Volume 15 No. 4/2012

Comparative Economic Research

Central and Eastern Europe

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

INSTITUTE OF ECONOMICS • UNIVERSITY OF ŁÓDŹ

Volume 15 **No. 4/2012**

Comparative Economic Research

Central and Eastern Europe

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

ŁÓDŹ 2012

The online version available at: *www.versita.com*

Reference version: printed version

TECHNICAL EDITOR
Aleksandra Przybył

Printed directly from camera-ready materials provided to the Łódź University Press

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego
Wydanie I. W.06104.13.0.C

ISSN 1508-2008

Wydawnictwo Uniwersytetu Łódzkiego
90-131 Łódź, ul. Lindleya 8
www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
tel. (42) 665 58 63, faks (42) 665 58 62

Oprawa i Druk: Quick Druk

10.2478/v10103-012-0022-x

ELŻBIETA CHĄDZYŃSKA*

Analysis of Spatial Diversification of Economic Growth Level in Powiats of Dolnośląskie Voivodship

Abstract

Intensive economic and urban processes which take place in Polish urbanized area have a significant impact on the changes of structure of territorial units, their functional development and demographic processes. Development of individual units is dependent on many factors like the location, demographic potential and economics. In the area of one voivodship – there are powiats with a various potential of development.

As a subject of research powiats in the area of the Dolnośląskie Voivodship are selected. Taxonomic methods were used for the classification and grouping of powiats as a spatial objects tested in the multidimensional space of characteristics. The aim of the paper is to present the spatial variation in the level of development in the light of selected factors.

1. Introduction

Intensive economic and urban processes which take place in the Polish urbanized space from more than 20 years have a significant impact on the basic spheres of human activity: environment, social system and economy. Under the influence of these processes occur changes in the structure of territorial units, their functional development and demographic processes. Development of individual units is dependent on many factors. There are the factors associated with the location, demographic potential and economics. In the area of one

* Ph.D., Wroclaw University of Technology

voivodship – there are powiats with a various potential of economic development.

As a subject of research powiats and cities with powiat status in the area of the Dolnośląskie Voivodship are selected. Taxonomic methods was used for the classification and grouping of powiats as spatial objects tested in the multidimensional space of characteristics. The aim of the paper is to present the impact of selected factors on the spatial variation in the level of economic development.

2. The characteristics of the area under research

As a subject of research the area of Dolnośląskie Voivodship is selected. Taxonomic methods was used for the classification and grouping of powiats as spatial objects tested in the multidimensional space of characteristics.

Figure1. Location of dolnośląskie voivodship

Source: author's own.

According to nomenclature of NUTS-4 Dolnośląskie Voivodship were the 26 powiats and 3 cities with powiat status: Jelenia Góra, Legnica and Wrocław (as of 31 Dec. 2010). Dolnośląskie Voivodship is situated in the southwestern part of Poland and is bordered by three Voividships: Lubuskie, Wielkopolskie and Opolskie, and with the Czech Republic and Germany. As of 31 DEC. 2010 area of the Voivodship is 19947 km^2 (6,4% of the territory of the country; 7 place among provinces). The population of the Voivodship is 2877,8 persons, representing 7.5% of the population of Poland. In terms of number of population dolnośląskie is located at the 5 spot in the country. In 2010, in Dolnośląskie Voivodship, in contrast to previous periods, there was a slight positive growth of population. This is largely an effect of migratory movements. The State belongs to the most urban area in Poland. Lower Silesian Voivodeship is the nature of the industrial-agricultural and well developed economy.

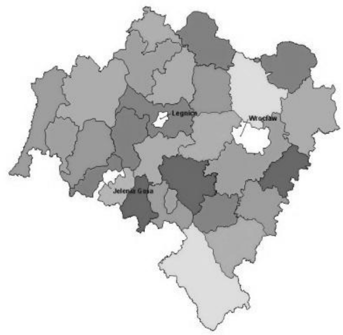
The level of economic development in the area of voivodship is determined in various ways. Influential is the position of the unit in the system of communications involving closer and further neighborhood (the availability of communication), which is linked directly to the quality of the infrastructure to develop the necessary contacts, population and age structure, investment attractiveness of individual units, the structure of employment and the characteristics of the labor market.

To examine the status of and prospects for economic development in the County, you need to select the set of features that characterize the best studied phenomenon. Area of Dolnośląskie Voivodship is highly diverse, both in terms of the size of individual units (provinces), their functional development, position in the hierarchy of administrative, rehabilitation of the environment and many others. Carried out numerous studies and analysis of opportunities for economic development and its determinants in the area of Dolnośląskie Voivodship evaluate three basic types of areas. They are: city districts – Wrocław, Legnica and Jelenia Góra, which have the strongest position during layout, with the most distinguished capital region - the city of Wrocław.

The second type of areas represent municipalities that make up the counties within the immediate vicinity of the listed cities, which benefit from the position in a broad-ranging contacts and areas located away from main roads, which to a large extent hinders their development (Chądzyńska, Iwaszko 2012, pp. 33 – 36, Litwińska 2007, Mlek, Zipser 2007).

The characteristics selected for testing the diversification level of economic development of powiats focus on topics concerning the demographic characteristics, infrastructure, income and expenditure and the labor market. These thematic areas are described using the characteristics, which figures in the form of indicators, which enables the comparison of territorial units of different sizes. Values of the indicators relate to the year 2010.

Figure 2. Poviats in dolnośląskie voivodship



Source: author's own.

Ratio of population density is one of the characteristics of the most commonly used in research. Lower Silesian Voivodship in 2010, it was an average of 144 persons per km^2, which is in fourth place in the country (average for Poland is 122 persons per km^2.) The coefficient of feminization (number of women per 100 men) amounted to 109, which gives a third place in the country. Natural growth for the Lower Silesian Voivodship was in 2010, negative and amounted to minus 172 people. According to data for 2009 the estimated value of the gross domestic product remained in the lower Silesian Voivodeship at 110448 million, representing an 8.2% of GDP in the country. This was the second, after the Masovian Voivodeship, largest value among the provinces. Expenditure on investments and investment purchases are the most important group of expenditure. They increase the potential of the region and foster the creation of advantageous conditions for development (Statistical Office in Wrocław 2012, report).

3. Course of analyze

Numeric Taxonomy provides the methods of grouping and classification of data in spatial studies. Taxonomic methods enable to order a collection of objects, the division into disjoint subsets, such as a group, the classes or concentrates containing elements similar, from the point of view of selected characteristics and simultaneously different elements of other subsets. As a result we get subsets of elements like inside the group and different from the other groups (Hellwig 1994, Mynarski 1992, pp. 117-146, Suchecki 2010, pp. 56-57). In comparative and classification studies the important is possibility to compare the objects described by many variables. For this purpose, the most widely used taxonomic methods, which enable the analysis of objects in multidimensional space. In the taxonomic analysis decisive step is the selection of diagnostic variables.

The potential set of diagnostic features chosen as the characteristics of level of economical development of dolnośląskie powiats, included the following variables:

1. population density,
2. index of feminization (woman per 100 men),
3. natural increase per 1000 population,
4. non-working age population per 100 persons of working age,
5. quantity of roads (in km) per 100 km^2 of area,
6. revenue own per capita,
7. expenditure of investment per capita,
8. entities of the National Economy register per 1000 population,
9. population employed in agriculture,
10. population employed in industry, (as a % of total employed)
11. population employed in services.

The vast majority of selected potential diagnostic variables are a stimulant. Only variable showing the number of persons in the not working age per 100 persons in working age is destimulant (for this characteristic negative values were taken). In order to reduce the number of diagnostic variables was verified, while the large variability, and the condition of weak correlate between variables. Assuming the value of the correlation coefficient equal to 0.65 as the threshold value, the final set of diagnostic characteristics contained the following variables:

1. population density,
2. natural increase per 1000 population,
3. non-working age population per 100 persons of working age,
4. expenditure of investment per capita,
5. population employed in agriculture,
6. population employed in industry, (as a % of total employed)
7. population employed in services,
8. revenue own per capita.

The values of the coefficients of correlation between diagnostic variables, presents table 1.

Table 1. Coefficient of correlation between diagnostic variables

	1	2	3	4	5	6	7	8
1	1,000	-0,106	0,083	0,106	-0,465	-0,458	0,410	0,518
2	-0,106	1,000	0,478	0,354	0,136	0,450	-0,479	0,206
3	0,083	0,478	1,000	0,325	-0,172	0,198	-0,278	0,206
4	0,106	0,354	-0,325	1,000	-0,176	0,175	-0,304	0,616
5	-0,465	0,136	-0,172	-0,176	1,000	-0,115	0,144	-0,431
6	-0,458	0,450	0,198	0,175	-0,115	1,000	-0,950	0,057
7	0,410	-0,479	-0,278	-0,304	0,144	-0,950	1,000	-0,123
8	0,518	0,206	0,206	0,616	-0,431	0,057	-0,123	1,000

Source: Own calculations on the basis of Statistical Office in Wrocław, Wrocław 2011.

Because of different units of characteristics, the standardization of diagnostic variables was made in accordance with the rule:

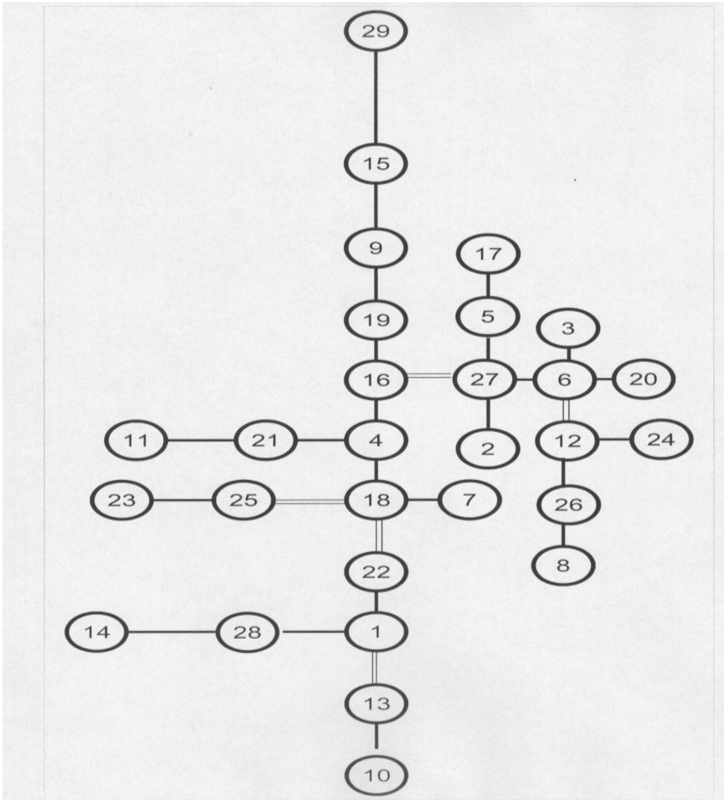
$$x_i^{'} = \frac{x_i - \bar{x}}{s} \qquad (1)$$

i = 1, 2, …, n – number of individuals.

Table of distances was calculating with accordance of Euclidean distance. On the basis of this table "wrocłwski" dendrite was constructed (Fig. 3). The focuses of the first row are marked by bold line, while the second-order focuses by double line.

Dendrite methods belong to the taxonomic hierarchical procedures using the concept in the theory of graphs, which are built based on the matrix of distance (D) between classified objects. At the stage of the construction of the graph they can be treated as the agglomeration procedures. Stage analysis leads to the procedures of subdivision (Mynarski 1992, pp. 140-143, Chądzyńska 2001, pp. 70 - 80).

Figure 3. "Wrocławski" dendrit of Dolnośląskie Powiats



Source: own drawing made by Autocad.

The concentrations are arranged by join each object with the nearest one. In this way we arrange separate groups containing minimum one element. Then we join each group of objects with the nearest group. The procedure is repeated until all objects are arranged into one group. We divide a set of objects into *k* parts by removing from the dendrite arrangement the *k* – 1 longest edge. The division into *k* parts above-mentioned has among all possible divisions into *k* parts the smallest total length of the shortest dendrite arrangements of particular parts (Steinhaus and others 1952, pp. 6-10).

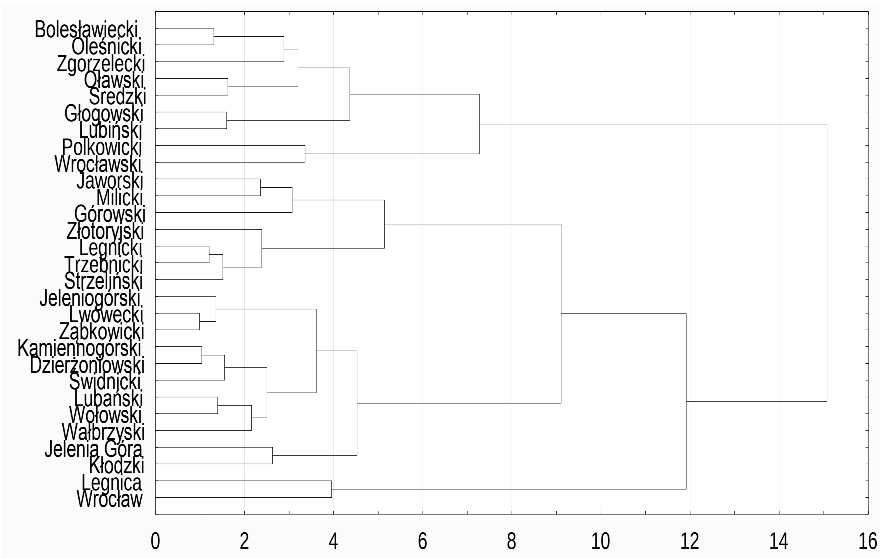
The number of groups was determined by calculating the successive quotients of adjacent elements in descending string of the edges of dendrite (d_i):

$$w_i = d_i/d_{i+1} \quad (i = 1, 2, ..., n-1). \tag{2}$$

The set of individuals dissolves in a "natural" into k elements if $w_k < w_{k+1}$.

In the analysis uses three methods: method of “wrocławski” dendrite (fig 3), the method of Ward (fig 4) and the method of full bindings (fig 5).The most clearly division of counties was obtained using the dendrite method and the method of Ward. These methods have allowed the arrangement of 9 groups of powiats - objects similar due to the level of economic development (as of 2010).

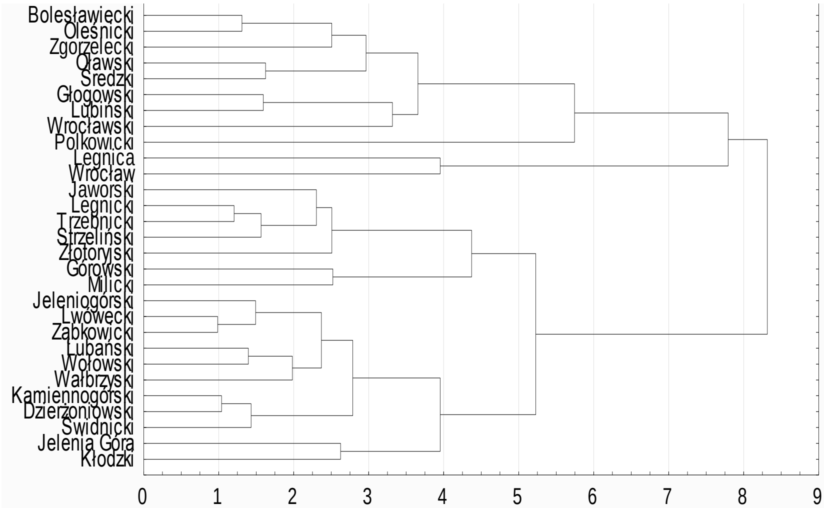
Figure 4. Classification made by Ward’s method. Distance of bindings



Source: Own elaborate in program Statistica.

As a result of the application of the method of Ward and the method of full bindings were obtained almost identical results, differing only in the classification of jaworski powiat. The method of "wrocławski" dendrite also shows a large coincidence with the results of both mentioned methods. Slight differences take place in some counties the "average" in terms of the level of economic development. The districts ' strengths ', city districts and counties "weak" show conformity of classification in all three methods. All of used methods gave a breakdown of the investigated objects on the 9 groups. Since the $w_8 < ... < w_9 < ... < w_7 < ...< w_{10}$ break into 9 groups is justified from the point of view of the natural breakdown.

Figure 5. Method of full bindings. Distance of bindings



Source: Own elaborate in program Statistica.

On the basis of the analysis of the Ward's chart you can say that the greatest similarity due to the tested characteristics can be observed in counties: lwówecki, ząbkowicki and kamiennogórski, dzierżoniowski. Finally, taking into account the received breakdown of the class the breaks as in table 2 of tested objects was accepted (the marks by the name means the number in dendrite).

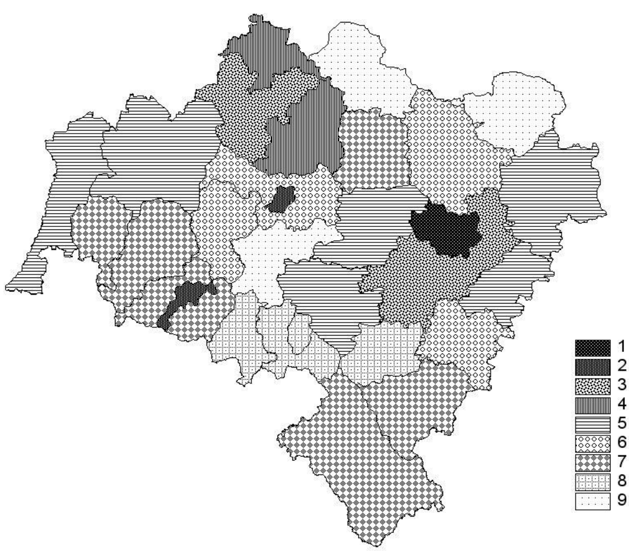
Table 2. Finally concentrations of powiats

Number of concentration	Powiats and cities with powiat status
1	Wrocław
2	Legnica
	Jelenia Góra
3	polkowicki
	wrocławski
4	lubiński
	głogowski
5	oławski
	oleśnicki
	średzki
	świdnicki
	zgorzelecki

	bolesławiecki
6	legnicki
	strzeliński
	trzebnicki
	złotoryjski
7	ząbkowicki
	lwówecki
	jeleniogórski
	lubański
	wołowski
	kłodzki
8	wałbrzyski
	kamiennogórski
	dzierżoniowski
9	górowski
	milicki
	jaworski

Source: Own elaboration on the base of mentioned above methods.

Figure 6. Spatial location of groups of poviats



Source: Own drawing made in Arcview.

4. Conclusions

The units (counties) the most distinguish in the detecting set – irrespective of the method used for grouping create identical groups. Make them the city districts where the level of economic development is relatively high and the counties of well-educated functions (powiats - lubiński, głogowski, polkowicki, wrocławski), and worse developed counties, as powiat górowski milicki and jaworski. Other units (with an average level of development) constitute a collection of diverse objects, as can be seen in some differences in assignment to groups using different methods. A lower level of economic development in counties confirmed the result obtained in many analyses show that counties less communicated, are develop slowly. Though account selected characteristics of economic development, such as expenditure on investment and employment in industry, polkowicki powiat creates in this case distinct quality.

The biggest differentiation among the adopted characteristics diagnostic show features: population density and investment expenditure. Broken down according to group terms adopted characteristics you can see the dependency from the position in the communications system and the neighborhood. The resulting classification of powiats is the reflection of the regional diversity of Dolnośląskie Voivodship.

References

Badania przestrzenne rynku i konsumpcji, Przewodnik metodyczny pod redakcją S.Mynarskiego, PWN, Warszawa 1992

Chądzyńska E. (2001), *Model empirycznych badań preferencji mieszkańców – wybrane metody matematyczne*, Oficyna Wyd. Politechniki Wrocławskiej, Wrocław

Chądzyńska E., Iwaszko K. (2012), *rozwój wybranych funkcji w małych miastach dolnośląskich – analiza wzorca rozkładu przestrzennego,* [in:] Krystian Heffner, Arkadiusz Halama (ed.) *Ewolucja funkcji małych miast w Polsce,* Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach, Katowice

Dolnośląskie Voivodship, Subregions, Powiats, Gminas. Statistical Office In Wrocław, Wrocław 2011

Ekonometria przestrzenna, Metody i modele analizy danych przestrzennych pod red. B. Sucheckiego, Wyd. C.H. Beck, Warszawa 2010

Florek K., Łukaszewicz J., Perkal J., Steinhaus H., Zubrzycki S. (1952), *Taksonomia wrocławska,* odbitka z Przeglądu Antropologicznego, tom XVII (1951), Poznań

Hellwig Z. (1994), *Taksonomia w konstrukcjach i ocenach strategii gospodarczych,* [in:] *Zastosowania metod taksonomicznych w gospodarce,* z. 1, Akademia Ekonomiczna, Wrocław

Identyfikacja i delimitacja obszarów wzrostu oraz obszarów problemowych w województwie dolnośląskim, Urząd Statystyczny we Wrocławiu, Raport, Wrocław 2012

Litwińska E. (2007), *Wybrane wskaźniki społeczno-eknomiczne do badania wpływu metropolii na małe miasta (przykład Wrocławia),* [in:] Krystian Heffner, Tadeusz Marszał (ed.) M*ałe miasta w obszarach metropolitalnych*, KPZK PAN, Warszawa

Mlek M., Zipser W. (2007), Szanse rozwoju małych miast w obszarze metropolitalnym a kształtowanie sieci komunikacyjnej, [in:] Krystian Heffner, Tadeusz Marszał (ed.) Małe miasta w obszarach metropolitalnych, KPZK PAN, Warszawa

Streszczenie

ANALIZA PRZESTRZENNEGO ZRÓŻNICOWANIA POZIOMU ROZWOJU GOSPODARCZEGO POWIATÓW DOLNOŚLĄSKICH

Intensywne procesy gospodarcze oraz urbanizacyjne, które mają miejsce w polskiej przestrzeni zagospodarowanej od przeszło 20 lat, wywierają znaczący wpływ na podstawowe sfery działalności człowieka, tj. środowisko, system społeczny oraz gospodarkę. Pod wpływem tych procesów zachodzą zmiany w strukturze jednostek terytorialnych, ich rozwoju funkcjonalnym oraz demograficznym. Rozwój poszczególnych jednostek uwarunkowany jest wieloma czynnikami. Można tu wyróżnić czynniki związane z lokalizacją, potencjałem ludnościowym i ekonomicznym. W obszarze, jakim jest województwo występują gminy o zróżnicowanym potencjale rozwojowym.

Jako przedmiot badań wybrano powiaty leżące w obszarze województwa dolnośląskiego. Zastosowano taksonomiczne metody klasyfikacji i grupowania powiatów, jako obiektów przestrzennych badanych w wielowymiarowej przestrzeni cech. Celem referatu jest przedstawienie przestrzennego zróżnicowania poziomu rozwoju gospodarczego powiatów dolnośląskich w świetle wybranych czynników.

10.2478/v10103-012-0023-9

BARBARA DAŃSKA-BORSIAK[*], IWONA LASKOWSKA[**]

The Determinants of Total Factor Productivity in Polish Subregions. Panel Data Analysis

Abstract

The significant role of TFP in stimulating the long-run economic development induces researchers to seek for the sources of the TFP growth.

The mail goals of the paper are: to estimate the level of TFP in the years 2003-2009 at the level of subregions, and to define the factors which determine this estimated TFP level. The first hypothesis being verified is, that the role of the quality of human capital in stimulating long-run economic growthis crucial and can be measured by the model. The second hypothesis is, that there are some factors affecting the TFP level which are common in all subregions.

1. Introduction

Division into subregions has been introduced in the research conducted by GUS [*Polish Central Statistical Office*] in connection with the necessity to adjust the Polish economy to the requirements of the European Union law regarding regional statistics. On the basis of the *Nomenclature of Territorial Units for Statistics* (*NUTS*), legally binding in the EU countries, in the year 2000 the Polish statistics was altered to include the Nomenclature of Territorial Units for Statistical Purposes [*pl. Nomenklatura Jednostek Terytorialnych do Celów Statystycznych*] (NTS) and a five-level hierarchical grouping of data (country, provinces – regions, subregions – groups of districts, districts, communes). The NUTS system is based on the binding administrative division of the member

[*] Ph.D., University of Łódź

[**] Ph.D., University of Łódź

countries. In exceptional cases, in order to improve comparability with regional levels within the EU, for statistical purposes new units are created which have no equivalents in the territorial division. In case of Poland, this pertains to subregions (NTS 3).

The economic and spatial analysis of subregions, performed using the synthetic indicators developed by GUS (gross domestic product - GDP and gross value added - GVA), characterizes the diversification of the level of development and the sector structure of the economy better than in case of the arrangement of 16 provinces.

The presented article sets two fundamental aims. The first of these is to estimate the level of TFP in the years 2003-2009 at the level of subregions, whereas the second one is to define the factors which determine the level of TFP in Polish subregions in the considered period, with particular attention paid to the role of human capital.

2. Methodology applied to measure TFP

In order to estimate the value of TFP we have taken advantage of the method applied in Tokarski's paper (2008), which consists in determining the estimation of parameter α on the basis of a two/double/bi-factor function by Cobb-Douglas: $Y = \alpha_0 K^{\alpha} L^{1-\alpha} e^{\varepsilon_t}$. This function is converted into an efficiency model in the form of:

$$\ln\left(\frac{Y_{it}}{L_{it}}\right) = \ln A + gt + \alpha \ln\left(\frac{K_{it}}{L_{it}}\right) + \varepsilon_{it} \tag{1}$$

where Y – gross value added in million of PLN, L – amount of labour (expressed in thousands of working people), K – gross outlays on property, plant and equipment in million of PLN, $\alpha_0 = Ae^{gt} > 0$ – total productivity of production factors (TFP) t – a time variable, g – rate of technological progress in the sense of Hicks, α – flexibility Y in relation to capital$K$.

After estimating model (6) the TFP values specific to individual subregions and years were calculated according to the following formula:

$$TFP_{it} = \frac{(Y_{it}/L_{it})}{(K_{it}/L_{it})^a} \tag{2}$$

where a is the estimate of parameter α of model (6).

In order to estimate the efficiency model and models which describe the shaping of TFP, alternative specifications were applied in case of panel data: fixed effects (FEM), random effects (REM), Swamy random-coefficients, Hausman-Taylor models. It appears that the first two constructions are commonly known, and therefore below we provide a brief description of the idea behind the last two approaches.

Random-coefficients models are more general than fixed- and random-effects models in that they allow each panel to have its own vector of slopes randomly drawn from a distribution common to all panels. A random-coefficients model has the form (Swamy, 1970)

$$y_{it} = \boldsymbol{x}_{it}^{T}\boldsymbol{\beta}_i + \varepsilon_{it} \tag{3}$$

where $\boldsymbol{\beta}_i$ is a k×1 vector of parameters specific to group i. The error term vector $\boldsymbol{\varepsilon}_i=[\varepsilon_{it}]_{Tx1}$ is distributed with mean zero and variance $\sigma_{ii}^2\mathbf{I}$. Each group-specific $\boldsymbol{\beta}_i$ is related to an underlying common parameter vector $\boldsymbol{\beta}$: $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{v}_i$ where $E[\mathbf{v}_i] = 0$, $E[\mathbf{v}_i\mathbf{v}_i^T] = \Sigma$, $E[\mathbf{v}_i\mathbf{v}_j^T] = 0$ for $j \neq i$, and $E[\mathbf{v}_i\boldsymbol{\varepsilon}_j^T] = 0$ for all i and j. The estimate of $\hat{\boldsymbol{\beta}}$ is a weighted average of the panel-specific OLS estimates of $\hat{\boldsymbol{\beta}}_i$.

Hausman and Taylor (1981) developed a method of estimation of models which contain variables that explain both the constants and the variables in time. In addition, irrespective of the above diversification, the method permits part of the variables not to be correlated, and part – to be correlated with group effects α_i. This means a combination of the assumptions of FEM and REM regarding the correlations between the group effects and explanatory variables. The estimation method is based on the Method of Instrumental Variables (IV).

3. Estimating the value of TFP according to subregions

Among all the 66 subregions there are eight specific ones, being large urban agglomerations. They have been isolated before commencing the analysis.They include: Warszawa, Poznań, Kraków, Wrocław, Trójmiasto, Łódź, Szczecin, Katowice. Model (6) has been estimated separately for the abovementioned urban subregions and others.

Table 1 presents the results obtained using the most optimal method for each subgroup, from the point of view of the subject matter and the statistical quality of the model. For urban subregions this involved the specification random effects model, while for the group of the remaining subregions – Swamy random-coefficients model.

Table 1. Results of the estimation of the model (1)

	Subgroup	
	Urban subregions	Remaining subregions
	Estimation method	
	Random effects model (REM)	Swamy random-coefficientsmodel
$\ln(K_{it} / L_{it})$	0.3489 [0.000]	0.1784 [0,039]
t	0.0252 [0.000]	0.0275 [0.000]
ln$A$	-1.6683 [0.000]	-2.0186 [0.000]
R^2	within = 0.8136 between = 0.8789 overall = 0.8259	-
Test of parameter constancy	-	chi2(171) = 11862.24 Prob> chi2 = 0.0000
Number of observations	56	406

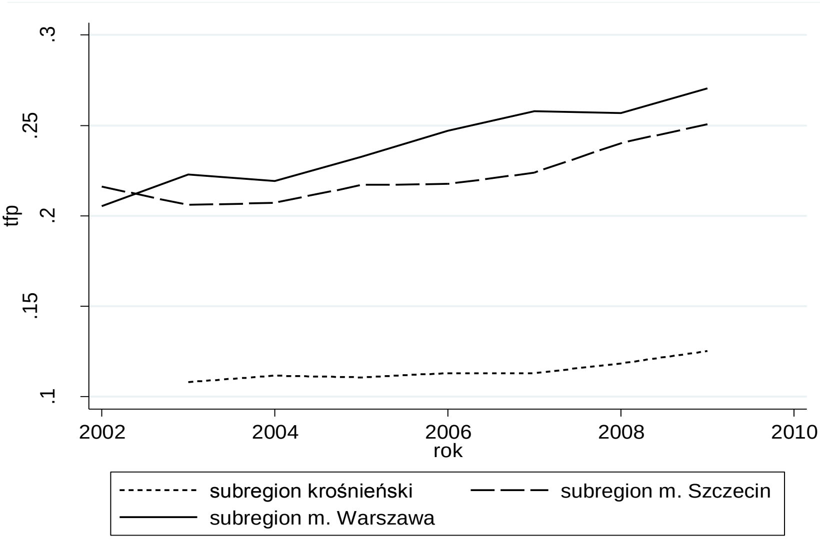
Note: In square brackets *p-values* have been given.

Source: author’s calculations.

The results of the estimation of the efficiency model appear to be satisfactory. The rate of technological progress in the understanding of Hicks, annually amounts to: 2.5% for urban subregions and 2.7% for the remaining ones. The flexibility of productivity in relation to the technical armament of work is significantly higher in case of urban subregions (0.35) than in case of the remaining ones (0.18), which may be due to a better utilization of capital in these subregions.

The next stage of the analysis involved calculating the total productivity of production factors for subregions. To this end, we have applied formula (7) with the value a equal to 0.3489 or 0.784, for “urban” subregions and the remaining ones respectively (see Table 1). The obtained values differ in a significant way. Overall, we have found that the TFP values for “urban” subregions are higher than for the remaining ones. A graphical representation of TFP values for all the subregions would render the graph illegible. Therefore, Graph 1 presents, as an illustration of the rate of TFP diversification, the shaping of this variable in subregions of extreme values.

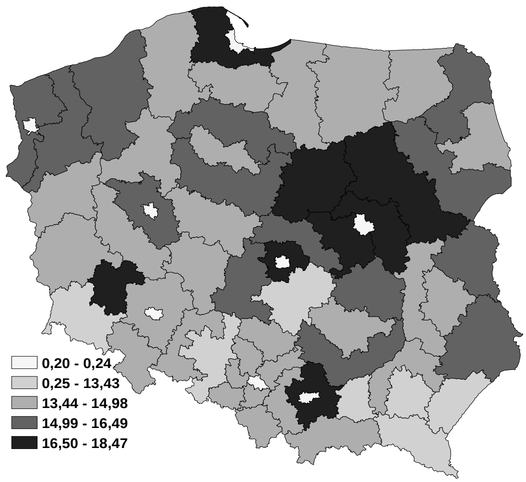
Graph 1. A comparison of the extreme values of TFP in subregions



Source author's calculations.

On the basis of graph 1 one can find a considerable diversification of TFP values between subregions. For this reason we have decided to perform the analysis of the determinants of this variable in smaller groups of subregions. Apart from the group of "urban" subregions, isolated at the previous stage of research, from among the 58 remaining subregions we have isolated four subgroups. The grouping criterion involved the average value of TFP for each subregion, calculated for the entire sample period: $\overline{TFP}_i = (\sum_t TFP_{it})/T$. In the first stage, on the basis of the obtained values we have calculated the national average: $\overline{TFP} = (\sum \overline{TFP}_i)/N$. The subregions have been divided into those in which $TFP_{it} > \overline{TFP}$ and those in which $TFP_{it} < \overline{TFP}$. For each of the two groups obtained in such a manner, procedure from stage one has been repeated, ultimately yielding four groups of regions (the fifth group includes "urban" subregions).

Graph 2. TFP level in subregions (TFP*100)



Source author's calculations.

4. Potential determinants of TFP in Polish subregions

Greater and greater interest of theoreticians and practitioners is aroused by the role of human resources in developing competitiveness of economies. Human capital is considered to be an important factor in regional and local development (Herbst 2009).

As regards the influence of human capital on the growth of the economy, two approaches are noticeable (Aghion, Howitt 1992). In the first of these, human capital is defined as an argument of the function of production. In the second approach, it is treated as a factor which has influence on developing innovations and assimilating new technologies – a factor which is indispensable for technological development. In accordance with the last of the approaches, human capital affects the growth of the economy in an indirect manner - by means of the total productivity of production factors.

The set of indicators which characterize the quality of human capital is extremely vast. Since they comprise the level of education, skills, health and migration opportunities (Herbst 2009; Kunasz 2010). Not all information is available at the level of NUTS3. The choice of variables has been dictated by the possibility to obtain data which ensure comparability with respect to space and time.

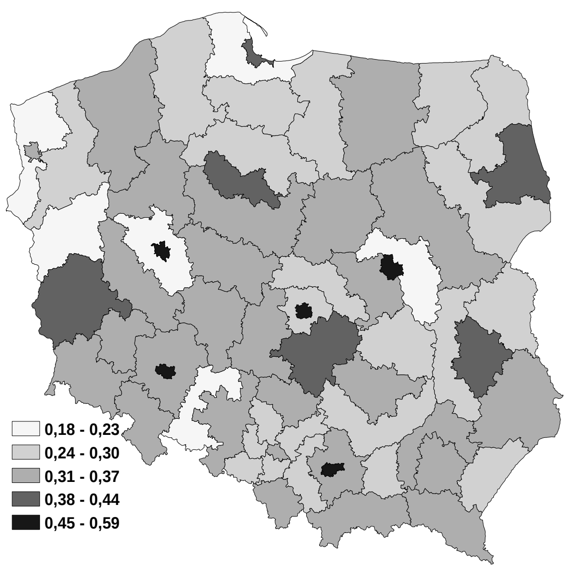
Therefore, the following indicators were selected to account for various human capital aspects:

- computer with Internet access per 10 thousand people,
- number of students per 10 thousand inhabitants,
- number of graduates of schools of higher education per 10 thousand people,
- gross scholarization coefficient in case of postsecondary schools (age: 19-21 years old),
- owing to the lack of information about the state of health (e.g. life expectancy) at the level of subregions, as an indicator of the state of health, we have used outlays on health measured in the number of consultations with physicians per 10 thousand inhabitants.

On the basis of selected indicators (diagnostic variables) we have created the synthetic variable *Z*. The aggregate variable Z_i^t for object *i* at time *t*is an unweighted sum of individual diagnostic characteristics after normalization (Panek 2009). The higher value of the variable Z_i^t, the better human capital level.

The average level of human capital (average value of the synthetic variable Z_i^t) in Polish subregions in the years 2003-2009 is illustrated by the map.

Graph 3. The average values of human capital measure in Polish subregions in the years 2003-2009



Source: Author's calculations.

Subregions with the highest level of human capital in the vast majority include urban subregions. On average, in the considered period the subregions with the highest level of human capital included: the city of Warszawa, city of Poznań, city of Łódź, city of Kraków. In the last year of the period, it was the most favourable in the city of Warszawa, city of Poznań, city of Kraków, city of Wrocław, city of Łódź. Whereas, the lowest average value of human capital in the considered period is observed in the following subregions: elbląskie, oświęcimskie, radomskie, skierniewickie.

Apart from the assessment of the role of human capital, one of the targets of the analysis involved examining the influence exerted on the shaping of TFP through research and development activity. The only available variable which measures the level of this last factor includes outlays on research and development (R&D). However, GUS does not provide its value at the level of subregions, but merely at the level of provinces. Despite this, we have made an attempt to include this indicator in the model. We have constructed an interactive variable, being the product of the estimated value of human capital in a subregion and the value of outlays on research and development in the province, to which the given subregion belongs. It appears that including a variable constructed in such a way in the model allows us to take into consideration the diversification of the effects of the R&D activity between subregions, in which the possibilities of their absorption are varied due to the unequal level of human capital.

Among the factors which can determine TFP investments are also taken into consideration. Such a variable has also been used in the presented study.

Research methodology and obtained results

This analysis uses a static panel model

$$\ln TFP_{it} = \mathbf{x}_{it}^{T}\boldsymbol{\beta} + \sum_{j=1}^{15} \alpha_j d_j + \varepsilon_{it} \tag{4}$$

where *i*is an object indicator and *t* is a time indicator,

$\boldsymbol{x}_{it} = [x_{kit}]_{K\mathrm{x}1}$ – a K-coordinate vector of explanatory variables,

$\boldsymbol{\beta}$ – a vector of parameters (*K*x1), identical for all *i*and *t*

d_j – a dummy variable indicating a voivodeship containing the subregion

The reason for including binary variables in the model is to examine whether the fact of belonging to a given province, and what follows from this, the influence of the directly superior economy, has influence on the level of TFP in the subregion.

In the analysis we have applied alternative specifications for panel data: fixed effects (FEM), random effects (REM), Hausman-Taylor models (HT). The tables below present the results yielded by the models with the best statistical properties and correct from the standpoint of economic theory.

Table 1 shows results of the estimation of TFP models for all the subregions along with urban subregions. Table 2 presents the results of estimation of TFP models in the remaining subgroups isolated due to the observed level of TFP.

Table 2. Results provided by the ln*TFP* models for 66 subregions and city-subregions

	66 subregions *RE model*	*City-subregions* *RE model*
	Parameter estimate [p-value]	
investment outlays per capita (in form of logarithm)	0.0649 [0.000]	0.054 [0.003]
interactive variable (in form of logarithm	0.1109 [0.000]	0.0746 [0.002]
dolnośląskie	-0.2107 [0.000]	-
kujawsko-pomorskie	-0.1678 [0.000]	
lubelskie	-0.2299 [0.000]	
małopolskie	-0.2870 [0.000]	-
pomorskie	-0.1377 [0.000]	
śląskie	-0.3069 [0.000]	-
wielkopolskie	-0.2848 [0.000]	-
łódzkie	-0.1872 [0.000]	
mazowieckie	-0.2887 [0.000]	-
podkarpackie	-0.2545 [0.000]	
const	1.2034 [0.000]	1.1905 [0.000]
R-sq within	0.4296	0.4296
R-sq between	0.462	0.462
R-sq overall	0.4561	0.4561
Breusch-Pagan test [*p*-value]	893.809 [0.000]	23.91 [0.000]
number of units	462	56

Note: a) all explanatory variables are given as the logarithms

Source: author's calculations.

Table 3. Selected results provided by the ln*TFP* models in isolated subgroups

	Group I RE model	*Group II HT model*	*Group III RE model*	*Group IV HT model*
investment outlays per capita (in form of logarithm)	0.0516 [0.001]	0.0564 [0.000]	0.0664 [0.000]	0.0534 [0.085]
interactive variable (in form of logarithm	0.1985 [0.000]	0.0961 [0.000]	0.0599 [0.000]	0.2260 [0.000]
kujawsko-pomorskie		0.0596 [0.000]	0.0679 [0.000]	
lubelskie		0.0774 [0.000]	0.0979 [0.000]	
lubuskie		0.2686 [0.000]		
łódzkie			0.0423 [0.000]	
małopolskie				-0.218 [0.000]
podlaskie		0.1889 [0.000]	0.1998 [0.000]	
pomorskie		0.0650 [0.000]		
mazowieckie				-0.476] [0.000]
opolskie	0.5951 [0.000]	0.3329 [0.000]		
podkarpackie	0.2736 [0.000]			
świętokrzyskie		0.2778 [0.000]	0.2417 [0.000]	
warminsko-mazurskie	0.4477 [0.000]	0.2348 [0.000]		
zachodniopomorskie			0.1443 [0.000]	
const	-0.1415 [0.000]	1.1010 [0.000]	1.5428 [0.000]	-0.0928 [0.000]
R-sq within	0.6477		0.413	
R-sq between	0.4483		0.811	
R-sq overall	0.5243		0.486	
Breusch-Pagan test [*p*-value]	55.41 [0.000]		2.17 [0.070]	
number of units	105	161	84	56

Note: a) all explanatory variables are given as the logarithms: group I- the highest level of TFP, group IV – the lowest level of TFP

Source: author's calculations.

In all the considered models the variable which has a significant influence on TFP includes investment outlays (for group IV at the significance level of 0.1). The effect of the influence of investments in the urban subregions and the subregions which belong to the four remaining groups is similar.

The interactive variable which is the product of the level of human capital and the outlays on R&D, is also of significant. It means, that in all the selected subgroups, the influence of the R&D absorption on the TFP level, measured by the parameter estimates, is important (and positive). The strength of this influence is greatly diversified and is the weakest in the subregions of urban nature).

5. Conclusions

The conducted study allows us to conclude that the diversified level of TFP values in the subregions is, to some extent, conditioned by the differences in the human capital. The impact of human capital on the level of TFP is observed both in the subregions with the highest level of this variable, as well as in those in which the level of TFP is relatively low. Thus, the investments in human capital can stimulate the competitiveness of the region.

Inclusion of the interactive variable into the model of human capital additionally allows one to estimate the possibility of using provincial outlays on research and development activities in subregions, depending on the level of human capital.

Investments constitute yet another factor which determines the shaping of TFP in subregions, where the strength of influence is similar for all subregions. The binary variables which determine the provincial membership are partially significant; however, it appears that the influence of the situation in the superior region is more significant in the provinces considered to be slightly worse developed with respect to the economy.

It seems that the results of the presented research can be helpful to the practitioners, especially in the field of social policy (mainly educational and health).

The conducted study allows us to conclude that the level of TFP in subregions is greatly diversified. The highest value of TFP characterizes the subregions of urban nature. In urban subregions we can also observe the highest level of human capital. The applied econometric models confirm the speculation about the positive role of human capital in the shaping of TFP, both in the subregions with the highest level of this variable, as well as in those in which the level of TFP is relatively low. Inclusion of the interactive variable into the model of human capital additionally allows one to estimate the possibility of using provincial outlays on research and development activities in subregions, depending on the level of human capital.

Investments constitute yet another factor which determines the shaping of TFP in subregions, where the strength of influence is similar for all subregions. The binary variables which determine the provincial membership are partially significant; however, it appears that the influence of the situation in the superior region is more significant in the provinces considered to be slightly worse developed with respect to the economy.

References

Aghion P., Howitt P. (1992), *A model of growth through creative destruction*, 'Econometrica', vol. 60(2), pp. 323-351

Hausman J.A., W.E.Taylor (1981), *Panel data and unobservable individual effects*, 'Econometrica', 49(6), pp. 1377–1398

Herbst M. (2009), *Tworzenie i absorpcja kapitału ludzkiego w polskich miastach akademickich w kontekście realizacji narodowej strategii spójności*, http://www.mrr.gov.pl/aktualnosci/fundusze_europejskie_2007_2013/Documents/3_8_M_Herbst_Tworzenie_i_absorpcja_kapitalu_ludzkiego_w_pols.pdf [access: 21.05.2012]

Kunasz, M. (2010), *Indeks gospodarowania kapitałem ludzkim*, raport przygotowany na zlecenie Centrum Badań Socjologicznych, VOLUMINA, Szczecin

Panek T. (2009), *Statystyczne metody wielowymiarowej analizy porównawczej*, Oficyna Wydawnicza SGH, Warszawa

Statistical Data of GUS: http://www.stat.gov.pl

Swamy P.A.V.B. (1970), *Efficient Inference in a Random Coefficient Regression Model*, 'Econometrica', vol. 38, pp. 311-323

Tokarski T. (2008), *Oszacowanie regionalnych funkcji produkcji*, 'Wiadomości Statystyczne', vol. 10, pp. 38-53

Streszczenie

DETERMINANTY ŁĄCZNEJ PRODUKTYWNOŚCI CZYNNIKÓW PRODUKCJI W PODREGIONACH W POLSCE

Znaczącą rolę TFP w stymulowaniu długookresowego rozwoju gospodarczego skłania badaczy do poszukiwania źródeł jej wzrostu.

Główne cele badań prezentowanych w artykule są następujące: oszacowanie wartości TFP w latach 2003-2009 w podregionach, a następnie określenie czynników determinujących łączną produktywność czynników produkcji. Z zastosowaniem modelu ekonometrycznego podjęto próbę weryfikacji hipotezy, iż jakości kapitału ludzkiego odgrywa istotną rolę w stymulowaniu długookresowego wzrostu gospodarczego. Zgodnie z kolejną hipotezą istnieją czynniki wpływające na poziom TFP, wspólne dla wszystkich podregionów.

10.2478/v10103-012-0024-8

EWA K. DŁUBAKOWSKA-PUZIO[*], MICHAŁ KARPUK[**], KAMIL PUZIO[***]

Analysis of Tourism Service Quality in Kołobrzeg Region by Means of Time Series Models

Abstract

The aim of the given paper is to present the analysis of tourism services by means of time series models and forecasting of evaluation of tourism services. Tourism services are analysed according to various parameters: hotel stay price, hotel services quality (such as cleanness of rooms, check-in, information provision etc.), catering quality and medical service quality.

The research has been undertaken on the basis of responses of foreign guests of 13 hotels in Kołobrzeg region taken from 1400 questionnaires divided according to age and sex of respondents. Various econometric models were used for the analysis of statistic regularities. First, customers evaluated quality of their stay in hotels. These data were examined during the 2006-2009 time period as a stochastic process. It was found that the processes are nonstationary, that is why the ARIMA (1,1,1) model was used in the study. On the basis of the analyses and prognoses one can deduce that models of time series make it possible to estimate a tendency that occurs for an analysed parameter, however misprediction is quite possible to appear (up to 30 %). Similar results were achieved during the analysis of evaluation of hotel stay quality on the basis of sex of respondents. One should point out that an average value of male evaluation rate is higher than an average value of female evaluation rate.

[*] Ph.D., University of Technology in Koszalin

[**] Ph.D., University of Technology in Koszalin

[***] Social School of Entrepreneurship and Management in Lodz, Branch Faculty in Kołobrzeg

Time series for analysed variables were integrated into level I(1). A certain co-integrational connection was found between the evaluation of hotel stay and number of stays in a given hotel, where tourists that have already visited a given hotel for several times presumably give a high evaluation rate of hotel stay quality. The result of the analysis of long-term relation between hotel stay and duration of check-in is presented in high mutual dependence of hotel evaluation rate on evaluation of check-in duration. A hotel stay price influences hotel stay evaluation in a negative way, but catering and medical services are considered as standard and do not have any particular influence on hotel stay evaluation.

The undertaken study shows that methods that take into account time series can be successfully used in analysis of parameters of tourist comfort and in evaluation of hotel services.

1. Introduction

Development of such economic branch as tourism is directly connected with development of hotel market and increase of quality of customer services. Hotel and tourism businesses must be closely related to a client and offer more services in order to achieve better financial results. Enterprise activity of tourism businesses occurs under pressure of strong competition and gets worse because of the world financial crisis (Woźniak 2009, pp. 127-138). Therefore broadening or provision of new services is of great importance for activity of tourism businesses.

Moreover, analysis of customers' opinions on quality of provided services and their feedback on relations between buyers and producers of tourism services has very significant positive influence on results of financial activity of tourism companies.

The given paper presents the analysis of tourism services according to various parameters: hotel stay price, hotel services quality (such as cleanness of rooms, check-in, information provision etc.), catering quality and medical service quality and quantity (Dłubakowska-Puzio et al. 2009, pp. 181-194). A questionnaire consisted of 16 questions which were grouped according to the following segments: hotel services quality, catering quality, medical services quality, and evaluation of hotel stay. Some of these questions are presented in the Table 1.

Table 1. Several questions from the questionnaire

No.	Question	Scale of evaluation rates (min-max)
X0a	Is it your first time at the hotel?	0 – 1
X1	Check-in duration.	0 – 5
X6	Breakfast offer.	0 – 5
X12a	Range of treatments: number of treatments offered.	0 – 5
X14a	Medical care: nurse attitude towards guests.	0 – 5
X15	Guests guide care	0 – 5
X16	How will you evaluate the given hotel during your conversation with friends and family?	0 – 5

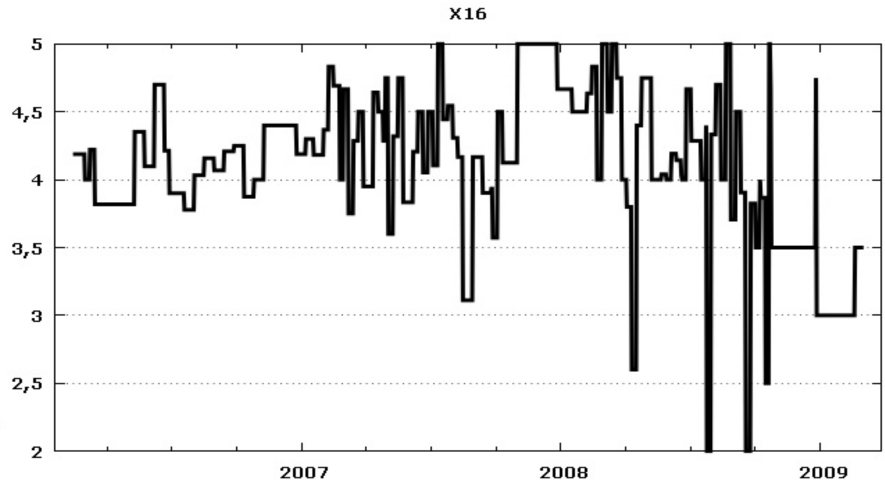
Source: own data.

Moreover, monitoring of hotel stay prices has been launched. Average price depends on time and can be described with dynamic variable SrCena.

The research has been undertaken in time period 2006-2009 on the basis of responses of foreign guests of 13 hotels in Kołobrzeg region taken from 1400 questionnaires divided according to age and sex of respondents.

Traditionally, statistical research uses methods of analysis of means and dispersions, and detection of correlational dependences between various data, on the basis of which conclusions about significant dependences between different phenomena are made (Maddala 2006, pp. 577-607). However, this kind of research does not take into account dynamism of concerned phenomena, their changes in time, which causes substantial problems with forecasting. For example, Figure 1 presents a time series of average evaluation rates of hotel service quality given by responding tourists during the time period 2006-2009. Similar time dependences are typical also for average evaluation rates made by tourists responding to other kinds of questions.

Figure 1. Average evaluation rate of hotel stay quality (question X16)



Source: own data.

Phenomena taking place in the tourism branch, evaluated by tourists in questionnaires, are considered as stochastic processes during the 2006-2009 time period. Therefore, the given paper uses methods of time series research. Hotel quality evaluation is correlated with a range of factors, such as hotel room prices, quality of services provided at the reception desk, catering quality. It is necessary to admit that one of special features of tourism in Kołobrzeg region is its health-improving character and provision of spa and rehabilitation services of high quality for tourists.

The given paper had several goals, which are the following:

1. Research of evaluation data of tourism services quality in Kołobrzeg region in the 2006-2009 time period as a stochastic process and detection of its regularities and features;
2. Forecasting of evaluation of tourism services quality on the basis of time-series methods;
3. Research of influence of other features (hotel room prices, quality of medical services, catering quality etc.) on evaluation of service quality by tourists;
4. Research of influence of sex of tourists on evaluation of hotel stay quality by tourists.

2. Statistical methods used in the article

ARIMA model for forecasting of evaluation of hotel stay quality

In econometric studies, ARIMA models (Autoregressive Integrated Moving Average), introduced by Box and Jenkins (Box, Jenkins, 1976) are usually used for time series analysis.

General view of ARIMA model (p,d,q) is the following:

$$x_t^d = \varphi_1 x_{t-1}^d + \varphi_2 x_{t-2}^d + ... + \varphi_p x_{t-p}^d + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \Theta_2 \varepsilon_{t-2} - ... - \Theta_q \varepsilon_{t-q} \qquad (1)$$

where x_t is the analysed time series (the given study analyses time series on the basis of answers to the question).

$$x_t^d = \Delta^d x_t = x_t - C_d^1 x_{t-1} + C_d^2 x_{t-2} - ... + (-1)^d C_d^d x_{t-d} \qquad (2)$$

ε_t – rest of the model for a time period t

$\varphi_1, \varphi_2, ..., \varphi_p, \Theta_1, \Theta_2, ..., \Theta_q$ - parameters of the model

a series $x_t^d, t = 1,...,T-d$, extracted from x_t, after d-fold method of successive differences was applied, can be described by ARMA model (p,q).

In ARIMA model (p,d,q) three parameters can be defined: autoregressive parameter p, time-series differentiation d, moving average parameter q. An input time-series for ARIMA method should be stationary – it must have a time-constant mean, dispersion and it must lack autocorrelation. Therefore a time-series usually requires time-series differentiation d until stationarity is achieved.

Methodology of ARIMA model modelling for examined time-series consists of the following fundamental stages:

- time-series stationarity test. In case of time-series nonstationarity, time-series differentiation d is necessary;
- identification of exploratory model;
- model parameters evaluations and diagnostic verification of model adequacy;
- use of model for forecasting.

First, we should acquire a stationary series. In order to reach this result, we must analyse an autocorrelation function (ACF) and a partial autocorrelation function (PACF) (see Figure 2. Horizontal axis *lag* presents delay, vertical axis presents values of (auto)correlation variables). Quick decay of ACF values is a simple stationarity test. Moreover, at the given stage statistical tests of the unit root (Dickey-Fuller test or Augmented Dickey-Fuller test) (Dickey, Fuller 1979, pp. 427-431) are used.

