variant I and (2) the matrix of the economic distance (different weights) – variant II. Then the quasi-congruent models describing the dependence between the investigated processes were built and the obtained results were compared.

For the investigated processes the following symbols were taken: $Y_t(\mathbf{p})$ – spatial process of unemployment (unemployment rate in the region at spatial location $\mathbf{p} = [p_1, p_2]$ in established year t), $X_t(\mathbf{p})$ – spatial process of investments (investment outlays per capita in the region at spatial location and time – as above), $Y(\mathbf{p}, t)$ – spatio-temporal process of unemployment, $X(\mathbf{p}, t)$ – spatio-temporal process of investments.

Investigation of the dependence in years – variant I

Investigation of the dependence between the unemployment rate and the investment outlays per capita across poviats in the successive years allows to ascertain the following:

1. The models of unemployment and investments contained spatial trends of the first degree in all the years and spatial autodependence of the first order, except for 2004 and 2005, where the spatial autodependence in investments appeared statistically insignificant.
2. The quasi-congruent models in their full version took into consideration the trend structure and the autoregressive structure of the separate processes. Thus, they were the models of the following form:
$$Y_t(\mathbf{p}) = \beta_{00} + \beta_{10} p_1 + \beta_{01} p_2 + \rho \mathbf{W} Y_t(\mathbf{p}) + \gamma X_t(\mathbf{p}) + \delta \mathbf{W} X_t(\mathbf{p}) + \varepsilon_t(\mathbf{p}) . \quad (4)$$
3. After reduction of the statistically insignificant components, for all the years the models with analogous structures were obtained, i.e. the models of the form:
$$Y_t(\mathbf{p}) = \beta_{00} + \beta_{10} p_1 + \beta_{01} p_2 + \rho \mathbf{W} Y_t(\mathbf{p}) + \gamma X_t(\mathbf{p}) + \varepsilon_t(\mathbf{p}) , \quad (5)$$
except for 2007, where the spatial trend was also reduced. The theoretical notation of the model for 2007 is following:
$$Y_4(\mathbf{p}) = \beta_{00} + \rho \mathbf{W} Y_4(\mathbf{p}) + \gamma X_4(\mathbf{p}) + \varepsilon_4(\mathbf{p}) . \quad (6)$$

In the investigation of the dependence between the unemployment and the investment outlays per capita across sub-regions in the successive years the following was established:

1. The fundamental models of unemployment and investments did not contain any significant spatial trend, but they had autoregressive components. Only in 2004 and 2009 the spatial autoregression in investments appeared statistically insignificant.

2. In the face of settlement 1. the full quasi-congruent models for 2005–2008 took the form:
$$Y_t(\mathbf{p}) = \beta_{00} + \rho \mathbf{W} Y_t(\mathbf{p}) + \gamma X_t(\mathbf{p}) + \delta \mathbf{W} X_t(\mathbf{p}) + \varepsilon_t(\mathbf{p}), \tag{7}$$
and for 2004 and 2009:
$$Y_t(\mathbf{p}) = \beta_{00} + \rho \mathbf{W} Y_t(\mathbf{p}) + \gamma X_t(\mathbf{p}) + \varepsilon_t(\mathbf{p}). \tag{8}$$
3. The reduction of the statistically insignificant components led to the models of the form (8), for all the years.

Table 1 presents specification of the values of the chosen parameters of the reduced quasi-congruent models for poviats and sub-regions. The comparison of the values leads to the statement, that in the successive years the regression coefficients estimated at the level of poviats are, as regards the absolute value, smaller than the ones obtained for sub-regions. In turn, the coefficients of autoregression for poviats are bigger than for sub-regions.

Table 1. Coefficients of regression and autoregression in the models estimated at the level of poviats and sub-regions – variant I

Level of aggregation	Years					
	2004	2005	2006	2007	2008	2009
Coefficients of regression γ						
Poviats	-1.6371	-1.6489	-1.3799	-0.8802	-0.6379	-0.5991
Sub-regions	-2.5394	-2.5007	-1.9284	-1.4969	-1.2207	-1.3018
Coefficients of autoregression ρ						
Poviats	0.6833	0.6747	0.6724	0.6851	0.6632	0.6381
Sub-regions	0.5112	0.5139	0.4353	0.3370	0.3544	0.4321

Note: All the calculations for the investigation presented in the paper were done with the use of R-CRAN.