If a series is nonstationary according to Dickey-Fuller statistic or rates, then, in order to achieve a stationary series, one should apply the successive differences operator which defines a value of parameter d (time-series differentiation). Thus, we become familiar with a value of one parameter in ARIMA(p,d,q).

After the achievement of a stationary series, behaviour pattern of selective ACF and PACF is investigated and hypotheses about values of parameters *p* (autoregressive scheme) and *q* (moving average scheme) are put forward. In this case several models with different *p* and *q* are used.

It is necessary to evaluate parameters after identification of the model. To this effect maximum likelihood estimation method (MLE) is used.

For every analysed model verification the number of residues is analysed. Number of residues of an adequate model should resemble a "white noise", which means that their PACF should not differ from zero. Moreover, for verification of hypothesis that observable data are a realization of the "white noise" the Q-statistic of Ljung and Box (Ljung, Box, 1978, pp. 297-303) is used. The statistic has an asymptotic distribution χ^2 with degrees of freedom *K*-*p*-*q*, where *K* is a maximum lag during the model examination.

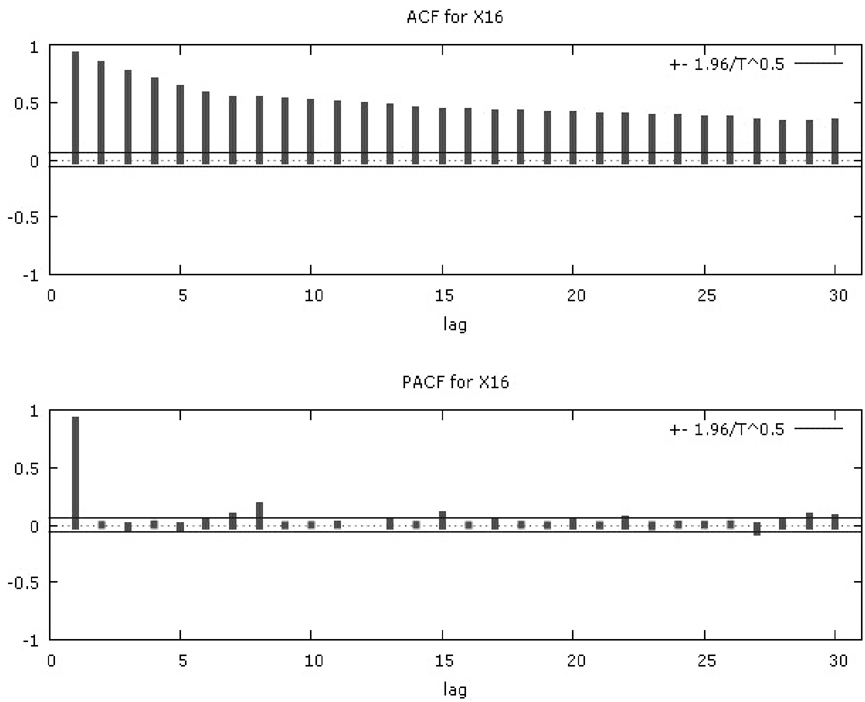
$$Q(K) = T(T+2)\sum_{i=1}^{K}\frac{r_i^2}{T-i} \tag{3}$$

where r_i is an evaluation of variable correlation coefficient x_t, x_{t-i}

In order to compare variants of models and to choose the best model we can use Akaike information criterion (Akaike 1974, pp. 716-723).

For the analysis we have used *gretl* free software (http://gretl.sourceforge.net/), which is actively used for the analysis of time series.

Figure 2. ACF and PACF function for the variable X16



Source: own data.

ACF and PACF functions for the variable X16 allow us to conclude that the variable X16 is a nonstationary variable, which means that an average value of stay quality evaluation in hotels in Kołobrzeg region changes with time under the influence of hotel clients' preferences. Therefore, human factor plays an important role in stay quality.

Similar conclusion can be made during the analysis of tourists' responses on the following questions: X1 Check-in duration., X6 Breakfast offer, X12 Range of treatments, X14 Medical care, X15 Guests guide care, and also an average price SrCena.

In order to create a stationary model for the X16 variable there were various differentiation levels of the X16 variable analysed. It was found out that after an application of ADF test (and additional KPSS test) (Kwiatkowski et al. 1992, pp. 159–178) we can achieve stationarity of time series for series differences d=1.

Analysis of correlograms of ARIMA model (p,1,q) for different values of p, q, and use of Ljung-Box and Akaike tests brings us to conclusion that the best model is the ARIMA model (1,1,1) where p=1 and q=1. Model can be presented as follows

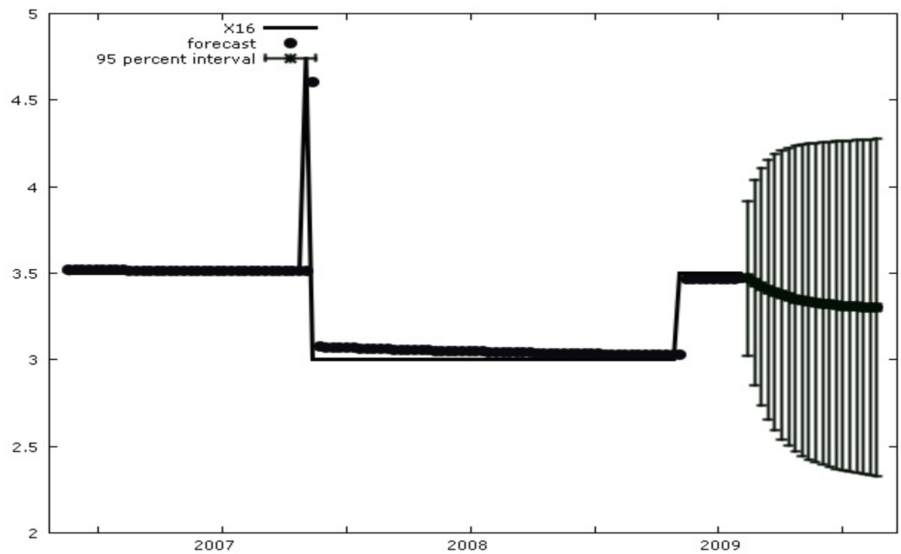
$$x_t^1 = 0.857560 x_{t-1}^1 - 0.980884 \varepsilon_{t-1} \tag{4}$$

where $x_t^1 = \Delta x_t = x_t - x_{t-1}$

Mean absolute percentage error (MAPE) makes 2.22%, coefficient of determination (R-squared) R^2=0,982 shows that the created model (4) explains 98,2% of changeability of the analysed X16 variable.

Similar results (optimal ARIMA model (1,1,1)), high values of R-squared and low values of MAPE) were received for other variables presented in Table 1.

Figure 3. Results of forecast by ARIMA model (1,1,1) for the X16 variable for 20 series



Source: own data.

It is necessary to admit that the ARIMA methodology was also used for analysis of X1, X6, X12, X14, X15 variables depending on sex. As a result, it specified that coefficients of male and female models differ for maximum 10%, even though average male evaluation value for the whole period of analysis is higher than average female evaluation value of the above-mentioned variables. It allows us to conclude that male and female models are very similar, which can be interpreted as a tiny difference in hotel stay evaluation and hotel service evaluation.

Created ARIMA models ARIMA(1,1,1) were used to forecast average evaluation values for the above-mentioned variables. Figure 3 presents forecast results for 20 days for the X16 variable on the basis of the model (4). Results of modelling show that the average value for the X16 variable decreases from 3,5 to 3,3. Relative forecast error (APE) increases from 12 % to 30%.

Similar results are achieved for other variables.

The given study shows that use of ARIMA method allows to forecast data received from tourists. The data is described by nonstationary time series with a small amount of error, which is impossible while using other methods.

3. Cointegration of processes in the tourism branch.

Analysis of economic empirical data often faces problems of nonstationary series presence and/or series with trends. As it was mentioned above, time series that are used to process data received from tourists are also nonstationary. In this case it is necessary to use differences in order to achieve stationarity and analyse the results using Box-Jenkins method.

Moreover, establishing dependences between a few nonstationary processes is lso of great importance.

In a regression model

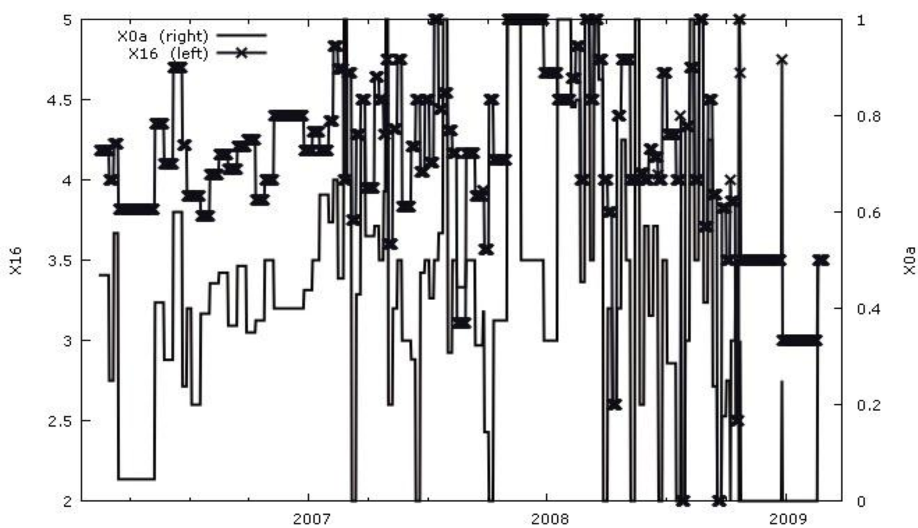
$$y_t = b \cdot x_t + \varepsilon_t \tag{5}$$

we assume that white noise ε_t is stationary. The assumption might not be right if *y* and *x* were integrated. Generally, if two processes are integrated into two different degrees, their linear combination's integration scheme will be equal to the biggest degree. Therefore, if *x* and *y* are integrated into the degree I(1), and both series include trend, then linear combination x and y will be integrate into the degree I(1) regardless of value b. If both *y* and *x* have their own trend of a similar sign, then, in case that there is no dependence between the trends, their difference will also have a trend. On the other hand, if both series are integrated into the degree I(k), then difference between them should vary around the fixed value. It means that series are characterised by similar growth rates. Two series that meet this demand are called cointegrative, and vector [1,-b] is called a cointegration vector. In this case we can notice a difference between a long-term dependence between *y* and *x*, and a short-term dynamism of deviations of *y* and *x* from their long-term tendencies (Charemza, Deadman, 1997, pp.103-144).

Notion of cointegration allows to analyse long-term relations in case of nonstationary variables. This can be connected with a notion of long-term equilibrium with the help of Granger theorem (Engle, Granger 1987, pp.251-276). Granger theorem bounds the notion of cointegration with the error correction mechanism (ECM). Error connection mechanism describes a way mentioned variable is adjusted to a long-term relation. Granger theorem allow us to interpret cointegration vector a a long-term relation between variables.

Gretl programme was also used for the cointegration analysis. Time series for analysed variables had an integration degree I(1).

Figure 4. Integrated time series X16 and X0a



Source: own data.

There was a significant cointegrational relation between variables X16 *How will you evaluate the given hotel during your conversation with friends and family?* and X0a *Is it your first time at the hotel?* (Figure 4) established, which can be interpreted as follows: tourists that have already visited the hotel (X0a→1) generally give higher evaluation rates of the hotel stay quality (X16 → 5).

Table 2 presents a very interesting result of the analysis of long-term hotel stay dependence (X16 variable) on the X1 variable (Check-in duration). Determination coefficient makes 0.990, which shows high interrelation between stay evaluation and check-in duration evaluation.

Table 2. Engle–Granger test results for the variables X16 and X1

Cointegrating regression - OLS, using observations 2006/02/16-2009/02/28 (T = 1109)
Dependent variable: X16

	coefficient	std. error	t-ratio	p-value	
X1	0.955752	0.00286437	333.7	0.0000	***

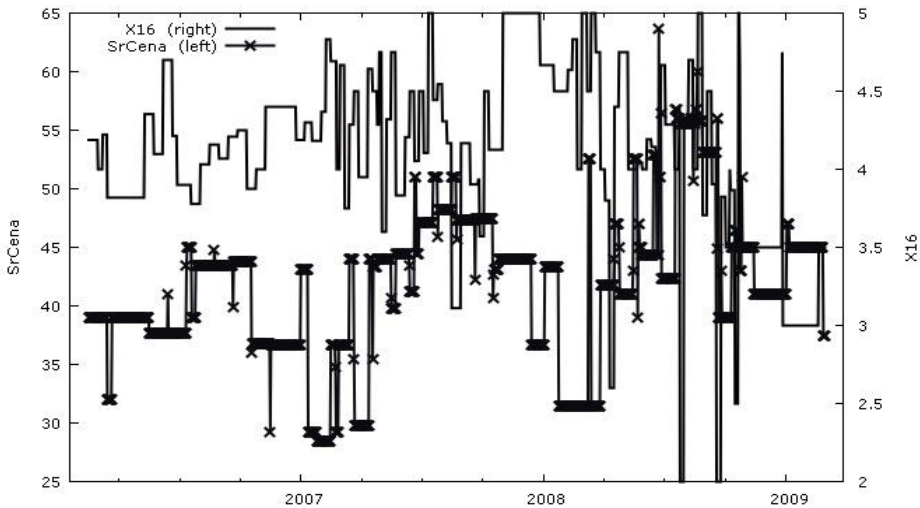
Mean dependent var	4.102426	S.D. dependent var	0.563983
Sum squared resid	187.3889	S.E. of regression	0.411246
R-squared	0.990146	Adjusted R-squared	0.990146
Log-likelihood	-587.6865	Akaike criterion	1177.373
Schwarz criterion	1182.384	Hannan-Quinn	1179.268
Rho	0.888808	Durbin-Watson	0.225046

Source: own research.

Figure 5 presents time series of the X16 and SrCena variables. During the analysis of nonstationary time series for the X16 and SrCena variables there was a very significant long-term negative dependence $X16_t = 4{,}906 - 0{,}019 \cdot SrCena_t + \varepsilon_t$ determined.

The dependence can be interpreted as an decrease of hotel stay evaluation by 0,02 along with an increase of price by 1 euro. Increase of price will always be connected with dissatisfaction with hotel stay and smaller stay evaluation value.

Figure 5. Integrated time series X16 and SrCena



Source: own data.

4. Conclusions

The undertaken analysis of data received from hotel guests concludes the nonstationarity of time series, which describe indicators from tourism branch. For real analysis and data forecast it is necessary to use methods of analysis of nonstationary time series (Box-Jenkins methodology). Moreover, it is required to pay attention to cointegration of processes which describe data received from tourists.

During our research we were interested in influence of various factors on the evaluation of hotel stay quality by foreign guests. It was found out that check-in duration at the reception desk and hotel stay price is of great importance. Spa and rehabilitation services and catering services have less influence. Therefore we can deduce that services that are considered as standard services by foreign tourists should be offered on a high level. Hotel stay evaluation made by males and females does not have any significant difference, but males' evaluation rates are higher to a little degree, which can be interpreted as smaller demands of males concerning standards of stay in a hotel.

References

Akaike, H. (1974), *A New Look at the Statistical Model Identification*, 'I.E.E.E. Transactions on Automatic Control', 19

Box, G.E.P., Jenkins G.M. (1976), *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco

Charemza W.W., Deadman D.F. (1997), *Nowa ekonometria [New econometrics]*, PWE, Warszawa

Dickey D.A., Fuller W.A. (1979), *Distribution of the Estimators for Autoregressive Time Series with a Unit Root*, 'Journal of the American Statistical Association', 74

Dłubakowska–Puzio E.K., Karpuk M., Puzio K. (2009), *Przyszłość uzdrowiska Kołobrzeg w obliczu ogólnoświatowych zjawisk kryzysowych [The future of Kołobrzeg resort in the face of global crisis phenomena]*, [in:] M.Malicki, S.Bilan (ed.) *Polityka gospodarcza, społeczna i ekologiczna w dobie kryzysu światowego [The economic, social and ecological policy in times of global crisis]*, Wydawnictwo Politechniki Szczecińskiej, Szczecin-Kijów

Engle R. F., Granger C. W. J. (1987), *Co-integration and error correction: Representation, estimation, and testing*, 'Econometrica', 55

http://gretl.sourceforge.net/

Kwiatkowski D., Phillips P. C. B., Schmidt P., Shin Y. (1992), *Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root*, 'Journal of Econometrics', 54

Ljung G.M., Box G.E.P. (1978), *On a Measure of Lack of Fit in Time Series Models*, 'Biometrika', 65

Maddala G.S. (2006), *Ekonometria [Econometrics]*, PWN, Warszawa

Woźniak R. (2009), *Pozycja konkurencyjna Kołobrzegu w świetle najnowszych badań empirycznych [The competitive position of Kołobrzeg in the light of recent empirical research]*, [in:] J. Chotkowski (ed.) *Marketing w rozwoju turystyki [Marketing in tourism development]*, Wydawnictwo Politechniki Koszalińskiej, Koszalin

Streszczenie

ANALIZA JAKOŚCI USŁUG TURYSTYCZNYCH W REGIONIE KOŁOBRZESKIM Z WYKORZYSTANIEM MODELI SZEREGÓW CZASOWYCH

Celem niniejszej pracy jest analiza jakości świadczonych usług w zależności od szeregu czynników: ceny pobytu w hotelu, jakości usług hotelarskich (czystość w pokoju, rejestracja pobytu, informowanie klientów itp.), jakości usług gastronomicznych (oferty gastronomiczne), ilości i jakości usług sanatoryjno-rehabilitacyjnych z wykorzystaniem metod szeregów czasowych oraz prognozowania oceny jakości świadczonych usług.

Badania były przeprowadzone na podstawie ankietowania gości zagranicznych z 13 hoteli regionu kołobrzeskiego, razem powyżej 1400 ankiet w zależności od wieku i płci ankietowanych. Dla ustalenia prawidłowości statystycznych zastosowano różne modele ekonometryczne. Po pierwsze, ocenianie jakości pobytu w hotelach przez klientów było rozpatrywane w przestrzeni czasowej 2006-2009 jako proces stochastyczny. Ustalono, że procesy są niestacjonarne, dlatego badanie bylo przeprowadzone na podstawie modelu ARIMA(1,1,1). Badania i prognozy pozwalają zrobić wniosek, że modele szeregów czasowych mogą oszacować trend, występujący dla badanego wskaźnika, natomiast błąd prognozowania jest dość wysoki (do 30%). Podobne wyniki były otrzymane rownież podczas badania ocen jakości pobytu w hotelach w zależności od płci badanych. Zaznaczymy, że średnia wartość ocen mężczyzn jest wyższa niż średnia ocen kobiet.

Szeregi czasowe dla badanych zmiennych miały stopien zintegrowania I(1). Została ustalona znaczna więź integracyjna między oceną pobytu w hotelu oraz ilością odwiedzin wybranego hotelu, kiedy turyści nie po raz pierwszy odwiedzający hotel przeważnie wystawiają wysokie oceny jakości pobytu w hotelu. Wynikiem badania zależności długookresowej pobytu w hotelu od czasu trwania procedury zameldowania jest bardza wysoka współzależność oceny pobytu od oceny czasu trwania procedury zameldowania. Cena pobytu w hotelu negatywnie wpływa na ocenę pobytu w hotelu, natomiast usługi sanatoryjne i gastronomiczne odbierane są jako standard i nieznacznie wpływają na ocenę pobytu w hotelu.

Przeprowadzone badania wskazują, że metody szeregów czasowych mogą być skutecznie zastosowane w badaniu wskaźników turystycznych i ocenie jakości usług hotelarskich.

10.2478/v10103-012-0025-7

ARTUR GAJDOS[*], EDYTA ŻMURKOW[**]

Skilled Personnel Supply and the Prospects for Regional Innovative Development in Poland

Abstract

The aim of this paper is to analyze the supply of highly qualified personnel in the context of prospects for the future innovative development of voivodships in Poland. Analysis of these problems and the relationship between them will be based on the studies on the educational profile of regions and on the analysis of potential trends and possibilities of creating a highly skilled labor force coming from higher education system, as well as on the research on the innovation level and profile of particular voivodships. Furthermore, analysis of possibilities and the level of knowledge diffusion will be conducted.

1. Introduction

New technologies and innovations are considered to be one of the most important factors in obtaining a competitive advantage, which leads to an economic growth, and thus to the improvement of socio-economic situation (Gaczek 2005, pp. 9-12). Simultaneously, there is a coexistence and mutual dependence noticed between the economic and educational development. Existence and the supply of well educated and highly qualified personnel determines scientific and technological progress, which is the source of deep changes in the economy (Grodzicki 2000, p. 22, 35).

[*] Ph.D., University of Łódź

[**] University of Łódź

The main objective of the paper is to analyze the supply of highly qualified personnel in the context of prospects for the innovative development of voivodships in Poland. The paper covers identification and analysis of the regional education profiles, analysis of potentional trends and possibilities of creating a highly skilled labor force by a regional higher education system and research on the innovation level of voivodships. Furthermore, the analysis of the relationship between the supply of highly qualified personnel and the level of innovation has been conducted. Research was performed for years 2004–2010.

2. Education profiles

The analysis includes nine profiles, which are based on a groups of fields of education according to International Standard Classification of Education (ISCED 97).

Table 1. Education profiles according to International Standard Classification of Education

Education profiles	Groups of the fields of education according to ISCED 97
education	teacher training and education science
humanities science and arts	humanities
	arts
social science	social
	journalism and information
	social welfare
economy and law	law
	economy and administration
health	health
science	biology
	physical science
	mathematics and statistics
	computing
engineering and technology	engineering and engineering trades
	manufacturing and processing
	architecture and building
	environmental protection

services	personal services transport services security services
agriculture	agriculture, forestry and fishery veterinary

Source: own research based on http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm, (20.05.2012).

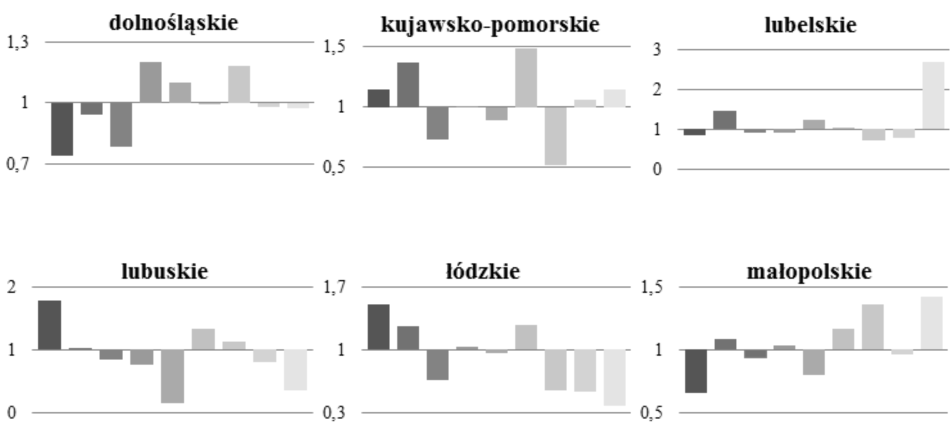
An identification of regional eductaion profiles has been conducted on the basis of location quotients calculated for the number of graduates in various fields of education *(i)* in particular voivodships *(v)*:

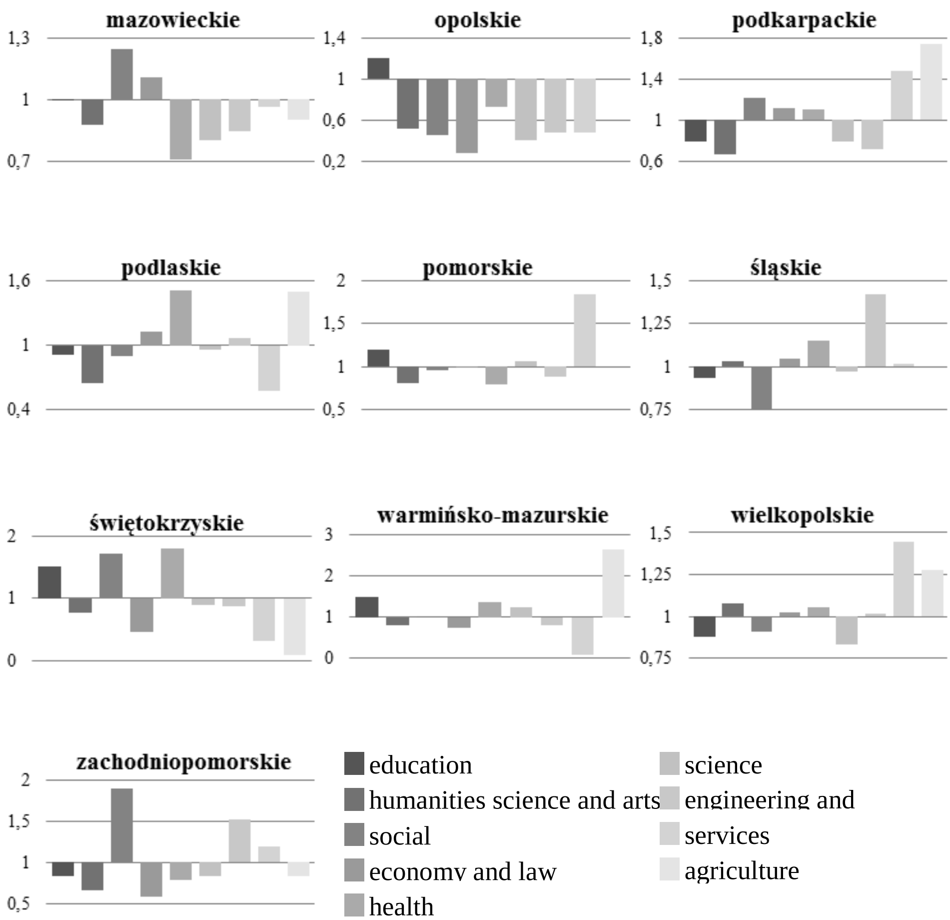
$$LQ_v^i = \frac{\text{Regional share of graduates in } i \text{ field of education}}{\text{National share of graduates in } i \text{ field of education}}$$

Value of location quotient greater than 1 indicates excess of share of the graduates of a particular field of education in the region in relation to the national average, while value less than 1 indicates shortage of this share.

The results presented in figure 1 confirms that the education profiles vary accross the regions.

Figure 1. Education profiles of voivodships in 2010





Source: own calculations using data from the Local Data Bank (http://www.stat. gov.pl/bdl).

What is more, those profiles are not constant over time. In most of regions the dominating profile of education have changed during the six years of analysis, which is marked in table 2 by arrows.

Table 2. Education profiles dominating in voivodships in 2004 and 2010

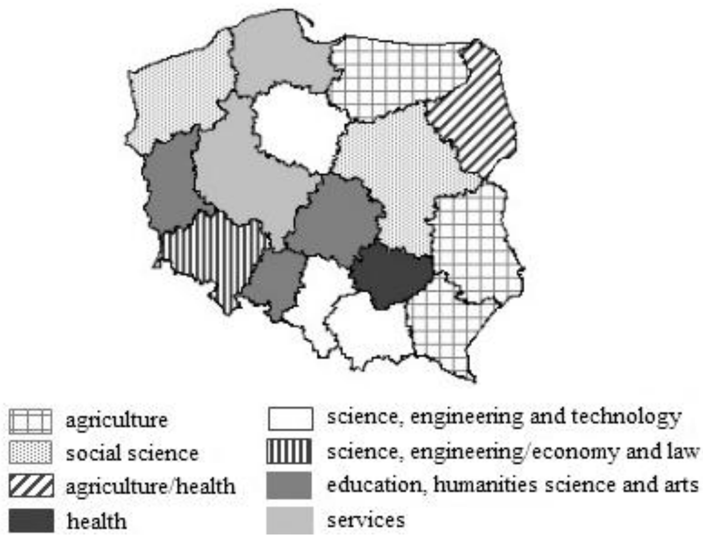
Voivodship	Dominant education profile in 2004		Dominant education profile in 2010
lubuskie	education		education
opolskie	agriculture	→	education
łódzkie	science	→	education
mazowieckie	services	→	social science
zachodniopomorskie	social science		social science
dolnośląskie	economy and law engineering and technology		economy and law engineering and technology
podlaskie	health agriculture		health agriculture
świętokrzyskie	social science	→	health
śląskie	services	→	science
kujawsko-pomorskie	education	→	science
małopolskie	agriculture	→	science
pomorskie	health	→	services
wielkopolskie	agriculture	→	services
lubelskie	agriculture		agriculture
podkarpackie	services	→	agriculture
warmińsko-mazurskie	services	→	agriculture

Source: own calculations.

According to the latest data (year 2010) voivodships are divided into seven main and two combined groups of profiles. A spatial location of those groups (figure 2) indicates that:

- science, engineering and technology profiles are dominating in the South of Poland while education,
- humanities science and arts profiles dominates in the south-west part of the country,
- profile connected with services is dominating in western and northern parts of the country,
- agriculture profile have a relatively highest share in the eastern Poland.

Figure 2. Spatial location of groups of education profiles in 2010



Source: own research.

3. Qualified personnel supply level

The next step of the analysis was determining the regional qualified personnel supply index, which has been done with use of linear arrangement method, where the aggregate measure (QPS_i) is a function of normalized values of input variables:

$$QPS_i = \frac{1}{m}\sum_{j=1}^{m} z_{ij} \text{ and } z_{ij} = \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}}, (i = 1,...,n,\ j = 1,...,m)$$

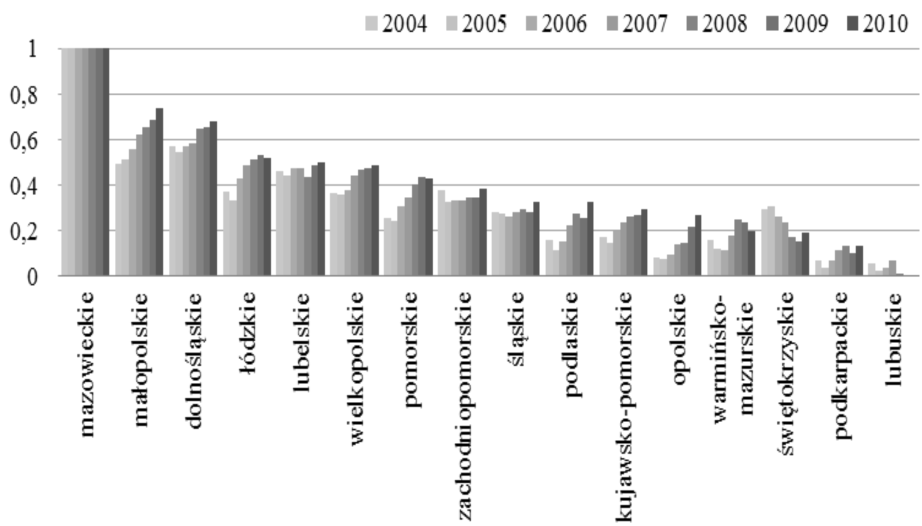
where z_{ij} is a normalized value of variable j in object i (voivodship) (Strahl 2008, p. 34). In order to increase the clarity of the results, index QPS_i have been normalized with the following formula:

$$QPS_i^{'} = \frac{QPS_i - QPS_{\min}}{QPS_{\max} - QPS_{\max}}, (i = 1,...,n)$$

Values of the index range from 0 to 1, where 0 indicates the lowest level of qualified personnel supply and 1 - the highest. Index is composed of four indicators:

- graduates per 1000 population aged 20-29,
- students (ISCED 5) per 10 thousand population aged 19-24,
- doctorate students (ISCED 6) per 1000 population aged 25-34,
- postgraduates per 1000 economically active population of working age.

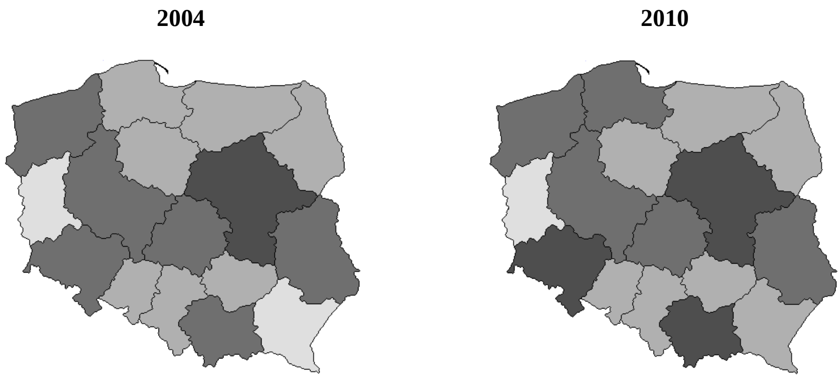
Figure 3. Qualified personnel supply level in voivodships over the years 2004-2010



Source: own calculations using data from the Local Data Bank (http://www.stat. gov.pl/bdl).

Value of the *QPS'*$_i$ index have been increasing over the years 2004-2010 in most of regions and the highest growth occurred in opolskie, podlaskie and podkarpackie. Only in two voivodships – lubuskie and świętkorzyskie – value of index decreased. After dividing regions into four groups characterized by a different level of qualified personnel supply with use of the technique based on the arithmetic mean and standard deviation (Czupich 2009, p. 40) it can be clearly seen, that the level of supply is generally rising (figure 4).

Figure 4. Qualified personnel supply level in voivodships in 2004 and 2010



Qualified personnel supply level:

High Medium-high Average Low

Source: own research.

4. Innovation level

Analysis of the innovation level basis on the innovation index, which has been determined with use of the same methodology as for the qualified personnel supply index. Selection of variables used in aggregate index was based mainly on the list of 25 indicators used in *the Innovation Union Scoreboard (IUS)* methodology (*European Innovation Scoreboard EIS* until 2009) and also on other analyses and publications on the subject of innovation.

Table 3. Indicators used in aggregate innovation index

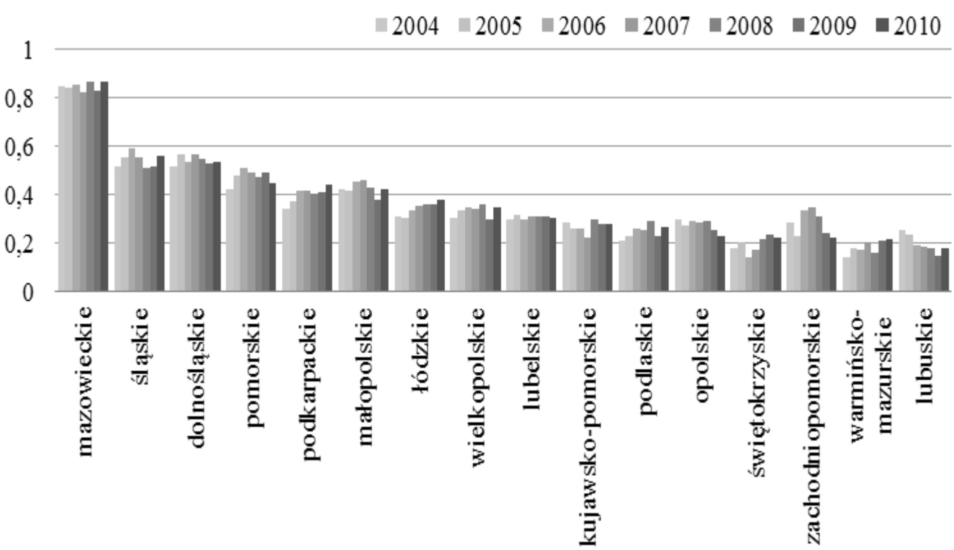
Main types of indicators	Indicators
Enablers	Human Resources in Science and Technology with higher education as a percentage of economically active population
	Percentage population aged 25-64 having completed tertiary education
	Population with higher educatiion (% of workforce)
	Percentage population aged 25-64 participating in life-long learning
	R&D expenditures (% of GDP)
Firm activities	R&D units per 100 thousand population
	R&D units in enterprises per 10 thousand enterprises
	Employed in R&D (in EPC) per 1000 economically active population
	Employed in R&D (persons) in manufacturing sector per 1000 employed in manufacturing sector

	Innovation expenditures (product and process innovation) of manufacturing enterprises per inhabitant
	Manufacturing enterprises introducting product and process innovations (% of all manufacturing enterprises)
	Number of patents applied for at the Polish Patent Officeper million population
Outputs	Firms using means of automation per 1000 firms form manufacturing sector
	Employment in medium-high & high-tech manufacturing (% of workforce)
	Employment in knowledge-intensive services (% of workforce)

Source: own research based on Innovation Union Scoreboard 2011, p. 10-11; Regional Innovation Scoreboard (RIS) 2009, p. 7-8; Regional Innovation Scoreboard 2012. Methodology report, p. 4-13.

The growth of the innovation level during the analyzed period is not so strong as the growth of qualified personnel supply index. Over the years 2004-2010 the biggest increase of the index occurred in łódzkie, podkarpackie, świętokrzyskie and warmińsko-mazurskie, while the bigest decrease in lubuskie and opolskie.

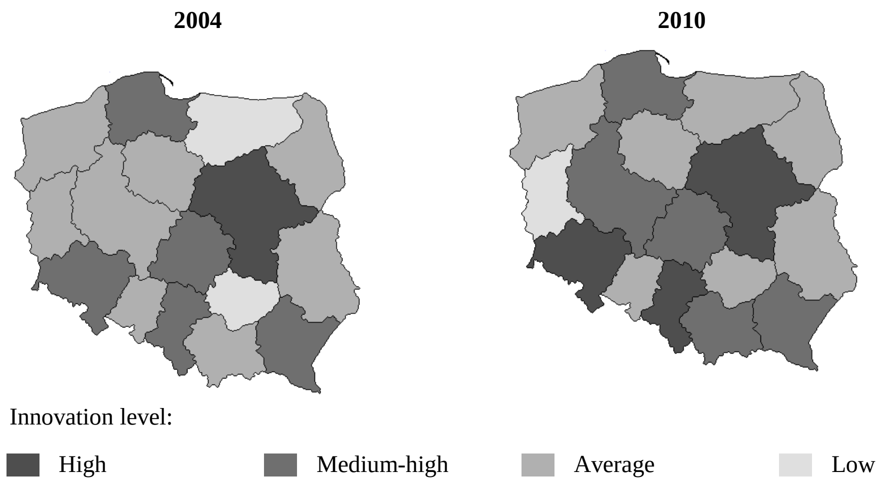
Figure 5. Innovation level in voivodships over the years 2004–2010



Source: own calculations using data from the Local Data Bank (http://www.stat. gov.pl/bdl) and Eurostat (http://epp.eurostat.ec.europa.eu).

Nevertheless, innovation level is also systematically rising - in year 2004 there was only one voivodship characterized by a high level of innovation, while in 2010 there were already three highly innovative voivodships. Also the number of voivodships with medium-high level of innovativeness increased over the six years of analysis.

Figure 6. Innovation level in voivodships in 2004 and 2010



Source: own research.

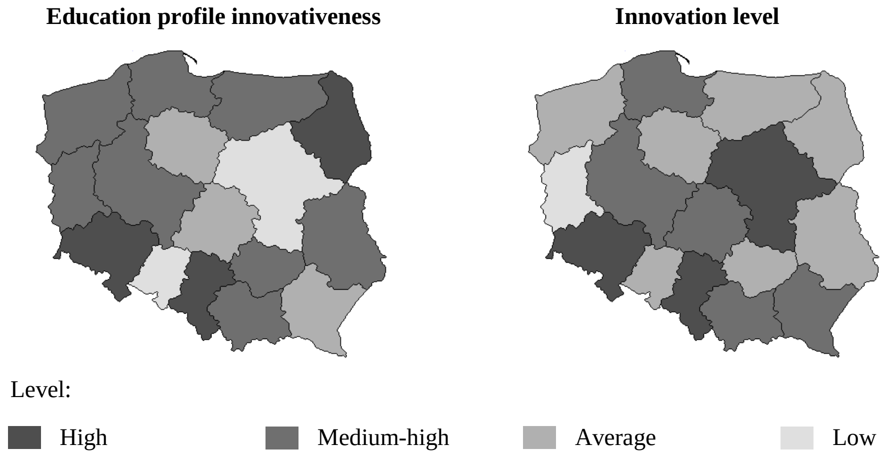
5. Relationship between innovation and qualified personnel supply

The analysis of relationship between innovation and qualified personnel supply, which is a key stage of research, covers an identification of education innovative potential, analysis of interdependence between innovation and qualified personnel supply and cluster analysis.

A level of innovativeness of particular regional education profiles was determined for year 2010 as the sum of shares of graduates in fields of study considered as a highly innovative, which are: health, science and engineering and technology. According to this, the highest innovative potential of education profile occurs in podlaskie, dolnośląskie i śląskie, and the lowest in mazowieckie and opolskie (figure 7, map on the left).

Comparison of these results with the level of innovation shows, that a high innovative potential of education profile does not correspond with the high level of innovation and, what is more, it is just the opposite. Therefore, there is a assumption that the relationship between these two characteristics is negative. It may indicate that the regional higher education system does not boost a regional resources for innovation, and so there is a low or limited knowledge diffusion to R&D units or business sector form education system.

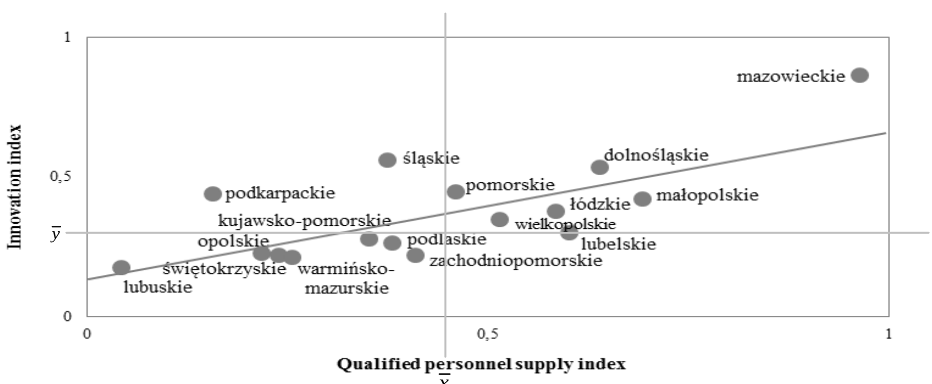
Figure 7. Innovative potential of education profile and innovation level in 2010



Source: own research.

On the other side, a scatter plot of innovation and qualified personnel supply indexes (figure 8) implies a positive association between those two characteristics – small values of one index tend to associate with small values of the other one and, similarly, large values of both indexes also tend to associate. Moreover, a positive value of correlation coefficient (*r*=0,75) confirms relatively strong relationship between those two characteristics.

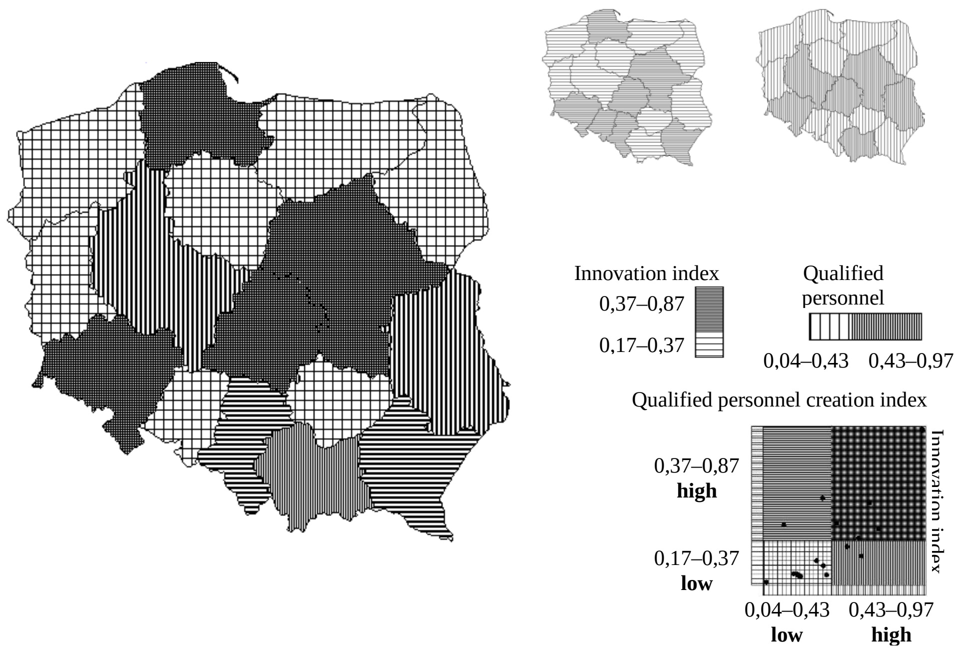
Figure 8. Spatial analysis of interdependence between innovation and qualified personnel supply in 2010



Source: own calculations.

Results of scatter plot analysis are presented on a bivariate map (figure 9), where two variables (indexes) are displayed on a single map by combining two different color scales (or different patterns) (Leonowicz).

Figure 9. Innovation and qualified personnel supply level in voivodships in 2010

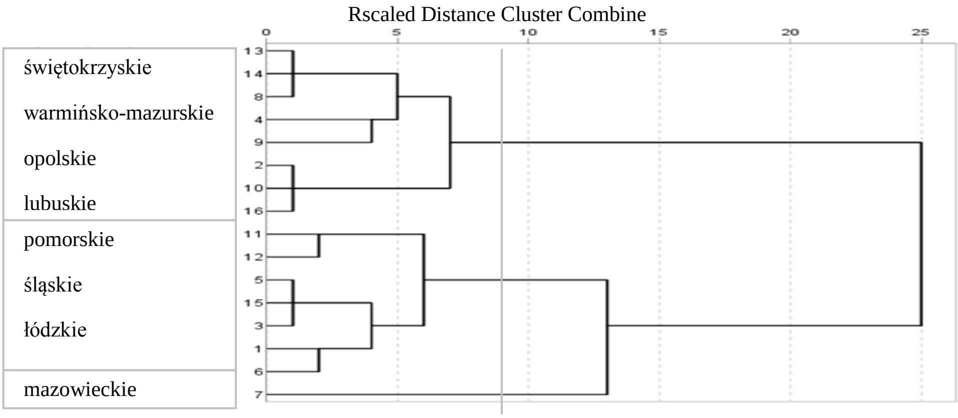


Source: own research.

Values of innovation index are marked with horizontal lines and qualified personnel creation index with vertical lines. Regions marked with patterns assigned to the first and third quarter of the chart are characterized by low or high level of both characteristics, and those marked with patterns from second and fourth quarter indicates the coexistence of low values of one index and high values of the other one. As can be seen in figure 9, a positive relationship between analyzed phenomena confirms.

In the last step cluster analysis have been conducted, where Ward's method and the Euclidean distance have been used. A large gap between joining at the distance from approximately 7 to 13 indicated a three-cluster solution (figure 10). What is more, this classification is generally consistent with the results of scatter plot analysis.

Figure 10. Cluster analysis - dendogram using Ward's Method



Source: own calculations using *IBM SPSS Statistics 19*.

6. Conclusion

As a conclusion to this research it should be noted, that education profiles vary accross regions of Poland and are not constant, but changes over time. Secondly, results of research confirms, that both qualified personnel supply and innovation level tend to increase in most of voivodships. Analysis of the relationship between these two characteristics provides two main conclusions: there is no correspondence between innovative potential of education profile and level of innovation, which leads to assume that the regional higher education system does not boost a regional innovation resources and the knowledge diffusion is low or limited, but on the other side there is a relatively strong positive relationship between the level of innovation and qualified personnel supply.

References

Czupich M. (2009*), Innowacyjność regionów Europy Środkowo-Wschodniej*, [in:] Nowakowska A. (ed.), *Innowacyjność regionów w gospodarce opartej na wiedzy*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź

Gaczek W. M. (2005), *Innowacyjność jako czynnik podnoszenia konkurencyjności gospodarki regionu,* [in:] Gaczek W. M. (ed.) *Innowacje w rozwoju regionu*, Wydawnictwo Akademii Ekonomicznej, Poznań

Grodzicki J. (2000), *Edukacja czynnikiem rozwoju gospodarczego*, Wydawnictwo Adam Marszałek, Toruń

Innovation Union Scoreboard 2011, (http://ec.europa.eu/enterprise/policies /innovation/files/ius-2011_en.pdf, 20.05.2012)

Leonowicz A., *Prezentacja zależności zjawisk metodą kartogramu złożonego*, Instytut Geografii i Przestrzennego Zagospodarowania PAN, Warszawa (http://globus.igipz.pan.pl/ksig/aleon/aleon_artykul.htm, 22.05.2012)

Regional Innovation Scoreboard (RIS) 2009, UNU-MERIT and Joint Research Center of the European Commission, December 2009, (http://www.proinno-europe.eu /page/european-innovation-scoreboard-2009; 20.05.2012)

Regional Innovation Scoreboard 2012. Methodology report, (http://www.merit.unu.edu/about/profile.php?id=31, 20.05.2012)

Strahl D., Markowska M. (2008), *Innowacyjność europejskiej przestrzeni regionalnej – ocena zmian w czasie*, „Ekonomia i Międzynarodowe Stosunki Gospodarcze”, no. 17

Targeting regional economic development. A joint initiative of the regional rural development centers, January 2010, (http://nercrd/psu.edu/tred/index.html, 22.05.2012)

Streszczenie

PODAŻ WYKWALIFIKOWANEJ KADRY A PERSPEKTYWY INNOWACYJNEGO ROZWOJU WOJEWÓDZTW

Celem niniejszego opracowania jest analiza podaży wykwalifikowanej kadry w kontekście perspektyw przyszłego rozwoju innowacyjnego województw w Polsce. Analiza powyższych zjawisk oraz związku między nimi przeprowadzona zostanie m.in. na podstawie badania profilu edukacyjnego województw oraz analizy potencjalnych kierunków i możliwości kreowania zasobów pracy wysoko wykwalifikowanej, pochodzących ze szkolnictwa wyższego, a także badania profilu innowacyjnego oraz poziomu innowacyjności poszczególnych województw. Jednym z elementów analizy będzie również badanie możliwości i poziomu dyfuzji wiedzy ze środowisk naukowych do gospodarki.

10.2478/v10103-012-0026-6

ROMAN GAVULIAK[*]

Exploring Social Exclusion in The European Union, a Quantitative Approach

Abstract

Social exclusion is a widely debated issue. Its definitions and perceptions vary. Within the paper we identify the underlying factors of social exclusion within the EUfor years 2005-2009 through the use of factor analysis. These factors are as such immeasurable by common indicators. Through factor scores we compare the severity of these factors in each EU 27 and suggest five categories of types of social exclusion.

1. Introduction

Social exclusion as a concept is perceived differently, however when it comes to analysing and aiming for results based on objective criteria, it always comes down to the measurability of its dimensions. Social exclusion isa multidimensional concept and focuses on deprivation in different areas: economic, social, and on the processes and mechanisms that exclude people (Haan 1998).

When analysing social exclusion across Europe, we will use the dataset from European Sustainable Development Strategy (Eurostat 2006), also overlapping with Europe 2020 indicators and targets (Eurostat 2010).

Not all of the dimensions of social exclusion can be captured or measured objectively by indicators. However if these underlying – hidden factors do have

[*] Ing., University of Economics in Bratislava

a significant impact on social exclusion, they should be reflected in the measurable indicators in some way. We aim to extract these underlying factors. First we will identify the major factors behind social exclusion in Europe using factor analysis. We will then calculate and transform their factor scores into a measurable form that is suitable for comparison. Using the transformed factor score values for the year 2009 we will use cluster analysis in order to categorize the countries of European Union based on the underlying factors.

1.1. Indicators of social exclusion and poverty in Europe

We will focus on measurements of poverty and social exclusion included in the Sustainable Development Strategy (SDS) of the European Union (EU) created in 2006. The hierarchy of indicators spans three levels. The headline indicators at the top, representing the monitored area, theme-related indicators on the second level that serve as operational objectives and targets supported by a third level of actions/explanatory variables divided into sections related to the theme-related indicators on the second level. Each of the themes can have contextual indicators that transcend the second and third level (Eurostat 2006) (Figure 1).

Figure 1. Hierarchy of social inclusion indicators within the SDS

Headline indicator	**Operational objectives and targets Actions/explanatory variables**	**Operational objectives and targets Actions/explanatory variables**
Population at risk of poverty or exclusion	***Monetary poverty and living conditions***	
	Persons at risk of poverty after social transfers	Persistent at risk of poverty rate
		Persons at-risk-of-poverty after social transfers, by gender
		At risk of poverty rate, by age group
		At risk of poverty rate, by household type
	Severely materially deprived persons	Relative median at risk of poverty gap
		Inequality of income distribution
	Access to labour market	
	Persons living in households with very low work intensity	In work at risk of poverty rate
		Total long-term unemployment rate
		Gender pay gap in unadjusted form

	Education	
	Early leavers from education and training	At risk of poverty rate, by highest level of education attained
		Persons with low educational attainment, by age group
		Life long learning
		Low reading literacy performance of pupils
		Individuals' level of computer skills
		Individuals' level of internet skills
Contextual indicator	Public expenditure on education (for sub-theme Education)	

Source: http://epp.eurostat.ec.europa.eu/portal/page/portal/sdi/indicators.

The indicator *Population at risk of poverty or exclusion* is constructed as the union of the three second level indicators *Persons at risk of poverty after social transfers*, *Persons living in households with very low work intensity*, *Early leavers from education and training* not featuring intersections.

The threshold of poverty according to Eurostat is defined as 60 % of the national median of the equivalised disposable incomes in an economy. The indicator *Persons at risk of poverty after social transfers* is calculated as the ratio of persons with equivalised disposable incomes below the poverty threshold.

Severe material deprivation is a share of population with an enforced lack of at least four out of nine material deprivation items. The nine items are defined in the EU SILC methodology (Eurostat 2009).

Households in low work intensity translate to the share of population aged 0-59 living in households where the working age members worked less than 20% of their total work potential during the past year.

The Early leavers from education and training indicator is defined as the percentage of the population aged 18-24 with at most lower secondary education and not in further education or training.

The Europe 2020 strategy aims for reduction of poverty by aiming to lift at least 20 million people out of the risk of poverty or exclusion across EU. Our aim is to extract additional dimensions of social exclusion that could serve as a basis for comparative policy analysis that could help achieve this goal. We extract the factors of social exclusion from the third level of the system of indicators – the explanatory variables. All of the indicators included in the area are obtained either through the EU SILC survey or the Labour Force Survey.

1.2. Methodology

Factor analysis can be used to analyze interrelations among a large number of variables and to explain these variables in terms of their common underlying dimensions. (Hair et. al 2009).Exploratory factor analysis is one of the commonly used evaluation tools. As for the usage of factor analysis in the field of poverty and social inclusion, one of the current examples is (Vojtková M. 2009), where factor analysis was used to extract the factors in order to map and evaluate social cohesion within the EU.

Cluster analysis is a tool commonly used for classification of objects and for developing meaningful subgroups of individuals and objects (Hair et. al 2009).

2. Factor analysis results

Factor analysis can result in in a number of different results based on the selected method and rotation, in our casethe same pattern kept recurring across all possible approaches. The results interpreted in this chapter come from the varimax rotation of the classical factoring method.

The data used have been obtained from the Eurostat database. From the 16 available indicators on the third level two have been discarded for data unavailability - *Persistent-at-risk-of-poverty rate* and *Low reading literacy performance of pupils*. We used country specific panel data from years 2005 – 2009, for this period, all the values for all EU 27 countries are available (Eurostat, 2006). All of the values of the indicators in the analysis were standardized. Some indicators offer different variants; we chose the variants that represent the most vulnerable and/or influenced population groups. These choices come from our previous analyses. Variants are listed in Table 1.

Table 1. Variants of chosen indicators

Indicator	Variant
Persons at-risk-of-poverty after social transfers, by gender	Female population
At-risk-of-poverty rate, by age group	Aged 65 and above
At-risk-of-poverty rate, by household type	Single female
At-risk-of-poverty rate, by highest level of education attained	at most ISCED 2
Persons with low educational attainment, by age group	25 – 64 years
Individuals' level of computer skills	Lowest level
Individuals' level of internet skills	Lowest level

Source: Author’s research.

Based on the rule of thumb for significance of factor score coefficients (Hair J. F. et. Al 2009) we set the minimum significance threshold of a factor score coefficient at 0.5 In order to interpret the factors. Table 3 represents the simplified rotated matrix of factor score coefficients.

The factor score coefficients of the first two factors *Unfavourable living conditions* (U) and *Deprivation of education* (D)feature only positive scores on variables where an increase of the indicator means a negative development and negative scores on indicators where a decrease translates to a negative development. This is not the case with the third factor - *Gender inequality persisting above low educational attainment*(G), where the decrease of the decrease of the *Persons with low educational attainment, by age group* indicator is a positive development on its own. We will have to take this into account when using the factor scores for cluster analysis (Hair J. F. et. al 2009).

Table 2. Simplified rotated matrix of factor loadings (factor score coefficients)

↓Indicator / Factor →	**U**	**D**	**G**
At risk of poverty rate, by household type	0.89		
Relative median at risk of poverty gap		0.68	
Inequality of income distribution	0.67		
In work at risk of poverty rate		0.63	
Total long-term unemployment rate		0.66	
Persons at-risk-of-poverty after social transfers, by gender	0.86		
At risk of poverty rate, by age group	0.88		
Gender pay gap in unadjusted form			0.61
At risk of poverty rate, by highest level of education attained	0.91		
Persons with low educational attainment, by age group			-0.82
Life long learning		-0.72	
Individuals' level of computer skills			0.74
Individuals' level of internet skills			0.51
Public expenditure on education		-0.74	

Source: Author's research.

The living conditions of individuals influence their options and motivations for the future. The first factor - *Unfavourable living conditions* represents an antagonistic process. *U* increases the risk of poverty for all population groups in the analysis excluding working poverty as well as includes the reflection of poverty of these groups on the total inequality in a society. While one of the reason for affecting more of the poverty groups are the overlaps between these groups (a single female over 65 years of age for example), the influence of the poverty of people aged 65 and above also hints to the existence of an intergenerational transfer of poverty.

The *Deprivation of education* factor describes the impact of reducing public expenditure on education, which results in the decrease of accessibility of life long learning, the effects of this process on the society decreases the competitiveness of a part of the population resulting in the affected population to either in low wages (increase of the *In work at risk of poverty rate* indicator) or in the increase of the long term unemployment. This further influences the inequality in society by increasing the *Relative median at risk of poverty gap*.

Based on the UN statistics of the ratio of estimated female to male earned income, EU countries still show a significant income gap between genders (UN,2009).*Gender inequality persisting above low educational attainment* factor represents a more advanced form of gender inequality. It is noticeable in countries where higher educational attainment is a standard. While there is a decrease in *Persons with low educational attainment, by age group* and an increase in basic level of computer and internet literacy, there is an increase in *Gender pay gap in unadjusted form* which points out to the fact, that even with a higher educational attainment, the gender pay gap still persists, possibly even in creases. This affects the overall inequality in a society, reflected in the *Inequality of income distribution* measures.

3. Cluster analysis results

By applying cluster analysis, we aim to classify the EU 27 countries based on the three factors we discovered. We believe such a classification is necessary as countries often look for inspiration in the area of social policies in countries with a similar setting. The socio-economical background of today calls for effective policies for combating social exclusion. Applying what has been successful in another country can only work in a similar setting. We came up with five groups of types of social exclusion that should serve as a basis for such comparison.

In order to improve the readability of the factors we aim for the factors to achieve only values above zero where higher values means a higher influence in a country. Usually factor scores are negative for countries (observations) with a below average influence of a factor and positive values for countries (observations) with an above average influence of a factor. First two factors could be transformed by one, very simple step – adding the value of the minimal factor score of a factor to each factor score of the same factor. This will retain the informational value of comparing the values of the factor to their mean. Adding a constant to every factor score results in the mean increasing by the same constant which can be expressed as follows:

$$\mu + c = \frac{\sum_{1}^{n} c + F_i}{n} \qquad (1)$$

where: i – country specific observation, n – number of observationsc – constant, c = F_{min}μ – mean, F – factor score, F_{min}– minimum of the factor scores

Both of these factors now have factor scores of zero and above and the higher their values are, the higher the influence of the respective factor which means a negative development.

Transforming the third factor score - *G*is different. Originally this factor included three factor score coefficients of indicators that meant a positive development and one with a negative. We aim for the factor scores to reflect negative development just as the two previous factors for easier comparability. Before calculating the factor score coefficient of this factor, we transform the single indicator with the negative development (*Gender pay gap in unadjusted form*) to an opposite, negative value.At this point all of the developments of the indicator of the factor are positive; in order to transform them into negative we will multiply all factor scores calculated after the transformation of the factor score coefficient by the value of -1.

$$\mu + c = \frac{\sum_{1}^{n} c \cdot F_i}{n} \qquad (2)$$

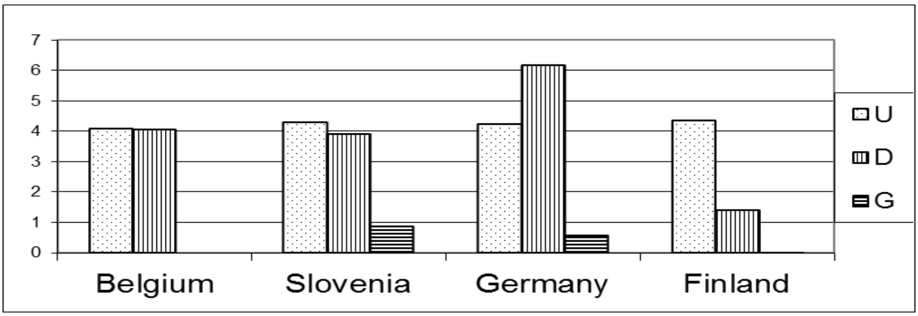
This leaves us with*Gender inequality persisting above low educational attainment*values that are negative for countries (observations) with a below average influence of a factor and positive values for countries (observations) with an above average influence of a factor. The final transformation is similar to the one applied to the previous two factors – adding the value of the minimal factor score of a factor to each factor score.

Results of the cluster analysis

Cluster analysis conducted through Ward's method with Euclidean distance measure, resulted in five country groups. All of the comparisons are not made against an absolute benchmark;yet, the resulting groups seem to reflect the current situation in European Union quite well.

The first group of countries attains low values of all of the factors and especially at*Gender inequality persisting above low educational attainment* being close to zero. We can conclude there are high levels of gender equality in these countries as well as lower levels of social exclusion compared to the EU 27 average (Figure 3).

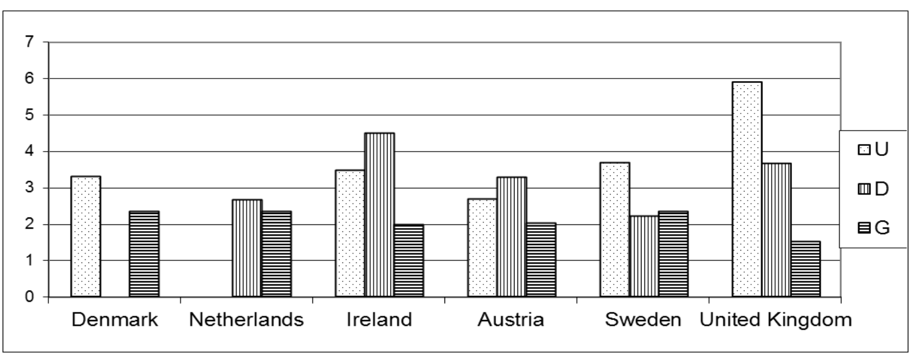
Figure 3. Factor score values: Belgium, Slovenia, Germany and Finland



Source: Author's calculations.

While Finland and Germany are often considered one of the most socio-economically stable countries, it is interesting that Slovenia ranks among such countries. Germany ranks above average in the *Deprivation of education* factor score. The next group of countries, while still achieving below average scores of factor scores for social exclusion experiences higher values of the *Gender inequality persisting above low educational attainment* factor (Figure 4).

Figure 4. Factor score values: Denmark, Netherlands, Ireland, Austria, Sweden, and United Kingdom

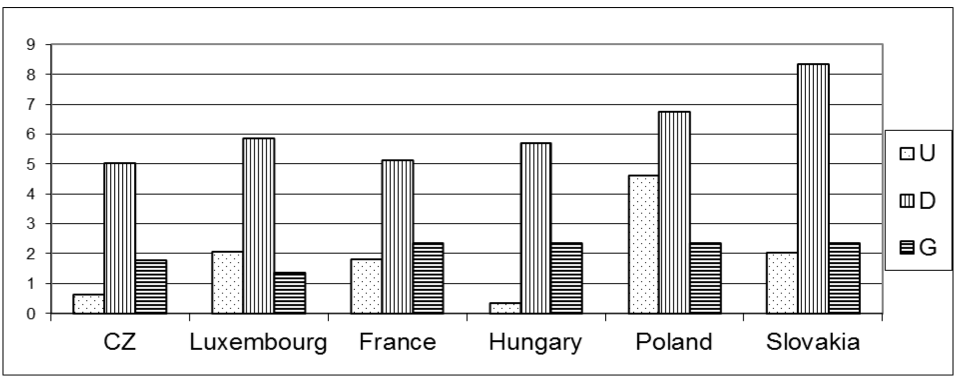


Source: Author's calculations.

Denmark has, compared to other EU 27 countries almost non-existing deprivation of education and Netherlands have a very good position on *Unfavourable living conditions*. Ireland, a country often mentioned with the connection with the current EU economical crisis is, compared to other affected countries much better off in the terms of social exclusion. The United Kingdom has however an above average value of the *Unfavourable living conditions* factor score which is disproportionate towards the two remaining factors.

The third group of countries contains France, Luxembourg and all Vysegrad countries. All of the countries have a below average value of *Unfavourable living conditions* and *Gender inequality persisting above low educational attainment* (Figure 5). For Vysegrad countries this could be caused by their communistic past where the equality of a society was higher (even though on the negative side). All of the countries experience disproportional values of the *Deprivation of education* factor scores which can lead to negative externalities connected with the deprivation of human capital.

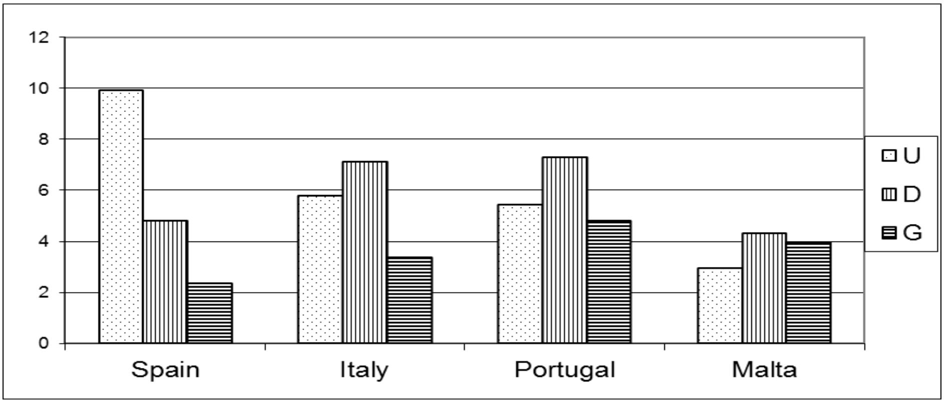
Figure 5. Factor score values: Czech Republic (CZ), Luxembourg, France, Hungary, Poland and Slovakia



Source: Author's calculations.

The fourth group features mostly Southern Europe. All countrieshave above average *Gender inequality persisting above low educational attainment* factor scores. Spain ranks high in the *Unfavourable living conditions* factor score. Both Italy and Portugal achieve values above average. The values of *Deprivation of education* are above average also for Italy and Portugal. Malta has the best values of the factors of social exclusion from this group of countries (Figure 6).

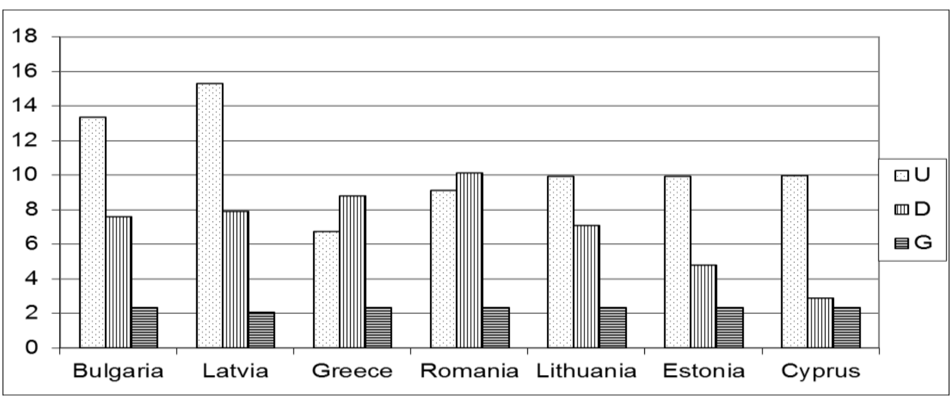
Figure 6. Factor score values: Spain, Italy, Portugal and Malta



Source: Author's calculations.

Last group of countries is the one ranking the worst on factor scores for social exclusion. Besides the last two admissions into the EU – Bulgaria and Romania it includes also countries from North-Eastern Europe, Cyprus and Greece. All of the factor score values of these countries are above average. For *Unfavourable living conditions* the highest values belong to Bulgaria and Latvia. The *Deprivation of education* is most prevalent in Romania and Greece. Factor scores of *Gender inequality persisting above low educational attainment* are very similar for all the countries (Figure 7).

Figure 7. Factor score values: Bulgaria, Latvia, Greece, Romania, Lithuania, Estonia and Cyprus



Source: Author's calculations.

4. Conclusion

With the use of exploratory factor analysis weidentified three major factors of social exclusion - *Unfavourable living conditions*, *Deprivation of education* and *Gender inequality persisting above low educational attainment*. These factors reflect crucial issues that need to be addressed across Europe. The transformed factor scores for the three factors behind social exclusion can serve further as indicators measuring the progress of the EU countries in combating social exclusion as well as variables when it comes to modeling social exclusion (such as regression or correlation analysis).

The extent of each factor varies in different countries; however there are similarities among the countries. We identified five groups of EU 27 countries that can be used as a basis for comparative approach when it comes to implementing policies aimed at combating social exclusion.

The groupings have a potential to serve as the basis for comparative policy analysis. When it comes to social exclusion, there is no dividing line between the founders of the European Union and the newcomers.

It is also interesting to note, that countries most associated with the ongoing budget crisis vary in the extent of the influence of the three factors behind social exclusion and thus we can’t expect the same measures to have the same effect for them.

A basis for further research could be the influence of cultural or regional dimension on the rankings of the factor scores. When we look at the values of the factor scores, we see the following:

- No Nordic country ranks above average on any factor score
- Group 4 is formed exclusively by countries from the south of Europe and has an above average influence of the *Gender inequality persisting above low educational attainment* factor
- No country from south of Europe belongs to any of the first two groups except for Slovenia
- All Vysegrad countries belong to the third group; this may be caused by a very similar socio-economical past

This paper was produced with the support of the IGPM funding framework under the project number 2317122/10: Modelling of chosen indicators of Sustainable development in the context of the European Union.

References

Council of the European Union, (2006) Renewed EU Sustainable Development Strategy Annex. Eurostat, Luxembourg (available from http://sd- network.eu/ pdf/quarterly%20reports /EU_SDS_final_ version_2006_June26_en.pdf)

Eurostat, (2010) Europe 2020 targets and indicators, Eurostat: Luxembourg (available from http://ec.europa.eu/europe2020/index_en.htm)

Haan, A. (1998). Social exclusion: An alternative concept for the study of deprivation? IDB Bulletin, 29(1), 10-19

Hair J. F. et. al, (2009) Multivariate data analysis, New Jersey,Pearson Education: USA

UN, (2009) United Nations Development Fund. Human Development Report, New York: UNDP, (available from http://hdr.undp.org/en/media/HDR_2009_EN_complete.pdf)

Vojtková, M. (2009) *A Multidimensional Evaluation of Social Cohesion*[in] J. Pociecha(ed.), *Statistical Analysis of the Economic and Social Consequences of Transition Processes in Central-East European Countries*, Cracow University of Economics Press

Streszczenie

ANALIZA WYKLUCZENIA SPOŁECZNEGO W UNII EUROPEJSKIEJ: PODEJŚCIE ILOŚCIOWE

W artykule określono podstawowe determinanty wykluczenia społecznego w Unii Europejskiej w latach 2005-2009, na podstawie analizy czynnikowej. Czynniki wykluczenia stanowią cechy niemierzalne, które w badaniu zostały zoperacjonalizowane za pomocą zmiennych mierzalnych. Porównanie wpływu poszczególnych czynników w krajach UE 27 pozwoliło na wyodrębnienie pięciu kategorii wykluczenia społecznego.

10.2478/v10103-012-0027-5

DANIEL A. GRIFFITH*

Selected Challenges from Spatial Statistics for Spatial Econometricians

Abstract

Griffith and Paelinck (2011) present selected non-standard spatial statistics and spatial econometrics topics that address issues associated with spatial econometric methodology. This paper addresses the following challenges posed by spatial autocorrelation alluded to and/or derived from the spatial statistics topics of this book: the Gaussian random variable Jacobian term for massive datasets; topological features of georeferenced data; eigenvector spatial filtering-based georeferenced data generating mechanisms; and, interpreting random effects.

1. Introduction

Geography experienced a quantitative revolution in the 1950s and 1960s (Curry 1967). Work generated by this movement initially analyzed distances from locations of privilege as well as attribute variables whose observations were distinguished merely by a locational index. Especially statistical decisions spawned by these analyses proved to display far more variability than indicated by classical statistical distribution theory; this increased variability is attributable to positive spatial autocorrelation (SA) latent in almost all georeferenced data.

* Ph.D., University of Texas at Dallas

Addressing these inadequacies, Cliff and Ord[1] (1969) and Besag (1974), among others, commenced a formal development of autoregression-based spatial statistics that popularized model specifications accounting for latent SA. This line of work soon eclipsed the point pattern analysis work that, until then, typified much of quantitative spatial analyses. Meanwhile, parallel spatial econometric developments flourished after the introduction of Paelinck and Klaassen's (1979) seminal book, followed by Anselin's (1988) classic book. Paelinck (2012) addresses this historical trajectory.

The purpose of this paper is to highlight selected challenges posed by SA alluded to and/or derived from the spatial statistics literature and contextualized in Griffith and Paelinck (2011). One challenge arises from the increasing size of georeferenced datasets, some of which are massive today. Calculating maximum likelihood estimates (MLEs) requires computing the determinant of an n-by-n spatial covariance matrix—the Jacobian of a transformation in calculus terms—which becomes excessively numerically intensive or even infeasible for massive georeferenced datasets. This paper outlines an alternative MLE solution to nonlinear regression, which is new, couched in the existing spatial statistics literature about approximating the Jacobian term. A second challenge stems from topological considerations accompanying georeferenced datasets. This paper focuses on a mistake appearing in the earlier literature, and describes a modified version of the well-known matrix powering algorithm that successfully computes the principal eigenfunction for a periodic matrix. A third challenge concerns georeferenced data generating mechanisms involving eigenvector spatial filtering, and further develops contributions in Griffith (2011a,b). A fourth challenge furnishes additional insight into the meaning of spatially structured random effects. Successful engagement of these challenges poses a potential to improve both spatial statistical and spatial econometric work.

2. The spatial statistical Jacobian term for Gaussian model specifications

In part because normal curve theory was the best developed probability model-based analysis of the time, most early spatial statistics assumed a bell-shaped curve. Gaussian spatial autoregressive model specifications to describe n georeferenced sample values include a Jacobian term, which is: (1) the determinant of an n-by-n matrix; (2) the normalizing constant ensuring that the

[1] *Geographical Analysis* celebrated the major contributions to science of this cluster of research with a special issue in 2009.

probability density function integrates to 1; and, (3) a function of the SA parameter(s). Computational difficulties introduced into calculating MLEs of model parameters by the logarithm of this determinant has generated a body of literature addressing its simplification and approximation (Ord, 1975; Griffith, 1992, 2004a; Barry and Pace, 1999; Smirnov and Anselin, 2001, 2009; Pace and LeSage, 2004; Zhang and Leithead, 2007; Walde et al., 2008).

The likelihood function is equivalent to a multivariate normal probability density function with a sample of size 1 and n variables:

$$L = (2\pi)^{-n/2} \,|\mathbf{V}|^{1/2}\, (\sigma^2)^{-n/2} e^{-(\mathbf{Y}-\mu\mathbf{1})^T \mathbf{V}(\mathbf{Y}-\mu\mathbf{1})/(2\sigma^2)} \qquad (1)$$

where $\mathbf{V}^{-1}\sigma^2$ is the SA variance-covariance matrix that is a function of the spatial autoregressive parameter ρ in a single-parameter model specification, Y is a normally distributed random variable, **Y** is an n-by-1 vector of random variable values, **1** is an n-by-1 vector of ones, T denotes the matrix transpose operation, and μ and σ^2 respectively are the constant mean and the variance of Y. When ρ = 0, **V** = **I**, the n-by-n identity matrix.

But all of the more recent literature overlooks the useful simplicity of the approximation developed by Griffith (1992, 2004a), with a special case for regular square tessellations (Griffith, 2004). The appeal of this latter approximation is that it can be employed efficiently and effectively with a dataset whose size is in the millions or billions—a massive dataset. For a symmetric distribution of eigenvalues, such as that for a regular square tessellation, the Jacobian approximation given by Griffith (2004) reduces to

$$-\sum_{i=1}^{n} \mathrm{LN}(1-\rho\lambda_j)/n \approx 2\omega\,\mathrm{LN}(\delta) - \omega\,\mathrm{LN}(\delta+\rho) - \omega\,\mathrm{LN}(\delta-\rho) \qquad (2)$$

where λ_j is the j[th] eigenvalue of matrix **V**, and ω and δ are coefficients to be calibrated. When ρ = 0, both $2\omega\,\mathrm{LN}(\delta) - \omega\,\mathrm{LN}(\delta+\rho) - \omega\,\mathrm{LN}(\delta-\rho) = 0$ and $-\sum_{i=1}^{n} \mathrm{LN}(1-\rho\lambda_j)$ = 0. Griffith (2004b) shows that the Jacobian term associated with a regular square tessellation forming a complete rectangular region also can be approximated by

$$-\sum_{i=1}^{n} \mathrm{LN}(1-\rho\lambda_j)/n \approx \mathrm{LN}(1+q_2\rho^2 + q_4\rho^4 + q_{20}\rho^{20}) \qquad (3)$$

where q_2, q_4, and q_{20} are coefficients to be calibrated. When $\rho = 0$, $LN(1+q_2\rho^2+q_4\rho^4+q_{20}\rho^{20}) = 0$.

MLEs for the three parameters of equation (1) are

$$\hat{\mu} = \mathbf{1}^T\mathbf{V}\mathbf{Y}/(\mathbf{1}^T\mathbf{V}\mathbf{1}), \tag{4}$$

$$\hat{\sigma}^2 = (\mathbf{Y}-\mu\mathbf{1})^T\mathbf{V}(\mathbf{Y}-\mu\mathbf{1})/n, \tag{5}$$

and for a spatial simultaneous autoregressive (SAR; the spatial error model in the spatial econometrics literature) model specification, for which $\mathbf{V} = (\mathbf{I}-\rho\,\mathbf{W})^T(\mathbf{I}-\rho\,\mathbf{W})$, where $\mathbf{W}$ is an n-by-n geographic weights matrix, ρ may be calculated by solving the differential equation

$$\begin{aligned}\partial LN(L)/\partial\rho = \left[\sum_{j=1}^{n}(-\lambda_j)/(1-\rho\,\lambda_j)\right]\Big/ n\, + \\ (\mathbf{Y}-\hat{\mu}\mathbf{1})^T(\mathbf{W}^T+\mathbf{W}-\rho 2\mathbf{W}^T\mathbf{W})(\mathbf{Y}-\hat{\mu}\mathbf{1})/(2n\,\hat{\sigma}^2) = 0\end{aligned} \tag{6}$$

where λ_j are the n eigenvalues of matrix $\mathbf{W}$.

Equations (2) and (3) respectively result in $\left[\sum_{j=1}^{n}(-\lambda_j)/(1-\rho\,\lambda_j)\right]\Big/ n \approx$

$$-2\,\alpha\,\rho/(\delta^2-\rho^2), \text{ and}$$

$$-(2q_2\rho+4q_4\rho^3+20q_{20}\rho^{19})/(1+q_2\rho^2+q_4\rho^4+q_{20}\rho^{20}).$$

These two substitutions dramatically simplify equation (6).

For a regular square tessellation forming a complete rectangular region (i.e., a remotely sensed image whose data may be important for an environmental economics analysis), with P > 3, Q > 3, and PQ ≤ 5,625, numerical experiments yield the following large sample results:

$\omega \approx 0.16361 - 0.00457(1/P + 1/Q) - 0.47594/(PQ)$

$\delta \approx 1.17583 - 0.33691[1/(P+1) + 1/(Q+1)] - 1.08316/[(P+1)(Q+1)]$, and

$q_2 \approx 0.11735 + 0.10091(1/P^{5/4} + 1/Q^{5/4}) + 0.42844/(PQ)$

$q_4 \approx 0.07421 + 0.05730(1/P^{2/3} + 1/Q^{2/3}) - 0.66001/(PQ)$

$q_{20} \approx 0.05221 + 0.52467(1/P^{7/4} + 1/Q^{7/4}) + 2.48015/(PQ)$.

The computation of Table 1 results utilized these numerical generalizations for a massive 3,000-by-5,000 pixels georeferenced dataset collected for the Florida Everglades[2]. Relatively few computational resources are needed to analyze n = 15,000,000 observations in this case.

A challenge for spatial econometricians suggested by this spatial statistical work is the generalization of coefficients for equation (2) for massive georeferenced datasets based upon the type of irregular surface partitioning that characterize administrative units. The popular autoregressive response (AR; the spatial error model in the spatial econometrics literature) model specification has the following MLEs:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{I}-\rho\,\mathbf{W})\mathbf{Y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} - \rho\,(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Y}$$

$$\hat{\sigma}^2 = [(\mathbf{I}-\rho\,\mathbf{W})\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}]^T[(\mathbf{I}-\rho\,\mathbf{W})\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}]/\mathrm{n}\text{ , and}$$

$$\left[\sum_{j=1}^{n}(-\lambda_j)/(1-\rho\,\lambda_j)\right]\Big/\mathrm{n} + [\mathbf{Y}^T(\mathbf{W}^T+\mathbf{W}-\rho 2\mathbf{W}^T\mathbf{W})\mathbf{Y} - 2\mathbf{Y}^T\mathbf{W}^T\mathbf{X}\hat{\boldsymbol{\beta}})/(2\mathrm{n}\hat{\sigma}^2) = 0\,.$$

Because $(\mathbf{X}^T\mathbf{X})^{-1}$ needs to be inverted only once, this model specification involves relatively little increase in computational intensity vis-à-vis the constant mean case. Consequently, timing results appearing in Table 1 remain informative for the nonconstant mean case.

3. The topology of georeferenced data

SA and autoregression analyses frequently articulate the topological structure of georeferenced data with a simple binary 0-1 n-by-n geographic weights matrix **C** based upon connectivity/contiguity. The row and column labels of matrix **C** are the ordered locations in a geographic landscape, with this ordering being the same for both the rows and the columns for the sake of convenience. The common definitions of contiguity for surface partitioning are based upon analogies with chess moves: the rook when non-zero length common boundaries, and the queen when both zero (i.e., points) and non-zero length common boundaries, determine contiguity. If a row and a column location are

---

[2] A January 1, 2002, 28.5-meter resolution LANDSAT 7 Enhanced Thematic Mapper Plus (ETM+) image.

contiguous, then the corresponding matrix cell is coded 1; otherwise, it is coded 0. Consequently, matrix **C** is sparse and symmetric. Often the preceding matrix **W** is a row-standardized version of this matrix **C**.

Table 1. Spatial autocorrelation parameter estimation

band	estimated eigenvalues				coefficient equations			
	equation (2)		equation (3)		equation (2)		equation (3)	
	$\hat{\rho}$	CPU time[1]	$\hat{\rho}$	CPU time[1]	$\hat{\rho}$	CPU time[1]	$\hat{\rho}$	CPU time[1]
1	0.9248	0.03	0.9236	0.04	0.9249	0.03	0.9231	0.01
2	0.9072	0.01	0.9097	0.03	0.9073	0.01	0.9094	0.04
3	0.8452	0.03	0.8621	0.04	0.8454	0.03	0.8639	0.03
4	0.4060	0.03	0.3990	0.03	0.4067	0.03	0.3991	0.01
5	0.4773	0.01	0.4672	0.03	0.4776	0.03	0.4683	0.03
7	0.6299	0.01	0.6255	0.03	0.6302	0.03	0.6288	0.03

Notes: [1] measured in seconds.

Source: own calculations.

Consider the case where matrix **C** is irreducible (i.e., it cannot be permuted into disjoint block diagonal submatrices). Powers of matrix **C** can be interpreted as follows (Maćkiewicz and Ratajczak 1996):

> Matrix $\mathbf{C}^k$ yields a count in cell c_{ij} that indicates the number of ways of moving from row location i to column location j crossing exactly k boundaries.

This combinatorial interpretation motivated the use of the row sums of a power sum of matrix **C**, standardized by the largest element in the resulting n-by-1 vector, say $\mathbf{E}_A$, as an index of topological accessibility for a network or surface partitioning. Relatively large values denote locations that are better connected (directly and indirectly) and more centrally located (i.e., more accessible) within the topology represented by the graph associated with matrix **C**, whereas relatively small values denote topologically peripheral locations within the graph (frequently those positioned on the boarder of the associated geographic landscape). The diameter of the graph counterpart of matrix **C** (i.e., the maximum number of links to be crossed when moving from any of the n nodes—areal units in the case of spatial analysis—to reached any of the other nodes) is a common stopping exponent for the power sum. Paths between nodes become redundant beyond the diameter, but do account for some detail in terms of topological structure. All entries in the summation matrix for this exponent have non-zero entries.

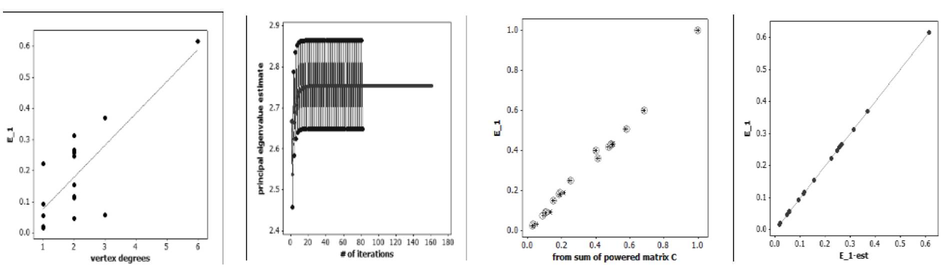
Vector $\mathbf{E}_A$ relates to the principal eigenvector of matrix $\mathbf{C}$, say $\mathbf{E}_1$, and tends to converge upon it; Maćkiewicz and Ratajczak (1996, p. 78) argue that computing vector $\mathbf{E}_1$ is the proper way to define topological accessibility. Mass (1985) criticizes some of the earlier discussion by geographers concerning this accessibility index, challenging conjectures about the relationship between the principal eigenfunction and the row/column sums of matrix $\mathbf{C}$ (Figure 1a). Cvetković and Rowlinson (1990) echo Maas' discussion, but without contributing to resolving the controversy. Mass employs the geographic connectivity matrix for the 1929 Uganda road network reported by Gould (1967, p. 67). But this graph is periodic (the well-know matrix powering algorithm oscillates between 2.64892 and 2.86332; Figure 1b), and hence Maas reports incorrect eigenvalues, having obtained the solution for the lower bound in the oscillation (he reports only the first 10 of the 18 eigenvalues):

Table 1a. Maas' and actual eigenvalues

eigenvalue	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
from Maas	2.652	1.828	1.618	1.447	0.9170	0.7035	0.6180	0.4419	0	0
actual	2.754	1.839	1.639	1.414	1.0000	0.8718	0.6787	0.3525	0	0

Source: Mass C. (1985).

Figure 1. Accessibility index scatterplots



Notes: Left (a): the principal eigenvector $\mathbf{E}_1$ versus the number of neighbors. Left middle (b): trajectories of the old and new algorithm estimates. Right middle (c): the principal eigenvector $\mathbf{E}_1$ versus the sum of the powers of matrix $\mathbf{C}$ through its diameter (open circle) and through 200 (*). Right (d): the principal eigenvector $\mathbf{E}_1$ versus its estimate produced by the new algorithm.

Source:own calculations.

The problematic periodicity of the 1929 Uganda road network graph can be resolved by modifying the well-know matrix powering algorithm to estimate the first eigenfunction so that it includes an iteration lag:

$$\lim_{k\to\infty} [\mathbf{1}^T(\mathbf{C}^k + \mathbf{C}^{k+1})\mathbf{1}/\mathbf{1}^T(\mathbf{C}^{k-1} + \mathbf{C}^k)\mathbf{1}] \to \lambda_1 .$$

This modification produces the proper convergences of the principal eigenvalue (Figure 1b) and its corresponding eigenvector (Figure 1d).

A challenge for spatial econometricians here is to determine the value of eigenvector $\mathbf{E}_1$ for empirical analyses. Is $\mathbf{E}_1$ a useful spatial analysis covariate? Is the correspondence between the row/column sums of matrix **C** sufficiently close to $\mathbf{E}_1$ that their vector is a useful spatial analysis covariate?

4. Georeferenced data generating mechanisms

Griffith and Paelinck (2011) present salient features of georeferenced data, including how variance inflation occurs through, and how correlation coefficients are impacted by, SA. They incorporate eigenvector spatial filters (ESFs) into georeferenced data generating mechanisms (i.e., a selected probability model that includes both random and SA components that combine together to yield individual observations) in some of their demonstrations.

Eigenvector spatial filtering methodology employs the eigenvectors extracted from a modified version of the geographic connectivity matrix **C**, namely $(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)\,\mathbf{C}\,(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n) = \mathbf{MCM}$, where **M** is the standard projection matrix commonly encountered in multivariate statistics, and ${}_C\mathbf{E}_j$ is its jth eigenvector. This matrix expression comes from the numerator of the Moran Coefficient (MC), whose matrix version for response vector **Y** adjusted only for its constant mean is given by

$$MC = [n/(\mathbf{1}^T\mathbf{C}\mathbf{1})][\mathbf{Y}^T\mathbf{MCMY}/(\mathbf{Y}^T\mathbf{MY})] .$$

Substituting the eigenvectors into this expression results in a Rayleigh quotient, with vector ${}_C\mathbf{E}_1$ maximizing the expression. Accordingly, these n eigenvectors can be interpreted as follows:

> the first eigenvector, say ${}_C\mathbf{E}_1$, is the set of real numbers that has the largest MC achievable by any set for the geographic arrangement defined by the spatial connectivity matrix **C**; the second eigenvector is the set of real numbers that has the largest achievable MC by any set that is orthogonal and uncorrelated with ${}_C\mathbf{E}_1$; and so on through ${}_C\mathbf{E}_n$, the set of real numbers that has the largest negative MC achievable by any set that is orthogonal and uncorrelated with the preceding (n – 1) eigenvectors.

As such, these eigenvectors furnish n distinct map pattern descriptions of latent SA in geographically distributed variables because they are mutually orthogonal and uncorrelated. An ESF is constructed from some linear combination of a subset of these eigenvectors, and serves as a spatial proxy variable capturing SA effects in a model specification. This control variable embeds stochastic spatial dependencies among location-indexed observations into the parameters of a probability density/mass function.

All but one of the matrix **MCM** normalized eigenvectors have means of 0 and variances of 1/n. In other words, all of their 1st and 2nd moments match. Consider the following linear combination of these vectors:

$$W\sum_{j=1}^{K}(2q-1)\,a_j\mathbf{E}_j/\sqrt{\sum_{j=1}^{K}a_j^2} = W\mathbf{E}_K\mathbf{B}, \tag{7}$$

where scaling coefficient W is some positive real number, binary 0-1 variable q is a Bernoulli RV, and a_j is a positive coefficient for eigenvector j. The term **B** describes the nature, whereas the scalar W describes the degree (i.e., the relative amount of variance accounted for), of SA. The mean of linear combination (7) is 0, whereas its variance is W^2/n, and the term (2q – 1) makes no difference because each eigenvector $\mathbf{E}_j$ is unique to a multiplicative factor of –1.

The central limit theorem governs expression (7), which implies that the ESFs described by it are approximately normally distributed.

Lemma 1: K << n eigenvectors of matrix **MCM** are independent and are not necessarily identically distributed RVs. As K goes to infinity, expression (7) converges on a normal distribution.

PF: Because all of the K means are 0, and

$\lim\limits_{K\to\infty} a_j^2/\left(n\sum_{j=1}^{K}a_j^2\right)\to 0$ (Bentkus et al., 1996; Chaidee and Tuntapthai, 2009).

Simulation experiments furnish evidence corroborating this lemma (Table 2); it can be tested for in practice with a normal quantile plot. Therefore, an individual observation of a georeferenced Gaussian random variable may be written as

$$Y_i = Y_i^* + W\mathbf{e}_{iK}\mathbf{B}, \tag{8}$$

where the n Y_i^* are iid N(μ , σ^2), and the $W\mathbf{e}_{iK}\mathbf{B}$ is an ESF.

Equation (8) comprises two normal components, one of which is equivalent to a spatially structured random mean response (i.e., $W\mathbf{e}_{iK}\mathbf{B}$ creates random deviations about μ), and furnishes the data generating mechanism in terms of the following parametric mixture distribution:

$$\left.\begin{array}{r} Y \mid SA \sim N(\mu, \sigma^2) \\ SA \sim N(0, W^2/n) \end{array}\right\} \Rightarrow Y \sim N(\mu, \sigma^2 + W^2/n)\,.$$

In other words, the distribution at each location i is conditional on $W\mathbf{e}_{iK}\mathbf{B}$, and SA inflates variance while not affecting an average mean response across a map.

A Manly transformation[3] modifies the geographic distribution of population density across Poland (Figures 2a and 3a), based upon communes, so that it approximates a bell-shaped curve (Figure 2b). The ESF (a linear combination of 22 of 591 candidate vectors) for this geographic distribution closely conforms to be bell-shaped curve (Figures 2c and 3b), and accounts for roughly 27% of the geographic variation in the transformed population density. The random variable Y* approximates a normal distribution, deviating from a bell-shaped curve with one heavy tail (Figures 2d and 3c). The estimated mixture distribution is

$$\left.\begin{array}{r} Y \mid SA \sim N(0.93211, 0.03911^2) \\ SA \sim N(0, 0.02417^2) \end{array}\right\} \Rightarrow Y \sim N(0.93211, 0.04598^2)\,.$$

The variance inflation factor is 1.38192; SA introduces an additional nearly 40% geographic variability into the transformed population density.

Table 2. Comparisons of eigenvectors and of ESFs: Poland surface partitionings

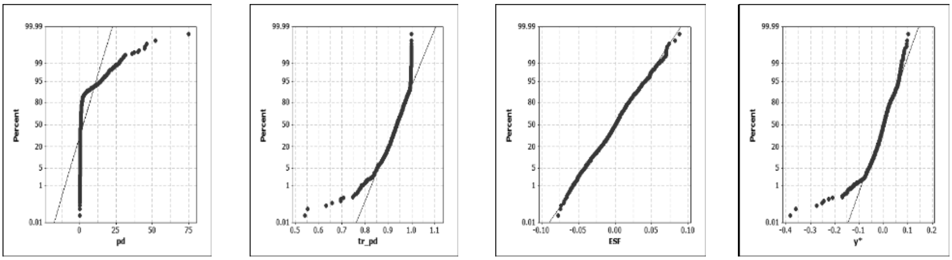
moments	communes	counties	viovodeships
n	2,468	369	16
mean	0	0	0
standard deviation	0.02013	0.05206	0.25000
skewness: mean	0.05806	0.09079	0.20614
standard deviation	0.20666	0.14942	0.21966
excess kurtosis: mean	0.63105	0.17750	-1.14832
standard deviation	1.07388	0.42391	0.33501
# eigenvectors with MC > 0.25	591	85	4
eigenvectors: % with Pr(normality) < 0.01	0	18.82	0
simulated ESFs: % with Pr(normality) < 0.01	0	16.62	0.62

Source: own calculations.

---

[3] The Manly transformation completes the family of Box-Cox power transformations. Here the empirically calibrated transformation is $\exp[-0.3319/(\text{population/area})^{0.7800}]$.

Figure 2. Normal quantile plots



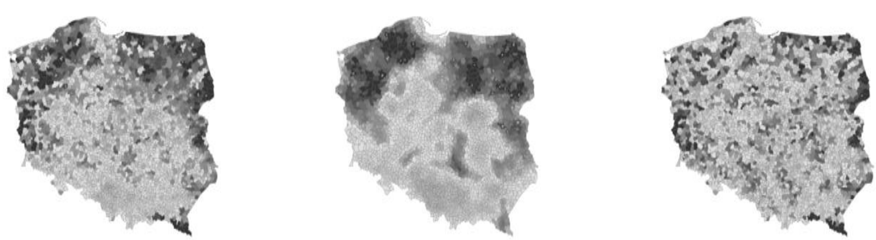
Notes: Left (a): population/area Middle left (b): transformed population density. Middle right (c): ESF. Right (d): Y*.

Source: own calculations.

One advantage of the ESF specification is that it very accurately captures spatial structure as reflected in map pattern (compare Figures 3a and 3b). Thus, ESF supports simulation experiments for which Y* can be determined with a pseudo-random number generator, and then added to the ESF, resulting in the spatial structure being held constant across simulation replications—the parametric mixture distribution specifies the geographic distribution of Y as being conditional on this map pattern.

One challenge for spatial econometricians suggested by this spatial statistical conceptualization is the establishment of ESF properties vis-à-vis spatial autoregression model specifications. Another is to fully develop the mathematical statistical theory associated with ESFs.

Figure 3. The geographic distribution of population density



Notes: scale from light gray to black is proportional to density. Left (a): population/area. Middle (b). ESF. Right (c): Y*.

Source: own calculations.

5. Explicating spatially structured random effects

Random effects model specifications address samples for which observations are selected in a highly structured rather than random way. Frequently random effects can be estimated in one of two ways: employing repeated measures in a frequentist context, or employing priors in a Bayesian context. This first conceptualization directly links random effects to means of time series for individual locations. Because the random effects term is a constant through time, its spatial structure can be captured by an ESF. In turn, this ESF can be estimated with only one slice of time (i.e., a map for a specified point in time), as in the preceding section, revealing that an ESF model specification is able to uncover at least part of a random effects term without repeated measures. Priors in a Bayesian analysis also allow the estimation of a random effects term with only one slice of time.

An average exists for each time series in a space-time dataset. This average ignores both spatial and serial correlation in the space-time series. A random effects model essentially works with these averages, adjusting them in accordance with the correlational structure latent in their parent space-time series, as well as their simultaneous estimation. The random effects model specification achieves this by fitting a distribution with a few parameters (e.g., a mean and a variance for a bell-shaped curve), rather than n individual means (fixed effects) for the n locations. Consequently, a relationship exists between the time series means and the random effects. This random effects specification relates to a fixed effects specification that includes n indicator variables, each for a separate district specific local intercept (one local intercept is arbitrarily set to 0 to eliminate perfect multicollinearity with the global mean).

A challenge for spatial econometricians suggested by this spatial statistical analysis concerns a need to better understand the number of degrees of freedom associated with a random effects term. Another challenge is to better understand random effects terms in the presence of covariates. Given that estimation of the spatially structured part of a random effects term is possible with a single map, a third challenge is to investigate whether or not estimation of the spatially unstructured component of a random effects term can be simplified.

6. Implications and conclusions

Spatial statistics and spatial econometrics are kindred spirits in terms of empirical analysis methodologies. Problems and challenges faced by one of these fields reveals parallel problems and challenges already faced, or to be faced, by the other field. Griffith and Paelinck (2011) present a number of contemporary non-standard sources and treatments of such problems and challenges. This paper builds on their work, extending and identifying other problems and challenges. No doubt the future will produce new problems and challenges, too. Paelinck (2012) points out that spatial econometrics seeks to obtain a better understanding of the workings of spatial economies. Similarly, spatial statistics seeks to obtain a better understanding of the workings of geographic landscapes, some of which constitute space economies. This paper crystallizes the following challenges for spatial econometricians suggested by contemporary spatial statistical work: (1) formulating efficient and effective spatial autoregressive implementations for massive georeferenced datasets; (2) determining the utility of the principal eigenvector of a geographic weights matrix for empirical analyses; (3) casting georeferenced data generating mechanisms in terms of parametric mixture models involving ESFs; and, (4) improving our understanding of spatially structured and unstructured random effects terms that may appear in spatial statistical/econometric model specifications. New insights about these issues offer the potential to improve both spatial statistical and spatial econometric work.

Recerence

Anselin L. (1988), *Spatial Econometrics*, Kluwer, Dordrecht

Bentkus V., Bloznelis M., Götze F. (1996), *A Berry-Esséen bound for Student's statistic in the non-i.i.d. case*, 'J. of Theoretical Probability', Springer, New York, 9

Besag J. (1974), *Spatial interaction and the statistical analysis of lattice systems*, 'J. of the Royal Statistical Society B', Wiley, New York, 36

Chaidee N., Tuntapthai M. (2009), *Berry-Esséen bounds for random sums of non-i.i.d. random variables*, 'International Mathematical Forum', m-Hikari, Ruse, 4

Cliff\ A., Ord J. (1969), *The Problem of Spatial Autocorrelation*, [in:] A. Scott (ed.) *London Papers in Regional Science*, Pion, London

Curry L. (1967), *Quantitative geography, 1967*, 'The Canadian Geographer', Wiley, New York, 11

Cvetković D., Rowlinson P. (1990), *The largest eigenvalue of a graph: a survey*, 'Linear and Multilinear Algebra', Taylor & Francis, Abingdon, 28

Gould P. (1967), *On the geographical interpretation of eigenvalues*, 'Transactions, Institute of British Geographers', Wiley, New York, 42

Griffith D. (1992), *Simplifying the normalizing factor in spatial autoregressions for irregular lattices*, 'Papers in Regional Science', Wiley, New York, 71

Griffith D. (2004a), *Extreme eigenfunctions of adjacency matrices for planar graphs employed in spatial analyses*, 'Linear Algebra & Its Applications', Elsevier, Amsterdam, 388

Griffith G. (2004b), *Faster maximum likelihood estimation of very large spatial autoregressive models: an extension of the Smirnov-Anselin result*, 'J. of Statistical Computation and Simulation', Taylor & Francis, Abingdon, 74

Griffith D. (2011a), *Positive spatial autocorrelation, mixture distributions, and geospatial data histograms*, [in:] Y. Leung, B. Lees, C. Chen, C. Zhou, and D. Guo (eds.), 'Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM 2011)' , IEEE, Beijing

Griffith D. (2011b), *Positive spatial autocorrelation impacts on attribute variable frequency distributions*, 'Chilean J. of Statistics', Sociedad Chilena de Estadística, Valparaiso, 2 (2)

Griffith D., Paelinck J. (2011), *Non-standard Spatial Statistics and Spatial Econometrics*, Springer-Verlag, Berlin

Maćkiewic, A., Ratajczak W. (1996), *Towards a new definition of topological accessibility*, 'Transportation Research B', Elsevier, Amsterdam, 30

Mass C. (1985), *Computing and interpreting the adjacency spectrum of traffic networks*, 'Journal of Computational and Applied Mathematics', Elsevier, Amsterdam, 12&13

Ord J. (1975), *Estimation methods for models of spatial interactions*, 'Journal of the American Statistical Association', Taylor & Francis, Abingdon, 70

Pace R., LeSage J. (2004), *Chebyshev approximation of log-determinants of spatial weight matrices*, 'Computational Statistics and Data Analysis', Elsevier, Amsterdam, 45

Paelinck J. (2012), *Some challenges for spatial econometricians*, paper presented at the 2nd International Scientific Conference about Spatial Econometrics and Regional Economic Analys is, University of Lodz, Poland

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

Smirnov O., Anselin L. (2001), *Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach*, 'Computational Statistics and Data Analysis', Elsevier, Amsterdam,, 35

Smirnov O., Anselin L. (2009), *An O(N) parallel method of computing the log-Jacobian of the variable transformation for models with spatial interaction on a lattice*, 'Computational Statistics and Data Analysis', Elsevier, Amsterdam, 53

Walde J., Larch M., Tappeiner G. (2008), *Performance contest between MLE and GMM for huge spatial autoregressive models*, 'J. of Statistical Computation and Simulation', Taylor & Francis, Abingdon, 78

Zhang Y., Leithead W. (2007), *Approximate implementation of the logarithm of the matrix determinant in Gaussian process regression*, 'J. of Statistical Computation and Simulation', Taylor & Francis, Abingdon, 77

Streszczenie

WYBRANE WYZWANIA STATYSTYKI PRZESTRZENNEJ DLA EKONOMETRYKÓW PRZESTRZENNYCH

Artykuł prezentuje wybrane, niestandardowe statystyki przestrzenne oraz zagadnienia ekonometrii przestrzennej. Rozważania teoretyczne koncentrują się na wyzwaniach wynikających z autokorelacji przestrzennej, nawiązując do pojęć Gaussowskiej zmiennej losowej, topologicznych cech danych georeferencyjnych, wektorów własnych, filtrów przestrzennych, georeferencyjnych mechanizmów generowania danych oraz interpretacji efektów losowych.

10.2478/v10103-012-0028-4

MACIEJ JEWCZAK*

Determinants for Spatial Location of Pharmacies

Abstract

The topic of drug reimbursement is an important subject when one makes a decision on the construction of the reform of the health sector. Any change in the reimbursement list ends with a hot debate in the media and in everyday life. Incomprehensible pricing strategies used by pharmacies, force patients to seek those places that offer the necessary medication at the lowest possible price. Recognizing the economic opportunities for a profitable business, in recent years, a significant increase in the number of pharmacies is observed, and therefore, the number of these entities makes the process of selling drugs, especially those from the reimbursement list, almost impossible to control.

The article aims to reveal the spatial dependence for the pharmaceutical market on the example of pharmacies in poviat districts of Poland. An attempt is made to assess the prevalence of spatial dependence between the number of pharmacies and other determinants indicating health care resources, ageing process and the state of health of Poles. The summary of the study is to build a spatial model with its diagnosis for the number of pharmacies according to various socio-economic factors.

1. Introduction

In recent studies, both theoretical and empirical, the distinguishing between health economics and health care economics should be considered. As far as the health economics are concerned it is the scientific discipline that treats health as an economic issue. Following that definition, health economics relates

* University of Łódź

to the process of manufacturing, exchange and consumption of health services. The issues of health economics continuously evolve under the influence of both internal and external surroundings of health care system and the health consideration of the population. On the other hand, the health care economics interests in the analyses of the health care functioning and the manner of financing of the health services in different organizational health care system types. Contemporary research in the field of health care economics, in major part, considers mostly the characteristics of the connection between the condition of national economy or/and the area of health care and health (in a broad meaning). Due to the diversity of the health care system problems the need for the interdisciplinary research and the use of appropriate research instruments induces.

The health economics is the science of allocating resources to the health system and within the system. In another words, health economics determines the subject of interest for economists working in this field, as well as, the methods of application of economic principles in health care.

In practice, several characteristic approaches to analysis can be used. One can highlight many important attributes of the economy, but in health economics three should be noted.

First of all, the scarcity of social resources. The classical economic analysis is based on the assumption that individuals must resign from a certain part of one resource in exchange for another. This means that on the national level, the growth in health expenditure to GDP results in reduction of other expenses. The opportunity cost (cost of giving up to get something else) of health care can be substantial. While most, pay attention to the monetary costs of goods and services, economists consider time as the most important scarce resource. Individuals sell time in exchange for remuneration, and most probably would refuse to work overtime, even if offered the rate of pay higher than normal, because it is not profitable. In a similar vein, many individuals resign from the use of admission free health care services because the costs of arrival at the health establishment and waiting for the service are too high.

Rational decision-making is another important attribute. Typically, economists examine economic problems of human behavior, assuming that the individual makes a rational decision. While rationality is defined as making the best possible choice to meeting the objectives of the limitations of resources, some of the individual’s behavior in the health care system may seem irrational. But when it comes to disputes about the rational behavior, economists often point out that the so-called irrational behavior often makes sense, but only if the achieved benefits are properly understood. The important characteristic of recent research in health care is the use of models in the analyses. In economics, models are developed to illustrate the ongoing or future possible processes, though should be understood as a reflection of reality. However, the models can be useful.