Investigation of the dependence in years – variant II

The influence of investment outlays on unemployment rate was also investigated by using the matrix of economic distance to define spatial autodependence in the considered processes. The matrix with differentiated weights, taking into account similarity of poviats/sub-regions established on the ground of the values of unemployment and investment outlays in the connected regions (i.e. in poviats, and then in sub-regions) was used. Thus, the economic distance was defined as follows:
$$d_{ij} = \begin{cases} |y_i - y_j|^{\delta_1} + |x_i - x_j|^{\delta_2}, & \text{for } i \neq j \\ 0, & \text{for } i = j, \end{cases} \tag{9}$$
where $\delta_1 = \delta_2 = 0.9$[5].

In this case, for poviats, it was established, that:

[5] The value has been established on the ground of the separated investigations.

1. The fundamental models of unemployment and investments identified at the level of poviats, for all the years, contained spatial trends of the 1st degree and the autodependence of the 1st order. It is the analogous result as that one obtained in variant I.
2. Also, likewise in variant I of the analysis the full quasi-congruent models contained spatial trends, spatially lagged unemployment, current investments and spatially lagged investments (see formula (4)).
3. After reduction of the statistically insignificant components, for 2004–2007 the spatial autoregressive regression models were obtained (see, formula (8)). For 2008 from the full model only the spatial trend was eliminated, yet for 2009 the full model did not need any component to be reduced.

The analysis of the data at the level of sub-regions led to the following settlements:

1. The unemployment and investments in all the years were described with the help of the pure autoregressive spatial models (without trend).
2. It led to the quite simple structure of the full quasi-congruent models, i.e.:
$$Y_t(\mathbf{p}) = \beta_{00} + \rho\mathbf{W}^* Y_t(\mathbf{p}) + \gamma X_t(\mathbf{p}) + \delta\mathbf{W}^* X_t(\mathbf{p}) + \varepsilon_t(\mathbf{p}). \qquad (10)$$
3. For all the years there were obtained the reduced models of the form:
$$Y_t(\mathbf{p}) = \beta_{00} + \rho\mathbf{W}^* Y_t(\mathbf{p}) + \gamma X_t(\mathbf{p}) + \varepsilon_t(\mathbf{p}). \qquad (11)$$

Table 2. Coefficients of regression and autoregression in the models estimated at the level of poviats and sub-regions – variant II

Level of aggregation	Years					
	2004	2005	2006	2007	2008	2009
Coefficients of regression γ						
Poviats	-1.2893	-1.3225	-1.1025	-0.7173	-0.4401	-0.4075
Sub-regions	-2.3358	-2.2648	-1.7279	-1.3171	-1.0979	-1.1736
Coefficients of autoregression ρ						
Poviats	0.8158	0.8026	0.7954	0.7727	0.7506	0.7110
Sub-regions	0.6170	0.5938	0.5506	0.5046	0.4829	0.5472

Table 2 presents specification of the values of the chosen parameters of the reduced quasi-congruent models for poviats and sub-regions. Comparing the results of the analysis for poviats and sub-regions, which are presented in Table 2, it should be stated that in all the years the regression coefficients estimated at the level of poviats, are as regards the absolute value smaller than their estimates at the level of sub-regions. However, the coefficients of autoregression in the considered years are always bigger for poviats than for sub-regions.

Spatio-temporal models – variant I

Constructing the spatio-temporal models the analogous structure of components as in the case of the spatial models was established. In the analysis the deterministic spatio-temporal trends were taken into consideration. Identification of the spatio-temporal autodependence was done with the use of the matrix of the form:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_6 \end{bmatrix}, \tag{12}$$

where $\mathbf{W}_1 = \mathbf{W}_2 = ... = \mathbf{W}_6$ – standard matrixes of the spatial connections, the same for all the considered years.

With specification of the fundamental models it was confirmed, that both the unemployment rate and the investment outlays showed the spatio-temporal trends of the 1st degree and spatial autodependence of the 1st order. The components were observed as well at the level of poviats as at the level of sub-regions and they were taken into account in the quasi-congruent spatio-temporal models constructed afterwards. Since the full models contained insignificant variables they were eliminated and in this way the reduced models were obtained. Table 3 contains the characteristics of the quasi-congruent models obtained in the investigation at the level of poviats, yet Table 4 presents the analogous specification for sub-regions.