This article focuses mainly on the quantitative analyses of the health care system characteristics and attempts to apply the spatial statistics and spatial model in health economics.

2. Methods

The rapid development of the methodological principles and their application enabled the extensive use of the spatial econometrics methods and models in economic research in many other scientific fields, for instance: labor market economic growth, social interactions, environmental protection and health care.

In the economical spatial studies, the impact of exogenous variables on the endogenous variable must also include an interactive combination between the observations. This follows from the fact that space is not consisting of mutually insulated units. The spatial interaction between two objects may also affect other objects. It should be noted that according to the Tobler’s law, the closer the objects are geographically, the spatial interactions are more significant.

2.1. Testing the spatial dependence

The term of spatial autocorrelations refers to spatial clustering of similar values and their interdependence or interactions in reference to the geographical location of the objects. The study of interdependence relationships in geographic space requires the assumption on the existence of the functional relationship between the values of observed variables (Anselin 1988, p. 11). By definition, this means a lack of independence between the observations and the direct application of the Tobler’s law.

Spatial autocorrelation is a degree of correlation of the observed values of variable in a given location with the values of the same variable in another location. This means that the tested variable at the same time determine and is determined by its implementation in other locations. When testing for spatial dependence, two types of relations are considered: positive and negative autocorrelation. Confirming the positive autocorrelation, means in terms of location, the spatial accumulation: high or low values of observed variables. On the other hand, negative autocorrelation can be interpreted as the reverse of the positive autocorrelation - high values of observed variables adjoin to low and vice versa (Suchecki 2010, pp. 103-105).

There are several types of indicators for testing the spatial dependence. The most commonly used statistic is Moran's I (Cliff and Ord 1973), It is used to test the presence of global spatial autocorrelation according to the scheme described by standardized weights matrix **W**. Let us consider the variable x of observed values x_i in n different regions (i = 1, 2, ..., n). Then the value of the Moran's I statistics can be describe as follows (Suchecki 2010, p. 113).

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \qquad (1)$$

where: n – number of observations, x_i, x_j – values of x variable in locations i and j, $\bar{x}$ – mean value of x variable, w_{ij} – elements of spatial weights matrix **W**.

While testing for spatial autocorrelation, a structure of hypothesis is examined, the null hypothesis for lack of spatial dependence against the alternative hypothesis for occurrence of spatial dependence. If the adjoined spatial objects are similar in reference to the descriptive characteristics, forming spatial clusters, then the value of Moran's I statistics is positive. If the adjoined spatial objects are varied (the spatial structure is regular, no clusters are formed) then the value of Moran's I statistics is negative. The Moran's I statistics ranges from (-1) to 1. For better visualizing of the type of Moran's I statistics, scatterplots are created and statistical significance graph is analyzed, the percentage of permutations for spatially random layout of variables is calculated. On this basis, it can be concluded about the existence or absence of spatial autocorrelation. The value of probability (p-value) is called the pseudo-significance level and is the ratio of the number of permutations for which $I_i > I_0$ to the number of all permutations made plus 1. The greater the p-value is, the less likely is the actual presence of autocorrelation.

Apart from the need to study global spatial autocorrelation, the literature indicates to obtain a detailed picture of the phenomenon of spatial dependence. Therefore, local indicators for spatial association analysis (Anselin 1995, pp. 93–115) (LISA) should be performed. It involves the study of correlation between the values of the variable in particular location in comparison to locations adjoined. Local Moran's I statistics I_i are calculated as follows (Suchecki 2010, p. 123):

$$I_i = \frac{(x_i - \bar{x})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \sum_{j=1}^{n} w_{ij}(x_j - \bar{x}) \tag{2}$$

here: n – number of observations, x_i, x_j – values of x variable in locations i and j, $\bar{x}$ – mean value of x variable, w_{ij} – elements of spatial weights matrix $\mathbf{W}$.

2.2. Spatial weights matrix W

In the construction of the measures of the spatial interactions, the spatial weights play a fundamental role. The spatial weights form the spatial weights matrix **W** and are calculated on the basis of distance or neighborhood matrices. The weight matrices can be constructed with the assumptions of different types and orders of contiguity (Suchecki 2010, pp. 33-34). The weight matrices are usually symmetric and in the analysis of spatial interaction, row standardization of the elements is assumed. This involves creating the transformed matrix, in which the sum of the elements in each row equals to 1. The values of the matrix elements are standardized with a closed interval <0,1>. For the construction of statistical measures, the standardization of the elements of the weights matrix is highly desirable, because of the possibilities of comparing different spatial processes and different models, therefor it is easier to interpret the processes of spatial autocorrelation and autoregression.

2.3. Spatial modeling

Spatial modeling improves the construction of the econometric model. If the analysis starts from the simple linear regression model, the application of cross-sectional sample as a localized data requires taking into account the spatial interactions that may occur between the studied units, which are expressed through the introduction of the model matrix of weights **W**. The interactions may relate to the endogenous variables – spatial autoregression is assumed, exogenous variables – cross spatial regression is assumed and random component – spatial autocorrelation of errors is assumed (Suchecki 2010,

p. 239). For the purpose of this article Spatial Autoregressive and Spatial Error Model are further described.

Spatial Autoregressive Model (or Spatial Lag Model) (Arbia 2006) assumes that the values of the endogenous variable in one location are dependent on spatially lagged mean values of the endogenous variable in adjoin locations. Formally, SLM model can be described as follows (Suchecki 2010, pp. 248-250):

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\ \ \boldsymbol{\varepsilon}{:}\mathrm{N}(\mathbf{0},\sigma^2\mathbf{I}) \tag{3}$$

where: $\mathbf{y}$ – endogenous variable, $\mathbf{X}$ – exogenous variables matrix, $\boldsymbol{\beta}$ – vector of structural parameters, $\mathbf{W}$ – spatial weights matrix, ρ – spatial autoregression parameter, $\mathbf{Wy}$ – spatially lag endogenous variable, $\boldsymbol{\varepsilon}$ – independent random component. In SLM models the significance of ρ parameter is tested.

Spatial Error Model can be tested, when in regression an autocorrelation in linear random component is assumed:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi},\ \ \boldsymbol{\xi} = \lambda\mathbf{W}\boldsymbol{\xi} + \boldsymbol{\varepsilon},\ \ \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0},\sigma^2\mathbf{I}) \tag{4}$$

where: $\mathbf{y}$ – endogenous variable, $\mathbf{X}$ – exogenous variables matrix, $\boldsymbol{\beta}$ – vector of structural parameters, $\mathbf{W}$ – spatial weights matrix, λ – spatial autocorrelation parameter, $\mathbf{W}\boldsymbol{\xi}$ – spatially lag error (mean error from adjoin locations), $\boldsymbol{\varepsilon}$ – independent random component. In SEM models the significance of λ parameter is verified. SEM model assumes the existence of spatial interactions (autocorrelation), caused by random factors (not included in modeling) or measurement errors (Suchecki 2010, p. 250).

3. Data set and research assumptions

3.1. Data source and specification

The source of data for the analysis is Local Data Bank of Central Statistical Office[1]. At the time of constructing the research the most current data was dated to 2010. The data used in the analysis was gathered on the NUTS 4[2]

[1] http://www.stat.gov.pl/bdlen/app/strona.html?p_name=indeks [day of access 14.07.2012].

[2] In accordance to the Nomenclature of Units for Territorial Statistics.

level – for 379 Polish poviats. As a subject of the research the number of pharmacies in Poland was accepted and a data set of potential explanatory variables was considered. The first group of explanatory variables was connected strictly with the health care and included:

- number of medical establishments,
- number of doctors, pharmacists.

The second group of explanatory variables included:

- gross monthly average income,
- number of people at age:
  - working,
  - pre-working,
  - post-working,
- number of people threatened with:
  - work environment,
  - work nuisance,
  - mechanical factors.

3.2. Main and specific objectives

The main objective of the article was to identify the spatial dependence on the example of the number of pharmacies in Polish districts. Apart from the main objective, some specific objectives were assumed as well. An attempt to verify the presence of spatial dependence between the number of pharmacies and explanatory variables[3] was made. As a summary of the research a spatial model for the number of pharmacies was designed and specified, depending on various factors.

3.3. Research hypotheses

For the purpose of the research hypotheses were formed. Firstly, if a patient goes to the doctor/medical establishment, then the pharmacy should be located close to that doctor/medical establishment. Secondly, people at post-

[3] Group consisting of data for: gross monthly average income, number of doctors, number of medical establishments, number of people at post-working age, number of people at risk because of the work environment.

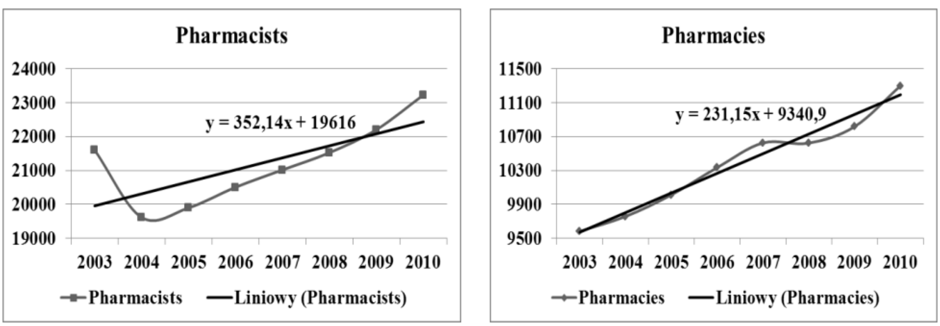
work age need an easy access to medicines. Thirdly, people who work and claim to be threatened because of their work environment need access to medicines.

4. Results

4.1. Baseline study

In the period 2003-2010, an increase in the number of pharmacies and their employees has been observed. The average growth rate noted for the number of pharmacies and the number of pharmacists amounted to 2.38% and 1.03%, respectively. The figure 1 below presents the tendencies observed in those time series and indicates that annually the number of pharmacies increased by 231,15 objects on average, and the number of pharmacists increased by 352,14 employees on average, ceteris paribus. Both parameters for the time variable *t* of the linear trend functions were significant.

Figure 1. Number of pharmacists and pharmacies



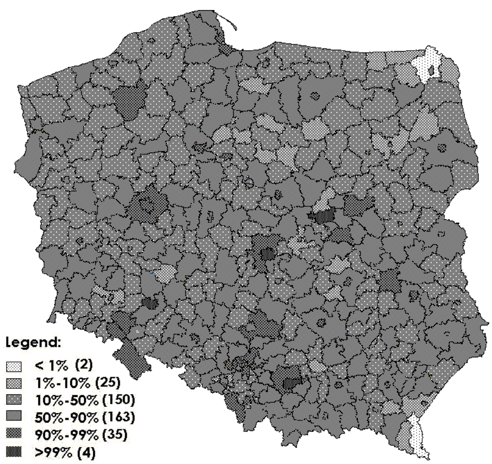
Source: developed by author, on the basis of CSO data in Microsoft Excel Software.

Comparing the ranks, a regularity is observed: on average, there are 2 pharmacists in a pharmacy, which is consistent with the guidelines of Ministry of Health.

4.2. Distribution of individual variables

Identifying a pattern for spatial autocorrelation is possible from the spatial distribution of analyzed variables. The figure 2 presents the geographical distribution of the number of pharmacies in Poland in 2010.

Figure 2. Spatial distribution of number of pharmacies variable – percentile map for 2010 data



Analyzing the distribution of number of pharmacies, areas of similar amounts of pharmacies can be identified. The brightest colors indicate that the lowest amounts of number of pharmacies can be observed in north-east and north-west Poland. In central and southern part of Poland, medium values were noticed. Outliers were possible to identify – dark and denser patterns specifies poviats of highest number of pharmacies.

Legend: dotted intensity indicates the scale of the phenomena – the higher values of the variable the darker and the denser pattern.

Source: developed by author, on the basis of CSO data in GeoDa 0.95.

4.3. Spatial autocorrelation

The spatial distribution map indicated that the location of pharmacies in Polish poviats was arranged in clusters of similar variable. To verify the hypothesis for spatial dependence, firstly, univariate Moran's *I* statistics were calculated and results are presented in table below.

Table 1. Univariate Moran's *I* for global spatial autocorrelation

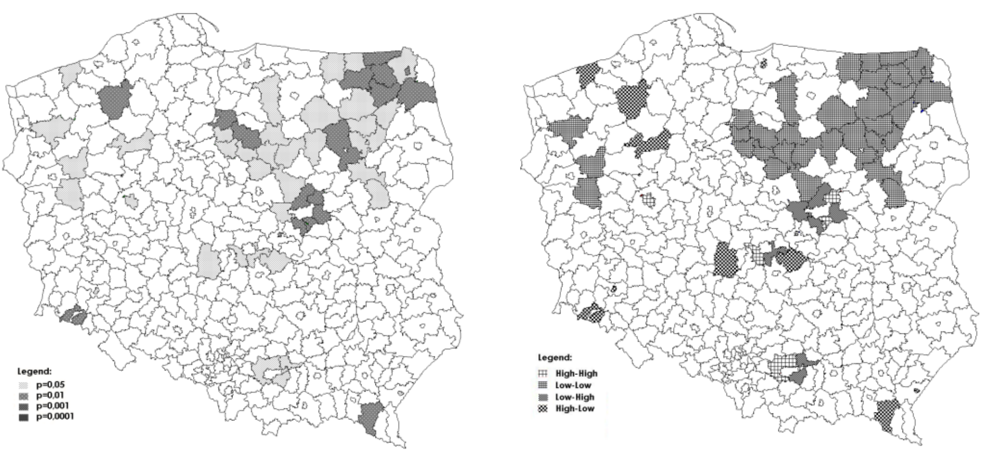
Variable	**Univariate Moran's I**	**p-value for α=0,05**	**Type of spatial autocorrelation**
number of pharmacies	0,0543	0,037	Positive
number of medical establishments	0,0465	0,084	None
number of doctors	-0,01	0,485	None
number of people at risk because of the work environment	0,207	0,001	Positive
number of people at post-working age	0,038	0,078	None
gross monthly average income	-0,037	0,152	None

Source: developed by author, on the basis of CSO data in GeoDa 0.95.

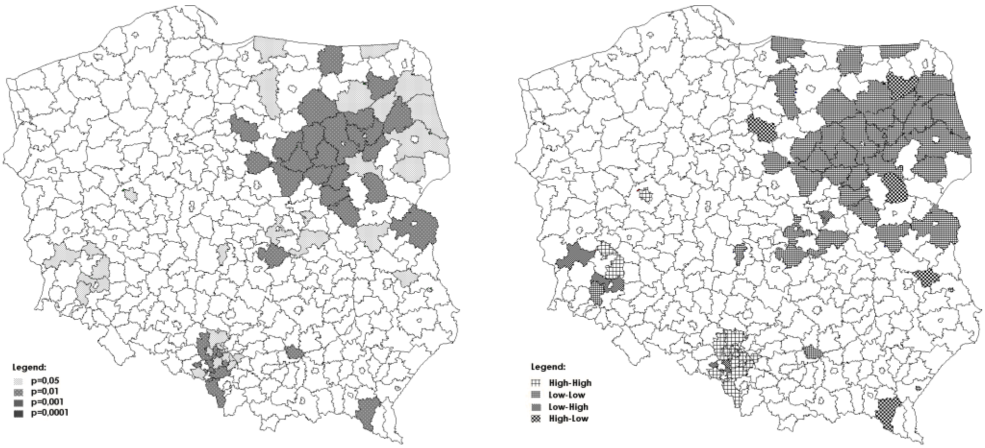
As the results show, only for the number of pharmacies and the number of people at risk due to the work environment the values of spatial autocorrelations were significant – assuming the 5% level of error, there were reasons for rejecting the null hypothesis in favor of the alternative. The results allowed to conclude that in both cases the correlations were positive, so there should be clustering of similar values of mentioned variables spatially observed. That led to further investigation and local indicators for spatial association *LISA* were calculated.

Figure 3. Local indicators for spatial association – significance and clusters maps

For number of pharmacies



For number of people at risk because of the work environment



Legend: maps on the left – significance maps – dotted intensity indicates areas of significant LISA statistics; maps on the right – clusters maps – checkered patterns indicate the neighboring areas of specific values of the variable.

Source: developed by author, on the basis of CSO data in GeoDa 0.95.

In both cases, it was possible to indicate the areas of specific values of the variables. The regions with a low number of pharmacies were pointed out in north-east Poland. The similar situation occurred for the number of people at risk due to the work environment. These results led to conclusion that between the endo- and exogenous variables a bivariate spatial dependence might occur. Following that bivariate Moran's *I* were calculated with reference to the number of pharmacies.

Table 2. Bivariate Moran's *I* for global spatial autocorrelation

Variable	Bivariate Moran's I	p-value for α=0,05	Type of spatial autocorrelation
number of medical establishments	0,04	0,25	None
number of doctors	0,00	1,00	None
number of people at risk because of the work environment	0,04	0,18	None
number of people at post-working age	0,04	0,11	None
gross monthly average income	-0,02	0,60	None

Source: developed by author, on the basis of CSO data in GeoDa 0.95.

As it turned out, none of the variables shown spatial autocorrelation of bivariate type with the endogenous variable – *p*-values for Moran's statistics indicated that there is no reason to reject the hypothesis of lack of spatial dependence.

4.4. Spatial modeling

Suggestion has been made that the methods of spatial econometrics can be used for researching geographic health care data. As a summary of the considerations of the possible determinants for the number of pharmacies, an initial model has been proposed, which was an attempt of a cross-spatial at time point analysis. The variables were selected on the basis of values of Pearson's correlation and the general model form was adopted as follows:

$$L_A = f(L_{ZOZ}, L_{LEK}, L_{ZSP}, L_{WPOP}, WB_{PM}, \varepsilon) \quad (5)$$

where:

Table 3. List of variables, their symbols and their impact on the endogenous variable

Symbol	Variable	Impact
L_A	number of pharmacies	N/A
L_{ZOZ}	number of medical establishments	+
L_{LEK}	number of doctors	+
L_{ZSR}	number of people at risk because of the work environment	+
L_{WPOP}	number of people at post-working age	+
WB_{PM}	gross monthly average income	+

Source: developed by author.

Having confirmed the existence of spatial dependence, it was advisable to assume the existence of the dependence in model construction. In this purpose the general model has been estimated with Ordinary Least Squares Method (OLS) and diagnosed for spatial dependence.

Table 4. Regression summary of output: Ordinary Least Squares estimation

Dependent Variable :	**LA**	Number of Observations:	379
Mean dependent var :	29.8074	Number of Variables :	5
S.D. dependent var :	42.8941	Degrees of Freedom :	374
R-squared :	0.970746	F-statistic :	3102.64
Adjusted R-squared :	0.970433	Prob(F-statistic) :	0
Sum squared residual:	20399.5	**Log likelihood :**	**-1293.07**
Sigma-square :	54.5442	**Akaike info criterion :**	**2596.15**
S.E. of regression :	7.3854	**Schwarz criterion :**	**2615.84**
Sigma-square ML :	53.8246		

S.E of regression ML: 7.33653

Variable	**Coefficient**	**Std,Error**	**t-Statistic**	**Probability**
CONSTANT	1,885	0,690	2,730	0,007
LZOZ	0,202	0,023	8,611	0,000
LLEK	0,019	0,003	5,822	0,000
LZSP	0,002	0,000	3,886	0,000
LWPOP	0,001	0,000	12,997	0,000

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 12.41284

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	1410.908	0.0000000

DIAGNOSTICS FOR HETEROSKEDASTICITY

TEST	DF	VALUE	PROB
Breusch-Pagan test	4	819.2369	0.0000000
Koenker-Bassett test	4	146.3062	0.0000000

SPECIFICATION ROBUST TEST			
TEST	DF	VALUE	PROB
White	14	250.2937	0.0000000

DIAGNOSTICS FOR SPATIAL DEPENDENCE			
TEST	**MI/DF**	**VALUE**	**PROB**
Moran's I (error)	**0.243092**	**7.2472706**	**0.0000000**
Lagrange Multiplier (lag)	**1**	**7.9233089**	**0.0048802**
Robust LM (lag)	**1**	**0.7817883**	**0.3765948**
Lagrange Multiplier (error)	**1**	**50.0300481**	**0.0000000**
Robust LM (error)	**1**	**42.8885274**	**0.0000000**

Source: developed by author, on the basis of CSO data in GeoDa 0.95.

After the first estimation, the coefficients of explanatory variables turned out to be significant – assuming the 5% level of error. The impact of each exogenous variable was consistent with previously made assumptions. As the results showed, the proposed model defined the complete changeability of endogenous variable in 97% and the standard error of regression indicated the +/- 7,38 miscalculation. What's more, the multicollinearity condition test did not indicate for the linearity problem, but unfortunately, the assumption for normality of errors was not fulfilled. Test for heteroskedasticity indicated problem of heterogeneity of variance. There is a possibility that the assumptions for normality of errors and homogeneity of variance could be granted by modifying the general model with spatial component.

As the diagnostics for spatial dependence module showed, introducing the spatial effects[4] to the model indicate the existence of spatial dependence. The Moran's *I* statistics was highly significant. This confirmed the necessity to use a model with spatial component. Looking at the values of Robust Lagrange Multiplier probability, it was highly recommended to select a Spatial Error Model (the Spatial Lag Model was insignificant). SEM model was constructed and estimated with Maximum Likelihood estimation (table 5).

[4] Spatial weights matrix **W** was used – row-standardized, first order, queen contiguity matrix.

Table 5. SEM summary of output: Spatial Error Model - Maximum Likelihood Estimation

Dependent Variable	: **LA**	Number of Observations	: 379
Mean dependent var	: 29.807388	Number of Variables	: 5
S.D. dependent var	: 42.894081	Degree of Freedom	: 374
Lag coeff. (Lambda)	: 0.443498		
R-squared	: 0.975087	R-squared (BUSE)	: -
Sq. Correlation	: -	**Log likelihood**	**:-1270.366036**
Sigma-square	: 45.836938	**Akaike info criterion**	**: 2550.73**
S.E of regression	: 6.7703	**Schwarz criterion**	**: 2570.419753**

Variable	**Coefficient**	**Std.Error**	**z-value**	**Probability**
CONSTANT	2,966	0,804	3,687	0,000
LZOZ	0,200	0,022	8,886	0,000
LLEK	0,023	0,003	7,900	0,000
LZSP	0,002	0,000	4,860	0,000
LWPOP	0,001	0,000	12,916	0,000
LAMBDA	**0,443**	**0,061**	**7,219**	**0,000**

REGRESSION DIAGNOSTICS

DIAGNOSTICS FOR HETEROSKEDASTICITY

TEST	**DF**	**VALUE**	**PROB**
Breusch-Pagan test	4	13020.8	0.0000000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

TEST	**DF**	**VALUE**	**PROB**
Likelihood Ratio Test	**1**	**45.41532**	**0.0000000**

Source: developed by author, on the basis of CSO data in GeoDa 0.95.

After the final estimation, the coefficients of the explanatory variables turned out to be significant, assuming the 5% level of error. The impact of each exogenous variable was consistent with previously made assumptions. The most important fact was the significance of λ parameter, which confirmed the existence of spatial dependence and indicated the influence of random factors or measurement errors on the number of pharmacies in Poland. Diagnosis for spatial dependence (Likelihood Ratio test) was significant, which indicated that application of SEM model for explaining the changes in the number of pharmacies in Poland in 2010 eliminated the problem of spatial autocorrelation in data, but did not deal with the problem of heterogeneity.

5. Discussion and Conclusions

The implementation of spatial interactions improved the fit of model:

Table 6. Criterions of fit for OLS and EM

Criteria[5]	**OLS**		**SEM**
Log likelihood	-1293,07	<	**-1270,37**
Akaike	2596,15	>	**2550,73**
Schwarz	2615,84	>	**2570,42**

Source: developed by author, on the basis of CSO data in GeoDa 0.95.

All of the criteria received for SEM model indicated the better usage of that model, comparing with the OLS. Unfortunately, due to the high heterogeneity of Polish poviats it was not possible to deal with the problem of heteroskedasticity. The model adjusted for spatial dependence (SEM) – eliminated the problem of spatial dependence in the data.

Apart from the strictly technical results and findings of the analyses, it should be emphasized that the methods of spatial econometrics can be widely used in health care analyses, for instance, in developing a tool/model for defining the determinants for the number of pharmacies in Poland. It was confirmed, that on average when five new medical establishments are found in a poviat, then a pharmacy appears in that region, as well. The methods for revealing the spatial dependence allowed to identify the areas of occurrence of spatial autocorrelation for the number of pharmacies and the number of people at

[5] It is not possible to compare the R^2 of OLS and SEM models, instead values of Log likelihood, Akaike and Schwarz criterions are used. Better model has higher values of Log likelihood, lower values of Akaike and Schwarz criterions.

risk due to the work environment. Moreover, the research confirmed the need to incorporate spatial interaction factor in the modeling of health care on the example of the number of pharmacies.

References

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

Anselin L., (1995), *Local Indicators of Spatial Association – LISA*, Geographical Analysis 27(2)

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

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

Suchecki B. (ed.), (2010), *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, C.H.Beck, Warszawa

Streszczenie

DETERMINANTY LOKALIZACJI PRZESTRZENNEJ APTEK

Tematyka refundacji leków to istotne zagadnienie przy podejmowaniu wszelkich decyzji dotyczących kształtu reform dla sektora opieki zdrowotnej. Każdorazowa zmiana wykazu leków refundowanych kończy się gorącą debatą w mediach i w codziennym życiu. Niezrozumiałe strategie cenowe, stosowane przez apteki, zmuszają pacjentów do poszukiwania tych miejsc, które oferują niezbędny lek w jak najniższej cenie. Dostrzegając możliwości dobrego biznesu ekonomicznego, na przestrzeni ostatnich lat, obserwuje się znaczący wzrost liczby aptek, a w związku z tym, liczebność tych podmiotów uniemożliwia pełną kontrolę w procesach sprzedaży leków, zwłaszcza tych z list refundowanych.

Celem artykułu jest wykazanie przestrzennych zależności obserwowanych na rynku farmaceutycznym na przykładzie liczby aptek w powiatach Polski. Podjęta została próba oceny występowania przestrzennych zależności pomiędzy liczebnością aptek a liczbą lekarzy, liczbą zakładów opieki zdrowotnej, jak również innych determinant wskazujących na proces starzenia się społeczeństwa i stanu zdrowia Polaków. Podsumowaniem badań jest próba budowy modelu przestrzennego i jego diagnoza dla liczby aptek w zależności od różnych czynników społeczno-gospodarczych.

10.2478/v10103-012-0029-3

EWA KUSIDEŁ[*], JOANNA GÓRNIAK[**]

Transport Availability vs. Development of Poland's Regions

Abstract

The article discusses the issue of transport availability, taking into account road and rail infrastructure, as compared with socio-economic development in specific regions of Poland. The issue is studied from the theoretical and empirical point of view. The nature of the of the research is to use of existing methods for new dataset, namely the transport development level is estimated based on a composite indicator of development and degree of convergence among regions is estimated using classical regional convergence tests. The main goal of the study is to assess the degree of transport availability diversification among regions (voivodships) of Poland in comparison with their socio-economic development. The hypothesis assumes that there is strong correlation between transport availability and development of the regions and therefore similarity in the convergence inference.

1. Introduction

The issue of transport development evaluation in regions is frequently raised as one of components of comprehensive studies on regional economy, including those commenced in 1993 by the Gdańsk Institute for Market Economics and concerning investment attractiveness of voivodships. Among considered evaluation criteria was transport availability of regions, including

[*] Ph.D., University of Łódź

[**] University of Łódź

their road and rail infrastructures, distance to airports, seaports, and border crossings.

The main objective of this study is to prove that there is a relationship between transport availability and economic growth (measured by the GDP per capita) in specific voivodships of Poland. If such a relationship exists, and taking into consideration numerous proofs of the lack of economic convergence among voivodships of Poland, an attempt will be made to verify a hypothesis that there is no convergence for a transport availability measure either. To that end, a synthetic indicator of transport availability of Poland's is built (using selected variables characterizing road and rail transport), which is discussed in the first chapter of this article. The verification of the convergence hypothesis applies classical convergence tests described in the second chapter and employed in the third chapter to examine spatio-economic cohesion.

The study covers years 2004 to 2009. The upper limit of the period is defined by data limitation of regional accounts – at the time of writing the paper the latest data on regional GDPs were available for 2009. The lower limit of the study period is the year of Poland's accession to the European Union, since when the cohesion policy has been carried out in Poland in order to attain economic, social, and spatial cohesion. Spatial cohesion is increased by eliminating barriers to accessibility of regions, in particular peripheral ones, by means of better communications, which is synthetically expressed by a transport availability measure constructed in this article.

2. Transport Availability – the Essence, Issues, and Structure of a Composite indicator

Transport infrastructure can be analysed from three points of view. Firstly, it ensures connections among territorial units. Secondly, it provides material conditions for carrying out production activity without which such activity would be impossible. Thirdly, it necessitates the presence of production activity preconditions associated with the presence of human capital, i.e. appropriate supply of labour, properly qualified and educated labour force (Wojewódzka-Król 2002, pp. 14–15). The role of transport infrastructure in economic development and its specificity result in the fact that any irregularities in the process of its development adversely affect the economy. Therefore, efficient and effective functioning of that area is vital from the economic point of view.

As a rule, transport structures of specific territorial units are influenced by numerous factors, which leads to considerable differences in their development levels. Elements that diversify the transport system arrangement include, among others: geographical situation, degree of urbanization, location of industrial and tourist centres, international cooperation, level of technical and technological development.

An analysis of the transport position may use specific or synthetic measures, such as a transport availability indicator enabling to assess a transport system from the spatial point of view. Simple indicators of provision with transport infrastructure include: the length of motorways and other roads (meant for the movement of motor vehicles) and railroads, number of railway stations and airports, time needed to reach the nearest transport hub of an interregional network. Complex indicators have a more complicated structure. The essence of complex transport availability indicators is that they take into account spatial interactions (i.e. distance to be covered, time and cost of travel). The attractiveness of a region can be assumed to increase along with its size and, in turn, decrease along with an increase in distance, time or cost of travel (Michałowska 2007, pp. 74–75).

Transport Availability Indicator for Regions (Voivodships) of Poland

The structure of a transport availability indicator is not unambiguous and may cause numerous problems. Researchers dealing with that phenomenon (e.g. Koźlak 2007; Rosik, Szuster 2008) tend to base their work on slightly different variables, which seems justified in both methodological and economic terms. In general, variables describing road, rail, air, and water transport may be verified.

This study attempts to build a synthetic indicator of transport availability for voivodships of Poland, taking into account different diagnostic variables (compare Table 1). Priority is given to variables related to road transport. The study considers one variable characterizing rail transport (the length of utilized railroads in km per 100 square km). Thus, such an approach will contribute to an increase in the transport availability indicator for the Śląskie voivodship, where the density of railroad network is the highest in Poland.

In general, variables were selected based on the following premises:

- road transport is the most common mode of transport in Poland,
- inland water transport cannot function effectively in Poland due to the insufficient traffic capacity of rivers and irregularity of river traffic flow throughout the year,
- sea transport applies only to two regions (Pomorskie and Zachodniopomorskie), hence, application of variables that characterize that

mode of transport would be irrational and discriminate against regions without access to the Baltic Sea,

- domestic air transport is of no particular importance as not all regions have passenger airports (the situation has slightly changed in the year of EURO 2012 – however, it is a year not covered by analyses in this article).

Table 1. List of diagnostic variables considered in the measurement of transport availability

No.	Name of variable	Conversion factor
1.	Public roads with hard surface	100 square km
2.	National roads	100 square km
3.	Provincial roads	100 square km
4.	Poviat roads	100 square km
5.	Communal roads	100 square km
6.	Motorways	100 square km
7.	Expressways	100 square km
8.	Bridge-type structures on public roads (bridges, flyovers)	100 square km
9.	Tunnels and underpasses on public roads	100 square km
10.	Utilized railroads	100 square km
11.	Regular bus transport lines (domestic)	100 square km
12.	Individual motorization indicator	1000 inhabitants
13.	Lorry number indicator	1000 inhabitants
14.	Paid passenger transport	kilometer

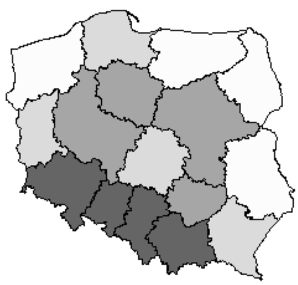
Source: own work based on Central Statistical Office (CSO) data.

Variables listed in the above table have features of stimulants, and thus, it is expected that their increase will positively affect the synthetic variable characterizing transport availability. Nevertheless, an excessive increase in levels of variables specified in the final lines of Table 1 may contribute to an increase in an adverse phenomenon of transport congestion[1] – hence, an increase in the specified diagnostic features will be desirable only to a certain limited extent. The calculation of the synthetic transport availability measure applied the development pattern method. The selection of diagnostic features was tested taking into account their universality, importance (Malina 2004, pp. 35–37), and variability. Moreover, two kinds of availability were considered: passenger transport availability and goods transport availability. In the case of both the indicators, several variables were identified characterizing both of them (those

[1] Transport congestion may be understood as overcrowding, overloading, accumulation. In other words, congestion is the occurrence of an excessive number of buyers or users of transport services at the same place and time (based on: Tundys B., 2008, pp. 128–129).

were variables 1–10 and 13 for goods transport availability and, additionally, variables 11, 12, and 14 for passenger transport availability). Maps provided below present diversification of the synthetic availability indicator over the period of six years (2004–2009).

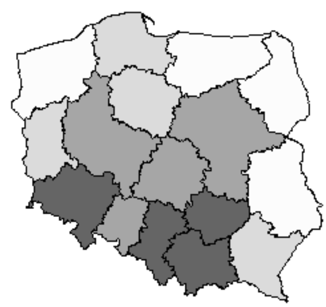
Map 1. Passenger transport availability indicator in 2004



Map 2. Passenger transport availability indicator in 2009

Map 3. Goods transport availability indicator in 2004

Map 4. Goods transport availability indicator in 2009



Source: own work based on Local Data Bank – Central Statistical Office (LDB – CSO) data.

The above maps indicate that regions with the highest indicator of transport availability include: Dolnośląskie, Małopolskie, and Śląskie (for both passengers and goods). The lowest passenger and goods transport availability indicator values characterize the following voivodships: Lubelskie, Podlaskie, Warmińsko-Mazurskie, and Zachodniopomorskie. At the same time, it can be observed that changes over the period of 2004–2009 (quartile-based) were minor – low mobility of the goods transport availability indicator provides premises to infer the absence of the indicator's convergence among the studied regions.

3. Regional Convergence – Types and Ways of Measurement

In free translation, convergence means similarity or gathering. Convergence is a process in which various units, initially bearing no similarity, become closer and more similar to one another. In studies, subjects tend to be countries or regions (thus, the name: regional convergence). A term opposite to convergence is divergence, i.e. lack of similarity.

The concept of convergence was formed based on the theory of systems in the 1940s and 1950s. It developed rapidly in the 1980s. Initially, convergence studies consisted in analysing the theory of the socialist and capitalist systems' development becoming similar. Over time, researchers began to refer to the concepts of economic growth, systems of financial markets or lifestyles, consumption, state administration (Woźniak et al. 2009, p. 53).

An important moment for the convergence theory was identification of classical convergence. The following three types developed within that kind of convergence:

- β-type convergence,
- σ-type convergence,
- γ-type convergence.

The fundamental element of β-type convergence is the study of the so called catch-up effect. Furthermore, β-type convergence can be considered from two different points of view: absolute and conditional. Absolute β convergence assumes that units pursue the same state of long-term balance, while the essence of conditional β-type convergence is the fact that specific units aim to achieve their own convergence levels.

σ-type convergence occurs when diversification of values of an analysed variable among countries or regions decreases in consecutive units of time. Diversification is examined by means of measures of variation (standard deviation, coefficient of variation) or concentration (e.g. Gini coefficient).

β convergence is a necessary but insufficient condition for σ-type convergence to occur. That means that the occurrence of β-type convergence is not tantamount to the presence of σ-type convergence (Wolszczak-Derlacz 2007, p. 11).

γ-type convergence aims to inform whether a given country or region has changed its ranking in respect of an examined variable. In order to perform verification, each territorial unit is assigned a rank. The examination of that kind of convergence may use Kendall's rank coefficient of concordance or Pearson's coefficient. Values of Kendall's coefficient range from 0 to 1. Convergence will occur when the value of Kendall's coefficient is close to zero (Boyle, McCarthy

1997, pp. 257-264). On the other hand, Pearson's coefficient ranges from -1 to 1. When values of the coefficient are negative or insignificant, convergence occurs.

4. Analysis of Spatio-Economic Convergence among Regions of Poland

When convergence is discussed, researchers usually mean its economic dimension reflected in decreasing disproportions in measures that are based on the gross domestic product. However, the issue of convergence may be extended to cover numerous other aspects of life. In particular, the cohesion policy carried out within the framework of the EU regional policy strives to increase the levels of not only economic but also social and territorial (spatial) convergence. Social convergence means decreasing differences in human and social capital, and standard of living[2]. Spatial cohesion is increased by eliminating barriers to accessibility of regions, especially peripheral ones, through better communications and connections with centrally located areas[3]. This article studies two kinds of convergence:

- economic – measured using one of the basic measures of national income (regional income, in this case) determining economic development, i.e. the gross domestic (regional) product,
- spatial – measured by means of the synthetic transport availability measure (described in Chapter 1).

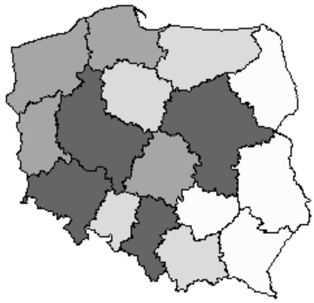
In the case of the GDP per capital, attention should be paid to significant differences in its value among specific voivodships of Poland – compare Maps 5–6. The gross regional product varies significantly among regions of Poland. A division into wealthier and poorer regions can be clearly observed. Regions characterized by significantly higher GDPs per capita are: Dolnośląskie, Mazowieckie, Śląskie, and Wielkopolskie provinces. Significantly lower GDPs per capita characterize regions located in the Eastern part of the country, i.e.: Lubelskie, Podkarpackie, Podlaskie, and Warmińsko-Mazurskie. Similarly to the transport availability indicator, there are no changes (quartile-based) in the arrangement of regions according to the GDP per capita – which is an argument against convergence of the GDP per capita among regions of Poland.

---

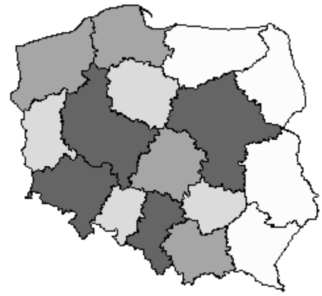
[2] They are measured by means of the unemployment rate or participation rate (employment rate) or indicators describing the standard of living, e.g. the HDI.

[3] Additionally, numerous other areas of convergence can be considered, of which the most important and very current, from the point of view of the EU strategy, is the convergence of sustainable development.

Map 5. GDP per capita in Poland in 2004



Map 6. GDP per capita in Poland in 2009



Source: own work based on LDB – CSO data.

There is a positive correlation of 0.5–0.6 between variables characterizing transport availability and economic growth measured by the GDP per capita. It can be assumed that the studied variables (the GDP per capita and transport availability indicator) will exhibit the same tendency as to their convergence (i.e. the actual convergence or divergence). Based on the mere analysis of maps it can be pointed out that regions characterized by a higher level of transport availability indicator achieved a higher level of economic growth than those having a lower transport availability indicator.

In order to formally examine the occurrence of cohesion of (passenger and goods) transport availability and economic growth among specific regions, σ-type and γ-type convergence was used. For σ-type convergence (for which the measure of diversification was the standard deviation), it was concluded that there is no such convergence (for both passenger and goods transport availability as well as economic growth). It is impossible, however, to state the occurrence of divergence as the parameter of the temporal variable in the trend model matching the diversification graphs of variables is statistically insignificant.

Figure 1. σ convergence – passenger transport availability

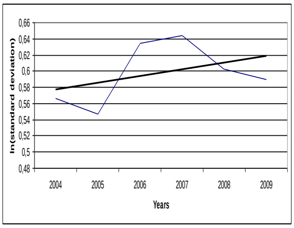


Figure 2. σ convergence – goods transport availability

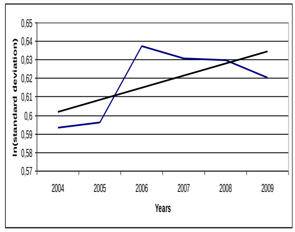
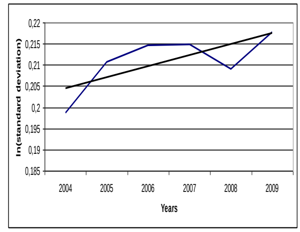


Figure 3. σ convergence – economic growth measured by GDP per capita



Source: own work based on CSO data.

In order to determine the occurrence of gamma-type divergence, Pearson's linear correlation coefficient was used. Correlation between the arrangement of voivodships in the first and last year of the analysis is high and statistically significant. That indicates the lack of changes in the arrangement of regions – i.e. the lack of gamma-type correlation – compare Table 2.

Table 2. Correlation for the arrangement of regions in two years of analysis: 2004 and 2009

Passenger transport availability indicator	Goods transport availability indicator	Economic growth measured by GDP per capita
0.994	0.997	0.968

Source: own work.

To sum up the methods applied to analyse the process of convergence, it can be said that the specified methods produced unanimous results based on which it can be stated that, taking into account both transport availability (for passengers and goods) and economic growth (measured by the GDP per capita), there was no convergence among regions of Poland in the 2004–2009 period. In consequence, the regions did not exhibit a tendency to become more similar in respect of the analysed variables.

5. Conclusions

Transport availability can be analysed at various geographical levels. This study takes into consideration the NUTS 2 territorial division – which, in the case of Poland, means the level of voivodships. The carried out study allowed to prove, in accordance with indications provided by literature, that the studied variables: transport availability indicators and economic growth measured by the GDP per capita show a mutual relationship. What is more, similar tendencies of changes in specific voivodships in respect of the analysed variables can be observed – regions characterized by high values of transport availability indicators (for passengers and goods) have higher GDPs per capita. A similar relationship occurs for less developed regions, i.e. those having lower values of passenger and goods transport availability indicators and low economic growth measured by the GDP per capita. The study proves the formulated hypothesis of the lack of convergence among regions in respect of (both types of) transport availability and GDP per capita.

Eventually, the conducted research indicates that there is a positive statistically significant correlation between economic well-being (measured by the GDP per capita) and transport availability (measured by the synthetic measure of development), while the analysis of convergence of both the variables proved that changes existing among regions in respect of those variables do not decrease. Particularly negative differences are observed for the Eastern Poland regions – whose weak position in terms of the GDP per capita and transport availability in 2004 did not improve in 2009. It ought to be recommended that special funds meant for the development of Eastern Poland in the 2007–2013 period be to a large extent allocated to infrastructural investments – due to their connection with economic growth as it is known that one of the main barriers to the development of less developed areas is the lack of modern transport infrastructure being among the fundamental conditions for development. Inadequately developed infrastructure adversely affects investment appeal of an area and the quality of life of its population. Therefore, appropriate development of transport infrastructure may prove to be the most important element of regional characteristics.

References

Bell G., Bowen P., Fawcett P. (1984). *The business of transport*. MacDonald & Evans Ltd, Plymouth

Boyle G.E., McCarthy T.G. (1997). *A simple measure of β–convergence,* Oxford Bulletin of Economics and Statistics, 59, 2

Michałowska M. (Ed.), (2009). *Efektywny transport – konkurencyjna gospodarka [Effective transport – competitive economy]*, Wydawnictwo Akademii Ekonomicznej im. Karola Adamieckiego w Katowicach, Katowice

Michałowska M. (Ed.), (2009). *Transport w gospodarce opartej na wiedzy [Transport in knowledge-based economy]*, Wydawnictwo Akademii Ekonomicznej im. Karola Adamieckiego w Katowicach, Katowice

Malina A. (2004). *Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarki Polski według województw [Multidimensional analysis of spatial diversification of Polish economy structure according to voivodships]*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków

Raport wykonany dla Ministerstwa Rozwoju Regionalnego [Report prepared for the Ministry of Regional Development]: T. Komornicki (Ed.), (2008). *Opracowanie metodologii liczenia wskaźnika międzygałęziowej dostępności transportowej terytorium Polski oraz jego oszacowanie [Methodology of computing the intramodal transport availability indicator for the territory of Poland and its estimation]*, Warszawa

Rosik P., Szuster M. (2008). *Rozbudowa infrastruktury transportowej a gospodarka regionów [Development of transport infrastructure and regional economy]*. Politechnika Poznańska, Poznań, Poznań

Spiekermann K., Neubauer J. (2002). *Accessibility and peripherality: concepts, models and indicators*, Nordregio

Wojewódzka-Król K. (Ed.), (2002). *Rozwój infrastruktury transportu [Development of transport infrastructure]*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk

Wolszczak-Derlacz J. (2007). *Wspólna Europa, różne ceny – analiza procesów konwergencji [Common Europe, different prices – analysis of convergence processes]*, Wydawnictwo CeDeWu, Warsaw

Woźniak M.G., Czuzhykov V.I., Lukianenko D.G. (2009). *Konwergencja modeli ekonomicznych. Polska i Ukraina [Convergence of economic models. Poland and Ukraine]*, Wydawnictwo Fundacja Uniwersytetu Ekonomicznego w Krakowie, Kraków

Zeliaś A. (2000). *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym [Dynamic taxonomic analysis of standard of living spatial diversification in Poland]*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków

Streszczenie

DOSTĘPNOŚĆ TRANSPORTOWA A ROZWÓJ POLSKICH WOJEWÓDZTW W ŚWIETLE ANALIZY KONWERGENCJI REGIONALNEJ

W artykule omówiono problem dostępności transportowej z wyróżnieniem infrastruktury drogowej oraz kolejowej na tle rozwoju społeczno-gospodarczego w poszczególnych województwach w Polsce. Zagadnienie to rozpatrywane jest zarówno w aspekcie teoretycznym, jak i empirycznym. Istotą badania jest wykorzystanie istniejących metod do nowego zestawu danych, a mianowicie poziom rozwoju transportu oszacowano na podstawie złożonego wskaźnika rozwoju, zaś stopień zbieżności pomiędzy regionami określono przy użyciu testów klasycznej konwergencji. Głównym celem badania jest ocena stopnia zróżnicowania dostępności transportowej pomiędzy województwami w Polsce w porównaniu do ich rozwoju społeczno-gospodarczego. Hipoteza badawcza zakłada, że istnieje silna korelacja pomiędzy dostępnością transportową i rozwojem regionów, co pociąga za sobą podobieństwo wniosków co do występowania konwergencji.

10.2478/v10103-012-0030-x

CHRISTIAN LE BAS[*], WILLIAM LATHAM[**], DMITRY VOLODIN[***]

Prolific Inventor Productivity and Mobility: A Western/Asian com-parison. Evidence from US Patent Data for 12 Countries[1]

Abstract

This paper provides new insights into the role of individual inventors in the innovation process. Individuals are central in this creative process because innovation is not simply a product of firms and organizations; it requires individual creativity (Rothaermel and Hess, 2007). We focus our analysis on prolific *inventors (a rich sub category of inventors) because they contribute so hugely to national invention totals (Le Bas* et al.*, 2010) and tend to produce inventions that have more economic value (Gambardella* et al.*, 2005; Gay* et al.*, 2008). Converging empirical evidence has established the significance of prolific inventors (Ernst* et al.*, 2000). Previous studies of prolific (or "key") inventors have focused more on the firms in which they work or on the industries in which the firms operate. Narin and Breitzman's (1995) seminal work on the topic is based on an analysis of only four firms in a single sector and a recent paper by Pilkington et al. (2009) uses only two firms. In contrast to these studies on small samples, we use a very large data set which includes thousands of inventors in thousands of firms from several countries.*

[*] Ph.D., Full Professor at the University of Lyon 2, France

[**] Ph.D., Full Professor at the University of Delaware, USA

[***] Ph.D., Full Professor, HDR Inc., Silver Spring, MD, USA

[1] An earlier draft of this paper was presented at the *Spatial Econometrics and Regional Economic Analysis* conference (4-5 June 2012, Lodz, Poland). The email address of the corresponding author is: Christian.Lebas@univ-lyon2.fr

1. Introduction

The core of the research is to investigate the role that mobility plays in the behaviour of prolific inventors. Labour mobility is a means for transferring knowledge (and newly created knowledge in particular) across countries and region (Saxenian, 2006). In general considerate seems that interregional mobility is weak. Breschi et al. (2010) find that inventors diffuse knowledge across social networks within regions but not across regions. Felsenstein, (2011) concludes that inventor mobility provides support for knowledge spillovers across agents and regions. The knowledge spillovers are important determinants of regional economic growth. Our approach is a little different: our goal is not to assess the rate and the direction of knowledge spillovers. Instead we attempt to account for the determinants of inventor mobility and measure its impact on inventor productivity. The scale, determinants and effects of inventor mobility have been analysed by Hoisl (2007 and 2009), Schankerman *et al.* (2006), Tratjenberg (2004) and Tratjenberg *et al.* (2006) among others. Hoisl, using European patents and a survey of 3049 German inventors, finds that an increase in inventor productivity, measured as the number of patents per inventor, decreases the number of moves. She tests the effect of inventor productivity on inventor mobility and finds that *more productive inventors are not more mobile*. For Hoisl (2007), a move increases productivity but an increase in productivity decreases the probability of observing a move. Schankerman *et al*. (2006) have studied the mobility of inventors using patents in the software industry in the US. Their findings are in accord with Hoisl's: they show that the very productive inventors have a decreasing probability of moving between assignees as their careers progress (Schankerman et *al.* 2006; 26).

We focus our research on prolific inventors. Previous papers have justified the identification of prolific inventors as those who have been issued at least 15 patents (Le Bas et al., 2010; Latham et al. 2011; Latham et al. 2012). In those papers we generally hypothesized that mobility of prolific inventors, as measured by their average numbers of inventions per year over their active inventive lives, affects both their productivity and the value of their inventions, measured as the numbers of citations a patent receives in the years after it is issued, positively. Our previous papers present evidence supporting these hypotheses for the five largest countries in terms of technological activity (the US, Japan, Germany, the UK, and France). Our data come from patents filed by inventors from each of the countries in the US Patent and Trademark Office during the period from 1975 to 2010. While we focus on the activities of prolific inventors, our data set includes all inventors so the unique characteristics of prolific inventors can be identified.

In Le Bas et al. (2010), Latham et al.(2011), and Latham et al.(2012) we extended the literature to prolific inventors in multiple countries, using alternative indicators for different kinds of mobility. In these papers we measure inventor mobility in two dimensions: across companies ("interfirm mobility") and across regions ("geographic mobility"). For each country we estimate equations for productivity, value and mobility. Our results for the determinants of inventor productivity, mobility and invention value in Germany, France, and the UK show (Latham *et al.* 2011): 1) In all three countries productivity is positively related to inter-firm mobility and temporal concentration of patenting is also positively related to productivity. However, for France, productivity is negatively related to geographic mobility, 2) For all three countries the value of inventions (as measured by citations per patent) is positively related to productivity. For UK and Germany the equations show consistent positive and significant relationships between value and inter-firm mobility (by contrast the coefficient is not significant for France), 3) The mobility equations show that productivity is positively associated with mobility and value is negatively associated with it. Inventor technological specialization is also negatively related to inter-firm mobility while the temporal pattern of inventing seems to be unrelated.

This paper extends the previous results in an important dimension. By focusing on Asian countries (China, Japan and Korea and Taiwan) in addition to North America and Western Europe, we are able to test whether the determinants and the effects of inventor mobility are the same in Asia as they are elsewhere. In the two last decades the three main Asian countries after Japan (China, Korea, and Taiwan) have caught up with the rest of the developed world by targeting the technologically most progressive industries (Fagerberg and Godinho 2006), and by creating R&D industrial clusters of sufficient size. They have established and developed significant domestic capabilities, first for imitation and then for innovation (Ernst, 2005; Lundvall et al., 2009). They have developed coherent national systems of innovation and are becoming important international contributors to innovation (Dodgson and Gann 2010). As a consequence, populations of researcher-inventors (including highly productive groups of prolific inventors) have been established in these countries.

2. Data, Variables and Models Data

Our data are from the NBER Patent Data Base (http://www.nber.org/patents/) which contains data for more than 5million patents granted to more than 2 million inventors by the USPTO from 1975 to 2010. For this paper we extract data for patents issued to inventors from eleven countries. For each patent we obtain the application and grant dates, the inventor's name and city of residence), assignee name and location, the US and international technical classifications, citations of prior patents, and the number of separate technical claims the patent makes. The data are compiled for individual inventors; we focus on as the prolific inventors, those who with at least 15 patents[2].

Primary Variables

Inventor productivity (PATENTS_PER_YEAR) is our most important variable**.** The simplest measure of an inventor's productivity is the number of patents he has obtained over a career. We adjust this for his career length to obtain the average number of patents per year as our productivity variable.

Value of inventor patents (CITATIONS_PER_PATENT)**.** For large patent data sets, many studies have accepted the number of citations as a proxy for the value of a patent (e.g., Gay and Le Bas 2005). The value of all of an inventor's patents can then be measured as the total number of citations they have received. The value of an inventor's patents might alternatively be measured as (a) his average number of citations per patent, (b) his average number of citations per year or (c) his average number of citations per patent per year, but we use the total number of career citations because it can be interpreted as capturing the concept of an inventor's potential.

Inventor technological specialization (TECH_CAT_CONC). Inventors may patent inventions in a few technological domains or in many. A small number of different technological fields might be a good proxy for inventor technological specialization. We use the Herfindahl-Hirschman Index (HHI) applied to the distribution of the inventor's for technological fields because of its emphasis (by squaring each field's percentage) on higher concentrations. We implemented the HHI at the level of the NBER's six broad technological fields. **Inter-firm mobility** (FIRMS_MOVES)**.** A simple way of identifying inter-firm mobility is to count the number of firms for which an inventor has worked and assume that the number of moves is this number minus one. . When it is the dependent variable we use FIRMS_MOVES/ CAREER_DURATION as a measure of the scale of inventor mobility over his career.

Regional and international mobility. The same principle applies for the geographic mobility. RES_MOVES_CITY describes the numbers of moves

[2] In some papers we have defined prolific inventors as those in the top 1 percent or top 5 percent of inventors by the number of patents in their corresponding countries.

between cities. RES_MOVES_INTL gives a measure of the scale of the inventor's international mobility. International moves do not duplicate inter-city moves.

Control Variables

In our dataset we observe that there are some inventors with careers of patenting that span many years and others whose patents are all produced in a very short period. To account for this variation we measure the duration of an inventor's career (years from first to last invention, inclusive= CAREER_DURATION). We wish to control for another phenomenon happening through an inventor career. When we look at the data we observe that inventors do not invent continuously. They seem interrupt the invention activity their activity over a more or less long time period. The variable is measured as the maximum number of years between two consecutive patent applications: CAREER_TIME_GAP. We also observe that the career patterns of inventing are highly variable from prolific inventor to prolific inventor with some inventors having most patents at the beginning, some having most at the end, some showing a pattern of increase followed by decrease and still others having multi-modal distributions. To determine whether particular types of patterns are associated with our measures of productivity, mobility and value, we create measures of the temporal skewness and peakedness (kurtosis) of each inventor's own temporal patenting distribution (the variables are respectively PATENT_TIME_SKEWNESS, PATENT_TIME_KUR). We observe from our data and for particular inventors a dispersion of patenting activity over the inventor's career. We decide to control for this phenomenon. The measure we use in our analysis is the inverse of dispersion; it is the Herfindahl-Hirschman Index for the time pattern of the number of patents in each year (PATENT_TIME_HHI. Hoisl (2007) uses a "time concentration" variable similar to ours. In technological fields for which patenting is an effective means of protecting inventions and where several patents are necessary for protecting a single invention inventors will tend to have more patents than in fields where these conditions do not hold. As a consequence inventor productivity differs across technological fields. We control for these differences by using dummy variables for the primary technological field in which each inventor patents. The control variables are TECH_CAT_i, where i = 1, …,6 for (1) Chemicals, (2) Computers & Communications, (3) Drugs & Medical, (4) Electrical & Electronic, (5) Mechanical and (6) Other (the omitted category in the regressions).

The way in which we deal with the *career truncation problem* are discussed in Latham et al. (2011)

In this paper our interest is in the relationships between interfirm and interregional (inter-city) mobility and the productivity of inventors. We estimate the parameters of three regression models for each country. The first model assesses the impacts of some determinants of inventor productivity (included mobility); the second model accounts for the scale of the inventor's interfirm mobility; and Table 1 the third examines the determinants of inter-city mobility. The dependent variables for the first and second models (patents per year and moves per year) are quantitative continuous variables so OLS is the method of estimation. For third model, where the dependent variable is a simple count, we fit a Poisson model. The parallel specifications of the equations are the result primarily of the limitations of our data. For example, while we are well-aware that there are both theories and empirical studies of productivity that highlight the roles of inventors' education and training, the capital available to them, the nature of the rewards system and the role of institutional constraints such as retirement ages and the nature of the patent system, we do not have those variables available to us. Consequently our work is not in the framework of those that attempt to propose and test comprehensive theories of the determinants of inventor productivity and mobility. Instead ours is a partial but coherent approach. We examine the ways in which productivity and mobility influence each other given our limited range of knowledge about other variables.

Table 1. Determinants of inventor productivity

Dependent Variable: PATENTS_PER_YEAR	USA	Canada	UK	France	Germany	Italy	Finland	Netherlands	Japan	Korea	Taiwan	China	All
	Coef.												
RES_MOVES_CITY	-0.074 ***	-0.004	0.000	-0.011	-0.047 ***	0.012	0.002	0.077 *	-0.016 ***	0.079 ***	-0.029 *	-0.161 *	-0.033 ***
RES_MOVES_INTL	0.006	-0.015	0.004	0.032	0.085 ***	0.013	0.116	-0.220 **	0.104 ***	-0.231 ***	0.215 ***	0.170	0.078 ***
CITATIONS_PER_PATENT	-0.006 ***	-0.005	-0.002	-0.013 ***	-0.023 ***	-0.011 *	0.001	-0.011 ***	0.001	-0.032 ***	0.018 *	0.009	-0.005 ***
TECH_CAT_CONC	0.192 ***	0.178 **	0.203 ***	0.230 ***	0.257 ***	-0.068	0.293 ***	0.076	0.218 ***	0.301 ***	0.427 ***	0.431	0.208 ***
FIRM_MOVES	0.069 ***	0.046 ***	0.036 ***	0.054 ***	0.052 ***	0.074 ***	0.062 ***	0.074 **	0.034 ***	0.025 ***	0.046 ***	0.112 ***	0.042 ***
PATENT_TIME_HHI	2.837 ***	3.731 ***	2.453 ***	2.059 ***	2.597 ***	3.220 ***	2.173 *	5.702 ***	1.697 ***	1.273	5.519 ***	3.658	2.706 ***
PATENT_TIME_SKEW	0.038 ***	-0.067	0.055 **	-0.018	0.020 *	0.009	-0.055	0.070	0.013 *	0.107 **	0.075	0.132	0.018 ***
PATENT_TIME_KURT	0.083 ***	0.046 ***	0.039 ***	0.079 ***	0.082 ***	0.061 ***	-0.002	-0.012	0.087 ***	0.125 ***	0.079 ***	0.027	0.088 ***
CAREER_DURATION	-0.029 ***	-0.027 ***	-0.042 ***	-0.035 ***	-0.020 ***	-0.035 ***	-0.047 ***	-0.052 ***	-0.021 ***	-0.039 ***	-0.028	-0.150	-0.025 ***
CAREER_TIME_GAP	-0.150 ***	-0.179 ***	-0.090 ***	-0.157 ***	-0.189 ***	-0.147 ***	-0.121 ***	-0.084 ***	-0.159 ***	-0.273 ***	-0.252 ***	-0.096	-0.169 ***
TECH_CAT_1	0.063 ***	0.324 ***	0.003	0.158 ***	0.208 ***	-0.261 **	-0.243	-0.260 *	-0.044 **	0.064	-0.285 *	-0.877	0.041 ***
TECH_CAT_2	0.195 ***	0.177	0.149 **	0.442 ***	0.148 ***	0.284 *	0.164	0.137	0.148 ***	0.419 ***	-0.050	-0.172	0.22 ***

Dependent Variable: PATENTS_PER_YEAR	USA	Canada	UK	France	Germany	Italy	Finland	Netherlands	Japan	Korea	Taiwan	China	All
	Coef.												
TECH_CAT_3	-0.107 ***	-0.198 *	-0.158 **	0.239 ***	0.095	-0.342 **	-0.453	-0.48 ***	-0.127 ***	0.021	-0.329	-1.657	-0.056 ***
TECH_CAT_4	0.270 ***	-0.053	0.023	0.159 ***	0.099 **	-0.087	-0.026	0.160	0.126 ***	0.209 *	0.253 **	-0.373	0.235 ***
TECH_CAT_5	0.047 **	0.076	0.008	0.105 *	0.050	0.041	0.142	0.714 ***	0.083 ***	0.571 **	0.094	0.649	0.061 ***
C	2.296 ***	2.147 ***	2.184 ***	2.247 ***	2.392 ***	2.175 ***	2.132 ***	1.821 ***	2.335 ***	3.059 ***	2.347 ***	4.369 ***	2.371 ***
R-squared	0.348	0.409	0.487	0.482	0.358	0.492	0.613	0.528	0.669	0.818	0.432	0.333	0.443
Number of Observations	54798	1111	1744	1847	7910	842	289	499	29225	2323	1803	157	106223

Notes: Sample: All prolific (15+) inventors. ***p-value<=0.01, **p-value<=0.05, *p-value<=0.10

Source: own calculations.

Table 2. Determinants of inventor interfirm mobility

DependentVariable:FIRM_MOVES/CAREER_DURATION	USA	Canada	UK	France	Germany	Italy	Finland	Netherlands	Japan	Korea	Taiwan	China	All
	Coef.												
PATENTS_PER_YEAR	0.136 ***	0.095 ***	0.116 ***	0.155 ***	0.092 ***	0.153 ***	0.349 ***	0.100 ***	0.531 ***	0.601 ***	0.262 ***	0.092 **	0.296 ***
CITATIONS_PER_PATENT	0.000	-0.001	-0.001	0.002 *	-0.001	0.006 **	0.002	-0.004 ***	-0.009 ***	0.01 *	-0.024 ***	-0.013	-0.002 ***
DependentVariable:FIRM_MOVES/CAREER_DURATION	USA	Canada	UK	France	Germany	Italy	Finland	Netherlands	Japan	Korea	Taiwan	China	All
	Coef.												
TECH_CAT_CONC	-0.231 ***	-0.184 ***	-0.174 ***	-0.088 ***	-0.245 ***	-0.053	-0.097	-0.073	-0.360 ***	-0.460 ***	-0.499 ***	-0.565 **	-0.371 ***
PATENT_TIME_HHI	-0.758 ***	-0.303	-0.608 ***	-0.035	-0.216 *	-0.870 ***	-1.732 ***	-0.872 **	-0.925 ***	-1.207 **	-1.051 *	-0.199	-0.902 ***
PATENT_TIME_SKEW	-0.026 ***	0.033 **	-0.021 ***	0.008	-0.021 ***	-0.009	0.054 **	-0.020	-0.038 ***	-0.042	-0.112 ***	-0.076	-0.041 ***
PATENT_TIME_KURT	0.012 ***	0.021 ***	0.007	0.009	0.016 ***	0.003	0.035 **	0.015 *	-0.002	-0.025 *	0.014	0.013	0.015 ***
CAREER_DURATION	0.007 ***	0.007 **	0.006 ***	0.008 ***	0.007 ***	0.003	-0.001	-0.001	0.040 ***	0.067 ***	0.081 ***	0.078	0.028 ***
CAREER_TIME_GAP	-0.020 ***	-0.019 ***	-0.016 ***	-0.010 **	-0.025 ***	-0.004	0.003	-0.008	-0.007	0.014	-0.120 ***	-0.131 **	-0.031 ***
TECH_CAT_1	0.007	-0.042	-0.037	0.050 **	0.027 **	0.078 **	0.122	-0.070	0.081 ***	-0.102	-0.122	0.156	0.029 ***
TECH_CAT_2	0.075 ***	-0.074 *	0.001	-0.048 *	0.072 ***	-0.070	0.133	-0.040	0.019	-0.168 **	0.265 ***	0.295	0.091 ***
TECH_CAT_3	0.181 ***	0.208 ***	0.144 ***	0.127 ***	0.089 ***	0.118 **	0.223 *	0.109 *	0.094 ***	0.054	0.252 **	0.820 **	0.176 ***

TECH_CAT_4	0.004	0.029	-0.043	0.022	0.054 ***	-0.004	0.157	-0.047	0.012	-0.027	0.258 ***	0.289	0.056 ***
TECH_CAT_5	-0.011 *	0.035	-0.069 **	0.058 *	0.007	-0.077 **	-0.020	-0.174 ***	-0.018	-0.447 ***	-0.055	0.202	-0.006
C	0.335 ***	0.371 **	0.378 **	-0.001	0.354 ***	0.157	0.127	0.418 ***	-0.824 ***	-1.305 ***	-0.138 ***	0.312	-0.128 ***
R-squared	0.241	0.173	0.185	0.199	0.170	0.269	0.474	0.176	0.691	0.869	0.525	0.241	0.441
Number of Observations	54798	1111	1744	1847	7910	842	289	499	29225	2323	1803	157	106223

Notes: Sample: All prolific (15+) inventors, ***p-value<=0.01, **p-value<=0.05, *p-value<=0.10

Source : own calculations.

Table 3. Determinants of inventor intercity mobility

Dependent Variable: RES_MOVES_CITY	USA	Canada	UK	France	Germany	Italy	Finland	Netherlands	Japan	Korea	Taiwan	China	All
	Coef.												
FIRM_MOVES/CAREER_DURATION	0.272 ***	0.456 ***	0.511 ***	0.471 ***	0.537 ***	0.395 ***	0.465 ***	0.311 ***	0.161 ***	-0.015	0.202 ***	0.352 ***	0.184 ***
RES_MOVES_INTL	0.330 ***	0.373 ***	0.219 ***	0.445 ***	0.241 ***	0.501 ***	0.573 ***	0.581 ***	0.044 ***	0.061 ***	0.083 ***	0.395 ***	0.201 ***
PATENTS_PER_YEAR	-0.134 ***	-0.020	-0.036	-0.022	-0.093 ***	0.010	-0.063	0.064	-0.068 ***	0.042 ***	-0.031 ***	-0.085 **	-0.097 ***
CITATIONS_PER_PATENT	-0.006 ***	-0.002	-0.001	0.005	-0.010 ***	-0.012	-0.010	0.002	-0.016 ***	-0.026 ***	-0.017 ***	-0.029 *	-0.019 ***
TECH_CAT_CONC	-0.464 ***	-0.153 ***	-0.210 ***	-0.25 ***	-0.42 ***	-0.219 **	-0.210	0.051	-0.368 ***	-0.318 ***	-0.225 ***	-0.387 ***	-0.559 ***
PATENT_TIME_HHI	-1.536 ***	-0.229	-0.926 ***	-0.243	-0.949 ***	-0.364	-0.345	-1.610 *	-3.566 ***	-5.359 ***	-3.265 ***	-0.772	-3.125 ***

PATENT_TIME_SKEW	-0.068 ***	-0.028	-0.043 ***	-0.018	-0.014	-0.034	-0.094	0.049	-0.067 ***	-0.039	-0.079 ***	0.006	-0.094 ***
PATENT_TIME_KURT	0.018 ***	0.000	0.018 ***	-0.012	0.026 ***	-0.014	0.042	0.031 *	0.055 ***	0.040 ***	0.02 ***	0.008	0.045 ***
CAREER_DURATION	0.046 ***	0.039 ***	0.032 ***	0.033 ***	0.047 ***	0.028 ***	0.038 ***	0.022 **	0.063 ***	0.057 ***	0.039 ***	0.075 ***	0.052 ***
Dependent Variable: RES_MOVES_CITY	USA	Canada	UK	France	Germany	Italy	Finland	Netherlands	Japan	Korea	Taiwan	China	All
	Coef.												
CAREER_TIME_GAP	-0.051 ***	-0.019	-0.024 ***	-0.003	-0.033 ***	0.002	-0.066 *	-0.012	-0.098 ***	-0.08 ***	-0.045 ***	-0.055	-0.092 ***
TECH_CAT_1	-0.005	-0.038	-0.024	0.012	0.094 ***	0.085	-0.076	0.068	0.026	-0.198 **	-0.044	-0.681	0.05 ***
TECH_CAT_2	0.151 ***	0.010	-0.103 *	0.087	0.255 ***	0.533 ***	0.101	-0.267	0.080 ***	0.145 *	0.085 *	-0.292	0.310 ***
TECH_CAT_3	0.046 **	-0.149 *	-0.119 **	-0.112	0.033	-0.112	-0.058	0.578 ***	-0.031	-0.344 *	0.040	-1.537 **	-0.007
TECH_CAT_4	0.036 **	-0.147 *	-0.136 **	0.072	0.129 ***	0.339 ***	-0.243	-0.276 *	0.059 ***	0.115	0.160 ***	-0.686	0.231 ***
TECH_CAT_5	-0.057 ***	-0.128	-0.062	0.080	0.082 **	0.245 **	-0.109	-0.270	0.048 ***	-0.086	-0.021	-0.833	0.105 ***
C	0.735 ***	0.046	0.891	-0.009 ***	0.360	-0.038	0.017	-0.163	1.425 ***	2.295 ***	1.779 ***	0.981	1.377 ***
R-squared	-0.094	0.446	0.571	0.409	0.233	0.189	0.310	0.401	0.733	0.801	0.719	0.833	0.336
Number of Observations	54798	1111	1744	1847	7910	842	289	499	29225	2323	1803	157	106223

Notes: Sample: All prolific (15+) inventors, ***p-value<=0.01, **p-value<=0.05, *p-value<=0.1

Source : own calculations.

3. Estimations Results and Findings

Table 1 give the estimated coefficients for inventor productivity relation, table 2 for the determinants of inventor interfirm mobility, and table 3 for inventor intercity mobility. For the productivity model we find that the coefficient for interfirm mobility is always positive, indicating that inventors with many moves are more productive and conversely. Of course we cannot infer any causal relation between the two. Our regressions simply show that the relationship between mobility and productivity well established by the literature is clearly confirmed. As to the sign of the coefficient related to intercity mobility we did not find consistent results; no relationship emerges from the results. The same is true for international mobility except for Korea, for which the coefficient is significantly positive. The coefficient related to the inventor degree of technological specialization is always positive when significant (for 9 countries on 12). It indicates that more specialized inventors are more productive than those less specialized. This result is in lines with the evolutionary view of the determinants of inventor productivity. In general temporal concentration of inventions has a positive effect on inventor productivity. This result appears in opposition with the finding by Hoisl ‘s (2007) for a population of German inventors. One reason for the difference may because we study only the the more productive inventors.. For this variable differences appear between Western and Asian countries. For instance the result is not valid for Korea and China. Career duration has a negative coefficient (except for Taiwan and China) expressing the idea that inventors with a longer career are less productive (to some extent this last result is in accordance with the result related to time concentration). Here we find again differences between Western and Asian countries. It might be that for China one reason for the difference is the very short time period in which we observe inventor productivity because of China’s late entry into patenting. Finally the variable CAREER_TIME_GAP matters as expected: inventors with a long time period without patented inventions have lower productivity (the two directions of causality are equally possible). This trend is pervasive and matches the situation of 11 countries out of 12 (the case of China is particular: fewer inventors and a shorter observation period).

For the interfirm mobility model we note that inventor productivity has a positive impact on the scale of interfirm mobility for all 12 countries of the sample. But we still cannot interpret this result in causal terms. Strong inventor technological specialization is related to less mobility. And conversely less specialized inventors are more mobile. This trend is true for the largest western countries but not for the smallest (Italy, Finland, and the Netherlands). By contrast it applied to all the Asian countries. Temporal concentration of patenting is always negative and very often significant. This result is partly due to a mechanical effect; if the inventor's patenting is really concentrated in a short time period he has fewer opportunities for moving. The opposite is true when we consider the variable CAREER_DURATION. A longer career generates many opportunities for moving. The coefficient related to CAREER_DURATION is positive and significant for 8 countries out of 12. For the small European countries (Italy, Finland, and the Netherlands) the variable has significant effects. The variable CAREER_TIME_GAP has negative and significant effects for many countries. Inventors with a long time period of time without patenting (all other things being equal) move less (we know from the first regression that they are less productive as well).

The determinants of intercity mobility are strongly linked to interfirm and international mobility. To put it in other terms: a great proportion of interfirm moves match geographic mobility (intercity or international). After controlling for different types of inventor mobility and career profile it appears that technological specialization matters significantly and for all the countries (the Netherlands excepted): the more specialized an inventor is, the less he moves geographically. Career duration has a trivial effect. The estimated coefficients related to CAREER_TIME_GAP are negative when significant. The same explanative reasons put forth for interfirm mobility can be applied here as well.

4. Conclusions

Two lessons can be drawn from this study. First the set of variables we have constructed and tested have been found to be highly relevant for explaining inventor mobility. For instance the new variable CAREER_TIME_GAP has significant explanatory power. One interesting finding is that the role played by inventor technological specialization that is not the same for inventor productivity and mobility. This variable is found to matter significantly in all the three regression models. Second, with respect to our goal of comparing the dynamics of inventor productivity and mobility according to the types of countries, the main finding is that there is not much difference between Western and Asian countries. The evolutionary laws determining inventor productivity apply generally, whatever the country. Moreover we have shown there are significant differences within the set of Western countries and within the Asian countries as well. As a consequence this second block of countries is not homogeneous. However, because the sizes of our samples of prolific inventors are very different across the countries, and are quite small in some cases, one must interpret the comparative results with caution.

References

Breschi S. *et al.* (2010), *The Geography of Knowledge Spillovers: the role of inventors'mobility across firms and in space.* in R. Boschma and R. Marton eds The Handbook of Evolutionary Economic Geography. Edward Elgar Cheltenham UK

Dodgson M. and D. Gann (2010), *Innovation. A very short introduction*. Oxford. University Press. Oxford and. New York

Ernst D. (2005), *Pathways to innovation in Asia's leading electronics exporting countries – a framework for exploring drivers and policy implications.* International Journal of Technology Management, Vol 29. N°1-2, pp. 6-20

Ernst H., Leptien C., Vitt J. (2000), *Inventors are not alike: The Distribution of Patenting Output among Industrial R&D Personnel*, in: IEEE Transactions on Engineering Management, Vol. 47, No. 2, pp. 184-199

Fagerberg F. and Godinho (2006), *Innovation and Catching up,* in Fagerberg F., Mowery D.C., Richard R. Nelson R.R eds The Oxford Handbook of Innovation. Oxford University Press; pp 514-542

Felsenstein D. (2011), *Human capital and labour mobility determinants of regional innovation*. In COOKE (Philip), ASHEIM (Bjor), BOSCHMA (Ron) et al. Handbook of regional innovation and growth. Edward Elgar

Gambardella A., Harhoff D., Verspagen B. (2006), *The value of Patents*. Paper presented at the NBER Conference. The Economics of Intellectual Property. Cambridge (MA), July 19th

Gay C. and Le Bas C. (2005), *Uses without too much abuses of patents citations or the simple Economics of Patents Citations as a measure of the value and origin of codified Knowledge*. The Economics of Innovation and New Technology, Vol. 14 (5), July, pp 333-338

Gay C., Latham B., Le Bas C. (2008), *Collective Knowledge, Prolific Inventors and the Value of Inventions: An Empirical Study of French, German and British Owned U.S. Patents, 1975-1998.* The Economics of Innovation and New Technology. Volume 17, Jan. -March. pp 5- 22

Hoisl K. (2007), *Tracing mobile inventors-The causality between inventor mobility and inventor productivity*, Research Policy, n° 36, pp. 619-636

Hoisl K. (2009), *Does Mobility Increase the Productivity of Inventors?* Journal of Technology Transfer, 34(2)

Latham W., Le Bas C., R. Bouklia-Hassane, C. Le Bas C., D. Volodine (2011), *Interregional mobility, productivity and the value of patents for prolific inventors in France, Germany and the U-K.* International Journal of Foresight and Innovation Policy. Vol. 7, N° 1/2/3. pp 92-113

Latham W., Le Bas C., Volodin D. (2012), *Value of invention, prolific inventor productivity and mobility: evidence from five countries, 1975-2002.* Economies et Sociétés. Série W, N°14. pp 11-40

Le Bas C., Cabagnols, A., Bouklia, R. (2010), *Prolific Inventors: Who are They and Where do They Locate? Evidence from a Five Countries US Patenting Data Set.* International Centre for Economic Research Working Paper No. 14/2010. Available at SSRN: hhttp://ssm.com/abstract=1625743

Lundvall Bengt-Ake *et al*. (2009*), Introduction. Handbook of Innovation Systems and Developing Countries.* Ed. Lundvall, Joseph, Chaminade, Vang. Cheltenham. Northampton: Edward Elgar. 2009

Narin F. and Breitzman A. (1995), *Inventive productivity.* Research Policy, 24, 507-519

Pilkington A. *et al.* (2009), *Defining key inventors: A comparison of fuel cell and nanotechnology industries.* Technological Forecasting and Social Change 76 (1), pp. 118-127

Rothaermel F.T. and Hess A. (2007). *Building dynamic capabilities: Innovation driven by individual, firm, and network-level effects.* Organization Science, 18 (6): 898-921

Saxenian A. L. (2006), *The New Argonauts: Regional Advantage in the Global Economy*. Harvard University Press, Cambridge

Schankerman M., Shalem R., Trajtenberg M. (2006), *Software Patents, Inventors and Mobility*, Working Paper

Trajtenberg M. (2004), *The names game: Using Inventors Patent Data in Economic Research.* Power-point Presentation. http://www.tau.ac.il/~manuel/

Trajtenberg M., Shiff G., Melamed R. (2006). *The "Names Game": Harnessing Inventors' Patent Data for Economic. Research*. NBER working paper series, No. w12479, September

Streszczenie

PRODUKTYWNOŚĆ I MOBILNOŚĆ KLUCZOWYCH WYNALAZCÓW: ANALIZA PORÓWNAWCZA DANYCH PATENTOWYCH 12 PAŃSTW AZJI, AMERYKI ORAZ EUROPY ZACHODNIEJ

Artykuł przedstawia nowe spojrzenie na rolę indywidualnych wynalazców w procesie tworzenia innowacji. Wynalazcy indywidualni stanowią element centralny procesu twórczego. Innowacja nie jest produktem firm i organizacji, wymaga indywidualnej kreatywności (Rothaermel i Hess 2007). Badanie koncentruje się na analizie płodnych wynalazców. Wynalazcy tej kategorii mają najwyższy udział w generowaniu ogółu wynalazków (Le Bas et al. 2010) o wysokiej wartości ekonkomicznej (Gambardella et al. 2005). Poprzednie badania kluczowych wynalazców skupiały się analizie firm, w których pracują lub w branżach, w których te firmy działają.

10.2478/v10103-012-0031-9

KAROLINA LEWANDOWSKA-GWARDA*

Spatio-temporal Analysis of Unemployment Rate in Poland

Abstract

The aim of this paper is to present results of spatio-temporal analysis of unemployment rate in Poland, with the usage of advanced spatial econometric methods. The analysis was done on data collected for 'powiat' level between 2006 and 2010. GIS and ESDA tools were applied for visualization of the spatio-temporal data and identification of spatial interactions between polish counties on labor market. Multi-equation spatial econometric models were used to describe unemployment rate in relation to selected social-economic variables.

1. Introduction

For a long time now unemployment has aroused interest among economists, sociologists and psychologists as, due to its consequences, it constitutes a serious socioeconomic problem. It produces adverse effects being of importance to both the unemployed and the entire economy. A rise in unemployment means not only a drop in the population's standard of living (loss of means of livelihood, growing social discontent, increased social pathologies and crime) but also the workforce not being fully utilized, and thus actual production being lower than the potential one (Milewski 2000, p. 532).

Poland's high unemployment rate has generated widespread interest in that issue. Specialist literature offers numerous studies whose authors try to find reasons for high values of that variable in Poland. Those studies, however, do

* Ph.D., University of Łódź

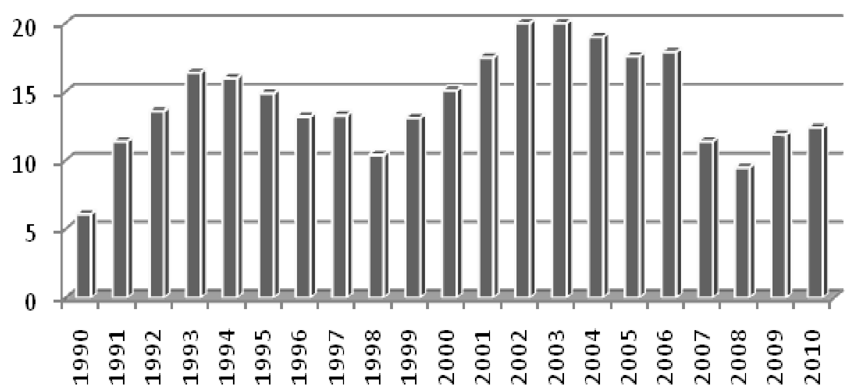
not include spatio-temporal analyses applying multi-eguation spatial econometric models.

The objective of the study is to analyse the unemployment rate in Poland from a spatio-temporal perspective, with the usage of advanced spatial econometric methods. The research uses GIS and ESDA tools for visualization of variables and identification of spatial interactions occurring among studied territorial units in the country's labour market. Seemingly Unrelated Regression models with spatial effects are employed to describe the impact of selected macroeconomic variables on the level of unemployment rate in Poland in specific poviats from 2006 to 2010.

2. Unemployment in Poland

There were considerable fluctuations in the unemployment rate in Poland after 1990. Figure 1 shows clearly two cycles in that variable with the first occurring in the 1990-1998 and the other – in the 1999-2008 period. The inflection point of the first cycle was observed in 1993 when unemployment hit 16.4%. As for the other, its inflection point was in 2002 and 2003 when unemployment reached a record high of 20%.

Figure 1. Unemployment rate in Poland in the 1990-2010 period



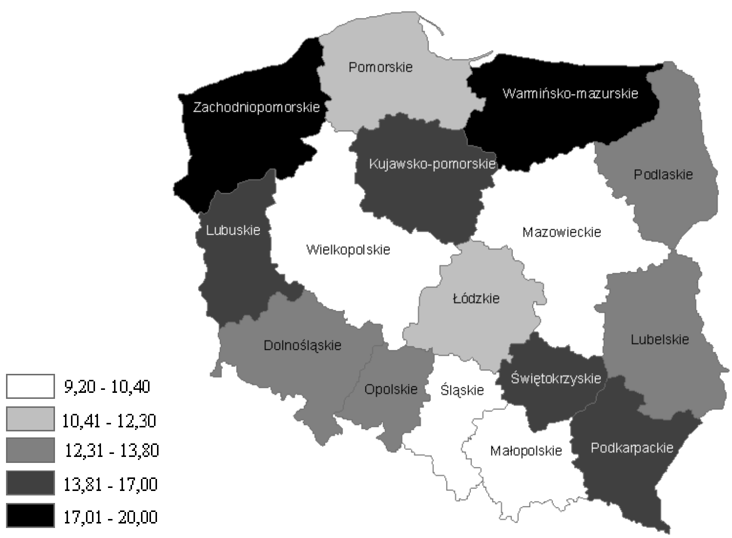
Source: own work based on statistical data of the Central Statistical Office.

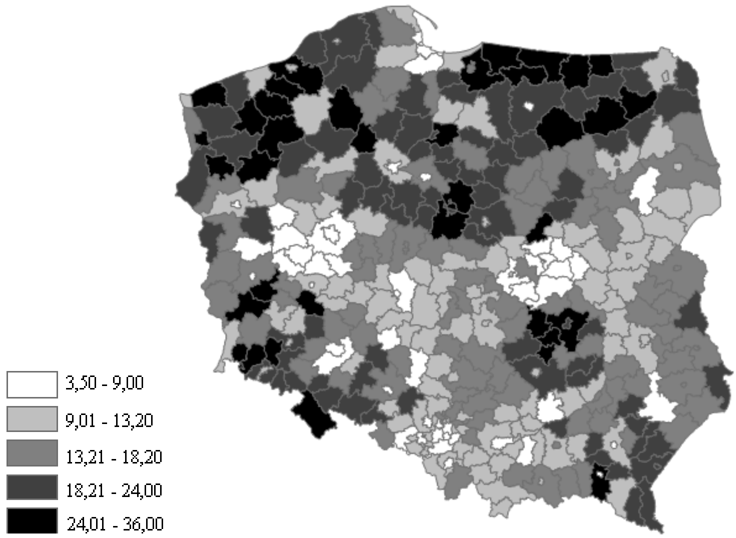
The first stabilization in the labour market appeared in 1994. The situation improved thanks to a high economic development rate and the slowing down of restructuring processes in some sectors of the economy. After 1998 unemployment started to rise again due to expired undertakings contained in privatization agreements of the mid-1990s. They obligated companies to maintain employment at specified levels. When they were no longer valid, companies began mass lay-offs. In consequence, the unemployment rate soared to as much as 20% in 2002 and 2003.

The second stabilization period lasted from 2004 to 2008. Mass emigration for economic reasons and rapid economic growth had a considerable impact on the fall in the unemployment rate. In 2008 the unemployment rate reached 9.5%, thus being the lowest since 1990. In 2009 the situation in the labour market everely deteriorated; the economic crisis caused a rise in the unemployment rate to almost 12% to subsequently hit 12.4% in 2010.

The unemployment rate in Poland is considerably spatially diversified. Figure 2 shows that in 2010 the highest level of the studied variable was noted for the Warmian-Masurian voivodship (20%), while it was the lowest for the Greater Poland (9.2%) and Masovian (9.7%) voivodships.

Figure 2. Unemployment rate in Poland in specific voivodships and poviats in 2010



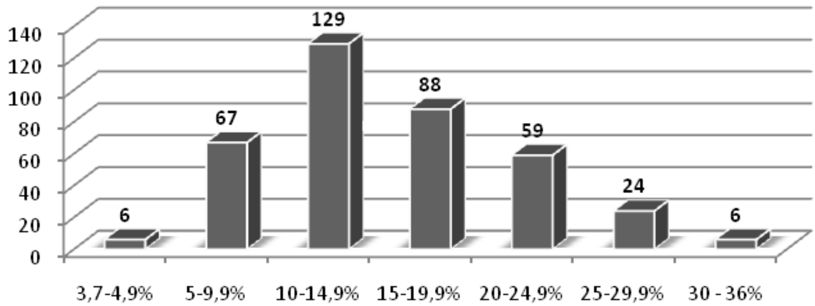


Source: own work based on statistical data of the Central Statistical Office.

The highest unemployment rate was observed for the following poviats: Szydłowiecki 36% (Masovian voivodship), Piski 31.5% (Warmian-Masurian voivodship) and Braniewski 30.9% (Warmian-Masurian voivodship). The lowest level of the variable occurred in the following poviats: Poznański 3.5% (Greater Poland voivodship), Warszawski 3.5% (Masovian voivodship) and cities with poviat rights – Poznań 3.6% (Greater Poland voivodship), Katowice 3.6% (Silesian voivodship) and Sopot 3.9% (Pomeranian voivodship). It should also be noted that the difference in unemployment rates between poviats with the lowest and highest levels of the variable accounted for as much as 32.5 percentage points.

Figure 3 presents results of the clustering of poviats according to their unemployment rates. In 2010 the lowest unemployment rate (below 5%) was noted for six poviats with a majority of those being cities with poviat rights, i.e. Poznań, Katowice, Sopot and Cracow. The largest cluster is formed by poviats with values of the variable ranging between 10% and 14.9%. It consists of one hundred and twenty-nine objects. Unemployment rates exceeding 20% were observed in as many as eighty-nine poviats, of which six showed levels of the variable above 30%.

Figure 3. Poviats clustered according to unemployment rates in 2010



Source: own work based on statistical data of the Central Statistical Office.

3. Determinants of unemployment

Unemployment is a multidimensional phenomenon determined by complex factors and mechanisms. There are numerous theories offered by specialist literature that specify variables causing an increase or decrease in the unemployment rate. The most frequently mentioned factors include, among others:

1. **Average Wages** – according to the concept by A. W. Phillips, higher unemployment rates are accompanied by a slower rise in nominal wages, while a fall in unemployment coincides with an increase in nominal wages.
2. **Gross Domestic Product** – A. Okun's law says that every drop by 2% in the real GDP as compared with the potential GDP results in a rise in the unemployment rate by 1 percentage point. The unemployment rate is very strongly connected with the GDP. When employment increases, the GDP rises, and thus economic growth occurs with a simultaneous fall in the unemployment rate.
3. **Investments** – the number of created jobs is largely dependent on the volume and type of investments. New development-oriented investments contribute to increased demand for labour. On the other hand, current investments aimed at property replacement allow to maintain existing jobs. It should be emphasized that not all investments contribute to creating or maintaining jobs as they increase the productivity of workforce (Bremond, Couet, Salort 2005, p. 134). However, greater workforce productivity enables enterprises to operate more effectively, hence, making them more competitive, which improves employment situation in the long run.

4. **Socioeconomic Position of Adjacent Areas, Migrations** – the economic situation of the area of origin is among important factors affecting the level of migration. The most frequently mentioned motives behind migration include high unemployment rates, low wages and high costs of living in the place of residence. Migration is encouraged by better socioeconomic conditions of adjacent areas – higher standard of living, employment opportunities (a greater number of offered jobs), higher wages, better working conditions. According to the world-systems theory[1] migrations result from an economic imbalance between the core, i.e. developed areas (countries, voivodships) and peripheries, i.e. developing ones that constitute workforce reserves for the core ones (Kuciński 2009, pp. 102-105). Since Poland's accession to the European Union over one million people have left the country, which resulted in the steadily declining unemployment rate from 2004 to 2008 (Figure 1). It ought to be emphasized, however, that the unemployment rate depends not only on external migrations but also, in large measure, on internal, temporal ones connected e.g. with commuting to another town every day.

The level of unemployment is also dependent on conditions in a specific labour market, for example, the number of people at the working age, economic activity of the population, number of registered economic entities and number of offered jobs.

4. Statistical database

The unemployment rate in Poland was analysed from a spatio-temporal perspective based on statistical database built on the basis of information available at the Local Data Bank of the Central Statistical Office published on the official website of the office. Statistical data were collected for 379 poviats of Poland in the 2006-2010 period. Major determinants of creating the database were both reasons pertaining to the subject matter (see Section 3) and availability of data at the time of carrying out the study.

Regrettably, not all variables needed for the analysis of the unemployment rate are available for specific poviats. Such variables include inflation and the GDP. Due to the fact that the GDP variable that describes economic

[1] It is a concept of social development in which the erstwhile analysed units, i.e. the state, economy, society, have been replaced by historical systems. The world is considered a spatio-temporal whole (Kuciński 2009, p. 105).

development is crucial for the study, it was replaced with one of local development measures, i.e. budgetary revenues of poviats per capita.

5. Preliminary Analysis of Statistical Data

Spatial data are characterised by a more complex structure than time series. When examining spatial objects (countries, regions, voivodships, povits), it should be kept in mind that they are not isolated in space and may be affected by other units. That may result in the spatial clustering of similar values of localized variables or their dispersion.

The first stage of the study used Exploratory Spatial Data Analysis tools in order to identify interactions occurring among unemployment rates in poviats. A spatial weight matrix **W** (row standardized) was generated based on a first degree contiguity matrix, in the queen configuration. Next, values of global Moran's *I* measure were calculated, being $I_{2006} = 0.5342$, $I_{2007} = 0.5198$, $I_{2008} = 0.5124$, $I_{2009} = 0.5022$ and $I_{2010} = 0.4514$ respectively. The received results allowed to conclude that there are certain spatial relationships concerning unemployment rates in Poland.

Apart from the need to examine global spatial autocorrelation, literature indicates that, in order to obtain a detailed picture of a studied phenomenon, it is necessary to perform the analysis of local spatial autocorrelation (LISA). The high value of global Moran's *I* is confirmed by the LISA analysis, that indicates positive spatial autocorrelation in all the periods (as high – low - values of unemployment rate adjacent to high - low - values of the variable in polish poviats). High unemployment rates are characteristic of poviats situated in the North of the country in the Warmian-Masurian and West Pomeranian voivodships. Low unemployment rates occur for central and Southern poviats located in the Masovian, Greater Poland and Silesian voivodships.

The carried out analysis led to conclusions being of the utmost importance from the point of view of econometric modelling. It was proved that there are certain spatial relationships concerning unemployment rates in Poland. It may be inferred that a rise (fall) in unemployment rate in poviats defined in the weight matrix as adjacent results in a rise (fall) in the level of the studied variable in poviat *i* in all the studied periods. Therefore, an econometric model ought to contain a variable that takes into account those relationships.

6. Modelling of Unemployment Rate

Three multi-equation models were applied to describe the impact of the selected socioeconomic variables on the level of the unemployment rate in Poland in specific poviats from 2006 to 2010 (Suchecki 2012, pp. 157-158).

1. Seemingly unrelated regression model (no spatial interactions).
2. Seemingly unrelated regression model with spatially lagged dependent variables.
3. Seemingly unrelated regression model with spatially lagged dependent variables and additionally introduced selected spatially lagged independent variables.

The multi-equation spatial econometric models used in the study were characterised by better goodness-of-fit and lower errors than the model not taking into account spatial interactions. They also allowed to eliminate the impact of spatial autocorrelation from the analysis. Therefore, it may be stated that spatial econometrics models are a better tool to analyse spatio-temporal data than models that do not consider spatial relationships among the studied geographical units, i.e. poviats. It was also observed that the Log-likelihood test value is the highest for the spatial model that additionally took into account spatial lags of selected independent variables (Table 1). The model exhibits lower root mean squared error (RMSE) values and higher significance of parameters; hence it explains the studied phenomenon in the most detailed manner. The received estimations are also consistent with the economic theory.

Table 1. Results of goodness-of-fit test

Model	Log-Likelihood
SUR (no spatial interactions)	-1070,74
SUR with spatially lagged dependent variables	- 653,29
SUR with spatially lagged dependent and selected independent variables	- 577,48

Source: own calculation.

The multi-equation unemployment model with spatially lagged dependent and selected independent variables takes the following form:

$$\begin{aligned}
U_{06} &= \alpha_{01} + \rho_1 \mathbf{W}_U_{06} + \lambda_{11}\mathbf{W}_W_{06} + \lambda_{21}\mathbf{W}_RE_{06} + \beta_{11}W_{06} + \beta_{21}BR_{06} + \\
&+ \beta_{31}I_{06} + \beta_{41}SA_{06} + \beta_{51}RE_{06} + \beta_{61}OJ_{06} + \beta_{71}IM_{06} + \beta_{81}FM_{06} + \varepsilon_{11} \\
U_{07} &= \alpha_{02} + \rho_2 \mathbf{W}_U_{07} + \lambda_{12}\mathbf{W}_W_{07} + \lambda_{22}\mathbf{W}_RE_{07} + \beta_{12}W_{07} + \beta_{22}BR_{07} + \\
&+ \beta_{32}I_{07} + \beta_{42}SA_{07} + \beta_{52}RE_{07} + \beta_{62}OJ_{07} + \beta_{72}IM_{07} + \beta_{82}FM_{07} + \varepsilon_{22} \\
U_{08} &= \alpha_{03} + \rho_3 \mathbf{W}_U_{08} + \lambda_{13}\mathbf{W}_W_{08} + \lambda_{23}\mathbf{W}_RE_{08} + \beta_{13}W_{08} + \beta_{23}BR_{08} + \\
&+ \beta_{33}I_{08} + \beta_{43}SA_{08} + \beta_{53}RE_{08} + \beta_{63}OJ_{08} + \beta_{73}IM_{08} + \beta_{83}FM_{08} + \varepsilon_{33} \\
U_{09} &= \alpha_{04} + \rho_4 \mathbf{W}_U_{09} + \lambda_{14}\mathbf{W}_W_{09} + \lambda_{24}\mathbf{W}_RE_{09} + \beta_{14}W_{09} + \beta_{24}BR_{09} + \\
&+ \beta_{34}I_{09} + \beta_{44}SA_{09} + \beta_{54}RE_{09} + \beta_{64}OJ_{09} + \beta_{74}IM_{09} + \beta_{84}FM_{09} + \varepsilon_{44} \\
U_{10} &= \alpha_{05} + \rho_5 \mathbf{W}_U_{10} + \lambda_{15}\mathbf{W}_W_{10} + \lambda_{25}\mathbf{W}_RE_{10} + \beta_{15}W_{10} + \beta_{25}BR_{10} + \\
&+ \beta_{35}I_{10} + \beta_{45}SA_{10} + \beta_{55}RE_{10} + \beta_{65}OJ_{10} + \beta_{75}IM_{10} + \beta_{85}FM_{10} + \varepsilon_{55}
\end{aligned} \tag{1}$$

where:

U – unemployment rate (2006-2010);

W – average wages in zlotys (2006-2010);

BR – budgetary revenues of poviats per capita (2006-2010);

I – investment outlays per capita (2006-2010);

SA – social assistance expenditures per capita (2006-2010);

RE – number of registered economic entities (2006-2010);

OJ – offered jobs per 1 thousand working age individuals (2006-2010);

IM – balance of internal migrations in poviats in Poland (2006-2010);

FM – balance of foreign migrations in poviats in Poland (2006-2010);

W – spatial weight matrix;

W_*SB*, **W**_*WYN*, **W**_*PG* – spatial lags of *U*, *W* and *RE* variables;

α, β, ρ, λ – structural parameters;

ε – errors.

Estimations of the model parameters applied the maximum likelihood method (Suchecki 2012, p. 159). Due to the fact that some of them appeared to be statistically insignificant, a part of variables were eliminated from the model. The received results (after sequential elimination of variables) are presented in Table 2.

Table 2. Results of estimations of unemployment model parameters

Independent variables	Dependent variable - *U*		
	Coefficient	*t*-ratio	Goodness-of-fit
2006			
Constant	15,945231	(17,31)	$pseudoR^2 = 0{,}5423$ *RMSE* = 2,96655
W_*U*	0,1276212	(19,76)	
W_*W*	-0,0007961	(-10,99)	
BR	-0,0011871	(-3,11)	
I	-0,0001734	(-2,35)	
SA	0,0023427	(2,32)	
2007			
Constant	12,26919	(14,43)	$pseudoR^2 = 0{,}5591$ *RMSE* = 2,73543
W_*U*	0,128756	(16,31)	
W_*W*	-0,0004412	(-8,21)	$pseudoR^2 = 0{,}5591$ *RMSE* = 2,73543
W_*RE*	-0,0003897	(-3,82)	
BR	-0,0007691	(-2,98)	
I	-0,0009192	(-2,21)	
2008			
Constant	9,987221	(16,43)	$pseudoR^2 = 0{,}5381$ *RMSE* = 2,89123
W_*U*	0,123231	(17,98)	
W_*W*	-0,0002971	(-7,44)	
W_*RE*	-0,0004967	(-3,98)	
BR	-0,0006941	(-2,67)	
I	-0,0014848	(-2,23)	
RE	-0,0012221	(-4,37)	
IM	-0,4551298	(-2,93)	
2009			
Constant	14,52855	(11,14)	$pseudoR^2 = 0{,}5131$ *RMSE* = 3,00721
W_*U*	0,137211	(11,51)	
W_*W*	-0,0003907	(-8,21)	
W_*RE*	-0,0004366	(-2,73)	
BR	-0,0007314	(-4,31)	
I	-0,0001141	(-2,51)	
RE	-0,000851	(-2,76)	

2010			
Constant	19,757000	(8,54)	$pseudoR^2 = 0{,}5259$ *RMSE* = 2,93810
W_*U*	0,5798206	(11,88)	
W_*W*	-0,0003996	(-4,19)	
W_*RE*	-0,0012519	(-3,42)	
BR	-0.0006549	(-2,34)	
RE	-0.0048212	(-3,46)	
IM	-0,0011247	(-2,57)	
n	379		

Source: own calculation.

The received results led to the following conclusions:

1. With other factors remaining constant, a rise (fall) in the unemployment rate in poviats defined in the weight matrix as adjacent resulted in a rise (fall) in the unemployment rate in the studied poviat in all the analysed periods. The unemployment rate in Poland is characterised by positive spatial autocorrelation. Thus, the geographic clustering of high and low values of the studied variable is observed. There are spatial clusters of objects with similar values of the variable.
2. It was noted for all the analysed periods that an increase in the poviats' budgetary revenues resulted in a fall in the unemployment rate *ceteris paribus*. Budgetary revenues of poviats per capita are among local development measures. If they rise, economic conditions in a poviat improve. Economic growth occurs and thus investments increase, companies employ more workers and, in consequence, the unemployment rate goes down.
3. With other factors remaining constant, a rise in the investment level caused a fall in the unemployment rate in 2006, 2007, 2008 and 2009. Conditions in the labour market depend in large measure on the level of investments made by enterprises. The number of created jobs is dependent on the volume and type of investments. New development-oriented investments contribute, to a large extent, to increased demand for labour. On the other hand, current investments aimed at property replacement allow to maintain existing jobs.
4. In 2008, 2009 and 2010 an increase in the number of economic entities led to a fall in the unemployment rate *ceteris paribus*. In 2008, when fast economic growth occurred in the country, a substantial increase was noted in the number of economic entities, with the trend continuing in 2009 and 2010 as well. New economic entities create new jobs, hence, their impact on the fall in the unemployment rate is obvious and clear.

5. In 2006 it was observed that, with other factors remaining constant, increased social assistance expenditures resulted in the increased unemployment rate. Unemployment adversely affects the condition of the society, leads to the loss of means of livelihood, loss of qualifications, increased crime and social pathologies. Therefore, the state intervenes in the labour market by using instruments of the so called passive labour market policy. Excessive protectionism, however, does not solve the problems as it may contribute to a further fall in the society's activity in the labour market, which results in increased unemployment.
6. An impact of the increased balance of internal migrations on the fall in the unemployment rate, *ceteris paribus*, was noted in 2008 and 2010. Higher internal migrations mean increased movement of the population within the country. In most cases, population movement is forced by poor economic situation of the area of origin – high unemployment, low wages and poor working conditions. A better socioeconomic position of adjacent areas encourages the so called migration for economic reasons, often even temporary (population movement connected with everyday commuting to another poviat), which causes a fall in the unemployment rate.
7. In all the analysed periods it was observed that, *ceteris paribus*, a rise in average wages in adjacent poviats resulted in a fall in the unemployment rate in the studied poviat. In 2007, 2008, 2009 and 2010 it was also noted that an increase in the number of economic entities in adjacent poviats led to a fall in the unemployment rate in the studied poviat *ceteris paribus*. It was emphasized in the preceding point that a better socioeconomic condition of adjacent areas, including also higher wages or greater demand for labour resulting from an increased number of economic entities, encourages migrations of the population. The movements of the population in search of work contribute to reducing a disproportion between supply and demand in local labour markets, which causes the unemployment rate to fall.

The received results also indicate that average wages, number of offered jobs and balance of foreign migrations did not significantly affect the unemployment rate in the studied period.

7. Conclusions

The study attempted to analyse the unemployment rate from a spatio-temporal perspective.

The carried out research indicates that multi-equation spatial econometric models are a better tool to analyse spatio-temporal data than models that do not take into account spatial relationships. The spatial SUR models used in the study were characterised by better goodness-of-fit and lower errors than the model not taking into account spatial relationships; they also allowed to eliminate the impact of spatial autocorrelation from the analysis. The models made it possible to obtain a lot of valuable information about the impact of selected macroeconomic variables on the level of unemployment rate in Poland from 2006 to 2010. It was proved that there are spatial clusters of areas with similar values of the unemployment rate in Poland. Neighbouring poviats have an influence on each other. A rise in the unemployment rate in adjacent areas results a rise in that variable in the studied poviat. Moreover an increase in average wages and number of economic entities in neighbouring areas led to a fall in the unemployment rate in the studied poviat. This are very important information for local governments, which should administer their areas in collaboration with each other.

References

Bremond J., Couet J.F., Salort M.M. (2005), *Kompendium wiedzy o ekonomii*, PWN, Warszawa

Kuciński K. (red.) (2009), *Geografia ekonomiczna*, Wolters Kluwer Polska, Kraków

Milewski R. (red.) (2000), *Podstawy ekonomii*, Wydawnictwo Naukowe PWN, Warszawa

Suchecki B. (red.) (2012), *Ekonometria Przestrzenna II. Modele zaawansowane.* Wydawnictwo C.H. Beck, Warszawa

Streszczenie

PRZESTRZENNO-CZASOWA ANALIZA STOPY BEZROBOCIA W POLSCE

Celem opracowania jest przestrzenno-czasowa analiza poziomu stopy bezrobocia w Polsce, z wykorzystaniem zaawansowanych metod ekonometrii przestrzennej. Badanie przeprowadzono na danych statystycznych zebranych na poziomie powiatu, w latach 2006-2010. Narzędzia GIS i ESDA zostały wykorzystane w celu wizualizacji zmiennych oraz identyfikacji interakcji przestrzennych zachodzących pomiędzy badanymi jednostkami terytorialnymi na rynku pracy. Wielorównaniowe modele o równaniach pozornie niezależnych zastosowano do opisu wpływu wybranych zmiennych makroekonomicznych na kształtowanie się poziomu stopy bezrobocia w Polsce w badanym okresie.

10.2478/v10103-012-0032-8

ELŻBIETA LITWIŃSKA[*]

Spatial Analysis of the Labour Market by Using Econometric Tools. The Case of Lower Silesia Region (Dolnośląskie voivodship)

Abstract

This paper presents the application of econometric techniques to examine the labour market in Lower Silesia. First the analysis was performed on a data set for variables connected with labour market recorded in poviats (NUTS 4). In order to determine the existence of spatial autocorrelation Moran's statistics I was calculated. Then the spatial regression model was used to describe the relationship between the rate of unemployment and other variables. Next, LISA cluster maps were generated for units at NUTS 5 level. The results indicate the spatial dimension of the unemployment and its tendency to creating concentrations.

1. Introduction

Studies on European regional labour markets show a significant degree of spatial dependence among them. These investigations use Moran's statistics I and spatial regression models. Both regions marked by high unemployment as well as areas characterized by low unemployment have tendency to cluster in space (Niebuhr 2003). Overman and Puga (2000) underline that existence of clusters of units with similar level of unemployment rate corresponds with polarization processes observed in Europe. However this fact is not often taken into consideration by European policy-makers.

[*] Ph.D., Wroclaw University of Technology

Poland has experienced changes of labour market both as a result of the transformation in Polish economy and the integration process with the European Union.

The paper aims to present an analysis of relationship among different variables connected with labour market and intra-regional unemployment disparities in Lower Silesia, one of the Polish region. The spatial analysis was made by means of GeoDa software (Anselin 2005). The source of statistical data is GUS (Główny Urząd Statystyczny, Central Statistical Office) (*Local Data Bank*, *Województwo Dolnośląskie* 2011).

2. Lower Silesia

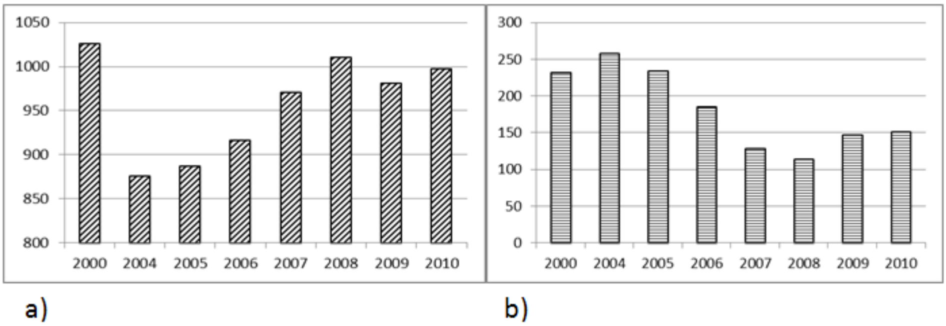
Lower Silesia (Dolnośląskie voivodship) is situated in the south-western corner of Poland. It neighbours Germany and Czech Republic. The region is strategically located in Poland and in Europe, at the intersection of transport routes leading from the east to the west, and from the south to the north. It covers an area of 19,947 sq.km and has a population of nearly 2.9 million.