The reduced models do not contain the spatially lagged investments. This means that the unemployment in the given poviat does not depend on the investment outlays in the neighbouring poviats. The current unemployment in the given poviat is influenced by the current investment outlays in the same poviat (each thousand PLN spent causes the unemployment rate to decrease on the average by 0.93 percentage point) and by the level of unemployment in the neighbouring poviats (the change of the unemployment rate in the given poviat by about 0.69 percentage point is connected with the one percentage change of the unemployment rate in the neighbouring poviats).

Likewise, the unemployment in the given sub-region does not depend on the investment outlays in the neighbouring sub-regions. The current unemployment rate in the given sub-region is influenced by the current investment outlays in this sub-region (each thousand PLN spent causes the unemployment rate to decrease by about 1.74 percentage point) and by unemployment rate in the neighbouring sub-regions (the increase of the unemployment rate in the neighbouring sub-regions by one percentage point is connected with the increase of the unemployment rate in a given sub-region by about 0.39 percentage point).

The coefficient which measures the influence of the investment outlays on the unemployment rate in a sub-region differs from the analogous coefficient which is estimated at the level of poviats. The main reason of the difference observed is the spatial autocorrelation in unemployment and also in investments. The value of coefficient γ for sub-regions is bigger than for poviats. In turn, autoregression coefficient ρ, which measures the connections among the unemployment rates in the neighbouring areas at the level of sub-regions is visibly smaller than the analogous parameter estimated for poviats. The results are analogous to the ones obtained previously for the spatial models constructed for the successive years. They are also convergent with the results of the previous analyses, pointed out in the introduction, which were obtained on the basis of the generated data (especially, when $\rho_y > \rho_x$, the coefficient of regression, estimated on the basis of the aggregated data, is overestimated).

Table 3. Characteristics of quasi-congruent models for poviats – variant I

Full model:

$$Y(\mathbf{p},t)=\beta_{000}+\beta_{100}p_1+\beta_{010}p_2+\beta_{001}t+\rho\mathbf{W}Y(\mathbf{p},t)+\gamma X(\mathbf{p},t)+\delta\mathbf{W}X(\mathbf{p},t)+\varepsilon(\mathbf{p},t)$$

Parameters	Estimates of parameters	Standard deviations	Statistics z	Pr(>|z|)
β_{000}	7.6305	0.6915	11.0347	0.0000
β_{100}	-0.1892	0.0652	-2.9025	0.0037
β_{010}	0.3472	0.0664	5.2255	0.0000
β_{001}	-0.3433	0.0665	-5.1612	0.0000
γ	-0.9315	0.0504	-18.4788	0.0000
δ	-0.0332	0.1085	-0.3060	0.7596

ρ = 0.6872
Test LR: 967.48; p-value: 0.0000

Wald statistic: 1409.3; p-value: 0.0000
AIC: 13592 (AIC for lm: 14557)

Residual autocorrelation
Test LM: 4.2911; p-value: 0.0383;
Moran statistic: -0.8026; p-value: 0.2111

Reduced model:

$$Y(\mathbf{p},t)=\beta_{000}+\beta_{100}p_1+\beta_{010}p_2+\beta_{001}t+\rho\mathbf{W}Y(\mathbf{p},t)+\gamma X(\mathbf{p},t)+\varepsilon(\mathbf{p},t)$$

Parameters	Estimates of parameters	Standard deviations	Statistics z	Pr(>|z|)
β_{000}	7.5202	0.6092	12.3439	0.0000
β_{100}	-0.1816	0.0613	-2.9644	0.0030
β_{010}	0.3464	0.0664	5.2171	0.0000
β_{001}	-0.3472	0.0647	-5.3681	0.0000
γ	-0.9334	0.0499	-18.6968	0.0000

ρ = 0.6890
Test LR: 1090.4; p-value: 0.0000

Wald statistic: 1538; p-value: 0.0000
AIC: 13590 (AIC for lm: 14679)

Residual autocorrelation
Test LM: 3.7569; p-value: 0.0526
Test Moran statistic: -0.8867; p-value: 0.1876