The Lower Silesian natural resources (including copper ores and lignite) form a base for the development of mining and other industries. Modern industrial processing and services are also developing. In addition, the region's geographical location, environmental conditions, medicinal waters and its rich cultural heritage make tourism and spa healthcare one of the leading industries.

A significant part of Lower Silesia is or was in the past heavily industrialized. During the transformation many production facilities have been liquidated as a consequence of technological, and market changes, as well as conflicts with environmental requirements. The largest fall has affected the textile industry and the coal mining. However key to the Lower Silesian economy is the extraction of copper, silver and lignite. In the northern part of the region there is the Legnica-Głogów Copper Mining District. The lignite mine is located in the gmina Bogatynia, in the south-western corner of the region.

Changes in economy have effected labour market, especially in the beginning of 1990s. Fluctuations in number of jobs and unemployed people in period 2000-2010 are presented on Figure 1. In 2000, the industry provided jobs to 343 thousand people, whereas in 2010 only 248 thousand. It results from a decay in many branches of industry, as well as obsolete technologies and production facilities.

Figure 1. Lower Silesia



Note: umber of work places (a) and unemployed persons (b) (in thousand) in the period 2000-2010

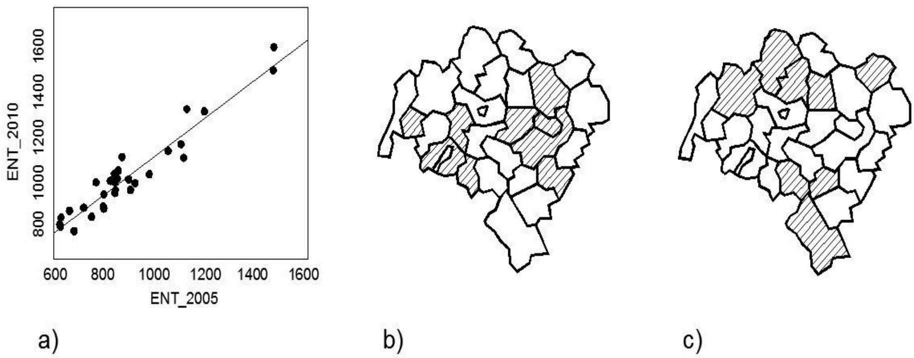
Source: Own elaboration based on GUS data.

3. NUTS 4 level (poviats)

In the first part of analysis it was taken into consideration the division of Lower Silesia into 29 poviats (NUTS 4). The following variables connected with labour market were taking into account: population (POP), population in working age (W_AGE), work places (JOBS), jobs in agriculture (AGR), industry (IND) and services (SER), registered rate of unemployment (UNEMP), the number of entities of national economy recorded in REGON register (ENT) and wages (WAGE). Wages were expressed by average monthly gross wages to the national average (Poland = 100). Data set was for 2005 and 2010.

Changes between 2005 and 2010 may be presented by scatter plots (Figure 2.). In the period 2005-2010 the number of entities increased about 12%. Selection of poviats in which the increase of the variable was noticed (points lying above line) allows to show chosen units on the map. The growth was noticed in Wroclaw and neighbouring poviats and in the group of units near Jelenia Góra (with exception of this city). The decline was recorded in Kotlina Kłodzka and in the northern part of voivodship. The images reflect the real economic situation.

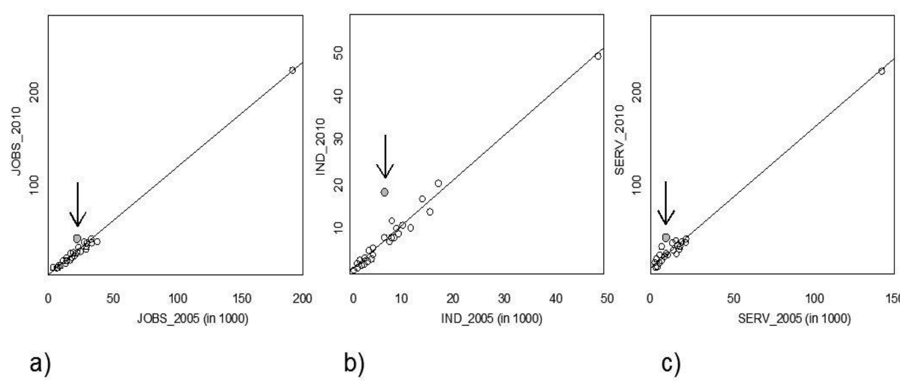
Figure 2. Lower Silesia (NUTS 4). The number of entities in 2005 and 2010. Scatter plot (a) areas with growth (b) decreased areas (c)



Source: Own elaboration by using GeoDa.

Scatter plots for the number of all work places, jobs in industry and in services show that in the period 2005-2010 considerable growth was recorded only in wroclawski poviat what demonstrates the impact of Wrocław on economy of neighbouring units (Figure 3).

Figure 3. Lower Silesia (NUTS 4). Scatter plots for all work places (a) jobs in industry (b) jobs in services (c) in 2005 and 2010



Note: The arrows point the unit with the biggest growth

Source: Own elaboration by using GeoDa.

The first law of geography according to Tobler (1970) is: "All places are related but nearby places are more related than distant places”. In other words, spatial autocorrelation means that dependency exists between values of a variable in neighbouring locations.

The spatial autocorrelation analysis is handled by means of Moran's I statistics (Moran 1950) which requires creating a weight matrix describing a local neighbourhood of each geographic unit. Most common is using binary connectivity based on contiguity: $w_{ij} = 1$ if units i and j are adjacent, $w_{ij} = 0$ otherwise.

The statistics I is based on cross-products of the deviations from the mean and is calculated for observations on a variable x at n locations:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (1)$$

where $\bar{x}$ is the mean of the x variable, w_{ij} are the elements of the weight matrix, i, j are number of units, and S_0 is the sum of the elements of the weight matrix: $S_0 = \sum_i \sum_j w_{ij}$ ($i, j = 1, 2, \ldots n$).

The Moran's statistics I is a single value which applies to the entire data set. It varies from -1 to +1. High negative spatial autocorrelation (-1) means dispersed pattern, and high positive spatial autocorrelation (+1) – clustered pattern. A random arrangement gives a value that is close to 0. It can be interpreted as the correlation between variable *X* and the "spatially lagged" of *X* formed by averaging all the values of *X* for the neighbouring units (*W_X*).

The Moran scatter plot is a plot with the variable *X* on the *x*-axis and the spatially lagged on the *y*-axis. The variables are standardized so that the units in the graph correspond to standard deviation. The slope of the regression line equals the Moran's I statistics.

The Moran's I statistics were generated for following variables: population, population in working age, work places, jobs in agriculture, industry and services, rate of unemployment, wages and the number of entities for 2005 and 2010 as well (Table 1.). Inference for Moran's I was based on a permutation approach, in which a reference distribution was calculated for spatially random layouts with the same values as observed variables (Anselin 1986). 49999 permutations were used in generating the reference distribution.

Weights were constructed based on contiguity to poviat's boundaries (the first order rook contiguity.

Table 1. Poviats. Moran's I statistics

Variable	2005		2010	
	Moran's I	p-value	Moran's I	p-value
Population (POP)	-0.0008	0.73	0.0095	0.22
Working age population (W_AGE)	-0.0047	0.71	0.0099	0.22
Work places (JOBS)	-0.0175	0.66	0.0138	0.15
Jobs iIn agriculture (AGR)	0.1288	0.09	0.1185	0.11
Jobs in industry (IND)	-0.0362	0.53	0.0748	0.12
Jobs in services (SER)	-0.0291	0.59	-0.0171	0.71
Rate of unemployment (UNEMP)	0.3050	**0.005**	0.2267	**0.024**
Wages (WAGE)	-0.0529	0.46	-0.0166	0.630
Entities (ENT)	0.1235	0.09	0.2177	**0.028**

Source: Own calculation based on GUS data and by using GeoDa.

The Moran's I statistics is statistically significant only for the unemployment rate in 2005 and for the unemployment rate and the number of entities in 2010. Further analysis was conducted for 2010.

In the next step it was decided to apply multivariate regression analysis (Anselin 1988; Anselin, Bera 1998; Suchecki 2010) with the rate of unemployment as the dependent variable and following independent variables: the number of jobs in agriculture, industry and services, the level of wages and the number of entities. The remained variables were excluded from the set because of the high correlation among them.

The regression linear function is:

$$y = x\beta + \varepsilon, \tag{2}$$

where dependent variable y (n×1) is vector of observed dependent variable in n spatial units, x (n×k) represents k independent, explanatory variables. Vector β (k×1) contains unknown coefficients, and vector ε (n×1) represents error term.

In the 1st variant of calculation unknown coefficients were solved by using standard linear regression (Ordinary Least Squares - OLS). Not all coefficients were significant (Table 2.) and in addition multicollinearity condition number was greater than 20 what is not acceptable. AGR_10 (employment in agriculture) was rejected from the set of independent variables.

In the 2nd variant the value of multicollinearity condition number diminished but it was still bigger than 20. So the next independent value was excluded. It was SER_10 (employment in services).

In the 3rd variant there were three following independent variables: IND_10 (employment in industry), WAGE_10 (wages) and ENT_10 (the number of entities). The multicollinearity condition number was proper (18). A test for non-normality (Jarque-Bera) said that errors had normal distribution (p=0.40) and three diagnostics for heteroskedasticity (Breusch-Pagan, Koenker-Bassett, and White) showed a constant variance of errors (p=0.95, 0.87, and 0.46). But tests did not confirm spatial dependence. Therefore it was decided to limit a set of independent variables. Only the level of wages and the number of entities were taken into consideration. The first variable measures the attractiveness of poviat for employees, the second one – expresses the level of economic activity of local population.

In 4th variant of modeling both coefficients were significant. The multicollinearity condition number was 14.3. Test Jarque-Bera said that errors had normal distribution (p=0.51). Homoskedasticity was confirmed (p=0.91, 0.84 and 0.29) for three tests. Both Lagrange Multipliers, M-Lag and LM-Error statistics, rejected the null hypothesis (p=0.020 and 0.0197), what told about the spatial dependence. Also Moran's test statistics I confirmed spatial autocorrelation (p=0.0037).

Table 2. Poviats. Regression analysis. Significance of independent variable coefficients (*p*-value)

	Variant 1 OLS	Variant 2 OLS	Variant 3 OLS	Variant 4 OLS	Variant 5 Spatial Lag Model
W_UNEM_10	-	-	-	-	0.004
AGR_10	0.245	-	-	-	-
IND_10	0.011	0.08	0.03	-	-
SER_10	0.007	0.17	-	-	-
WAGE_10	0.052	0.06	0.065	0.014	0.007
ENT_10	0.082	0.14	0.5	0.035	0.002

Source: Own calculation based on GUS data and by using GeoDa.

Then it was applied the spatially lagged model (5th variant):

$$y = \rho Wy + x\beta + \varepsilon. \tag{3}$$

The spatially lagged rate of unemployment W_UNEM_10 appeared as the additional indicator. Its coefficient parameter (ρ=0.504) reflected the spatial dependence, measuring the average influence on observation in neighbouring units. Breusch-Pagan test detected hetroskedasticity. The final diagnostics was for spatial autocorrelation. Likelihood Ratio Test was significant (p=0.019).

Finally the equation has the form:

$$\text{UNEM\_10} = 0.504 \cdot \text{W\_UNEM\_10} - 0.064 \cdot \text{WAGE\_10} - 0.000499 \cdot \text{ENT\_10} + 15.17$$

what confirms spatial dependence of the unemployment rate and, what it is obvious, the negative relationship between the unemployment and the level of wages and the number of entities.

4. NUTS 5 level (gminas)

One of the problems with areal data is scale effect – spatial data at different scales produce different results. In contrast, the next analysis examined the spatial aspects of Lower Silesian labour market using smaller units of observation – municipalities (gminas - NUTS 5). There are 169 gminas (36 urban gminas, 54 urban-rural and 79 rural ones).

For the year 2010 there were taken into account: population (POP), population in working age (W_AGE), work places (JOBS), rate of unemployment (UNEMP), budget revenue *per capita* (REV), the number of entities (ENT) and the variable describing attraction of labour market (ATR). The last variable was equal the number of people coming to work from other gminas divided by the number of inhabitants working outside their own gmina in 2006 (*Dojazdy do pracy*… 2010).

In the first place, three types of weight matrices were generated using both rook and queen criteria, and 5 nearest neighbours (5nn) as well. Next Moran's I statistics was calculated (Table 3.).

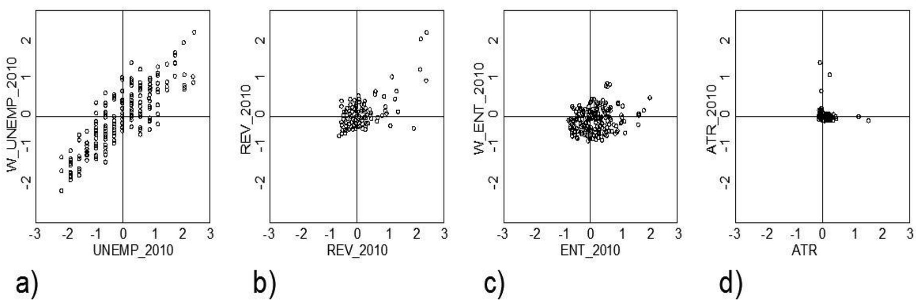
Table 3. Gminas. Moran's I statistics for different weight matrices

W_matrix	POP	W_AGE	JOBS	UNEMP	REV	ENT	ATR
Rook p- value	0.0094 (0.148)	0.0072 (0.162)	0.0151 (0.105)	**0.7285** **(0.00002)**	**0.5097** **(0.00002)**	**0.1342** **(0.0042)**	**0.0198** **(0.0039)**
Queen	0.0087	0.0067	0.0154	0.7154	0.5072	0.1414	0.0210
5nn	0.0025	0.0031	-0.030	0.6717	0.2895	0.1295	-0.0844

Source: Own calculation based on GUS data and by using GeoDa.

The highest value of Moran's statistics was for the unemployment rate (0.7285). Kossowski and Hauke (2012) calculated Moran's coefficient for all Polish gminas and got value 0.7426. Besides the unemployment rate Moran's statistics is statistically significant for the budget revenue *per capita*, the number of entities and the attraction. But the last ones had value near zero (Figure 4.).

Figure 4. Lower Silesia (NUTS 5). Moran scatter plots for unemployment rate (a), budget revenue *per capita* (b), entities (c) and attraction of gmina (d)



Source: Own elaboration by using GeoDa.

In order to determine whether or not geographic patterns exist it was performed LISA (the local indicator of spatial association), the local version of Moran's I (Anselin 1995). For each unit i the value of indicator is calculated according to formula:

$$I_i = z_i \sum_j w_{ij} z_j \tag{4}$$

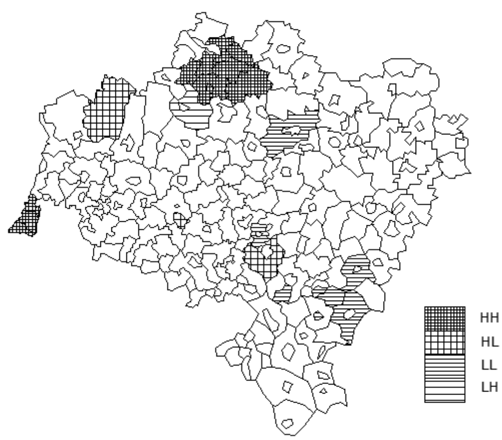
where z_i is the standardized form of original variable x_i, w_{ij} is the element of the spatial weight matrix and j describes neighbours of unit i.

This measure can be used to identify local clusters or spatial outliers – areas different from their neighbours.

To illustrate the local spatial autocorrelation there were created LISA cluster maps. This map shows a statistically significant relationship with neighbours and four types of spatial autocorrelation. The positive spatial association is when units with high values is surrounded by ones with high values (high-high, H-H) or units with low values are bordered by similar units (low-low, L-L). The negative autocorrelation says that units are different from their neighbours (high-low, HL or low-high, LH).

On Figure 5, the LISA cluster map is reported for the budget revenue *per capita*. The adopted level of significance was p=0.05. Two clusters of the highest budget revenues overlap with Legnica-Głogów Copper Industrial District in the northern part of region and gmina Bogatynia situated in the west. These areas have industrial character. The municipalities (Strzelin, Ciepłowody, Niemcza, Piława Górna and Ziębice) with the lowest budget create cluster in the southern east part of the region.

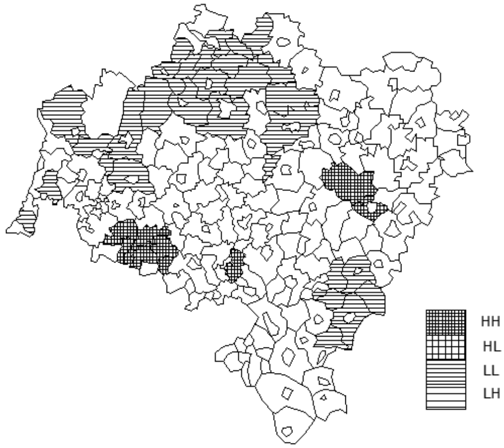
Figure 5. LISA cluster map for budget revenue *per capita*



Source: Own elaboration by using GeoDa.

The LISA map for the number of entities (Figure 6.) shows two groups of low values. The first cluster is located in the northern part of region, where either there are less developed municipalities or municipalities have little entities, but employing a lot of employees (Legnica-Głogów Industrial District). The second one consists of following gminas: Strzelin, Kondratowice, Przeworno, Ciepłowody, Ziębice and Kamieniec Ząbkowicki. Clusters of units with high values is created by gminas near Jelenia Gora where tourism is well developed and Wrocław with gmina Święta Katarzyna.

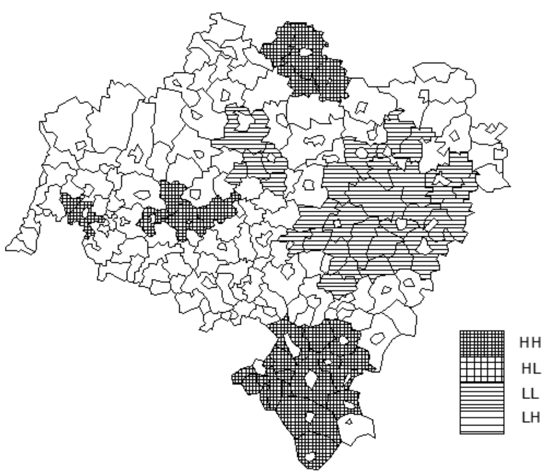
Figure 6. LISA cluster map for the number of entities



Source: Own elaboration by using GeoDa.

In the next step LISA cluster map was created for the rate of unemployment (Figure 7.). There are two clusters of low unemployment, the big one covers Wroclaw Metropolitan Area and smaller one is situated near Lubin. High level of unemployment is recorded in Kotlina Kłodzka, in the northern part of the region (Góra, Niechlów, Wąsosz and Jemielno), and in two groups in the western part.

Figure 7. LISA cluster map for rate of unemployment

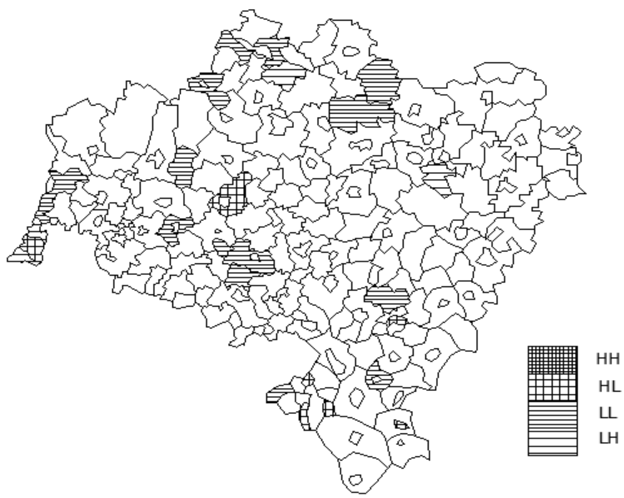


Source: Own elaboration by using GeoDa.

This image shows the spatial structure of regional unemployment disparities. On the one hand it is seen the important role of Wrocław as pole of growth and on the other hand it is easy to notice that despite the great touristic potential municipalities in Kotlina Kłodzka are not able to use it sufficiently.

The attraction of the gmina was defined as the number of people coming to work from other gminas divided by the number of inhabitants working outside their own gmina. The variable defined in such a way does not create concentrations (Figure 8).

Figure 8. LISA cluster map for attraction of gmina



Source: own elaboration by using GeoDa.

5. Conclusions

In this paper it was presented the spatial analysis of the labour market in Lower Silesia. The first part of research was run on a data set for poviats (NUTS 4). The Moran's statistics was calculated and then the spatial regression model for the unemployment was obtained.

In turn, division of Lower Silesia into gminas (NUTS 5) was taken into account. Among different variables connected with labour market it were chosen four of them with statistically significant Moran's statistics I and then LISA cluster maps were created. The biggest clusters were found for the unemployment and the budget revenue *per capita*.

All these calculations confirm that the unemployment is spatially dependent and tends to concentrate. The rate of unemployment seems to be one of the most important indicators which characterize the labour market.

References

Anselin L. (1986), *MicroQAP: a Microcomputer Implementation of Generalized Measures of Spatial Association*, Working Paper, Department of Geography, University of California, Santa Barbara

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

Anselin L. (1995), *Local Indicators of Spatial Association — LISA*, Geographical Analysis, 27: 93-115

Anselin L. (2005), *Exploring Spatial Data with GeoDa*™: *A Workbook*, Center for Spatially Integrated Social Science, Urbana, IL

Anselin L., Bera A. K. (1998), *Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics*. [in:] A. Ullah and D. Giles (eds), *Handbook of Applied Economic Statistics*. New York: Marcel Dekker, pp.237-289

Dojazdy do pracy w Polsce. Terytorialna identyfikacja przepływów ludności związanych z zatrudnieniem [2010] Urząd Statystyczny w Poznaniu, Poznań

Moran P.A.P. (1950), *Notes on Continuous Stochastic Phenomena* Biometrika 37:17-23

Kossowski T. Hauke J. (2012), *Analysis of the Labour Market in Metropolitan Areas: a Spatial Filtering Approach*, Quaestiones Geographicae, 31[2], Bogucki Wydawnictwo Naukowe, Poznań, pp. 39-48

Local Data Bank. Central Statistics Office of Poland < http://www.stat.gov.pl/>

Niebuhr A. (2003), *Spatial Interaction and Regional Unemployment in Europe* European Journal of Spatial Development, 5 <http:www.nordregio.se/EJSD/>

Overman H.G. Puga D. (2000), *Unemployment clusters across European Regions and Countries* Centre for Economic Performance, Discussion Paper 434

Suchecki B. (2010), *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H. Beck, Warszawa

Tobler W. (1970), *A Computer Movie Simulating Urban Growth in the Detroit Region* Economic Geography, 46[2]:234-240

Województwo Dolnośląskie. Podregiony. Powiaty. Gminy (2011) Urząd Statystyczny we Wrocławiu, Wrocław

Streszczenie

PRZESTRZENNA ANALIZA RYNKU PRACY PRZY UŻYCIU NARZĘDZI EKONOMETRYCZNYCH. PRZYKŁAD WOJEWÓDZTWA DOLNOŚLĄSKIEGO

Artykuł prezentuje wykorzystanie technik ekonometrycznych do badania rynku pracy na Dolnym Śląsku. Najpierw analiza dotyczyła zbioru danych związanych z rynkiem pracy na poziomie powiatów (NUTS 4). W celu zbadania istnienia autokorelacji przestrzennej policzono statystykę Morana I, potem został użyty model regresji przestrzennej do opisu zależności pomiędzy stopą bezrobocia a innymi zmiennymi. W następnym kroku wygenerowano mapy obrazujące lokalny wskaźnik zależności przestrzennej LISA dla gmin (NUTS5). Wyniki wskazują na przestrzenny wymiar bezrobocia i jego tendencję do tworzenia koncentracji.

10.2478/v10103-012-0033-7

ZOFIA MIELECKA-KUBIEŃ*

The Analysis of the Life Expectancy and the Selected Causes of Deaths in Poland with the Use of Spatial Statistics Methods

Abstract

The goal of the presented research was to test the spatial autocorrelation of the life expectancy and the age-standardized mortality rates for selected causes of death in Poland according to gender in 2010. It was assumed that in the above mentioned cases the positive spatial autocorrelation in populations of men and women appears, and the spatial diversity of mortality depends on the standard of living of the population in question and on the level of industrialization of the region and its consequences. It has been stated that most of the considered coefficients show positive spatial autocorrelation, but differences between populations of men and women were observed. Agricultural capacity of the voivodeship shows positive effect on life expectancy and the level of some of the mortality rates of both genders.

1. Introduction

With the development of the methods of spatial statistics, as well as of the geographic information system (GIS), it became possible to conduct the deepened research on spatial variation of mortality characteristics. The aim of the presented study was to verify the supposition – using the basic spatial statistics methods - whether life expectancy (e_0) and mortality rates from selected causes of deaths in Poland show spatial autocorrelation. The especially interesting question was, whether the pattern of spatial autocorrelation was the

* Ph.D., Professor at the University of Economics in Katowice

same for men and women (living in the same country, at the same time). Finally, an attempt was made to find factors possibly influencing the observed spatial differences and spatial autocorrelation patterns of the coefficients taken into account.

2. The applied method

The research was conducted with the use of the basic spatial statistics methods, i.e.:

1. Moran's global statistic *I*, defined as follows (Kopczewska 2006, p.72):

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2} \tag{1}$$

where: n – denotes the number of spatial units indexed by i and j (here voivodeships or in the case of life expectancy, also subregions of Poland), y – the considered coefficient, w_{ij} – element of the matrix of spatial weights W, constructed according to numbers of the first order neighbours, row-standardized. Expected value of statistic I is:

$$E(I) = \mu = \frac{-1}{n-1} \tag{2}$$

Assuming that the sample comes from the independent random variables normally distributed, test statistic I_s :

$$I_S = \frac{I - E(I)}{D(I)} \tag{3}$$

is asymptotic normal distributed[1] $N \approx (0,1)$. On this basis the hypothesis $H_o : I = 0$ against the hypothesis $H_1 : I \neq 0$ is tested.

[1] See for instance (Ekonometria przestrzenna 2010, p.109).

Local Moran statistic I_i is defined as:

$$I_i = \frac{(y_i - \bar{y})\sum_i w_{ij}(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2 / n} \tag{4}$$

with the expected value (Anselin 1995) given by:

$$E(I_i) = \frac{-1}{n-1}\sum_{j=1}^{n} w_{ij} \,. \tag{5}$$

The local Moran test, based on the conditional randomisation or permutation (see: Anselin 1995), detects the local spatial autocorrelation. There can be two interpretations of the local Moran statistics: as indicators of the local spatial clusters (regions where adjacent areas have similar values) and as a diagnosis for the outliers in global spatial patterns (areas distinct from their neighbours). The Local Moran statistic I_i decomposes Moran's I into contributions for each location:

$$I = \sum_i \frac{I_i}{n} \tag{6}$$

Additionally the Moran scatterplots and maps were applied. Moran scatterplot allows (Anselin 2005) to explore the global patterns of autocorrelation in space. The graph depicts the standardized variable (here life expectancy or age-standardized mortality rates) in the *x*-axis versus the spatial lag of that standardized variable, where the spatial lag shows the effects of the neighboring spatial units. Moran scatterplot presents the relation of the variable in the location *i* with respect to 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* statistic. If that slope is positive it means that there is the positive spatial autocorrelation: 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. If the slope in the scatter plot is negative it means that high values in a variable in location *i* tend to be co-located with lower values in the neighboring locations.

3. Empirical results

The presented research was conducted[2] on the basis of 16 voivodeships of Poland (spatial units NUTS2); in the case of the life expectancy for men and women smaller units (66 subregions, NUTS3) could be taken into account. The data for the year 2010 come from the Chief Statistical Office in Warsaw. All the considered mortality rates were standardized with regard to age. The following variables were subjects of the study for men (m) and women (k): life expectancy in voivodeships (Y_{om},Y_{ok}) and subregions (X_{om},X_{ok}), general mortality rates (Y_{1m},Y_{1k}), cancer mortality rates (Y_{2m},Y_{2k}), circulatory system diseases mortality rates (Y_{3m},Y_{3k}) and respiratory system diseases mortality rates (Y_{4m},Y_{4k}). While circulatory system diseases and cancer were the most frequent causes of death in 2010 in Poland (tab.1), respiratory system diseases were chosen with regard to their specific spatial differences pattern.

Table 2 presents the results of the testing of the hypothesis about the absence of the spatial autocorrelation on the basis of Moran global statistic *I* (as described above). It can be observed that not all of the considered variables show significant spatial autocorrelation.

Table 1. Percent of deaths for chosen causes in Poland, 2010

DISEASE	PERCENT OF DEATHS
Circulatory system	46.0
Cancer	25.4
Respiratory system	5.1

Source: author's own.

As can be observed (fig.1, tab.2) the life expectancy for men in voivodeships does not show spatial autocorrelation – different values of life expectancy are randomly distributed across the country. The shortest life expectancy for men in Poland in the year 2010 was observed in the voivodeship Lodzkie; on the contrary, the longest life expectancy for men was in the voivodeships Malopolskie and Podkarpackie, in the southern part of the country.

[2] For calculation there were used computer programs R and EXCEL, for visualisation – programs EXCEL and Statistica.

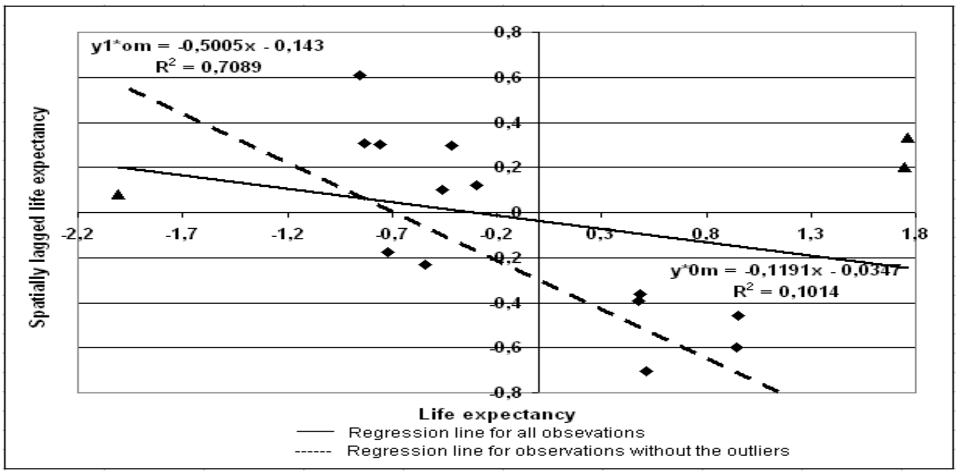
Table 2. Results of testing hypothesis of absence of spatial autocorrelation

COEFFICIENT	MORAN *I* STATISTIC		TEST STATISTIC I_S	
	MEN	WOMEN	MEN	WOMEN
Life expectancy	-0.1191	**0.2771**	-0.6163	**4.0433**
Life expectancy, subregions	**0.4002**	**0.6095**	**4.5198**	**6.7966**
General mortality rates	-0.1311	**0.1908**	-0.7577	**3.0285**
Cancer mortality rates	**0.4194**	**0.4602**	**5.7161**	**6.1962**
Circulatory system diseases mortality rates	-0.0790	**0.1044**	-0.1453	**2.0123**
Respiratory system diseases mortality rates	**0.1901**	**0.2502**	**3.0201**	**3.7268**

Remark: cases of rejected null hypothesis (α = 0.05) are marked in bold.

Source: author's own.

Figure 1. Moran plot for life expectancy (e_0), men

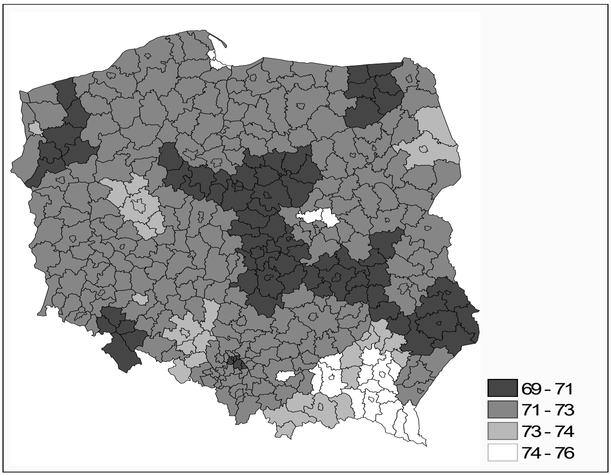


Source: author's own.

However, the theoretical elimination of the three outliers (voivodeships: Lodzkie, Malopolskie, Podkarpackie, marked on fig.1 with black triangles) indicates, that apart from the three voivodeships, the life expectancy for men is characterized by negative spatial autocorrelation (y1*om, dashed line), which means, that in the case of the remaining voivodeships it is revealed that the neighboring values are more dissimilar than expected by chance.

A considerably different pattern emerges from the men life expectancy considerations based on the subregions of Poland (tab.2, fig.2) – here the spatial autocorrelation is more positive than expected at random, which indicates the clustering of similar values across smaller items in geographic space. The longest life expectancy is observed (tab.3) in the big cities: Warsaw, Cracow, subregion trojmiejski containing the cities of Gdansk, Sopot and Gdynia and in the south-eastern corner of Poland (subregions: rzeszowski and tarnowski). The lowest values of men life expectancy can be seen (as in fig.2) in the subregions of the voivodeship Lodzkie, and subregion stargardzki in the northern part of Poland.

Figure 2. Spatial differences in men life expectancy (in years), subregions



Source: author's own.

For women both spatial patterns of life expectancy (the one based on the voivodeships as well as the one based on the subregions), are different than those for men (fig.3 and 4). Apparently the Polish women living, generally speaking, in the eastern part of the country enjoy longer life than the ones living in the westren part. In this case both spatial differences patterns are similar, however the more detailed analysis (subregions) uncovered some significant exceptions: Warsaw, Wroclaw and subregion trojmiejski, but, as indicated in table 3, the longest women life expectancy is observed in the eastern subregions of Poland (and not in the big cities as in the case of men). The shortest women life expectancy was observed in the subregions belonging to the voivodeships: Lodzkie and Slaskie.

A strong positive spatial autocorrelation can be observed for the men and women mortality of cancer (tab.3, fig.5 and 6). Apparently the inhabitants of the north-western part of Poland are more at risk from cancer, and, in the case of women, especially those living in the voivodeship kujawsko-pomorskie. The lowest values, for both genders, can be observed in the eastern part of Poland – with generally much higher level of age-standardized cancer mortality rates in the population of men. For the age-standardized mortality rates for the diseases of the respiratory system the spatial differences pattern as well as the level of positive spatial correlation (tab.3) are similar for both genders, but the voivodeship Warminsko-mazurskie takes the strongly exceptional position (fig.7) – the mortality rates are very high (especially for men). The theoretical elimination of that outlier could change the slope of the regression line in the Moran scatter plot for men from a_1 = 0.1901 to a_2 = 0.6484 (fig.8, y1*4m, dashed line), and for women from a_1 = 0.2502 to a_2 = 0.6762.

Table 3. Subregions of highest and lowest life expectancy (in years) according to gender

HIGHEST LIFE EXPECTANCY, MEN		HIGHEST LIFE EXPECTANCY, WOMEN	
75.3	Warsaw	82.1	bialostocki
75.1	Cracow	82.0	tarnobrzeski
74.6	trojmiejski	81.8	łomzynski
74.4	rzeszowski	81.7	rzeszowski
74.2	tarnowski	81.7	suwalski
LOWEST LIFE EXPECTANCY, MEN		LOWEST LIFE EXPECTANCY, WOMEN	
70.0	lodzki	79.1	grudziadzki
70.0	Lodz	79.1	sosnowiecki
70.0	skierniewicki	79.0	Lodz
70.0	stargardzki	78.8	lodzki
69.9	piotrkowski	78.5	katowicki

Source: author's own.

Figure 3. Spatial differences in women life expectancy (in years), voivodships

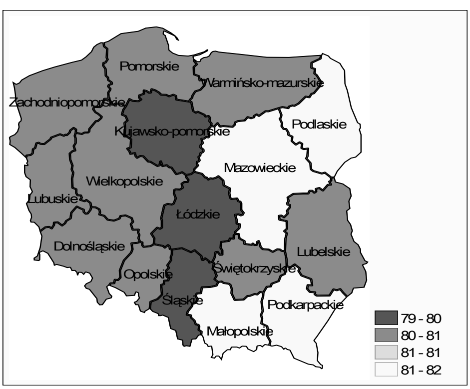
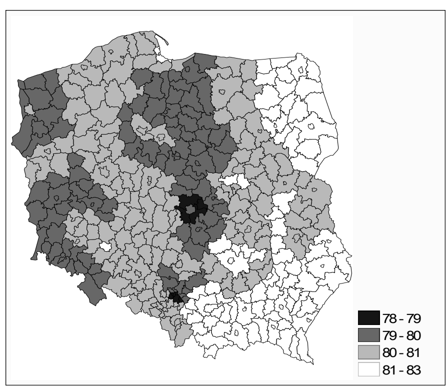


Figure 4. Spatial differences in women life expectancy (in years), subregions



Source: author's own.

Figure 5. Spatial differences in cancer mortality rates*, men, voivodships

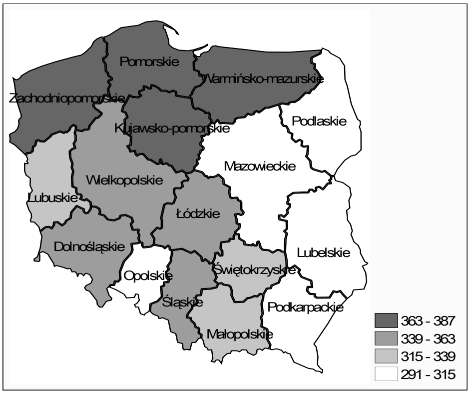
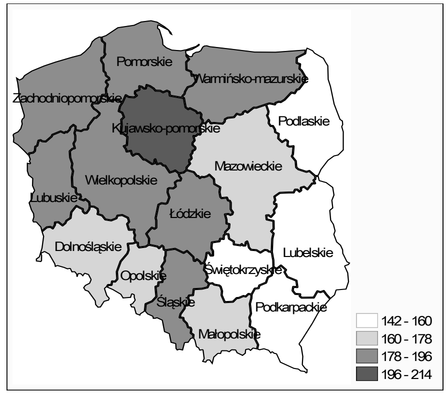


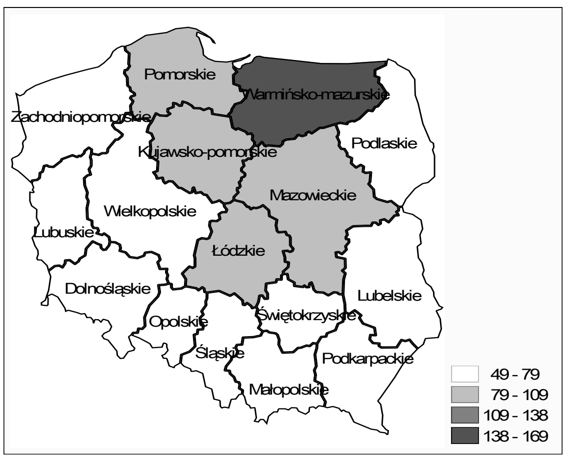
Figure 6. Spatial differences in cancer mortality rates*, women, voivodships



Note: * – per 100 000 population

Source: author's own.

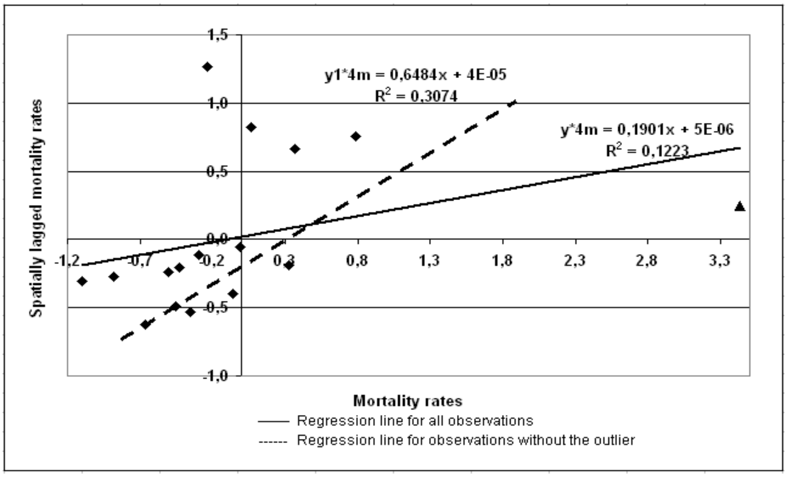
Figure 7. Spatial differences in respiratory diseases mortality rates*, men



* – per 100 000 population

Source: author's own.

Figure 8. Moran plot for respiratory diseases mortality rates, men



Source: author's own.

For women, the positive spatial autocorrelation was observed also for the general mortality rates – the pattern is similar to the one for the life expectancy and for the circulatory system diseases mortality rates, where the women living in the north-eastern part of Poland are in the most favourable position; the highest level of circulatory system diseases mortality rates can be observed in voivodeships: Lodzki, Swietokrzyski, Lubelski – in the central part of the country. In the case of men both adequate global Moran statistics I were not significant ($\alpha = 0.05$) though in the case of general mortality rates, similarily, as in the case of life expectancy, this situation is due to the same three outliers.

Table 4 presents the results of the testing of the significance of Local Moran Statistic I_i. Significant ($\beta = 0.9$) negative local statistic occurred only in one case – for men general mortality rates in voivodeship Kujawsko-pomorskie, which is surrounded by voivodeships with lower level of the mortality rates. Most of all the significant positive Local Moran Statistics I_i can be observed in the case of cancer mortality rates for men (Y_{2m}) and women (Y_{2k}) which supports the previous conclusion that in case of cancer clusters of items of similar values of mortality rates are distinctly visible.

The question arises what factors influence the values of the considered coefficients and cause the specific spatial autocorrelation patterns? To look for a possible explanation the set of 16 diagnostic variables was applied. The variables are indicators of factors often associated with the mortality level such as: the standard of living, the health care level, the industrialization and the connected phenomena including air pollution and population density, some stressful social occurrences (unemployment). They were: population per 1 km^2. (Z_1), share of urban population (Z_2), voivodeship revenues per 1 inhabitant (Z_3), voivodeship expenditures for health care per 1 inhabitant (Z_4), district revenues per 1 inhabitant (Z_5), district expenditures per 1 inhabitant (Z_6), district expenditures for education per 1 inhabitant (Z_7), district expenditures for culture per 1 inhabitant (Z_8), district expenditures for health care per 1 inhabitant (Z_9), share of forest land in land area (Z_{10}), sold production of industry per 1 inhabitant (Z_{11}), average monthly wages and salaries (Z_{12}), unemployment rate (Z_{13}), emission of air pollutant particulates per 1 km^2 (Z_{14}), emission of air pollutant sulphur dioxide per 1 km^2 (Z_{15}) and emission of air pollutant nitrogen oxides per 1 km^2 (Z_{16}).

On the basis of the diagnostic variables Z_1-Z_{16} the following synthetic variables were constructed[1]: $V_1 = (Z_1, Z_2)$ – characterizing the demographic features of the voivodeship, $V_2 = (Z_4, Z_9)$ – describing its level of health care, $V_3 = (Z_5, Z_6, Z_7, Z_8)$; the purpose of introducing the variable was to determine the factors influencing the social standard of living, $V_4 = (Z_{14}, Z_{15}, Z_{16})$ – characterizes the level of pollution in the voivodeships.

The coefficients of the correlation between the life expectancy, the above considered mortality rates and the synthetic or the diagnostic variables are presented in table 5.

[1] In the form of unweighted averages, the values of diagnostic variables Z_j standardized according to formulae: $v_i = \dfrac{z_i - \bar{z}}{s(z)}$.

Table 4. Local Moran statistics (I_i), test statistics (I_l) and *p-values* for Y_{0m}-Y_{4m} and Y_{0k}-Y_{4k}

VOIVODSHIP	Y_{0m}			Y_{1m}			Y_{2m}			Y_{3m}			Y_{4m}		
	I_i	I_l	*p-value*	I_i	I_l	*p-value*	I_i	I_l	*p-value*	I_i	I_l	*p-value*	I_i	I_l	*p-value*
DOLNOSLASKIE	-0.046	0.049	0.480	-0.001	0.154	0.439	-0.135	-0.157	0.562	-0.087	-0.048	0.519	0.227	0.843	0.200
KUJAWSKO-POMORSKIE	-0.364	-0.932	0.824	**-0.521**	**-1.423**	**0.923**	-0.144	-0.240	0.595	-0.249	-0.571	0.716	0.252	1.113	0.133
LODZKIE	-0.036	0.097	0.461	-0.035	0.100	0.460	**0.469**	**1.662**	**0.048**	0.225	0.910	0.181	0.037	0.362	0.359
LUBELSKIE	-0.123	-0.110	0.544	-0.091	-0.048	0.519	-0.020	0.090	0.464	-0.085	-0.037	0.515	0.410	1.189	0.117
LUBUSKIE	0.127	0.529	0.298	0.154	0.605	0.273	**0.929**	**2.689**	**0.004**	-0.238	-0.467	0.680	**0.603**	**2.144**	**0.016**
MALOPOLSKIE	-0.521	-1.066	0.857	-0.546	-1.129	0.870	**1.096**	**2.688**	**0.004**	-0.180	-0.266	0.605	0.000	0.192	0.424
MAZOWIECKIE	-0.157	-0.282	0.611	-0.236	-0.530	0.702	0.018	0.262	0.397	-0.010	0.177	0.430	-0.064	0.009	0.496
OPOLSKIE	0.583	1.269	0.102	**0.623**	**1.353**	**0.088**	0.539	1.165	0.122	-0.361	-0.573	0.717	0.251	0.792	0.214
PODKARPACKIE	-0.437	-0.867	0.807	-0.620	-1.301	0.903	**1.316**	**3.196**	**0.001**	0.148	0.503	0.307	0.059	0.361	0.359
PODLASKIE	0.125	0.374	0.354	0.198	0.520	0.301	-0.024	0.081	0.468	0.129	0.381	0.351	0.183	0.623	0.267
POMORSKIE	-0.176	-0.388	0.651	-0.015	0.182	0.428	**0.434**	**1.770**	**0.038**	-0.229	-0.576	0.718	0.085	0.571	0.284
SLASKIE	-0.565	-1.169	0.879	-0.628	-1.320	0.907	-0.426	-0.831	0.797	0.009	0.176	0.430	0.333	1.149	0.125
SWIETOKRZYSKIE	0.354	0.821	0.206	0.285	0.690	0.245	**1.408**	**2.835**	**0.002**	0.090	0.305	0.380	0.032	0.247	0.402
WARMINSKO-MAZURSKIE	-0.256	-0.443	0.671	-0.294	-0.535	0.704	0.312	0.875	0.191	**0.484**	**1.288**	**0.099**	**0.866**	**2.678**	**0.004**
WIELKOPOLSKIE	-0.185	-0.232	0.592	-0.167	-0.198	0.578	0.243	0.596	0.276	-0.160	-0.182	0.572	-0.296	-0.572	0.716
ZACHODNIOPOMORSKIE	-0.228	-0.315	0.624	-0.210	-0.282	0.611	**0.663**	**1.404**	**0.080**	-0.759	-1.350	0.911	0.117	0.458	0.323

VOIVODSHIP	Y_{0k}			Y_{1k}			Y_{2k}			Y_{3k}			Y_{4k}		
	I_i	I_l	*p-value*	I_i	I_l	*p-value*	I_i	I_l	*p-value*	I_i	I_l	*p-value*	I_i	I_l	*p-value*
DOLNOSLASKIE	0.062	0.300	0.382	0.173	0.562	0.287	-0.319	-0.586	0.721	0.191	0.598	0.275	0.265	0.928	0.177
KUJAWSKO-POMORSKIE	-0.017	0.156	0.438	-0.116	-0.154	0.561	0.008	0.231	0.409	-0.311	-0.761	0.777	0.287	1.221	0.111
LODZKIE	0.091	0.491	0.312	0.008	0.235	0.407	**0.547**	**1.912**	**0.028**	**0.413**	**1.495**	**0.067**	0.132	0.686	0.246
LUBELSKIE	0.199	0.516	0.303	0.097	0.320	0.375	-0.013	0.103	0.459	-0.037	0.058	0.477	0.406	1.143	0.127
LUBUSKIE	0.264	0.898	0.185	0.265	0.906	0.182	**1.018**	**2.944**	**0.002**	-0.465	-1.082	0.860	**0.697**	**2.398**	**0.008**
MALOPOLSKIE	**0.642**	**1.654**	**0.049**	0.264	0.777	0.219	**1.560**	**3.784**	**0.000**	0.018	0.197	0.422	0.055	0.341	0.367
MAZOWIECKIE	**0.440**	**1.582**	**0.057**	**0.618**	**2.143**	**0.016**	0.073	0.436	0.331	0.252	0.994	0.160	-0.017	0.172	0.432
OPOLSKIE	0.361	0.831	0.203	0.201	0.523	0.300	0.289	0.689	0.245	-0.647	-1.125	0.870	0.373	1.064	0.144
PODKARPACKIE	-0.141	-0.174	0.569	-0.330	-0.619	0.732	**0.880**	**2.203**	**0.014**	**0.648**	**1.664**	**0.048**	**0.400**	**1.306**	**0.096**
PODLASKIE	0.365	0.838	0.201	0.302	0.722	0.235	0.306	0.721	0.235	-0.001	0.126	0.450	0.146	0.514	0.304
POMORSKIE	0.105	0.609	0.271	0.238	1.082	0.140	**0.341**	**1.445**	**0.074**	-0.141	-0.263	0.604	0.039	0.393	0.347
SLASKIE	0.177	0.569	0.285	-0.138	-0.167	0.566	-0.250	-0.425	0.665	0.127	0.451	0.326	0.298	1.023	0.153
SWIETOKRZYSKIE	**1.239**	**2.536**	**0.006**	**0.773**	**1.644**	**0.050**	**1.784**	**3.585**	**0.000**	0.323	0.754	0.225	0.038	0.254	0.400
WARMINSKO-MAZURSKIE	-0.113	-0.109	0.543	0.018	0.200	0.421	0.341	0.947	0.172	**1.034**	**2.560**	**0.005**	**1.280**	**3.771**	**0.000**
WIELKOPOLSKIE	**0.594**	**1.283**	**0.100**	0.609	**1.322**	**0.093**	0.242	0.598	0.275	0.352	0.810	0.209	-0.467	-0.968	0.833
ZACHODNIOPOMORSKIE	0.167	0.453	0.325	0.101	0.328	0.371	0.577	1.247	0.106	-0.063	0.007	0.497	0.016	0.200	0.421

Remark: Bold denotes significant values of statistic I_l ($\beta = 0.1$, $1\text{-}\beta = 0.9$).

Source: author’s own.

The most important factors determining the life expectancy and the mortality pattern in the voivodeships of Poland seem to be the ones connected with the agricultural character of the voivodeship (Z_{10}, V_1) – less industry (Z_{11}) and its consequences (population density, share of urban population) respond to the more favourable values of the mortality characteristics of the voivodeships; on the contrary the factors contributing to the affluence of a voivodeship are positively correlated with some of the mortality rates and negatively with the life expectancy for men. The life expectancy for women and some kinds of their mortality rates are sensitive to pollution (V_4) – the more polluted the voivodeship the less favourable the women mortality characteristics. Some of the both genders mortality rates are positively correlated with the unemployment rate, which seems to be an important factor determining the mortality pattern, and the average wages and salaries are, at least in the profile of the voivodeships, not important. The above remarks can be indicators for the directions of the more detailed analysis of the mortality patterns in Poland and the factors influencing them.

Table 5. Correlation between coefficients Y_{0m}-Y_{4m}, Y_{0k}-Y_{4k}, synthetic variables V_1-V_4 and diagnostic variables Z_{10} –Z_{13}

VARIABLES	COEFFICIENTS OF CORRELATION r(Y,V), r(Y,Z):	
MEN	POSITIVE VALUES	NEGATIVE VALUES
Y_{0m}	Z_{10}(0.57)	V_3(-0.4)
Y_{1m}	V_3(0.41)	Z_{10}(-0.48)
Y_{2m}	V_1(0.32). Z_{13}(0.31). V_3(0.41)	Z_{10}(-0.38)
Y_{3m}	Z_{13}(0.33)	-
Y_{4m}	V_3(0.59). Z_{13}(0.54)	-
WOMEN	POSITIVE VALUES	NEGATIVE VALUES
Y_{0k}	Z_{10}(0.40)	V_1(-0.51). V_4(-0.42)
Y_{1k}	V_1(0.44). V_4(0.41)	Z_{10}(-0.39)
Y_{2k}	V_1(0.48). Z_{11}(0.44)	Z_{10}(-0.35)
Y_{3k}	-	-
Y_{4k}	V_3(0.61). Z_{13}(0.56)	-

Remark: For 16 observations coefficient of correlation is significant (according to *t*-Student's test) beginning from the value r^* : □ = 0.05, r^* = 0.50; □ = 0.1, r^* = 0.42; □ = 0.2, r^* = 0.33.

Source: author's own.

4. Conclusions

1. The selected causes of death as well as the life expectancy in most cases show the positive spatial autocorrelation, which may indicate that there are some common determinants of their level in bigger parts of the country.
2. In many cases the spatial autocorrelation pattern is different for men and women, which means that there are some special determinants of the level of the discussed coefficients, apart from coming out from the geographical position of the items.
3. In the author's opinion special attention should be paid to the outliers (identified with use of Moran's scatter plots as well as with Moran's local statistic *Ii*), because the differences (e.g. socio-economic) among the outliers and their surroundings may explain their causes. The occurrence of the outliers may also strongly influence the spatial correlation pattern of the considered variable.
4. The results of the study could be applied in the process of formulating the social policy within the scope of the population's health.