Table 4. Characteristics of quasi-congruent models for sub-regions – variant I

Full model: $Y(\mathbf{p},t)=\beta_{000}+\beta_{100}p_1+\beta_{010}p_2+\beta_{001}t+\rho\mathbf{W}Y(\mathbf{p},t)+\gamma X(\mathbf{p},t)+\delta\mathbf{W}X(\mathbf{p},t)+\varepsilon(\mathbf{p},t)$				
Parameters	Estimates of parameters	Standard deviations	Statistics z	Pr(>|z|)
β_{000}	17.2312	2.0420	8.4384	0.0000
β_{100}	-0.0459	0.0156	-2.9479	0.0032
β_{010}	0.0482	0.0151	3.2023	0.0014
β_{001}	-0.5223	0.1628	-3.2082	0.0013
γ	-1.6817	0.1431	-11.7497	0.0000
δ	-0.6643	0.3645	-1.8225	0.0684
ρ = 0.3541 Test LR: 30.415; p-value: 0.0000				
Wald statistic: 35.928; p-value: 0.0000 AIC: 2279.5 (AIC for lm: 2308)				
Residual autocorrelation Test LM: 6.2037; p-value: 0.0127 Moran statistic: 0.6251; p-value: 0.2660				
Reduced model: $Y(\mathbf{p},t)=\beta_{000}+\beta_{100}p_1+\beta_{010}p_2+\beta_{001}t+\rho\mathbf{W}Y(\mathbf{p},t)+\gamma X(\mathbf{p},t)+\varepsilon(\mathbf{p},t)$				
Parameters	Estimates of parameters	Standard deviations	Statistics z	Pr(>|z|)
β_{000}	14.8309	1.5208	9.7522	0.0000
β_{100}	-0.0321	0.0139	-2.3164	0.0205
β_{010}	0.0522	0.0149	3.4839	0.0005
β_{001}	-0.6181	0.1565	-3.9504	0.0000
γ	-1.7460	0.1365	-12.7951	0.0000
ρ = 0.3883 Test LR: 40.419; p-value: 0.0000				
Wald statistic: 50.843; p-value: 0.0000 AIC: 2281 (AIC for lm: 2319.5)				
Residual autocorrelation Test LM: 0.2836; p-value: 0.5943 Moran statistic: -0.1386; p-value: 0.4449				

Spatio-temporal models – variant II

In the discussed variant for the purpose of quantification of the spatio-temporal connections there was used the matrix of the form:

$$\mathbf{W}^* = \begin{bmatrix} \mathbf{W}^*{}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^*{}_2 & \cdots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}^*{}_6 \end{bmatrix}, \tag{13}$$

where: $\mathbf{W}^*{}_1 \neq \mathbf{W}^*{}_2 \neq \ldots \neq \mathbf{W}^*{}_6$ – matrixes of spatial connections, taking into account the economic distance between regions, different for the successive years.

Various specifications of the fundamental models of unemployment and investments and the models of the dependence between the investigated processes, which resulted from them, were used. Tables 5–6 present the results of the investigation, in which the time lags of the spatial dependence are not taken into account. Tables 7–8, on the contrary, refer to the investigation, in which such the lags were considered.

The structure of the full models results from the structures of the individual processes described by the fundamental models. In the first specification the fundamental models with the spatio-temporal trend and with the spatial autoregression were established. In the full model identified at the level of poviats all the components were statistically significant. Other than in variant I, the measurement of spatial (auto)dependence caused the spatially lagged investments had not been removed from the model. It denotes some connection of the current unemployment rate in a given poviat with the investment outlays in the neighbouring poviats. In turn, in the analysis at the level of sub-regions the reduced model without the spatially lagged investments was obtained.

Like in variant I of the investigation, the difference in the estimation of the coefficient measuring the influence of investments on unemployment at the level of poviats and sub-regions, was observed. For poviats its value amounted to circa -0.69, while for sub-regions to circa -1.52. Also the coefficient of autoregression, as in variant I, is bigger for poviats (about 0.76) than for sub-regions (about 0.53).