References

Anselin L. (1995). *Local Indicators of Spatial Association-LISA*, Geographical Analysis, Vol. 27, No.2, pp.93-115

Anselin L. (2005). Exploring Spatial Data with GeoDaTM: A Workbook. Spatial Analysis Laboratory, https://geodacenter.asu.edu/system/files/geodaworkbook.pdf

Ekonometria przestrzenna (Spatial Econometrics). Suchecki B. (ed). (2010). Wydawnictwo C.H.Beck, Warszawa

Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna* (Spatial Econometrics and Statistics). CeDeWu.pl, Wydawnictwa Fachowe, Warszawa

Mielecka-Kubień Z., Dziembała M. (2010). *Przestrzenna autokorelacja wybranych przyczyn zgonów w woj. śląskim w latach 2004-2006* (Spatial Autocorrelation of Selected Causes of Death in Voivodship Slaskie in the years 2004-2006). Śląski Przegląd Statystyczny, 8(14), pp. 55-80

Streszczenie

BADANIE DŁUGOŚCI ŻYCIA ORAZ WYBRANYCH PRZYCZYN ZGONÓW W POLSCE Z ZASTOSOWANIEM METOD STATYSTYKI PRZESTRZENNEJ

Celem prezentowanego badania było testowanie hipotezy o braku autokorelacji przestrzennej w odniesieniu do przeciętnego dalszego trwania życia oraz standaryzowanych ze względu na wiek współczynników zgonów dla wybranych przyczyn zgonów w Polsce według płci w 2010 r. Przypuszczano, że w wyżej wymienionych przypadkach występuje dodatnia autokorelacja przestrzenna w populacjach mężczyzn i kobiet oraz, że przestrzenne zróżnicowanie umieralności zależy od poziomu życia danej populacji i stopnia industrializacji regionu i jej konsekwencji. Okazało się, że większość z rozważanych współczynników wykazuje dodatnią autokorelację przestrzenną; zaobserwowano też różnice między populacjami mężczyzn i kobiet. Rolniczy charakter województwa wykazuje pozytywne oddziaływanie na przeciętne dalsze trwanie życia i wartości niektórych współczynników zgonów dla obu płci.

10.2478/v10103-012-0034-6

ANNA OJRZYŃSKA[*], SEBASTIAN TWARÓG[**]

Dynamics of Change in Spatial Dependencies in Blood Donation System in Poland

"Blood is to health care as oil is to transportation"
Arthur Caplan

Abstract

Blood donation allows to obtain blood and its components from healthy people in a bid to help treatment of anonymous individuals, relying on timely and sufficient supplies of matching blood. Being a social initiative, it depends on multiple factors. Those factors are possible to be shaped and are subject to research. This paper aims to present the dynamics of change in spatial dependencies determining development of blood donation in Poland from 2005 to 2010. Spatial analysis of data enables identification of similarities and differences between voivodeships in a given period. Testing of hypothesis concerning spatial autocorrelation was carried out using tools of spatial statistics. This paper's subject matter concentrates on pointing towards the direction and extent of changes illustrated with an example of analysis investigating the number of blood donations per hospital bed in wards with high

[*] University of Economics in Katowice

[**] University of Economics in Katowice

This content is the result of implementing research project N N115 411040, concerning *Logistics aspects of blood supply chain management in Poland*, funded by the National Science Centre.

demand on blood and its components. The number of blood donors per 1000 residents in 18 – 65 age was also analysed.

1. Introduction

Could oil be the most precious of all liquids? At the current time, when oil prices rocket this is a popular perception. In reality though, that idea is far from truth. We each carry in our bodies several litres of much more valuable liquid - blood. It remains a medicine which have not been synthesised thus far and blood demand increases 6-10% year-on-year. Keeping blood availability at levels necessary to efficiently carry out complex medical procedures is one of factors determining national health security and citizens' initiative is key here. Determinant of that security is a reliable blood donation system. The Polish blood donation system consists of two concurrently operating subsystems: civil and uniformed services (Szoltysek, Twarog 2009, p.14). The civil part, which is an area of interest for authors of this paper comprises 1) donors of blood and its components (residents of Poland, satisfying provisions stipulated by the Act[1] and regulation[2] concerning age: 18 - 65 years and health), 2) blood recipients, 3) hospitals managing allocation and administration of blood components, and 4) 21 Regional Blood Donation and Hemotherapy Centres (RBDHC) managed independently, responsible for collecting, processing, storing and relocating blood and its components.

This paper set out to **identify voivodeships with high degree of proximity which could be grouped into clusters. This could be the first stage in creating the Polish blood donation network.** The impulse for exploring spatial dependency dynamics of Polish blood donation between 2005 and 2010 were results of analysis concerning 2009 (Ojrzynska, Twaróg 2011, pp. 129 - 141). Authors of this paper would like to address scientific circles and practitioners - managers of blood donation and hemotherapy system in Poland (Ministry of Health, National Blood Centre, regional blood donation and hemotherapy centres), and other parties with interest in health care (blood donation). This paper's outcomes can trigger and give grounds to modify the national blood donation and hemotherapy policy in Poland.

[1] Act from 22 August 1997 concerning public blood service.

[2] Regulation by the Minister of health from 18 August 2005 Journal of Laws, No. 79, item 691, 2005) concerning blood collection condition for candidates applying to become blood donors and blood donors, amended on 31 December 2009 (Journal of Law Journal of Laws No. 7 item 50).

2. Data and methodology

Data for this research came in form of information provided by the *Journal of Transfusion Medicine*, which concerns all Regional Blood Donation and Hemotherapy Centres (RBDHCs) operating in the Polish blood donation system between 2008 and 2010 as well as from analyses provided by RBDHC for 2005-2007 concerning the total number of donors and packed cells units (PC) - most commonly used in medicine blood component. Data on hospital bed numbers came from *Statistical Bulletin* issued by the Ministry of Health. All data is voivodeship specific.

In order to determine spatial dependency between voivodeships spatial analysis was employed. Key assumptions for autocorrelation investigation draw on presumption the intensity of phenomena subject to dependency within given location depends on intensity of that phenomenon in proximal locations. No spatial autocorrelation means the phenomenon is spatially variable. Parameters observed for given area do not depend on parameters observed in proximal locations and observed spatial pattern is equally as probable as any other pattern. Positively autocorrelated values cluster across a space and neighbouring areas are similar. Negative autocorrelation means proximal locations are dissimilar to a greater extent than would be have been dictated by random distribution (Kopczewska 2007, p. 14). Fundamental to all spatial analyses is determining structure of proximity through spatial weights. Spatial weights could be determined using the criterion of either tangency or distance. In this paper it was assumed that voivodeships sharing a common border are correlated. Hence a binary matrix is created, where voivodeships sharing a common border represent 1 and if there are no shared borders they are represented by 0. This matrix is then row-standardised in order to maintain weights comparable (Suchecki 2010, pp. 105-107).

Global and local spatial statistics are parameters of spatial autocorrelation. Global Moran's *I* is used to test for global spatial autocorrelation. It is given by (Kopczewska 2007, p.72):

$$I = \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_i \sum_j w_{ij}} \qquad (1)$$

where x_i is the number of observation within location *i*, x_j $\bar{x}$ is mean of all analysed locations, *n* is the number of locations, and w_{ij} j is the element of spatial weights matrix.

To analyse dimensions of significant clusters of similar values grouping around particular location used are measures of local autocorrelation, which are derived for each observation and they measure the relationship between investigated regions and their neighbours (Suchecki 2010, p. 123). **Moran's I statistics $\boldsymbol{I_i}$**, measures whether any specific region is surrounded by neighbouring regions with similar and dissimilar values of observed variable relative to random spatial distribution of those values. It is given by the formula (Kopczewska 2007, p. 90):

$$I_i = \frac{(x_i - \bar{x})\sum_{j=1}^{n} w_{ij}(x_j - \bar{x})}{\sum_{i=1}^{n}(x_j - \bar{x})^2} \tag{2}$$

3. Results

Autocorrelation relationships between voivodeships were identified in respect of the number of blood donors[3] and the number of packed cells units[4]. For significance level $\alpha = 0,05$ global statistics for number of blood donors presented in table 1, calculated for period 2005-2007 are statistically significant and are indicative of weak positive spatial autocorrelation. Between 2008-2010 that statistic is not statistically significant i.e. there no statistically significant tendency for voivodeships with similar number of blood donors to group into clusters.

Table 1 Moran's I statistics for number of blood donors

YEAR	MORAN'S I STATISTICS	P-VALUE
2005	**0.204**	**0.033**
2006	**0.171**	**0.050**
2007	**0.228**	**0.024**
2008	0.166	0.061
2009	0.042	0.224
2010	-0.030	0.389

Source: own research.

[3] The variable is: total number of blood donors / 1000 residents aged between 18-65.

[4] The variable is: number of PCUs / hospital beds at words with highest PCU demand (cardiologic, surgical, intensive care, nerosurgical, haematological, obstetric and gynaecological)

The global Moran's I statistics has been graphically represented by scatter plot. It enables to visualise local spatial relationships (clusters). The relationship between given region and its neighbours is determined by distribution of spatial areas across coordinate system across the OX axis where standard value of given variable is marked. Standard value of spatially lagged variable is marked across the OY axis. The plot is divided into quadrants relative to the origin (0, 0)

In order to identify spatial regimes, different voivodeships were plotted on the cartograms 1-12 according to quadrants of Moran's scatter plot. Figures 1-6 show distribution of voivodeships according to the number of blood donors.

I – **HH** – High surrounded by High
II – **HL** – High surrounded by Low
III – **LL** – Low surrounded by Low
IV – **LH** – Low surrounded by High

Figure. 1 Distribution of regions according to the quadrants of Moran's scatter plot for the number of blood donors in 2005

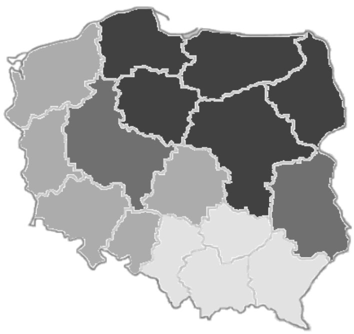


Figure. 2 Distribution of regions according to the quadrants of Moran's scatter plot for the number of blood donors in 2006

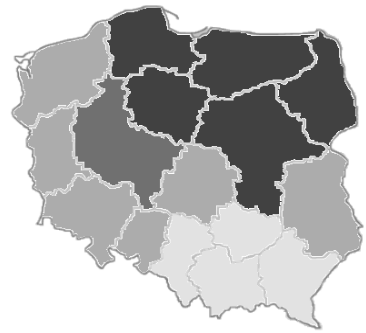


Figure 3. Distribution of regions according to the quadrants of Moran's scatter plot for the number of blood donors in 2007

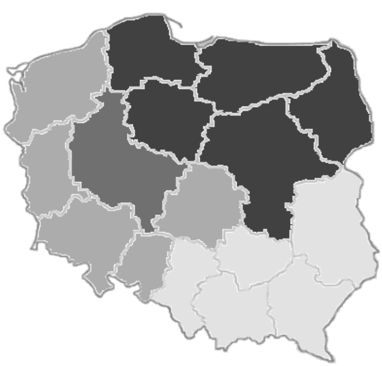


Figure 4. Distribution of regions according to the quadrants of Moran's scatter plot for the number of blood donors in 2008

Figure 5. Distribution of regions according to the quadrants of Moran's scatter plot for the number of blood donors in 2009

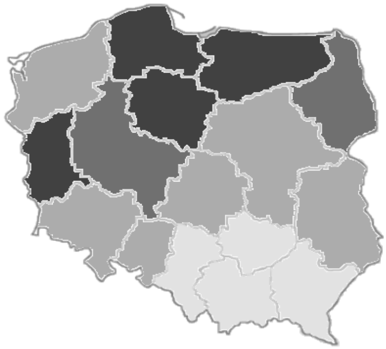


Figure 6. Distribution of regions according to the quadrants of Moran's scatter plot for the number of blood donors in 2010

Source: own development.

Voivodeships enjoying the highest number of blood donors grouped into high value clusters are marked by the darkest shade. Between 2005 and 2007 those were voivodeships located North and North-East of the country. Voivodeships with low values surrounded by similar regions are marked by the lightest shade. Between 2005 and 2007 those were voivodeships located South and South-East of Poland. Voivodeships located within the band spanning from North-West to South-West are also clusters with low number of blood donors, however, neighbouring with regions displaying high value of that variable. Figures 5 and 6 are a confirmation there is no spatial dependency between variables describing the number of blood donors between 2009 and 2010, because there are no explicit clusters. This means that the number of blood donors observed in that area is independent of similar numbers in proximal locations.

At the next stage the hypothesis there is spatial autocorrelation between the number of PCUs was verified. For significance level $\alpha = 0{,}05$ global statistics for that variable (table 2) calculated for the period between 2005 and 2008 are statistically insignificant i.e. values observed for given region are independent of similar values in neighbouring areas. Between 2009 and 2010, however, those statistics indicate there is a positive autocorrelation i.e. the proximal locations are similar.

Table 2. Moran's I statistics for number of packed cell units

YEAR	Moran's I statistics	P-VALUE
2005	0.169	0.057
2006	0.151	0.071
2007	0.114	0.112
2008	0.138	0.085
2009	**0.232**	**0.022**
2010	**0.221**	**0.027**

Source: own calculations.

Voivodeships enjoying the highest number of PCUs grouped into high value clusters are marked by the darkest shade.

I – **HH** – High surrounded by High
II – **HL** – High surrounded by Low
III – **LL** – Low surrounded by Low
IV – **LH** – Low surrounded by High

Figure 7. Distribution of regions according to the quadrants of Moran's scatter plot for the PCU in 2005

Figure 8. Distribution of regions according to the quadrants of Moran's scatter plot for the PCU in 2006

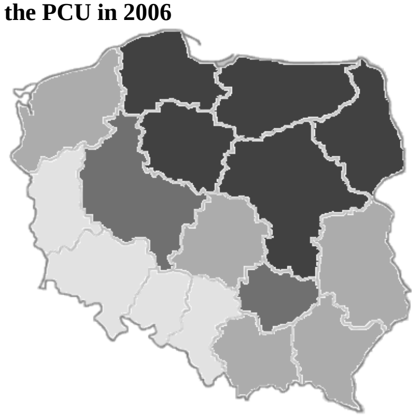


Figure 9. Distribution of regions according to the quadrants of Moran's scatter plot for the PCU in 2007

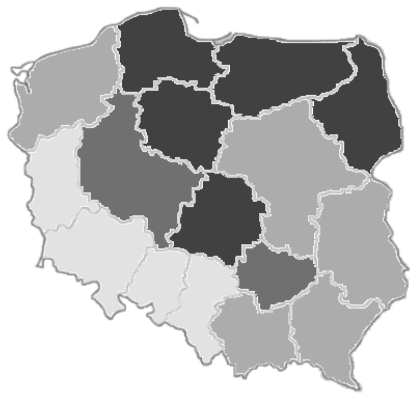


Figure 10. Distribution of regions according to the quadrants of Moran's scatter plot for the PCU in 2008

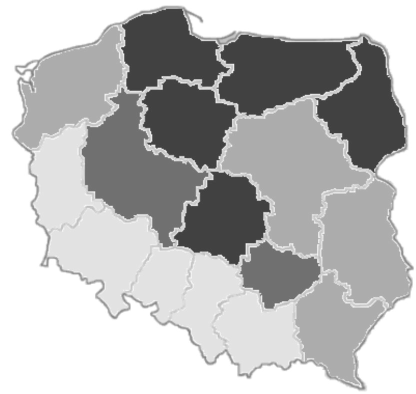


Figure 11. Distribution of regions according to the quadrants of Moran's scatter plot for the PCU in 2009

Figure 12. Distribution of regions according to the quadrants of Moran's scatter plot for the PCU in 2010

Source: own development.

Between 2009 and 2010 those were voivodeships located in Central part of the country. Voivodeships with low values surrounded by similar regions are marked by the lightest shade. Between 2009 and 2010 those were voivodeships located anywhere between West and South-East of Poland. An exception is the Silesian voivodeship which in 2010 had a high number of packed cells units.

Spatial data analysis requires not only investigating global autocorrelation, but also to analyse local spatial autocorrelation which gives better insight into how investigated variable is spatially distributed across given area and it also enables to identify unusual observations. Tables 3 and 4 show analytical local statistics -empirical significance levels are given in brackets - for the two variables analysed.

Table 3. Moran's local statistics for the number of blood donors

VOIVODESHIP	MORAN'S LOCAL STATISTICS I_i					
	2005	**2006**	**2007**	**2008**	**2009**	**2010**
Wielkopolskie	-0.289 **(0.786)**	-0.221 **(0.709)**	-0.153 **(0.621)**	-0.037 **(0.458)**	-0.096 **(0.542)**	-0.020 **(0.433)**
Kujawsko-Pomorskie	0.696 **(0.018)**	0.415 **(0.090)**	0.634 **(0.027)**	0.389 **(0.107)**	0.388 **(0.102)**	0.151 **(0.262)**
Małopolskie	0.586 **(0.098)**	0.436 **(0.157)**	0.581 **(0.102)**	0.538 **(0.119)**	0.409 **(0.168)**	0.229 **(0.262)**
Łódzkie	-0.039 **(0.465)**	-0.012 **(0.432)**	-0.007 **(0.426)**	0.009 **(0.407)**	-0.014 **(0.433)**	-0.006 **(0.421)**
Dolnośląskie	-0.077 **(0.508)**	-0.006 **(0.452)**	-0.061 **(0.496)**	-0.266 **(0.651)**	-0.188 **(0.597)**	-0.060 **(0.494)**
Lubelskie	-0.025 **(0.461)**	-0.020 **(0.456)**	0.013 **(0.425)**	0.074 **(0.371)**	-0.006 **(0.441)**	-0.058 **(0.491)**
Lubuskie	-0.031 **(0.471)**	-0.075 **(0.507)**	-0.009 **(0.455)**	0.088 **(0.381)**	0.012 **(0.437)**	-0.033 **(0.471)**
Mazowieckie	0.154 **(0.244)**	0.038 **(0.371)**	0.009 **(0.406)**	-0.044 **(0.472)**	-0.099 **(0.541)**	-0.083 **(0.521)**
Opolskie	-0.394 **(0.781)**	-0.112 **(0.543)**	-0.088 **(0.520)**	-0.101 **(0.532)**	-0.155 **(0.584)**	-0.105 **(0.539)**
Podlaskie	0.398 **(0.178)**	0.121 **(0.354)**	0.289 **(0.242)**	0.010 **(0.440)**	-0.817 **(0.935)**	-1.461 **(0.999)**
Pomorskie	0.954 **(0.008)**	0.610 **(0.052)**	0.672 **(0.041)**	0.502 **(0.091)**	0.410 **(0.125)**	0.236 **(0.221)**
Śląskie	0.220 **(0.248)**	0.267 **(0.212)**	0.351 **(0.163)**	0.343 **(0.169)**	0.121 **(0.325)**	-0.003 **(0.436)**
Podkarpackie	0.662 **(0.074)**	0.811 **(0.039)**	0.873 **(0.032)**	1.046 **(0.015)**	0.940 **(0.021)**	0.984 **(0.012)**
Świętokrzyskie	0.198 **(0.202)**	0.299 **(0.123)**	0.343 **(0.099)**	0.549 **(0.027)**	0.437 **(0.054)**	0.384 **(0.070)**
Warmińsko-Mazurskie	0.574 **(0.064)**	0.449 **(0.108)**	0.708 **(0.034)**	0.534 **(0.080)**	0.092 **(0.351)**	-0.328 **(0.747)**
Zachodniopomorskie	-0.323 **(0.695)**	-0.269 **(0.658)**	-0.503 **(0.804)**	-0.986 **(0.964)**	-0.765 **(0.921)**	-0.302 **(0.693)**

Source: own calculations.

Between 2006 and 2010 the podkarpackie voivodeship was the most correlated with its neighbours in terms of the number of blood donor, indicative of which were statistically significant positive Moran's local I_i statistics. In 2005 and 2007, on the other hand, the number of blood donors in pomorskie and kujawsko-pomorskie voivodeships was correlated with values observed in proximal locations. Unusual areas though, where those statistics were negative and statistically significant are voivodeships zachodniopomorskie (in 2008) and podlaskie (in 2010).

Table 4. Moran's local statistics for the number of PCUs

VOIVODESHIP	**MORAN'S LOCAL STATISTICS I_i**					
	2005	**2006**	**2007**	**2008**	**2009**	**2010**
Wielkopolskie	-0.024 **(0.440)**	-0.108 **(0.558)**	-0.041 **(0.464)**	-0.089 **(0.531)**	0.003 **(0.403)**	0.004 **(0.401)**
Kujawsko-Pomorskie	0.459 **(0.077)**	0.459 **(0.077)**	0.497 **(0.062)**	0.574 **(0.042)**	0.739 **(0.014)**	0.504 **(0.059)**
Małopolskie	-0.053 **(0.490)**	-0.103 **(0.5280)**	-0.101 **(0.527)**	0.044 **(0.416)**	0.218 **(0.291)**	0.368 **(0.197)**
Łódzkie	-0.095 **(0.536)**	-0.086 **(0.524)**	0.034 **(0.376)**	0.127 **(0.274)**	0.186 **(0.216)**	0.080 **(0.323)**
Dolnośląskie	0.897 **(0.031)**	0.842 **(0.039)**	0.783 **(0.049)**	0.704 **(0.069)**	0.781 **(0.051)**	0.798 **(0.045)**
Lubelskie	-0.166 **(0.592)**	-0.116 **(0.546)**	-0.047 **(0.482)**	-0.043 **(0.479)**	-0.056 **(0.490)**	0.012 **(0.427)**
Lubuskie	0.329 **(0.221)**	0.330 **(0.221)**	0.375 **(0.194)**	0.269 **(0.259)**	0.377 **(0.196)**	0.238 **(0.275)**
Mazowieckie	0.249 **(0.162)**	0.174 **(0.227)**	-0.113 **(0.558)**	-0.114 **(0.559)**	0.179 **(0.222)**	0.221 **(0.183)**
Opolskie	0.514 **(0.088)**	0.467 **(0.107)**	0.328 **(0.178)**	0.096 **(0.353)**	0.013 **(0.427)**	-0.193 **(0.617)**
Podlaskie	0.168 **(0.324)**	0.197 **(0.305)**	0.059 **(0.403)**	0.081 **(0.388)**	0.126 **(0.355)**	0.252 **(0.267)**
Pomorskie	0.819 **(0.020)**	0.814 **(0.020)**	0.785 **(0.023)**	0.934 **(0.010)**	0.834 **(0.018)**	0.718 **(0.033)**
Śląskie	0.042 **(0.400)**	0.057 **(0.387)**	0.017 **(0.423)**	0.007 **(0.433)**	0.026 **(0.415)**	-0.088 **(0.520)**
Podkarpackie	-0.201 **(0.603)**	-0.230 **(0.624)**	-0.297 **(0.673)**	-0.189 **(0.593)**	0.224 **(0.287)**	0.380 **(0.191)**
Świętokrzyskie	-0.809 **(0.990)**	-0.775 **(0.986)**	-0.820 **(0.991)**	-0.659 **(0.967)**	-0.332 **(0.796)**	-0.065 **(0.498)**
Warmińsko-Mazurskie	0.504 **(0.092)**	0.544 **(0.078)**	0.481 **(0.100)**	0.552 **(0.076)**	0.482 **(0.101)**	0.286 **(0.204)**
Zachodniopomorskie	0.069 **(0.396)**	-0.043 **(0.482)**	-0.111 **(0.535)**	-0.094 **(0.521)**	-0.086 **(0.515)**	0.026 **(0.428)**

Source: own computations.

In the investigated period, apart from 2009 the dolnośląskie and pomorskie voivodeships were the most correlated with its neighbours in terms of the number of PCUs, indicative of which were statistically significant positive Moran's local I_i statistics. An unusual area though, where those statistics were negative and statistically significant was the małopolskie voivodeship (between 2005 and 2007).

4. Discussion

1. The spatial dependency between the number of blood donors over the investigated period fluctuated between moderate and none.
2. The spatial dependency between the number of PCUs over the investigated period fluctuated none and moderate.

Hence, lower number of blood donors between 2005 and 2010 accompanied by concurrent increase in the number of PCUs is indicative of higher number of multiple donors. That change might be the result of the following events:

- Implementation of Nationwide health initiative promoting honorary blood donation.
- Higher social sensitivity.
- Higher number of mobile blood donation points, mobile units.

5. Conclusions

There are voivodeships in Poland, where current situation is satisfactory as well as unsatisfactory, what from perspective of logistics management will enable building an efficient network stage-by-stage (aggregation of voivodeships sharing the same blood donation situation). Furthermore, the presumptions articulated in the discussion section of this paper open a new research field - **recognising importance of logistics and marketing in terms of motivating blood "suppliers"- donors of blood and its components.**

References

Biuletyn statystyczny Ministerstwa Zdrowia 2005, Centrum Systemów Informacyjnych Ochrony Zdrowia, Warszawa (www.csioz.gov.pl)

Biuletyn statystyczny Ministerstwa Zdrowia 2006, Centrum Systemów Informacyjnych Ochrony Zdrowia, Warszawa (www.csioz.gov.pl)

Biuletyn statystyczny Ministerstwa Zdrowia 2007, Centrum Systemów Informacyjnych Ochrony Zdrowia, Warszawa (www.csioz.gov.pl)

Biuletyn statystyczny Ministerstwa Zdrowia 2008, Centrum Systemów Informacyjnych Ochrony Zdrowia, Warszawa (www.csioz.gov.pl)

Biuletyn statystyczny Ministerstwa Zdrowia 2009, Centrum Systemów Informacyjnych Ochrony Zdrowia, Warszawa (www.csioz.gov.pl)

Biuletyn statystyczny Ministerstwa Zdrowia 2010, Centrum Systemów Informacyjnych Ochrony Zdrowia, Warszawa (www.csioz.gov.pl)

Kopczewska K. (2007), *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa

Ojrzyńska A., Twaróg S. (2011) *Badanie autokorelacji przestrzennej krwiodawstwa w Polsce*, (in): J. Suchecka (ed.) *Ekonometria przestrzenna i regionalne analizy ekonomiczne*, Acta Universitatis Lodziensis, Folia Oeconomia, no. 253, Łódź

Rosiek A. i in. (2009), *Działalność jednostek organizacyjnych służby krwi w Polsce w 2008 roku*, „Journal of Transfusion Medicine”, tom 2, no. 4

Rosiek A. i in. (2010), *Działalność jednostek organizacyjnych służby krwi w Polsce w 2009 roku*, „Journal of Transfusion Medicine”, tom 3, no. 4

Rosiek A. i in. (2011), *Działalność jednostek organizacyjnych służby krwi w Polsce w 2010 roku*, „Journal of Transfusion Medicine”, tom 4, no. 4

Suchecki B. (2010), *Ekonometria przestrzenna*, Wydawnictwo C.H. Beck, Warszawa

Szołtysek J., Twaróg S. (2009), Gospodarowanie zasobami krwi jako nowy obszar stosowania logistyki, „Gospodarka Materiałowa i Logistyka”, no. 7

Streszczenie

DYNAMIKA ZMIAN ZALEŻNOŚCI PRZESTRZENNEJ SYSTEMU CYWILNEGO KRWIODAWSTWA W POLSCE

Krwiodawstwo jest sposobem pozyskiwania krwi i jej składników od osób zdrowych na rzecz anonimowych osób, których leczenie jest uwarunkowane podaniem właściwej krwi w odpowiednim czasie oraz ilości. Jako akcja społeczna uwarunkowana jest od wielu czynników, będących przedmiotem kształtowania oraz badania.

Celem niniejszego opracowania jest przedstawienie dynamiki zmian zależności przestrzennej w zakresie poziomu rozwoju krwiodawstwa w Polsce w latach 2005 – 2010. Analiza przestrzenna danych umożliwia określenie podobieństw i różnic między województwami w badanym okresie. Za pomocą narzędzi statystyki przestrzennej została zweryfikowana hipoteza o występowaniu autokorelacji przestrzennej. Przedmiotem opracowania jest wskazanie kierunku oraz zakresu zmian na przykładzie analizy zróżnicowania liczby donacji krwi przypadającej na łóżko szpitalne w oddziałach o wysokim zapotrzebowaniu na krew i jej składniki oraz liczby dawców przypadających na 1000 mieszkańców w wieku 18 –65.

10.2478/v10103-012-0035-5

WŁODZIMIERZ OKRASA*

Spatially Integrated Social Research and Official Statistics: Methodological Remarks and Empirical Results on Local Development[1]

Abstract

*This paper aims to elucidate some aspects of confluence of the developments in modern methods of spatial analysis and in the public statistics' data generating processes, along with empirical illustration of these interconnections from the development policy evaluation standpoint. Especially, how the growing availability of data at the finest level of territorial division (such as commune/*gmina*-level data within the Local Data Bank system) may improve decisional processes at the regional and local level, while paying special attention to allocation of resources assuming geographic system of targeting public support or intervention. Three interrelated questions that are empirically treated in the context of both* β-convergence *and* σ-convergence *issues at the local level involve (i) measurement of the level of local (under)development /local deprivation; (ii) assessment of how responsive and equitable are distributive policies in the domain of area-addressed public resources; and (iii) evaluation of their impacts in terms of convergence and social cohesion, which are among chief objectives of development policy (especially in the 'new' EU member states).*

* Ph.D, Professor at the University of Cardinal Stefan Wyszynski in Warsaw, Central Statistical Office of Poland

[1] Paper presented at the International Conference of Spatial Econometrics and Regional Economic Analysis; Łódź, 4-5 June, 2012.

Acknowledgements: The author would like to thank Prof. Janusz Witkowski for his comments on an earlier version of the paper. Grzegorz Gudaszewski and Beata Witek provided helpful assistance with calculations and editorial work, respectively.

1. Introduction

Space plays growingly important role in several social science disciplines along with methodological developments in statistics, geography, economics, and urban planning and community studies, to mention the few leading in this respect. Consequently, it became a natural candidate for providing a platform to integrate cross-disciplinary research efforts in social and behavioral sciences in terms of data, methods and problems (Goodchild et al. 2000, pp. 139-149). There are, however, significant differences between the types of spatial analysis employed in particular domains that need to be acknowledged at the outset when interdisciplinary approach is being adopted, especially from a socioeconomic perspective, as in this paper.

One may distinguish between more and less advanced disciplines in terms of the advancement in some or all of the overlapping areas of spatial analysis (SA) encompassing (i) data-driven emphasis of SA: *spatial statistics* (map-related, geometrical presentation of information, points, lines, patterns, and map testing patterns), *ESDA*/Exploratory spatial data analysis (pre-modelling exploration of geo-referenced data); (ii) method-driven emphasis of SA: *spatial econometrics* (tool of *regional science;* spatial *autocorrelation* and *regression* – GWR/geographically weighted regression), *geostatistics* (physical phenomena in spatial data environment – variograms, kriging); and (iii) problem-driven emphasis of SA: *interdisciplinary approach* (cross-disciplinary interaction and *spatially integrated social science)* – see Fischer and Getis (2010, pp. 1-24).

Despite the fact that all social phenomena take place in space, as Georg Simmel noted more than a century ago (e.g. Urry 2004, p. 5), such categories as 'space' and 'place' have, for a long time, been taken for granted, as their obvious attributes. For example, in such schools of thinking as *ecological analysis* in sociology, that is seen by others (mainly, economists) as impressionistic rather than systematic due to not being based on a regular type of geo-referenced data[2]. Things have changed over the past two decades. The question "how do places come to be the way they are?" (Gieryn 2000, p. 463) absorbs sociologists analyzing spatial effect of such phenomenon as inequality, power, politics, interaction, community, social movements, poverty, deviance, crime, life course, science, identity, memory and history.

---

[2] According to G. Simmel, the difference between an 'empty space' and something meaningful is due to five qualities which form uniqueness of a space: (i) the ways in which a space may be divided into pieces and activities spatially 'framed', (ii) the degree to which social interactions may be localized in space; (iii) the degree of proximity/distance, especially in the city; (iv) the role of the sense of sight; (v) and the possibility of changing location, and the consequences of the arrival of the 'stranger' (cf. Urry 2004, p. 5).

For economists, the regional and local growth and development are of main objects of interest, with special attention being paid – particularly within so-called ‘new economic geography’ – to endogenous factors, including such as *technological progress* and *knowledge creation* processes. And focusing on *innovation, learning process* and *scale economies* which generate increasing returns in factor productivity (offsetting the effect of diminishing marginal productivity assumed in the neoclassical conceptualization of the production function) - e.g., Capello (2009, pp. 33-43).

As regards statistics, specialization is progressing from methods of detecting autocorrelation, autoregression and spatial patterns or clusters offered within general spatial statistics and geostatistics, to methods of exploring complex forms of spatial dependence, heterogeneity and ‘causation’ in models of growth and dynamics employed in spatial econometrics, ecological or epidemiological studies (Fisher and Getis 2010), to hierarchical models for spatio-temporal data (Cressie and Wikle 2011). Public statistics both responds to increasing demand for geo-referenced data while employing new techniques of generating data for spatial analysis and benefits from the advancement in overall spatial methodology. In Poland, the most significant out of all recently organized activities in this matter was employment of GIS techniques (*digital mapping)* to Census 2011 alongside a new fieldwork technology (GPS-oriented *handhelds*) for obtaining spatial information on households and population. Among the products, in addition to maps of about 35 thousand of statistical subregions and about 200 thousand of enumeration districts it was planned to create vectors of geographic information (borders) for territorial units of all levels in digital version, in the shape format (Dygaszewicz 2007, p. 189-196); however, some of these products are still being under preparation.

So far, as no cartographic automation system has been established yet – at least for socio-economic data at the level of the lowest territorial units, i.e., *gmina* (in terms of the EUROSTAT’s NUTS5) – one of the way to perform spatial analysis is to use substitute for geo-referenced data and to explore the Local Data Bank’s files (formerly Bank of Regional Data) that is maintained by the CSO (GUS). Given the impressive advancement that has recently been made in broadly-conceived *spatial analysis oriented research*, the utilization of refined methods and constantly richer geo-coded data on a larger scale may require a coordination of a dual-track – academic and institutional (meaning greater involvement of official statistics) – in analysis of real problems.

General thesis of this paper, emphasizing reciprocal influence of the developments both in methods of spatial analysis and in the data generation processes in public statistics remains valid, but as a postulate only, without further elaboration here. The chief intention is to call attention to opportunity that is being brought about by the emerging *spatially integrated social science* framework and to enhance a bigger involvement of the public statistics in supporting the development of such an interdisciplinary approach.

The paper is structured along the sequence of three interconnected questions which design its contents. The next (second) section is devoted to the issue of measurement of local (under)development. In third section an evaluation exercise is performed for checking the effectiveness (in addition to equity) of the policies distributing public resources to territorial units at the smallest level, *gmina.* Finally, in fourth section, the issue of spatial disparities will be analyzed in the context of *σ-convergence* to check whether and how the inequalities had eventually been reduced over time (between the years 2008 and 2010) due to the employed policy of public resource allocation.

2. Community deprivation as an argument for local development – research problem operationalization and the measurement issues.

Since the choice of a measure of development (underdevelopment) as well as of observational unit in spatial analysis determine to a large extent its results, the typically employed solution is to use highly disaggregated data, both in the context of regional growth (see Abreu et al. 2005, pp.34-35) and of spatial inequality (Rey 2004, pp. 192-193). Therefore, all variables analyzed in this study are observed at the finest level, i.e., *gmina*.

The basic unit of analysis, *gmina,* is treated here as an operationalization of ‘local community’ (*commune*) being conceptualized holistically. That is, as a kind of a social space that – in addition to standard definitions as a unit of development (cf. Bhattacharyya 2004, pp. 5-12, Cnaan et al. 2008), or as a sociological category (for a survey of such meanings see Vaisey, 2007, pp. 851-60) – is characterized in terms of four types of community-constituting factors or dimensions, embracing: (i) integrated local capital; (ii) community wellbeing; (iii) local deprivation or risk factors; and (iv) community-building activities, agents and programs (see Okrasa 2011, pp. 268-272).

Research ***operationalization*** of the problem involves some assumptions and needs to formulate explicitly the main questions and hypotheses. The chief assumption can be stated briefly: *locality matters.* Along with conviction that *gmina* is the appropriate unit of analysis in studying local development. Especially if it is the end-user of programs or intervention despite the relevant sources might be distributed at other levels (as it is the case in EU-sponsored programs). It is complemented by both policy and methodological observations. The first underlies the usefulness of spatial analysis for policy design and evaluation of programs that involve distribution of public resources at the local community level. Second makes it clear that in order to assess policies employing geographic targeting mechanism of resource allocation suitable indicators to capture the complexity of a unit's wellbeing are indispensible.

The respective ***questions*** refer to:

i. What is the level of (under)development of *gmina*, taking into account most of the important aspects of local population's wellbeing of? This would require employment of a *multidimensional index of local deprivation.*

ii. How *responsive* is the area-targeting distributive policy to the 'needs for development', taking into account the *equity* issue? In particular, to what extent has the policy about distribution of public resources (represented here by 'total subsidies' to *gmina)* followed the principle of *horizontal equity*, according to which a greater amount of subsidies ought to go to poorer units.

iii. How *effective* is public allocation policy in terms of reduction of the local deprivation over time, specifically between 2008 and 2010?

iv. At which level of groupings of *gminas* – *powiat* or *voivodship* – is the process of changes toward a greater homogeneity more visible and advanced?

Hypothetically, we may expect that public *intervention* policy follows the principle of both *horizontal* and *vertical equity*, contributing to social cohesion and territorial convergence, while reducing disparities and demonstrating overall *efficiency* through providing relatively more for underprivileged localities. However, such an expectation is not obvious in the context of Central and Eastern European (CEE) countries – this point is discussed in the next section.

The ***measurement of local deprivation*** using Local Data Bank for the years 2008 and 2010 involved process that started with selection of the domains along with the appropriate sets of variables describing *gmina* in the relevant respect. Altogether 11 dimensions have been defined, each containing several dozens of variables, making it necessary to employ factor analysis (PCA) in order to reduce them. Actually, confirmatory version of the analysis was conducted as only the first factor was extracted for constructing the scale of deprivation, and factors loading were used as weights in calculating the composite index; the scale was transformed to have origin at zero and maximum at one hundred. In consequence, the constructed Multidimensional Index of Local Deprivation (MILD) embraces the following domains of deprivation: *ecology – finance – economy – infrastructure – municipal utilities – culture – housing – social welfare – labour market – education – health.* The reliability of the composite measure (MILD) was sufficiently high (alpha-Cronbacha above 0,8).

The index of ***local deprivation*** was used first to assess the ***responsiveness of policy*** about public resource distribution. How close (distant) is the actual distribution of total subsidies accrued to *gminas* to the one expected on the ground of the MILD (that supposes to reflect 'demand for development')? The latter distribution was simulated according to the proportionality formula (see Okrasa et al., 2006, p. 1058):

$$pS_i = \frac{P_i \cdot MILD_{i.}}{\sum_i P_i \cdot MILD_i} \cdot \sum_i S_i \qquad (1)$$

where i – subscript for territorial units, i= 1, 2, 3 . . . n (for *gminas* in Mazowieckie *voivodship* n = 314; in country, n = 2478); pS_i – subsidies accrued to i-th *gmina* according to the proportionality index; $MILD_i$ **-** Multidimensional Index of Local Deprivation used as index of proportionality; P_i - local population (number of *gmina's* residents); S_i - actual amount of subsidies accrued to i-th unit/*gmina*. The spatial distribution of the simulated and actual values of subsidies per person are visualized on figures 1a and 1b at the level of *gmina* in Mazowieckie *voivodship*; for the whole country the same types of distributions are presented for the level of *powiat* on figures 2a and 2b.

Figure 1a. **Allocation of subsidies to *gminas* proportionally to MILD, Mazowieckie voi. (2008)**

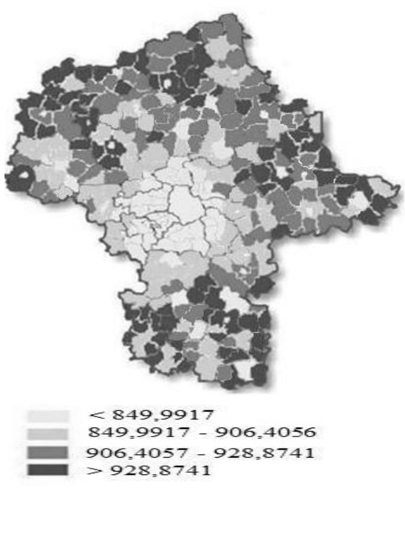
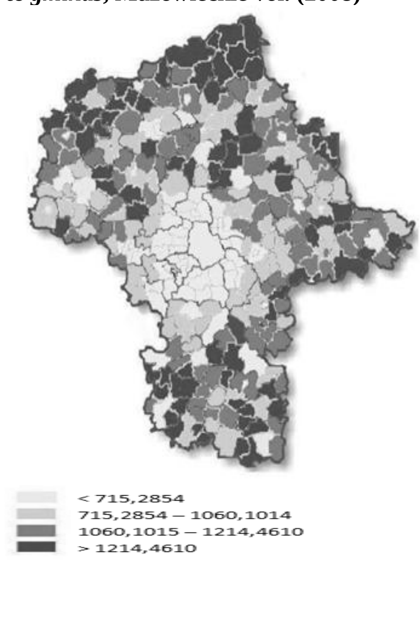


Figure 1b. **Actual distribution of subsidies to *gminas*, Mazowieckie voi. (2008)**



Source: author's calculations.

The evident similarity of actual distribution of subsidies to the ones expected under the local deprivation index (MILD) – except for north-east subregion – accords with the assumptions of convergence and social cohesion, as well as the equity principle. With clearly recognizable metropolitan and centre-periphery effects that exists in practically all of the eleven dimensions, analogous comparisons between actual and expected distribution can be made for each of them.

The country-wide distribution of resources – subsidies to *gminas* grouped in *powiats* – expected under the local deprivation-based criterion of allocation, despite showing some spatial resemblance to the actual distribution (in 2008 year), differ from it in several areas significantly. The most under-invested from this point of view are some eastern *voivodships* – especially Lubelskie voi., and central south (Świętokrzyskie voi.); while Małopolskie voi. seems to be 'disproportionally' beneficial.

Figure 2a. Allocation of subsidies to *gminas* grouped in *powiats*, proportionally to MILD in country (2008)

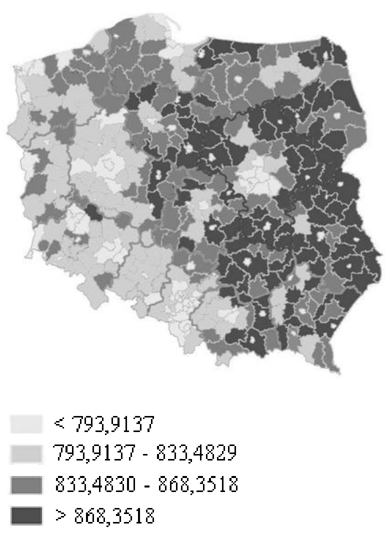
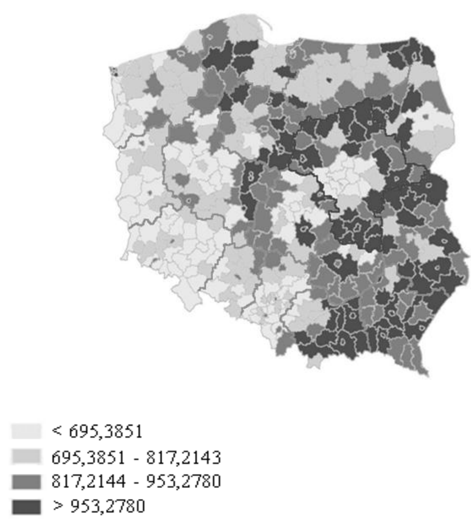


Figure 2b. Actual allocation subsidies to *gminas*, grouped in *powiats*, in country (2008)



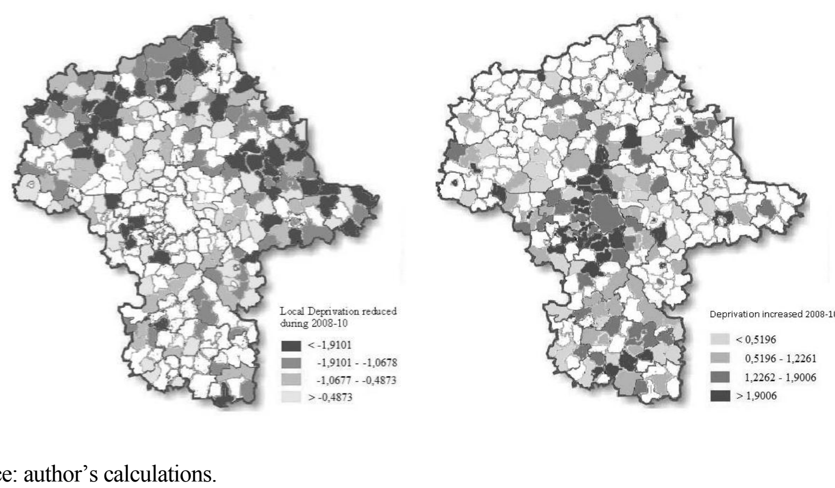
Source: author's calculations.

Slightly less, but still noticeably smaller than expected are amounts allocated to *gminas* in western *voivodships*, with Dolnośląskie, Lubuskie and Opolskie voi. showing relatively biggest shortfalls compared to 'deserved' on the local deprivation basis. In addition to the above presented empirical results which, in general, conform to the equity concern in the context of convergence and social cohesion policies, further analysis are needed to address the issue of how effective they are in this respect. This leads to the necessity of involving dynamic aspect.

3. Local development – dependence in deprivation and efficiency of resource allocation

Two aspects of the question of how *effective* is public allocation policy in the local development context – descriptive and inferential – are discussed below using the same type of data as previously, but for the years 2008 and 2010, to make some comparisons in time. Descriptively, an efficient use of public resources ought to be reflected in greater reduction of local deprivation in more disadvantaged units/*gminas* (i.e., higher on the MILD scale). In particular, in such a case as illustrated by figure 1a for Mazowieckie voi. we might expect a kind of 'photo-negative' effect: the lighter were the localities on this map marking smaller deprivation at the previous moment of observation (year 2008), the heavier they ought to be marked (the bigger changes recorded) at the next moment (year 2010); and *vice versa*. The figure 3 presents results of the changes for one *voivodship* – on the left map reduction of local deprivation is shown (darker means bigger reduction), while the right map depicts increased deprivation.

Figure 3. Changes in local deprivation (MILD) at the level of *gmina* during 2008-2010, Mazowieckie voi



Source: author's calculations.

Indeed, *gminas* located at the central metropolitan area (including the capital city) show relatively biggest increase in local deprivation, while *gminas* located at the periphery show the relatively biggest reduction in it, with significant exceptions in the south part of the *voivodship* where several underprivileged communes suffer from persistent deficiencies. The term ‘relatively’ is critical here because the increase in deprivation observed in *gminas* like Warsaw and nearby occurs at much lower level than its reduction at the border-periphery units.

In order to shed light on the relations between the levels of local deprivation over time a simple regression was run for the MILD in 2010 on its values in 2008. For the whole country (2 478 *gminas*), the slope was 0,7968082 (sign. at $p < 0{,}0001$); and RsqAdj = 0,82; there is high dependence of the deprivation, supporting the above observation of its persistence at the absolute level (deprived areas remains deprived over time, at least at the short time perspective). On the other hand, in a given dimension of deprivation changes could occur in one direction (for instance, reducing the level of deprivation in labour market) but in the opposite direction in another (for instance, increasing in the domain of health); therefore, such an exercise was performed for all domains. The results are illustrated below, for Mazowieckie voi. (all models significant).

Table 1. Selected parameters of the linear regression of local deprivation in 2010 on local deprivation in 2008

Domain of deprivation	Rsq_{Adj}	Slope
MILD Total deprivation	***0,87***	***0,773***
Ecology	0,76	0,661
Economy	0,96	0,990
Infrastructure	0,08	0,315
Culture	0,33	0,478
Municipal utilities	0,99	1,002
Housing	079	0,667
Social welfare	0,19	0,324
Labour market	0,43	0,653
Education	0,89	0,939
Health	0,93	0,994

Source: author’s calculations.

The scope of auto-determination in the different domains of local deprivation over time – that is pretty high for the joint measure (MILD) – varies considerably, towards a bipolar pattern: from very low ‘dependence’ in the case of infrastructure, social welfare, labour market and culture to high for the rest, peaking for economy, municipal utilities and health. This underlies analytical importance of ‘sector’ (all parameters of the above model significant).

In the context of ***efficiency of allocation*** policy analyzed from spatially oriented evaluation perspective, focused on the convergence and social cohesion issues, both aspects of changes – growth (β-convergence) and disparities (*σ-convergence)* – are of objects of research interest. The former is addressed below, the latter in the next section.

In a study of social cohesion and convergence policies across EU member states – for instance, as reported in the World Bank-organized paper (Kochendorfer-Lucis and Pleskovic, 2009) – a serious difference was stressed between the 'old' and 'new' member states. While the former demonstrated the β-convergence – pattern of growth (defined as a negative relationship between initial income levels and subsequent growth rates), in the CEE countries quite opposite tendency was observed: a positive relationship between the initial level of development and its dynamics, with capital and metropolitan regions developing faster than border-peripheral regions (Gorzelak 2009, pp. 259-264).

As noted by Quah (1993 p. 5) the term *convergence* refers to different things, even in the context of comparative analysis of income growth (with "Barro regression" as its core). But leaving aside its possible interpretations, this notion is being evoked here to simply validate the commonly employed and usually empirically confirmed pattern of income dynamics, checking at the same time the above conclusion at the level of *gmina* (other calculations were made at the NUTS3 level). To this aim, a common OLS approach was employed with the local deprivation (MILD) measure used instead of income – with converted value of 'growth rate' (since reduction in deprivation may indicate development), as follows:

$$[(Y_{t1} - Y_{t0})/Y_{t1}] = \alpha + \beta Y_{t1} + \varepsilon; \qquad (2)$$

where Y_{t1} and Y_{t0} stand for values of MILD in 2008 and 2010, respectively.

The obtained results: *Beta* = 0,56 (t=33,85; p <0,001; Rsq_{Adj} = 0,32) – despite using somewhat different, indirect concept of growth (decrease of local deprivation) – look rather surprising. The positive value of *beta* coefficient confirms the above quoted findings, indicating two things. First, that discussed earlier distribution of subsidies, despite being made according to 'demand for development', and apparently to the convergence and social cohesion policy objectives, may in fact contribute further to gaps between advanced and disadvantaged areas, also at the local level. Second, slightly weakening this argument, one may claim that the overall impact of subsidies is too small as a factor of development, while others influence the level of deprivation more effectively – in the direction of divergence rather than convergence and cohesion. This observation yields the need to look at the second aspect of convergence/divergence in terms of local deprivation, i.e., trends in spatial disparities over the same period of time.

4. Spatial disparity in local deprivation – does 'between' or 'within' inequality prevail?

Spatial disparities in regional and local growth measures become an increasing object of interest along with shift from international income dynamics to intranational dynamics in recent literature (Rey 2004, p. 193). When local deprivation is used instead of income, the main question concerns the level of territorial units – *gmina* or *powiat* or *voivodship* – at which a tendency to greater homogeneity (heterogeneity) can be expected in terms of a measure of overall development, such as the multidimensional index of local deprivation (MILD).

Inequality of local deprivation is here measured by the Theil index (Theil 1967) which is commonly used to describe spatial disparities (cf. Rey 2004, p. 194), according to formula:

$$T = \sum_{i=1}^{n} s_i \log(ns_i) \tag{3}$$

$$s_i = y_i / \sum_i y_i \tag{4}$$

where: *n* is the number of *gminas*; y_i is local deprivation in *i-th gmina* and: (i = 1, … , n).

The first impression from the results calculated for inequalities of *gminas'* deprivation for the years 2008 (T_{2008} = 0,00137) and 2010 (T_{2010} = 0,00102), respectively, was their low level. Partly at least, it might be attributed to very large number of the observational units (2 478 *gminas*). But, more importantly, the clearly declining tendency – by one-fourth of overall inequality in local deprivation between 2008 and 2010 – raises the question about the level of grouping at which this reduction is primarily taking place. Specifically, are there *powiats* or *voivodships* of that level of aggregation at which the apparent processes of homogenization is becoming more visible? For this, spatial decomposition of the T index – into intra- and inter-regional components – was performed.

The choice of the T index for ***spatial inequality decomposition*** is due to its convenient characteristics – as a member of *generalized entropy class of* *inequality measures* – such as being additively decomposable (Shorrocks and Wan 2004; Rey 2004, pp. 194-198). And there is a possibility to employ it to determine the extent to which total inequality *T* can be attributed to each of two types of sources of differentiations, one operating inter- and other intra-regions: $T = T_B + T_W$, where T_B is inequality "between" and T_W is inequality "within" a given set of appropriately partitioned territorial units. In particular, such as *powiats* encompassing *gminas* as their own units. Accordingly, T can be decomposed, as follows:

$$T = \sum_{g=1}^{\omega} s_g \log(n / n_g s_g) + \sum_{g=1}^{\omega} s_g \sum_{i \in g} s_{i,g} \log(n_g s_{i,g}) \tag{5}$$

That is, all *n gminas* (2 478) are partitioned into ω mutually exclusive and exhaustive sets of *powiats* (379). The results of such decomposition are listed below, in Table 2.

Table 2. Decomposition of the Theil index of local *(gmina)* deprivation into inequality 'between' and 'within' components, for (A) subregions/*powiats* and (B) regions/*voivodships* [values are rescaled by 10^{-2}]

Years	A. Inequality of local (*gmina)* deprivation decomposed into between (T_B) and within (T_W) *powiats;* country			B. Inequality of local deprivation decomposed into between (T_B) and within (T_W) *voivodships;* country		
	Total	T_B	T_W	**Total**	T_B	T_W
2008	0,1371	0,0708	0,0663	0,1371	0,0229	0,1142
2010	0,1029	0,0497	0,0532	0,1029	0,0127	0,0902

Source: author's calculations.

Figures suggest that generally diminishing inequality of local deprivation over time (confirming overall tendency to convergence and social cohesion) shows some differences in its structure – i.e., in proportion between and within components – at *powiat* and *voivodship* levels, as well as in pattern of changes, respectively. However, since the T–measure reaches maximum at log*n*, the number of units of observations counts for the value of the components. In other words, results of inequality decomposition are highly sensitive to the scale of the observational unit and their partitioning (see Rey, 2004, p. 198; Abreu et al., 2005, p. 34). Therefore, the problem of homogeneity/heterogeneity – and especially of spatial convergences in terms of the local deprivation – ought to be interpreted cautiously. But due to the fact that employed here decomposition involves the same units of observations, *gmina*, in both types of partitioning, this bias seems to be significantly less harmful. Hypothetically, one should expect that between units component will prevail over within component at the lower level of aggregation (e.g. NUTS4, compared to NUTS2). Some simple indicators proposed in the literature – such as: $P = T_B / T_W$ and $R(B) = T_B/T$ (see Rey, 2004, p 197-198) – may facilitate qualitative assessment of the relative importance of between and within inequality; results are in table 3.