Table 5. Characteristics of quasi-congruent models for poviats – variant II a)

Full model:

$$Y(\mathbf{p},t)=\beta_{000}+\beta_{100}p_1+\beta_{010}p_2+\beta_{001}t+\rho\mathbf{W}^*Y(\mathbf{p},t)+\gamma X(\mathbf{p},t)+\delta\mathbf{W}^*X(\mathbf{p},t)+\varepsilon(\mathbf{p},t)$$

Parameters	Estimates of parameters	Standard deviations	Statistics z	Pr(>|z|)
β_{000}	5.9703	0.5611	10.6413	0.0000
β_{100}	-0.1403	0.0548	-2.5583	0.0105
β_{010}	0.2307	0.0548	4.2073	0.0000
β_{001}	-0.1419	0.0537	-2.6408	0.0083
γ	-0.6937	0.0453	-15.3140	0.0000
δ	-0.4661	0.1219	-3.8231	0.0001

ρ = 0.7595
Test LR: 1467.6; p-value: 0.0000

Wald statistic: 2912.2; p-value: 0.0000
AIC: 12943 (AIC for lm: 1.4408)

Residual autocorrelation
Test LM: 68.725; p-value: 0.0000
Moran statistic: -2.9692; p-value: 0.0010

In the more extended specification of the structures of the investigated processes the time lags of the spatial autodependence were additionally taken into account. The characteristics of the full quasi-congruent model and of the reduced model for poviats are presented in Table 7, whereas for sub-regions – in Table 8.

Table 6. Characteristics of quasi-congruent models for sub-regions – variant II a)

Full model:

$$Y(\mathbf{p},t)=\beta_{000}+\beta_{100}p_1+\beta_{010}p_2+\beta_{001}t+\rho\mathbf{W}^*Y(\mathbf{p},t)+\gamma X(\mathbf{p},t)+\delta\mathbf{W}^*X(\mathbf{p},t)+\varepsilon(\mathbf{p},t)$$

Parameters	Estimates of parameters	Standard deviations	Statistics z	Pr(>|z|)
β_{000}	12.3451	1.7497	7.0555	0.0000
β_{100}	-0.0240	0.0143	-1.6530	0.0983
β_{010}	0.0315	0.0138	2.2901	0.0220
β_{001}	-0.3442	0.1466	-2.3488	0.0188
γ	-1.4485	0.1495	-9.6917	0.0000
δ	-0.4057	0.3925	-1.0337	0.3013

ρ = 0.5178
Test LR: 77.168; p-value: 0.0000

Wald statistic: 117.56; p-value: 0.0000
AIC: 2235.4 (AIC for lm: 2310.5)

Residual autocorrelation
Test LM: 37.529; p-value: 0.0000
Moran statistic: -0.7447; p-value: 0.2282

Reduced model:

$$Y(\mathbf{p},t)=\beta_{000}+\beta_{100}p_1+\beta_{010}p_2+\beta_{001}t+\rho\mathbf{W}^*Y(\mathbf{p},t)+\gamma X(\mathbf{p},t)+\varepsilon(\mathbf{p},t)$$

Parameters	Estimates of parameters	Standard deviations	Statistics z	Pr(>|z|)
β_{000}	11.1540	1.3366	8.3450	0.0000
β_{100}	-0.0161	0.0128	-1.2578	0.2085
β_{010}	0.0343	0.0135	2.5450	0.0109
β_{001}	-0.4002	0.1362	-2.9374	0.0033
γ	-1.5245	0.1266	-12.0426	0.0000

ρ = 0.5315
Test LR: 86.996; p-value: 0.0000

Wald statistic: 137.67; p-value: 0.0000
AIC: 2234.5 (AIC for lm: 2319.5)

Residual autocorrelation
Test LM: 14.549; p-value: 0.0001
Moran statistic: -1.1656; p-value: 0.1219

In the full quasi-congruent model of unemployment for poviats the following were taken into consideration: spatio-temporal trend, current unemployment rate in the neighbouring poviats, unemployment rate from the previous period in the neighbouring poviats, current investment outlays in the given poviat and in the neighbouring poviats and also investment outlays in the neighbouring poviats from the previous period. Most of the mentioned components were statistically significant. Only the spatial trend appeared insignificant. After elimination of it the reduced model was obtained.

Starting from the analogous structure of the full model for sub-regions, as a result of eliminating the statistically insignificant components the reduced model was obtained. In the model there are the following components: current unemployment in the neighbouring sub-regions, current investment outlays in a given sub-region, current and time lagged investment outlays in the neighbouring sub-regions.

Coefficient of regression (γ) estimated at the level of poviats amounted to circa -0.63, and at the level of sub-regions – to circa -1.21. The coefficients of autoregression amounted to circa 0.45 and 0.33 respectively.

Table 7. Characteristics of quasi-congruent models for poviats – variant II b)

Full model:

$$Y(\mathbf{p},t)=\beta_{000}+\beta_{100}p_1+\beta_{010}p_2+\beta_{001}t+\rho\mathbf{W}^*Y(\mathbf{p},t)+\rho^\bullet\mathbf{W}^*Y(\mathbf{p},t-1)+$$
$$+\gamma X(\mathbf{p},t)+\delta\mathbf{W}^*X(\mathbf{p},t)+\delta^\bullet\mathbf{W}^*X(\mathbf{p},t-1)+\varepsilon(\mathbf{p},t)$$

Parameters	Estimates of parameters	Standard deviations	Statistics z	Pr(>|z|)
β_{000}	-1.6307	0.8154	-1.9998	0.0455
β_{100}	0.1506	0.0632	2.3803	0.0173
β_{010}	-0.0589	0.0621	-0.9496	0.3423
β_{001}	0.4975	0.0794	6.2677	0.0000
ρ^*	0.4882	0.0253	19.2694	0.0000
γ	-0.6319	0.0461	-13.7052	0.0000
δ	-0.7915	0.1780	-4.4470	0.0000
δ^*	0.9709	0.1823	5.3247	0.0000

ρ = 0.4466
Test LR: 189.85; p-value: 0.0000

Wald statistic: 351.93; p-value: 0.0000
AIC: 10456 (AIC for lm: 10643)

Residual autocorrelation
Test LM: 3.4071; p-value: 0.0649
Moran statistic: 0.7442; p-value: 0.2284

Reduced model:

$$Y(\mathbf{p},t)=\beta_{000}+\beta_{001}t+\rho\mathbf{W}^*Y(\mathbf{p},t)+\rho^\bullet\mathbf{W}^*Y(\mathbf{p},t-1)+$$
$$+\gamma X(\mathbf{p},t)+\delta\mathbf{W}^*X(\mathbf{p},t)+\delta^\bullet\mathbf{W}^*X(\mathbf{p},t-1)+\varepsilon(\mathbf{p},t)$$

Parameters	Estimates of parameters	Standard deviations	Statistics z	Pr(>|z|)
β_{000}	-0.2661	0.5758	-0.4622	0.6439
β_{001}	0.4718	0.0775	6.0847	0.0000
ρ^*	0.4622	0.0233	19.8427	0.0000
γ	-0.6332	0.0462	-13.7195	0.0000
δ	-0.8660	0.1757	-4.9275	0.0000
δ^*	0.8658	0.1784	4.8538	0.0000

ρ = 0.4483
Test LR: 191.29; p-value: 0.0000

Wald statistic: 359.94; p-value: 0.0000
AIC: 10458 (AIC for lm: 10648)

Residual autocorrelation
Test LM: 7.136; p-value: 0.0076
Moran statistic: 1.0981; p-value: 0.1361

Table 8. Characteristics of quasi-congruent models for sub-regions – variant II b)

Full model:

$$Y(\mathbf{p},t)=\beta_{000}+\beta_{100}p_1+\beta_{010}p_2+\beta_{001}t+\rho\mathbf{W}^*Y(\mathbf{p},t)+\rho^{\bullet}\mathbf{W}^*Y(\mathbf{p},t-1)+$$
$$+\gamma X(\mathbf{p},t)+\delta\mathbf{W}^*X(\mathbf{p},t)+\delta^{\bullet}\mathbf{W}^*X(\mathbf{p},t-1)+\varepsilon(\mathbf{p},t)$$

Parameters	Estimates of parameters	Standard deviations	Statistics z	Pr(>|z|)
β_{000}	4.5273	2.8085	1.6120	0.1070
β_{100}	0.0047	0.0165	0.2861	0.7748
β_{010}	-0.0006	0.0165	-0.0371	0.9704
β_{001}	0.2194	0.2483	0.8838	0.3768
ρ^*	0.4392	0.0837	5.2446	0.0000
γ	-1.1887	0.1569	-7.5757	0.0000
δ	-1.6994	0.5483	-3.0992	0.0019
δ^*	1.7360	0.5679	3.0568	0.0022

$\rho =$ 0.3148
Test LR: 14.607; p-value: 0.0001

Wald statistic: 26.774; p-value: 0.0000
AIC: 1820.7 (AIC for lm: 1833.3)

Residual autocorrelation
Test LM: 0.2322; p-value: 0.6299
Moran statistic: 0.2175; p-value: 0.4139

Reduced model:

$$Y(\mathbf{p},t)=\beta_{000}+\rho\mathbf{W}^*Y(\mathbf{p},t)+\rho^{\bullet}\mathbf{W}^*Y(\mathbf{p},t-1)+\gamma X(\mathbf{p},t)+$$
$$+\delta\mathbf{W}^*X(\mathbf{p},t)+\delta^{\bullet}\mathbf{W}^*X(\mathbf{p},t-1)+\varepsilon(\mathbf{p},t)$$

Parameters	Estimates of parameters	Standard deviations	Statistics z	Pr(>|z|)
β_{000}	6.3379	1.4305	4.4306	0.0000
ρ^*	0.3855	0.0566	6.8075	0.0000
γ	-1.2131	0.1491	-8.1375	0.0000
δ	-1.7387	0.5353	-3.2479	0.0012
δ^*	1.8079	0.5220	3.4631	0.0005

$\rho =$ 0.3256
Test LR: 16.022; p-value: 0.0000

Wald statistic: 29.457; p-value: 0.0000
AIC: 1815.6 (AIC for lm: 1829.6)

Residual autocorrelation
Test LM: 1.1578; p-value: 0.2819
Moran statistic: 0.4223; p-value: 0.3364

Conclusions

In the modelling of the dependence of spatial processes and also spatio-temporal processes the principle of congruency of the structures of the separate processes should be applied. Limited efficiency of the spatial and spatio-temporal quasi-congruent models as the tools of discovering the real dependence between the processes (in particular the differences in estimating the dependence at different levels of data aggregation observed) result, first of all, from imperfection of the measurement of autodependence in the investigated processes.

References

Arbia, G. (1988), *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*, Kluwer Academic Press, Dordrecht.

Haining, R. (2005), *Spatial Data Analysis. Theory and Practice*, Cambridge University Press, 3th ed., Cambridge.

Pietrzak, M. B. (2010), Dwuetapowa procedura budowy przestrzennej macierzy wag z uwzględnieniem odległości ekonomicznej (Two-stage Procedure of Building a Spatial Weight Matrix with the Consideration of Economic Distance), *Oeconomia Copernicana*, 1, Wydawnictwo UMK, Toruń, 65–78.

Szulc, E., Müller-Frączek, I., Pietrzak, M. B. (2011), Modelowanie zależności między ekonomicznymi procesami przestrzennymi a poziom agregacji danych (Modeling of Dependence between Economic Spatial Processes and the Level of Data Aggregation), in Suchecka J. (ed.), *Ekonometria przestrzenna i regionalne analizy ekonomiczne* (*Spatial Econometrics and Regional Economic Analyses*), *Folia Oeconomica*, Wydawnictwo Uniwersytetu Łódzkiego, 327–344.

Szulc, E. (2011), Ekonometryczna identyfikacja struktur procesów przestrzennych wobec problemu agregacji danych (Econometric Identification of the Structures of Spatial Processes and a Problem of Data Aggregation), *Acta Universitatis Nicolai Copernici, Ekonomia XLII – Nauki Humanistyczno-Społeczne*, z. 392, Toruń, in press.

Identyfikacja struktur procesów przestrzennych i przestrzenno-czasowych wobec problemu agregacji danych

Z a r y s t r e ś c i. Artykuł dotyczy odkrywania zależności między ekonomicznymi procesami przestrzennymi a także przestrzenno-czasowymi, gdy są one mierzone na różnych poziomach agregacji danych. Rozważania nawiązują do badań, potwierdzających efektywność tzw. quasi-zgodnego modelu przestrzennego jako narzędzia pomiaru rzeczywistych zależności między procesami. Wykorzystuje się koncepcję modelu quasi-zgodnego także w odniesieniu do procesów przestrzenno-czasowych. Poszukuje się opisu powiązań przestrzennych i przestrzenno-czasowych adekwatnego do wyrażenia autozależności w badanych procesach. Proponuje się wykorzystanie odległości ekonomicznej, mierzącej podobieństwo regionów na podstawie wartości analizowanych procesów.

S ł o w a k l u c z o w e: przestrzenny i przestrzenno-czasowy model quasi-zgodny, autozależności, macierz sąsiedztwa, odległość ekonomiczna.