Table 3. Patterns of changes in local deprivation during 2008-2010 using ratio indicators

Years	**$P = T_B / T_W$**	**$R (B) = T_B/T$**
between and within ***powiats*** **,** country-wide		
2008	1,067	0,516
2010	0,934	0,483
between and within ***voivodships*** **,** country-wide		
2008	0,200	0,170
2010	0,140	0,120

Source: author's calculations.

In the light of these results, the above expectations are generally met: when county/*powiat* level aggregation is used, the inter-subregional inequality seems to be more important. The opposite can be said for higher level aggregation, *voivodship,* where within region differentiation plays more important role. Despite similarity of the patterns of changes at both levels, this suggests a tendency towards a less homogenous *voivodships* due to greater heterogeneity among *powiats*. To shed light on this question, the same type of analysis has been conducted for each *voivodship,* for the years 2008 and 2010 – see Figure 4a and b.

Figure 4. Decomposition of local deprivation inequality into between and within *powiats* components, by *voivodship*, for the years 2008 and 2010

Figure 4a - 2008

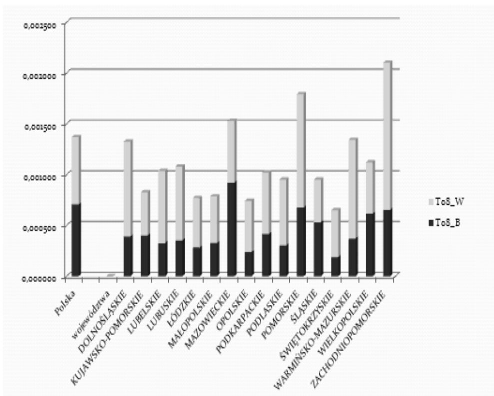
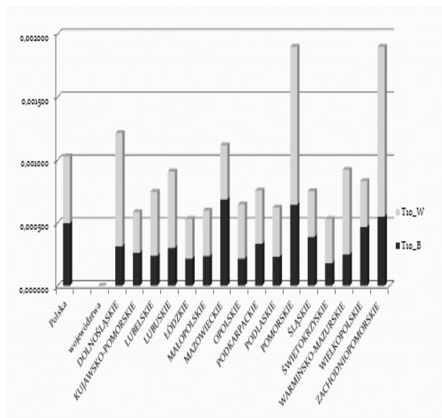


Figure 4b - 2010



Source: author's calculations.

In addition to generally smaller inequalities of local deprivation in 2010 than in 2008 – in all *voivodships* except for Pomorskie – the most impressive information carried out by the above figures is very big variation of disparities in local deprivation among *voivodships*. Some *voivodships* (Zachodniopomorskie, and Pomorskie) surpass 3 to 4 times the level of disparities in others (Kujawsko-Pomorskie, Łódzkie, Świętokrzyskie). There is also a tendency of more discernible differentiations among *powiats*, along with declining inequalities among *gminas* within them, meaning a likely shift of *powiats* (if this short time trend continue) toward a greater homogeneity in terms of overall measure deficiencies (MILD).

5. Conclusions

Official statistics has contributed a great deal in terms of the appropriate data to the progress of spatial analysis and its relevance for policy purposes. Especially in regional growth and local development areas of research that are increasingly involving micro-level aspects of such issues as convergence and social cohesion, along with evaluation-focused interest in patterns of spatial disparities. Some of these issues were analyzed in this paper starting with a holistic approach to measuring community (*gmina*) level of (under)development (with a multidimensional index of local deprivation (MILD) created for this purpose) and using data from Local Data Bank to address the question of *equity* and *efficiency* of public resource allocation (subsidies to *gminas*). Bearing in mind the complexity of the relationship between development and inequality, the Theil index was employed to check the impact (weak, as showed by data) of the resources accrued to *gminas* (subsidies) on reducing disparities in measures of local deprivation. Decomposition of the T index revealed a tendency to homogeneity of *powiats* but also to their greater differentiations within *voivodships* which, subsequently, are much more differentiated intraregionally than between themselves. In addition to some policy suggestions for public resource allocation policy, this results call for explicit involvement of 'sector' approach and for factor-based (by the dimensions) decomposition of inequality.

References

Abreu, M., De Groot H. L. F., Florax R.J.G.M.(2005), *Space and Growth, A Survey of Empirical Evidence and Methods.* Région et Dévelopment, N^0 21

Bhattacharyya, J. (2004), *Theorizing Community Development*, Community Development. Journal of the Community Development Society, Volume: 34, Issue: 2. Taylor & Francis, UK

Capello, R. (2009), *Space, growth, and development*, [in]: Roberta Capello and Peter Nijkamp (eds*), Handbook of Regional Growth and Development Theories*, Edward Elgar, Cheltenham, UK

Cnaan R.A., Milofsky C., Hunter A., *Creating a Frame for Understanding Local Organizations*, [w:] Cnaan R.A., Milofsky C. (ed.), *Handbook of Community Movements and Local Organizations*, Sudbury, MA

Cressie, N., Wikle, Ch. K. (2011), *Statistics for Spatio-Temporal Data*, John Wiley & Sons, Inc., Hoboken, New Jersey

Dygaszewicz, J. (2007), *GIS in the Public Statistics* (in Polish). Archiwum Fotogrametrii, Kartografii i Teledetekcji, Vol. 17a. Warsaw

Fischer, M. M., Getis A. (2010), *Handbook of Applied Spatial Analysis*, Springer-Verlag Berlin Heidelberg

Gieryn, T. F. (2000), *A Space for Place in Sociology,* Annual Review of Sociology, Vol. 26, Annual Reviews, USA

Goodchild, M.F., Anselin, L., Appelbaum, R .P., Harthorn, B. H. (2000), *Toward Spatially Integrated Social Science*, International Regional Science Review 23, (April 2000)

Gorzelak, G. (2009), *Coherence and Convergence: Synonyms or Two Different Notions*,[in]: Gudrun Kochendorfer-Lucis and Boris Pleskovic (eds.), *Spatial Disparities and Development Policy*, InVent, Berlin and The World Bank, Washington D.C.

Okrasa, W., Lapins J., Vremis M. (2006*), Measuring Community Deprivation for Geographic Targeting of Public Resources – case of Moldova*, Statistics in Transition, The Polish Statistical Association and Central Statistical Office, Warsaw. Vol. 7, Nr. 5 (June), Warsaw

Okrasa, W. (2011), *Public Administration and the Third Sector: Models of inter-sector interaction in the development context in light of empirical evidence* (in Polish), Administracja i Społeczeństwo, Wyższa Szkoła Administracyjno-Społeczna w Warszawie

Quah, D. (1993), *Galton's Fallacy and Tests of the Convergence Hypothesis,* Mimeo, LSE, London

Rey, S. J. (2004), *Spatial Analysis of Regional Income Inequality,* [in]: Michael F. Goodchild and Donald G. Janelle (ed.), *Spatially Integrated Social Science*, Oxford University Press, Oxford, UK

Shorrocks, A., Wan, G. (2004), *Spatial Decomposition of Inequality*, UNU-WIDER Discussion Paper No. 2004/01, Helsinki

Urry, J. (2004), *The Sociology of Space and Place,* [in] Judith R. Blau (ed.) *The Blackwell Companion to Sociology*. Blackwell Publishing, Malden, MA, USA

Vaisey S., 2007, *Structure, Culture, and Community: The Search for Belonging in 50 Urban Communes*, "American Sociological Review"*;* Dec; 72, 6.

Streszczenie

PRZESTRZENNA INTEGRACJA BADAŃ SPOŁECZNYCH I STATYSTYKI PUBLICZNEJ - UWAGI METODOLOGICZNE ORAZ WYNIKI BADAŃ ROZWOJU LOKALNEGO

Celem artykułu jest podkreślenie współzależności pomiędzy rozwojem nowoczesnych metod analizy przestrzennej a podażą odpowiednich danych generowanych w ramach statystyki publicznej. Przedmiotem rozważań jest sposób w jaki rosnąca dostępność danych na poziomie jednostek terytorialnych (gmin), może zwiększyć efektywność procesów decyzyjnych zachodzących na szczeblu lokalnym i regionalnym, ze szczególnym uwzględnieniem decyzji dotyczących alokacji publicznych środków rozwoju (lokalnego) poprzez tzw. geograficzne adresowanie interwencji publicznych.

10.2478/v10103-012-0036-4

ANATOLIY PILYAVSKYY[*], YURIY MATSIV[**], OLGA VOVCHAK[***]

Measuring Cost Efficiency of Ukrainian Banks in 2008

Abstract

The paper presents the results of a stochastic frontier analysis (SFA) of cost-efficiency of Ukrainian banks. As of lack of data on the personnel costs, we had to set limits to the year of 2008 only. To modeling banking activity, we apply the intermediary approach as one of the most commonly used in literature. Considering the results of statistical tests, we chose translog functional form of cost function and half–normal distribution of random inefficiency term. As a result of the research, we found out that efficiency of Ukrainian banks varies within 0.5224 and 0.9869 with an average value of 0.8734. Having checked a range of hypotheses, we discovered insignificant distinctions among banks by their size, type of owner and location.

1. Introduction

Present state of economy of Ukraine requires constant attention to banking system, conducting of a policy aimed at a creation of favorable conditions of stable and efficient functioning. Banking system plays a key role in the modern market economics. It is banks that attract deposits and give loans to the market participants, contribute to increasing competition and efficient re-distribution of money resources.

[*] Ph.D., Full Professor, Department of Management, Lviv Academy of Commerce

[**] Department of Management, Lviv Academy of Commerce

[***] Ph.D., Full Professor, Institute of Economics and Finance, Lviv Academy of Commerce

The economic crisis and post-crisis unstable political situation in the country predetermines the necessity of banking activity assessment and discovering of the causes of worsening financial position of commercial banks in order to preserve their financial stability. This is an important precondition of the country's coming out of the crisis, securing its economic rise and investment attractiveness. That is why information on bank's efficiency is rather important for the market participants. The problem is that none of the existing coefficients on banking activity (either absolute, or relative) give exhaustive information on bank's efficiency. Therefore in the modern practice of efficiency measurement along with classical analysis of financial coefficients more sophisticated methods of frontier analysis are used. One of the main advantages of these methods is a possible integral estimation of efficiency of banking activity. With such an approach the results of activity of a certain bank can be integrally compared with the results of the selected banks and at present are the best-practice ones (i.e. make the most of the existing technology), namely are on the so called frontier. The methods of frontier analysis can be parametric or non-parametric depending on the assumption used modeling a frontier.

In our previous research papers (see Pilyavskyy and Matsiv 2010, pp. 91-106, Pilyavskyy et al. 2010, pp. 16-22) we used a non-parametric method of frontier analysis, namely DEA, while in this very paper we use one of the parametric approaches, i.e. *stochastic frontier analysis* (SFA). SFA is widely used for bank's efficiency estimation in Central and Eastern Europe, in particular Russia (Byelousova 2009, pp.489-519, Styrin 2005, pp.1–29, Peresetsky 2010), Hungary (Hasan and Marton 2003, pp.2249–2271), Slovenia (Stavárek and Šulganová J. 2009), Czech Republic (Weill et al. 2006). As to Ukraine, we are acquainted only with one paper devoted to efficiency measurement of Ukrainian banks using SFA method (see Mertens and Urga 2001, pp. 292-308). That is why we consider research in this direction rather vital. In this research paper we suppose to check if:

- cost-inefficiency is present in the Ukrainian banking system;
- foreign banks are more efficient, than Ukrainian ones;
- efficiency of Ukrainian banks depends on their size;
- efficiency of Ukrainian banks somehow differs depending on their location.

The structure of our paper is as follows. In section 2 method and model of banking activity as well as data used for estimation of efficiency of Ukrainian banks are discussed. In section 3 we provide the main results of efficiency measurement and test some hypotheses. Finally, in section 4 we summarize.

2. Method, model, data

The foundations of the methodology of frontier analysis and modern efficiency estimation are in the paper by Farrell (Farrell 1957, pp. 253-290), who, in his turn, on the basis of the preceding works by Debreu (Debreu 1951, pp. 273-292) and Koopmans (Koopmans 1951, pp. 33-97), offered simple measure of economic efficiency of a firm and its decomposition onto allocative and technical. Depending upon the way a production frontier is built, methods of frontier analysis fall under: non-parametric, in which linear programming technique is used and parametric, where econometric analysis is applied. SFA method is the most widely used of the parametric methods.

SFA was introduced in the works by Aigner et al. (Aigner et al. 1977, pp. 21-37) and Meeusen and van den Broeck (Meeusen and van den Broeck 1977, pp. 435-444) independently from one another. In the approach to measurement of technical efficiency, econometric analysis is used to model production function, which contains two random components. One of them estimates random errors, while the other one deals with inefficiency measurement. Then firm's efficiency depends on a functional form for approximation of a production frontier and a distribution form of random components. Cobb-Douglas and translog are the two functional forms most often used for efficiency estimation, taking into consideration a multiplicative nature of efficiency and that Cobb-Douglas and translog can be linearized.

Having somewhat modified a model used for technical efficiency measurement, SFA also allows cost-efficiency estimation. Cost-function can be expressed as follows:

$$\ln C = f(y, w, z) + u + v \qquad (1)$$

here: C – costs, y – outputs (volume of output), w – prices for inputs (resources), z – so called netputs (fixed parameters), u – random inefficiency term, v – random error term. Distribution of random error term can be considered normal, while random inefficiency term - half-normal, truncated normal, exponential, gamma etc. There are no clear criteria for choosing a distribution of random inefficiency term. That is why, more often they choose either half-normal or truncated normal distributions of random inefficiency term. Then, having K banks, efficiency of bank k ($k = 1, \ldots, K$) of them (Eff_k) can be calculated as follows:

$$Eff_k = e^{-\hat{u}_k}, \qquad (2)$$

where $\hat{u}_k$ - estimate of parameter u_k.

Unfortunately, there exists no simple way of calculating $\hat{u}_k$ of u_k. It depends upon both distribution u_k, and a chosen method of estimation. For details see, e.g. Kumbhakar and Lovell (Kumbhakar and Lovell 2000).

Efficiency of the banking system on the whole (Eff) is arithmetic mean of measures of efficiency of individual banks:

$$Eff = \frac{\sum_{k=1}^{K} Eff_k}{K} \tag{3}$$

In our research, for efficiency measurement of Ukrainian banks we use data on the activity of Ukrainian banks in 2008[1] that are on the NBU's Web site[2]. Selecting data for the research we applied the intermediary approach to modeling banking activity (see Sealey and Lindley 1977, pp. 1251–1266).

According to the intermediary approach banks are considered as financial intermediaries between depositors and borrowers. Banks 'produce' intermediary services attracting deposits and other obligations and allocate them in earning assets (loans, securities, etc.). Loans and securities and other earning assets are outputs in our model. Prices of labor, borrowed funds and physical capital make price of inputs. We use an amount of banking capital as a netput (fixed input) (in details for the list of variables see table 1). However, independent variables that form a regression-equation may significantly correlate with each other, but that is undesirable, because of sensitivity of regressors even to inconsiderable data changes, so we calculated variance inflation factors (VIF) (see Gujarati 2004) to discover multicolinearity. For all independent variables VIF's values appear to be less than 10, so it can be considered that there is no multicolinearity.

[1] We use data of 2008, since the NBU ceased publishing data on personnel costs after 2008 and it is the key parameter for efficiency estimation.

[2] www.bank.gov.ua

Table 1. Variables and their definitions

Variable	Name	Definition
TC	Total costs	operative costs, interests and charges
TL	Total loans	personal or commercial loans but for the reserves under them
SOEA	Securities and other earning assets	securities (incl. state securities) and assets in other banks but for the reserves under them
PBF	Price of borrowed funds	interest and charge costs divided by all the types of borrowed funds
PL	Price of labour	personnel costs divided by assets[3]
PPC	Price of physical capital	total administrative costs divided by tangible and intangible assets
BC	Capital of bank	banking capital

Source: developed by the authors.

Consequently, we have data on activity of 151 Ukrainian banks in 2008. The following step is to choose a functional form and a distribution of a random inefficiency term. In order to choose between functional forms of either Cobb-Douglas or Trans-Log models, we used the Log-Likeliood Ratio Test (LR Test) (see Coelli et al 2005). According to the results of the test, on the level of significance equal to 0.05, a half-normal distribution is preferred. We also used the LR Test to choose a distribution of random efficiency term between half-normal and truncated normal. A half–normal distribution is preferred according to the results of the test on the significance level of 0.05. Consequently, the specification of our model is as follows:

[3] Let us note that the best approximation of labour costs is a ratio of personnel costs to a number of employees. Unfortunately, NBU do not publish data on a number of personnel.

$$\ln(\frac{TC}{PPC \cdot BC}) =$$
$$= \beta_0 + \beta_1 \ln\frac{TL}{BC} + \beta_2 \ln\frac{SOEA}{BC} + \beta_3 \ln\frac{PBF}{PPC} + \beta_4 \ln\frac{PL}{PPC} + \beta_5 \ln^2\frac{TL}{BC} +$$
$$+\beta_6 \ln^2\frac{SOEA}{BC} + \beta_7 \ln^2\frac{PBF}{PPC} + \beta_8 \ln^2\frac{PL}{PPC} + \beta_9 \ln\frac{TL}{BC}\ln\frac{SOEA}{BC} +$$
$$+\beta_{10} \ln\frac{TL}{BC}\ln\frac{PBF}{PPC} + +\beta_{11} \ln\frac{TL}{BC}\ln\frac{PL}{PPC} + \beta_{12} \ln\frac{SOEA}{BC}\ln\frac{PBF}{PPC} +$$
$$+\beta_{13} \ln\frac{SOEA}{BC}\ln\frac{PL}{PPC} + +\beta_{14} \ln\frac{PBF}{PPC}\ln\frac{PL}{PPC} + u + v \tag{4}$$

Random components are distributed in the following way:

$$v \sim N(0, \sigma_v^2), \quad u \sim N_+(0, \sigma_u^2) \tag{5}$$

It is known that the cost function has to be homogeneous. To satisfy this condition, we used one of the prices (PPC), namely *numeraire*, and divided total costs by it. In order to eliminate a heteroscedasticity effect, total costs and all outputs were divided by banking capital.

3. Results

To estimate the efficiency of Ukrainian banks, we applied R program, namely Benchmarking package (see Bogetoft and Otto 2011). The estimates of cost-function parameters (4) are given in, Table 2.

Table 2. Summary of estimation

Parameter name	Estimator of parameter	Std.err	t-value	Pr(>|t|)
β_0	-1.50712	0.08606	-17.5126	0.000
β_1	0.65071	0.05497	11.8368	0.000
β_2	0.50007	0.06704	7.4589	0.000
β_3	0.39522	0.05632	7.0176	0.000
β_4	0.61489	0.07581	8.1113	0.000
β_5	0.11808	0.01891	6.2449	0.000
β_6	0.07776	0.01102	7.0588	0.000
β_7	0.11479	0.01545	7.4289	0.000
β_8	0.11395	0.01562	7.2942	0.000
β_9	-0.20917	0.02032	-10.2942	0.000
β_{10}	0.00609	0.02680	0.2272	0.820
β_{11}	0.04712	0.02704	1.7424	0.083
β_{12}	0.06520	0.02262	2.8819	0.004
β_{13}	-0.05360	0.02698	-1.9868	0.048
β_{14}	-0.21266	0.02898	-7.3376	0.000
Λ	4.60382	1.30053	3.5400	0.000

$\sigma^2 = 0.035358$, $\sigma_v^2 = 0.001593$, $\sigma_u^2 = 0.033765$

log likelihood = 120.2094, $\lambda = \sqrt{\frac{\sigma_u^2}{\sigma_v^2}}$

Source: developed by the authors using the R program, Benchmarking package.

Having used the Wald test (see Coelli et al 2005), on the significance level of 0.05, we can affirm that inefficiency is present in the Ukrainian banking system. Moreover, taking into consideration the results of estimation, 95% of total variation can be explained by the inefficiency and only 5% - by random errors.

The average cost-efficiency of Ukrainian banks is rather high; it is 0.8734, while individual measures of cost-efficiency vary within 0.5224 to 0.9869. Within the framework of our research we also discuss cost-efficiency of Ukrainian banks by their size[4], type of owners (banks with foreign capital and Ukrainian ones) and their location (Kyiv or regional).

Table 3. Results of efficiency estimation

	N	mean	min	max	Std
All banks	151	0.8734	0.5224	0.9869	0.0885
Banks by size					
I (The Largest)	17	0.9153	0.8340	0.9869	0.0470
II (Large)	19	0.8785	0.7019	0.9612	0.0739
III (Medium)	21	0.8768	0.7044	0.9652	0.0730
IV (Small)	94	0.8640	0.5224	0.9797	0.0974
Banks by owner					
With foreign capital	44	0.8708	0.5224	0.9797	0.0877
Ukrainian	107	0.8744	0.5697	0.9869	0.0888
Banks by location					
In Kyiv	96	0.8765	0.5697	0.9869	0.0944
In regions	55	0.8678	0.5224	0.9709	0.0769

Source: developed by the authors.

We can see from table *3 that the larger the banks are, the higher is their efficiency*. Thus, the efficiency of the largest banks is 0.9153, while of the small ones – 0.8640. In the largest-bank-group the least efficiency variation is observed, while in the group of small banks it is the highest. However, having used an ANOVA to check a hypothesis on efficiency differences among the

[4] In the paper we use the NBU's methodology of differentiation of banks into groups. The methodology anticipates referring a certain bank to one of four groups by amount of their assets and regulatory capital.

bank groups by their size on the significance level of 0.05, we can affirm that there exist no differences in efficiency of banks by the groups.

As to the efficiency of banks by the type of owner, the average value of banks with foreign capital (0.8708), it hardly differs from that of Ukrainian banks (0.8744). The thing is quite the same with the banks located in Kyiv or regions (the average values respectively are 0.8765 and 0.8678). On the significance level of 0.05 the *t*- tests also point to the fact that efficiencies of foreign banks vs domestic ones, as well Kyiv banks vs regional ones do not differ.

4. Summary

The paper is a preliminary research of a possible application of stochastic frontier analysis to estimation of cost-efficiency of Ukrainian banks. Unfortunately, as of lack of data on the personnel costs, we had to set limits to the year of 2008 only. According to the results of efficiency measurement, we found out that the efficiency of Ukrainian banks varies within 0.5224 and 0.9869 with an average value of 0.8734.

Having checked a range of hypotheses, we discovered insignificant distinctions among banks by their size, type of owner and location.

Appendix A

Table A.1. Descriptive statistics of data used for estimation*

Variable	mean	min	max	Std
TC	544 081	7 396	10 000 821	1 194 851
TL	4 431 940	25 548	64 420 601	10 067 381
OEA	800 842	1 343	18 916 820	1 949 545
PBF	0.086	0.011	0.262	0.033
PL	0.027	0.003	0.143	0.017
PPC	0.332	0.027	0.978	0.221
BC	747 273	28 057	15 471 943	1 721 791

* Variables TC, TL, OEA and BC given in thousands of UAH

Source: developed by the authors.

References

Aigner, D., Lovell, C.A., Schmidt, P. (1977). *Formulation and Estimation of Stochastic Frontier Production Models.* Journal of Econometrics, no. 6

Bogetoft P., Otto L. (2011). *Benchmarking with DEA, SFA, and R.*, Springer Byelousova V.Yu. *Effektivnost' izdyerzhek odnorodnyh rossiyskih kommercheskih bankov: obzor problyemy I novyye ryezul'taty.* Ekonomichyeskiy zhurnal VShE, no. 4 // Белоусова В.Ю. (2009). Эффективность издержек однородных российских коммерческих банков: обзор проблемы и новые результаты. Экономический журнал ВШЭ, 4, сс. 489-519

Coelli T.J., Rao D.S.P., O'Donnell C.J. & Battese G.E. (2005). *An Introduction to Efficiency and Productivity Analysis* (Chapter 8, 9).Springer

Debreu G. (1951). *The Coefficient of Resource Utilisation.* Econometrica, l951, no. 19

Farrell M. J. (1957). *The Measurement of Productive Efficiency*. Journal of the Royal Statistical Society, l957, A CXX, Pt. no. 3

Gujarati D.N. (2004). *Basic Econometrics*, 4th Ed., The McGraw−Hill

Hasan I. and Marton K. (2003). *Development and efficiency of the banking sector in a transitional economy: Hungarian experience*. Journal of Banking & Finance, no. 27

Koopmans T. C. (1951). *Analysis of Production as an Efficient Combination of Activities.* // Koopmans T. C. (ed). Activity Analysis of Production and Allocation. Wiley, New York

Kumbhakar S. C. and Lovell C. A. K. (2000). *Stochastic Frontier analysis.* Cambridge University Press, Cambridge

Meeusen W. and van Den Broeck J. (1977). *Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error.* International Economic Review, no. 18(2)

Mertens A. and Urga G. (2001). *Efficiency, scale and scope economies in the Ukrainian banking sector in 1998.* Emerging Markets Review, no. 2

Peresetsky A. (2010). *Bank cost efficiency in Kazakhstan and Russia*. BOFIT Discussion Papers, 1/2010

Pilyavskyy A. and Matsiv Yu. (2010). *An Analysis of the Efficiency and Productivity of Ukrainian Banks.* Studia i Prace Uniwersytetu Ekonomicznrgo w Krakowie (Data Analysis Methods in Economics Research), Kraków, no. 11

Pilyavskyy A., Vovchak O., Matsiv Yu., Khoma T. (2010). *Efektyvnist' diyal'nosti ukrayins'koyi bankivs'koyi systemy (2005-2009). Metodolohiya analizu frontiv.* Visnyk NBU, no. 4 // Пілявський А. , Вовчак О., Маців Ю., Хома Т. (2010). Ефективність діяльності української банківської системи (2005-2009). Методологія аналізу фронтів. Вісник НБУ, 4, сс.16-22

Sealey C. W. and Lindley J. T. (1977). *Inputs, outputs and a theory of production and cost at depository financial institutions.* Journal of Finance, no. 32

Stavárek D. and Šulganová J. (2009). *Analýza efektívnosti slovenských bánk využitím Stochastic Frontier Approach*. Ekonomická revue – Central European Review of Economic Issues no. 12

Styrin K. (2005). *What Explains Differences in Efficiency Across Russian Banks?* Final Report (project no. 01-258), EERC

Weill L., Pruteanu-Podpiera A., Schobert F. (2006). *Banking Competition and Efficiency: A Micro-Data Analysis on the Czech Banking Industry* // Presentation at the Conference "Risk, Regulation and Competition: Banking in Transition Economies", Ghent, September 2006

Streszczenie

POMIAR EFEKTYWNOŚCI KOSZTOWEJ BANKÓW UKRAIŃSKICH W 2008

W artykule przedstawiono wyniki stochastycznej analizy granicznej (SFA) efektywności kosztowej banków ukraińskich. Ze względu na braki w danych dotyczących kosztów personelu, analizę ograniczono do roku 2008. W modelowaniu działalności bankowej, zastosowano podejście pośrednika jako jeden z powszechnie stosowanych w literaturze. Biorąc pod uwagę wyniki testów statystycznych, wybrano funkcjonalną formę funkcji kosztów I pół-normalny rozkład losowy.

10.2478/v10103-012-0037-3

ELŻBIETA SOBCZAK*

Smart Specialization of Workforce Structure in the European Union Countries – Dynamic Analysis Applying Shift-Share Analysis Method[1]

Abstract

The objective of the hereby paper is to present dynamic analysis and assessment of workforce structure in the European Union countries based on structural and geographical shift-share analysis. Workforce structure in economic sectors, distinguished based on R&D work intensity in the European Union countries in the period of 2008-2010, was the subject of diversification and transformations assessment.

Shift-share analysis enabled the decomposition of occurring changes into regional, structural and global effects as well as the identification of the, so called, allocation effect resulting in the classification of the studied countries with regard to combinations of local specialization and competitive advantages.

The performed research also allowed for the identification different kinds of workforce structure characterized by smart specialization (significant share of workforce in high-tech manufacturing sector or high-tech services sector) and the assessment of generated structural and competitive effects.

* Ph.D., Wrocław University of Economics

[1] The study was prepared within the framework of research grant provided by The National Centre of Science no.2011/01/B/HS4/04743 entitled: Classification of the European regional space in the perspective of smart development concept - dynamic approach.

1. Introduction

Contemporary socio-economic processes stimulate the importance of economy sectors based on knowledge and innovation. It results in the growing interest in sectors distinguished by research and development intensity level. Workforce level and structure and also their changes in sector and regional system constitute key elements of economy functioning analysis. They may also be used for comparative analysis of changes occurring in regions against the reference area.

In 2010 the European Union approved the *Europe 2020 Strategy* defining objectives aimed at providing support for member states to overcome economic crises successfully and ensure smart, sustainable and facilitating social inclusion development. The specified, by the strategy, smart development consists in knowledge-intensive economy and innovation development. Smart specialization of workforce structure constitutes one of the components of this development.

Innovations, as the major source of smart development, may manifest themselves in improvements occurring in particular sectors and inter-sector shifts. In relation to employment they may result in workforce structure transformation by shifts to more productive sectors and, in this way, exerting a long-lasting impact on economy. Therefore the growing interest, in EU countries research referring to changes in workforce level and structure in sectors distinguished by volume of R&D outlays, seems justified.

Workforce in high-tech manufacturing and knowledge-intensive services presents the domain focused approach covering production and services defined as high-tech in line with criterion of R&D outlays volume against added value. This relation is defined as R&D intensity.

Shift-share analysis (SSA) represents one of the research tools for investigating structural changes occurring in regional space in the given time range. SSA allows for analysing fluctuations in employment rate characteristic for a given country at the background of reference area (e.g. European Union), as well as their assessment in the context of the occurring structural and competitive changes. Additionally, SSA results provide information indispensable for the identification of key types of activities in a given region, as well as facilitate the typology of workforce structure with regard to different types of smart specialization and competitiveness.

2. Basic information and applied research methods

The domain focused approach is based on NACE – statistical classification of economic activities in the European Community. The division of high-tech sectors was first published in 1997 by OECD. From 1 January 2008 updated NACE classification (NACE Rev. 2) and the definition of high-tech manufacturing and knowledge-intensive services has changed. Therefore data comparisons before and after 2008 must be carefully performed or these changes must be referred to as a break in data continuity. Therefore it was accepted that the research time range will cover the period of 2008-2010 (in line with NACE Rev. 2).

Prepared by *Eurostat* and OECD workforce structure in the cross section of the following activities types by R&D intensity levels became the basis for conducting analysis: high-tech manufacturing (HTM), mid-high-tech manufacturing, mid-low-tech manufacturing, low-tech manufacturing, knowledge-intensive services (KIS), less knowledge-intensive services (LKIS), other sectors.

27 EU countries were covered by the study. Statistical data, necessary for workforce structure diversification analysis in EU space, were taken from Eurostat Internet database.

Structural and geographic workforce analysis by R&D intensity was conducted in EU member states applying classical and dynamic shift-share analysis and Esteban-Marquillas model using allocation effect[2].

Shift-share analysis is a research tool allowing to determine changes rate of total employment and R&D intensity sectors in each EU country on the background of reference area i.e. EU area.

Shift-share analysis of workforce changes rate in EU countries allowed for:

1. specifying structural and regional effects of workforce number changes in sectors distinguished by R&D intensity,
2. identification of key sectors for regional development,
3. classification of EU countries by positive and negative change effects values: structural and competitive,
4. classification of EU countries by components of allocation effects: specialization and competitiveness.

[2] For more information about the listed methods check, among others, the following publications: (Dunn 1960, pp. 97-112), (Barff, Knight 1988, pp. 1-10), (Esteban-Marquillas 1972, pp. 249-261), (Suchecki 2010, pp. 162-185), (Malarska, Nowakowska 1992, pp. 75-85).

3. Shift-Share Analysis of workforce in economy sectors distinguished by R&D intensity

Table 1 presents the results of classical SSA analysis for the period of 2009 and 2010 in relation to the previous year and referring to the overall result and structural effects of employment changes in the distinguished sectors. The general declining tendency of employment in the European Union is responsible, on average, for 1,82% of workforce size drop rate in every country and economy sector in 2009, as well as 0,5% in 2010. The global crisis diverted growth oriented tendencies in employment, which were observed in the EU since 2000, and one of its effects was the liquidation of many work places.

Individual structural effects for particular economy sectors are quite diversified. Structural gross effects refer to the average rate of employment changes in selected economy sectors in the EU countries. As the data presented in table 1 illustrate the highest average employment rate drop in 2009 was observed, respectively, in low-tech processing industry sector (-8,04%), mid-high tech (-6,66%) and high-tech (-5,84%) and also in mid-tech industry sector (-5,61%). In 2010 the observed employment rate decline in all sectors was lower. The only sector in which the employment rate increased, in both studied time periods, was the knowledge-intensive service sector – the increase by 1,5% in 2009 and by 0,6% in 2010. The lowest drop rate in workforce number was characteristic for less knowledge-intensive services sector (-1,97% and – 0,43%).

Table 1. Results of classic shift-share analysis with regard to effects of employment changes in sectors distinguished by R&D intensity

Effects of employment changes in EU countries (in %)		2009/2008	2010/2009
Total effect (growth rate of employment in EU)		-1,82	-0,50
Gross structural effect	1. high-tech manufacturing (HTM)	-5,84	-2,81
	2. mid-high-tech manufacturing	-6,66	-1,93
	3. mid-low-tech manufacturing	-5,61	-1,44
	4. low-tech manufacturing	-8,04	-2,62
	5. knowledge-intensive services (KIS)	1,50	0,60
	6. less knowledge-intensive services	-1,97	-0,43
	7. other sectors	-3,87	-1,68
Net structural effect	1. high-tech manufacturing (HTM)	-4,02	-2,31
	2. mid-high-tech manufacturing	-4,84	-1,43
	3. mid-low-tech manufacturing	-3,79	-0,94
	4. low-tech manufacturing	-6,22	-2,12
	5. knowledge-intensive services (KIS)	3,32	1,10
	6. less knowledge-intensive services	-0,15	0,07
	7. other sectors	-2,05	-1,18

Source: Author's estimations.

Net structural effects were defined by means of decreasing gross effects by workforce growth rate in the EU. Employment changes in knowledge-intensive services sector in 2009 resulted in higher workforce number in all EU countries, on average by 3,32% and in 2010 by 1,1%. Employment growth rate in less knowledge-intensive services sector in 2010 also influenced the slight growth of workforce size in the EU countries (0,07%). The employment in remaining sectors distinguished in line with R&D activities intensity was related to the drop of employment in the analysed countries. However, in 2009 the largest employment rate decline in the EU countries was caused by transformations occurring in low-tech industry sector (-6,22%). In 2010 employment rate decline in the EU countries (by 2,31%) was, to the largest extent, influenced by the changes taking place in high-tech industry sector.

In order to identify sectors responsible for the EU countries development table 2 presents Pearson linear correlation coefficients values for structural effects (distinguished in accordance with the classical shift-share analysis) and the share of workforce in particular economy sectors. The relations referring to both studied years turned out similar. Definitely, the strongest positive relation occurred between structural effects and the share of workforce in knowledge-intensive services sector (0,964 and 0,965). Positive relation, but of much lower intensity (statistically irrelevant for the accepted significance level $\alpha = 0,05$) was characteristic for structural effects and the share of workforce in less knowledge-intensive services sector and in high-tech industry sector. The remaining sectors featured negative influence on structural effects in the particular EU countries. Definitely the strongest negative correlation was registered in case of structural effects and the share of workforce in low-tech sector (-0,870 and 0,851).

Table 2. Correlation coefficient of structural effects and workforce share in the analyzed sectors

Workforce share in sectors	Structural effect	
	2009/2008	2010/2009
1. high-tech manufacturing (HTM)	0,008	0,070
2. mid-high-tech manufacturing	-0,393	-0,310
3. mid-low-tech manufacturing	-0,590	-0,534
4. low-tech manufacturing	-0,870	-0,851
5. knowledge-intensive services (KIS)	0,964	0,965
6. less knowledge-intensive services	0,289	0,229
7. other sectors	-0,736	-0,821

Source: Author's estimations.

Table 3 and picture 1 present the decomposition of overall workforce growth rate, ranked by R&D activities intensity in the period of 2010/2008, preformed in line with the shift-share dynamic analysis rules. Therefore further analysis covered aggregated structural and competitive effects calculated based on the effects for the years 2009/2008 and 2010/2009. Countries were ranked by the declining values of aggregated structural effects. The interdependence between aggregated net effect and aggregated structural effects and also the competitive ones measured by correlation coefficient was 0,385 and 0,995 respectively (both values are statistically significant). Therefore it may be concluded that the relevance of structural factor was much lower than in case of competitive factor.

In thirteen of the analysed EU countries a positive aggregated structural effect was observed, which means that workforce structure in these countries had a positive impact on workforce size changes. Only two countries form the, so called, new accession were included in this group, i.e. Malta and Cyprus. In the countries characterized by positive structural effects the share of workforce in knowledge-intensive services ranged from over 35% in Cyprus to almost 55% in Luxemburg.

Table 4 illustrates the classification of the EU countries with regard to positive and negative values of aggregated structural and competitive effects. The first group includes countries featuring positive influence of both structural and competitive factors on employment structure fluctuations, which indicates that workforce number changes in these countries may be more favourable for two reasons: because sectoral workforce structure has a positive impact on employment rate growth and also because economic sectors are characterized by higher dynamics of workforce size fluctuations than in other regions. This group covered six countries from EU 15 and 2 countries from EU 12.

Table 3. Dynamic shift-share analysis results of workforce number growth rate in economic sectors by R&D activities intensity in the period of 2010/2008

No	Country	Net total effect	Structural effect	Competitive effect	Workforce share in KIS
Positive structural effect					
1	Luxembourg	12,40	1,67	10,73	54,98
2	Sweden	1,31	0,99	0,32	50,70
3	United Kingdom	0,78	0,91	-0,14	48,01
4	Denmark	-2,35	0,75	-3,10	49,72
5	Netherlands	0,50	0,74	-0,24	45,64
6	Belgium	3,24	0,63	2,61	46,12
7	France	1,63	0,56	1,07	43,69
8	Ireland	-10,57	0,38	-10,96	44,71
9	Finland	-0,99	0,35	-1,34	42,22

10	Malta	4,90	0,31	4,59	40,49
11	Cyprus	2,82	0,24	2,58	35,24
12	Germany	2,16	0,04	2,12	40,00
13	Austria	2,48	0,02	2,47	37,11
Negative structural effect					
14	Greece	-1,48	-0,13	-1,35	33,29
15	Spain	-6,73	-0,13	-6,60	34,48
16	Italy	0,01	-0,29	0,30	33,69
17	Latvia	-14,55	-0,34	-14,22	34,34
18	Hungary	-0,22	-0,47	0,25	35,03
19	Lithuania	-9,66	-0,56	-9,10	33,93
20	Estonia	-11,12	-0,61	-10,50	35,25
21	Portugal	-2,02	-0,70	-1,32	30,05
22	Slovenia	-0,81	-0,87	0,06	33,51
23	Czech Republic	-0,09	-0,93	0,83	31,84
24	Slovakia	-2,53	-0,94	-1,58	32,35
25	Poland	3,34	-0,95	4,29	30,36
26	Bulgaria	-6,89	-1,13	-5,77	28,86
27	Romania	0,93	-1,86	2,79	19,95

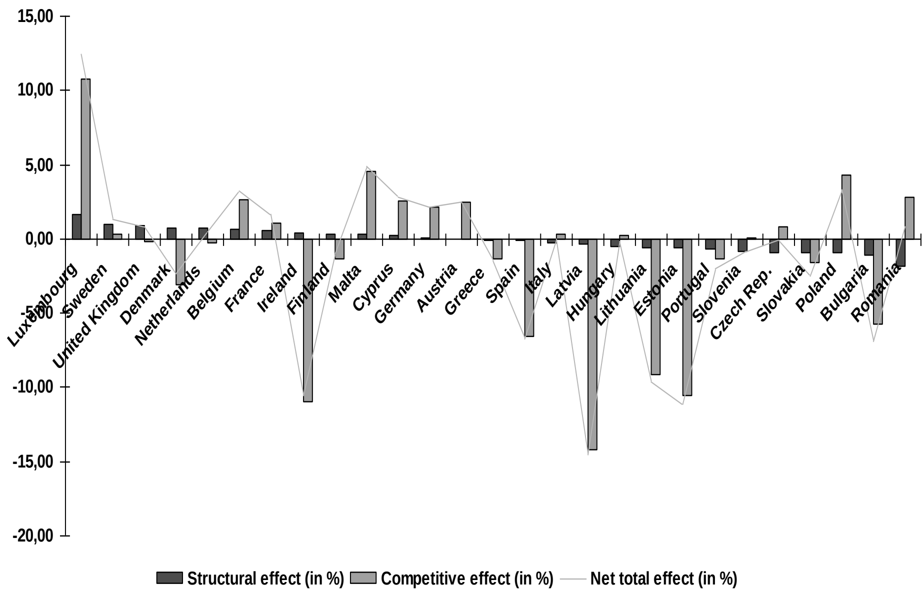
Source: Author's estimations.

Table 4. Classification of EU countries by positive and negative effect values: structural and competitive (dynamic SSA 2010/2008)

Group	Criterion of division	Countries	Number of countries
I	effects: structural (+) competitive (+)	Luxembourg, Sweden, Belgium, France, Germany, Austria, Malta, Cyprus	8 EU15 6 EU12 2
II	effects: structural (+) competitive (-)	United Kingdom, Denmark, Netherlands, Ireland, Finland	5 EU15 5 EU12 0
III	effects: structural (-) competitive (+)	Italy, Hungary, Slovenia, The Czech Republic, Poland, Romania	6 EU15 1 EU12 5
IV	effects: structural (-) competitive (-)	Greece, Spain, Portugal, Latvia, Lithuania, Estonia, Slovakia, Bulgaria	8 EU15 3 EU12 5

Source: Author's compilation.

Figure. 1. Dynamic shift-share



Source: Author's compilation.

The second group, characterized by positive influence of only the structural factor does not include any country from the, so called, new EU accession. The third group, featuring positive influence on employment changes of only the competitive factor, covered 5 new EU accession countries, including Poland. The forth group lists countries in which both the employment structure and internal regional development determinants exerted negative influence on workforce number changes in the period of 2008-2010. It covers 8 countries including Greece, Spain and Portugal of EU 15.

Tables 5 and 6 present the classification of EU countries with regard to allocation component effects: smart specialization or its absence as well as the advantage or disadvantage of competitiveness in high-tech industry and knowledge-intensive services sectors, respectively.

Table 5. Classification of EU countries by allocation effect components of workforce in HTM in 2010

Definition	Countries	Components of allocation effect	
		specialization (workforce share in HTM in %)	competitiveness (growth rate of employment in HTM in country less in EU in %)
Reference area	EU	1,08	-8,48
Smart specialization Competitive advantage	1. Ireland	3,10	2,13
	2. Hungary	2,77	5,15
	3. Malta	2,58	6,54
	4. Finland	1,89	6,53
	5. Slovenia	1,76	10,02
	6. Denmark	1,67	10,68
	7. Germany	1,51	1,36
	8. Czech Rep.	1,49	6,76
	9. Estonia	1,24	8,86
	10. France	1,09	1,07
Smart specialization Competitive disadvantage	1. Slovakia	1,46	-13,61
	2. Belgium	1,27	-0,73
Absence of smart specialization Competitive advantage	1. Italy	1,05	5,77
	2. Poland	0,78	7,14
	3. Sweden	0,71	2,65
	4. Romania	0,53	3,57
	5. Greece	0,46	4,81
	6. Lithuania	0,32	1,27
	7. Luxemburg	0,31	40,98
Absence of smart specialization Competitive disadvantage	1. United Kingdom	1,06	-2,47
	2. Austria	0,98	-0,16
	3. Spain	0,64	-12,97
	4. Netherlands	0,62	-11,07
	5. Bulgaria	0,59	-24,76
	6. Latvia	0,38	-7,39
	7. Portugal	0,37	-36,85
	8. Cyprus	0,20	-40,5

Source: Author's estimations.

Table 6. Classification of EU countries by allocation effect components of workforce in KIS in 2010

Definition	Countries	Components of allocation effect	
		Specialization (workforce share in KIS in %)	Competitiveness (growth rate of employment in KIS in country less in EU in %)
Reference area	EU	38,54	2,12
Smart specialization Competitive advantage	1. Luxembourg	54,98	9,73
	2. Denmark	49,72	0,97
	3. United Kingdom	48,01	1,48
	4. Belgium	46,12	0,63
	5. Malta	40,49	2,59
	6. Germany	40,00	0,69
Smart specialization Competitive disadvantage	1. Sweden	50,70	-0,92
	2. Netherlands	45,64	-4,27
	3. Ireland	44,71	-1,90
	4. France	43,69	-0,85
	5. Finland	42,22	-4,45
Absence of smart specialization Competitive advantage	1. Austria	37,11	4,53
	2. Cyprus	35,24	0,67
	3. Hungary	35,03	0,61
	4. Spain	34,48	0,49
	5. Slovenia	33,51	2,00
	6. Slovakia	32,35	2,04
	7. Czech Rep.	31,84	2,51
	8. Poland	30,36	6,12
	9. Romania	19,95	1,56
Absence of smart specialization Competitive disadvantage	1. Estonia	35,25	-3,90
	2. Latvia	34,34	-12,68
	3. Lithuania	33,93	-3,62
	4. Italy	33,69	-3,63
	5. Greece	33,29	-3,18
	6. Portugal	30,05	-0,78
	7. Bulgaria	28,86	-5,16

Source: Author's estimations.

A country is characterized by workforce structure featuring smart specialization in high-tech industry sector (knowledge-intensive services) if workforce share in this sector is higher than EU average. On the other hand, competitive advantage in high-tech industry sector (knowledge-intensive services) is present in the country in which employment changes rate in this particular sector is more favourable than sectoral changes rate in EU.

Based on the information presented in tables 5 and 6 the typology of workforce structure in EU countries was prepared with regard to smart specialization and the presence of competitive advantage, which was illustrated in table 7. As this analysis indicates, both smart specialization and competitive advantage, in both high-tech sectors in 2010, were characteristic for workforce structures in Denmark, Germany and Malta. Two-sectoral absence of smart specialization and competitive advantage occurred in Bulgaria, Latvia and Portugal.

Single-sectoral smart specialization in high-tech industry sector, as well as competitive advantage in this sector were registered in Estonia, The Czech Republic, Hungary and in Slovenia. Single-sectoral smart specialization in knowledge-intensive services sector and competitive advantage were present in this sector in 2010 in Luxemburg and Great Britain.

Table 7. Typology of employment structure by smart specialization and competitiveness in 2010

Smart specialization	Competitiveness			
	two-sector	single sector in HTM	single sector in KIS	absence
two-sector	Denmark, Malta Germany	Ireland, France, Finland	Belgium	-
single sector in HTM	Czech Rep., Hungary, Slovenia	Estonia	Slovakia	-
single sector in KIS	Luxembourg	Sweden	United Kingdom	Netherlands
absence	Poland, Romania	Greece, Italy, Lithuania	Spain, Cyprus, Austria	Bulgaria, Latvia, Portugal

Source: Author's compilation.

Poland and Romania were included in the group for which two-sectoral absence of smart specialization, as well as the occurrence of two-sectoral competitive advantage were identified which, while maintaining high employment rate growth in both high-tech sectors, may be the prognosis for workforce structure evolution in these countries towards smart specialization development.

4. Conclusions

The global crisis resulted in the fact that the average employment rate changes in EU were negative and equal -1,82% in the period of 2009/2008, -0,5% in the period of 2010/2009 and -2,31% in the period of 2010/2008. In the period of 2010/2008 the employment rate changes were better than average in UE in 13 countries (including four UE12 countries i.e.: Malta, Cyprus, Poland and Romania). These changes resulted mainly from internal changes occurring in the analyzed countries (competitive effect). The relationship between net effect and structural and competitive effects was respectively 0,385 and 0,995. Low-tech industry workforce number had the largest influence on employment rate decline in the period of 2009/2008. It resulted in employment rate drop in EU countries on average by 6,22%. In the period of 2010/2009 this influence was significantly lower, however, persisted negative (-2,12%). Number of workforce employed in knowledge-intensive services resulted in higher employment rate in EU countries in 2009 by 3,32% on average and in 2010 by 1,1%. Slight positive influence had, in 2010, the number of workforce in less knowledge-intensive services. The influence in remaining sectors was negative. 14 analyzed countries (including 10 from EU12, excluding Malta and Cyprus) were characterized by a negative structural effect, which confirms that in these regions workforce structure had negative impact on employment rate changes. In the countries featuring positive structural effect the workforce share in knowledge-intensive service ranged from 35% in Cyprus to almost 55% in Luxemburg. In the group of countries characterized by negative structural effect the lowest workforce share in knowledge-intensive services was registered in Romania at the level of 19,95%. Negative competitive effect occurred in 13 countries, which means that their sectors were characterized by lower than average dynamics of changes as compared to other EU countries. This group covered 5 EU12 countries – Latvia, Lithuania, Estonia, Slovakia and Bulgaria.

Two-sector smart specialization was identified in Denmark, Germany and Malta. These countries were characterized by both higher share and better employment rate changes in high-tech industry sectors and knowledge-intensive services than in EU. Bulgaria, Latvia and Portugal were included in the group of countries which featured the absence of both specialization and competitiveness in high-tech sectors. Poland and Romania constituted the target group characterized by the absence of smart specialization and competitiveness in both high-tech sectors which may open an opportunity for smart specialization development in the future.

The ongoing employment changes were related to economic crisis and their interregional diversification resulted mainly from internal conditions. It has to be emphasized that the competitive effect of employment rate changes was of dominating importance, which allows to assess favorably the positive competitive effects of less wealthy EU12 countries, including Poland.

References

Barff R.A., Knight III P.L. (1988), *Dynamic Shift-Share Analysis*, 'Growth and Change', no. 19/2

Dunn E.S., *A statistical and analytical technique for regional analysis*, Papers of the Regional Science Association, 1960, no. 6, pp. 97-112

Esteban-Marquillas J.M. (1972), *Shift and Share analysis revisited*, 'Regional and Urban Economics', vol. 2, no. 3

Hatzichronoglou T. (1996), *Revision of the High-Technology Sector and Product Classification*, OECD, Paris

Malarska A., Nowakowska B. (1992), *Metoda przesunięć udziałów w analizie dynamiki zmian strukturalnych [Shift-share method in the analysis of structural changes dynamics]*, 'Przegląd Statystyczny' *['Statistical Review']* no. 1

Nauka i technika w 2007r. [Science and technique in 2007], Informacje i opracowania statystyczne *[Statistical information and studies]*, GUS *[Central Statistical Office]*, Warsaw 2009

Suchecki B. (edit.) (2010), *Ekonometria przestrzenna. Metody i modele analizy przestrzennej [Spatial econometrics. Methods and models for spatial analysis]*, C.H. Beck Publishers, Warsaw

Streszczenie

INTELIGENTNA SPECJALIZACJA STRUKTURY PRACUJĄCYCH W KRAJACH UNII EUROPEJSKIEJ – ANALIZA DYNAMICZNA Z WYKORZYSTANIEM SHIFT-SHARE ANALYSIS

Celem referatu jest dynamiczna analiza i ocena struktury pracujących w krajach Unii Europejskiej w oparciu o strukturalno-geograficzną metodę przesunięć udziałów (Shift-Share Analysis). Ocenie zróżnicowania i przemian poddano strukturę pracujących w sektorach ekonomicznych wyodrębnionych wg intensywności działalności badawczo-

rozwojowej w krajach Unii Europejskiej w latach 2008-2010. Analiza przesunięć udziałów umożliwiła dekompozycję zachodzących zmian na efekty regionalne, strukturalne i globalne, jak również identyfikację tzw. efektu alokacji prowadzącego do klasyfikacji badanych krajów ze względu na występujące kombinacje specjalizacji lokalnej i korzyści konkurencyjności.

Przeprowadzone badania pozwoliły na identyfikację różnych rodzajów struktur pracujących cechujących się inteligentną specjalizacją (znaczący udział pracujących w sektorze przemysłu wysokiej techniki lub usług opartych na wiedzy) oraz ocenić generowane przez nie efekty strukturalne i konkurencyjne.

10.2478/v10103-012-0038-2

MAŁGORZATA MARKOWSKA[*], DANUTA STRAHL[**]

European Regional Space Classification Regarding Smart Growth Level[1]

Abstract

Europe 2020 strategy, as the successor of Lisbon strategy, represents the vision of social market economy for Europe of the 21st century covering three related priorities: 1/ smart growth: knowledge based economy and innovation growth; 2/ sustainable growth: support for effective and taking advantage of resources economy which is more environmentally friendly and more competitive; 3/ inclusive growth: support for high employment level economy providing social and territorial cohesion.

The paper presented classification results including European regional space positional statistics regarding advancement and smart growth level in NUTS 2 regions. Regional smart growth covers three pillars: innovation, creativity (Knowledge Based Economy) and smart specialization illustrated by characteristics available in Eurostat data resources. These qualities constituted the basis for constructing aggregate measures for specific pillars and also for the purposes of regional classification. The results of obtained classes were assessed in an overall European space, in groups of regions distinguished in line with integration processes chronology and also in capital regions and these including the country capital, as well as in the system of Polish regions.

[*] Ph.D., Wrocław University of Economics

[**] Ph.D., Full Professor at the Wrocław University of Economics

[1] The study was conducted within the framework of research grant NCN no. 2011/01/B/HS4/04743 entitled.: *European regional space classification in the perspective of smart growth concept – dynamic approach.*

1. Introduction

Three main components of Europe 2020 strategy (Europe 2010), i.e. smart growth, sustainable growth and inclusive growth, constitute the unity and reflect an integrated and holistic territorial approach towards the European economic development. It is required from regional managers to define substantive combinations of the five goals resulting from Europe 2020 strategy – namely: improved conditions for innovation, research and development; improved education levels; meeting goals with regard to climate change and energy; employment promotion or social inclusion promotion, especially by means of reducing poverty – which also reflects major challenges they face in local perspective. Preparing clear and measurable indicators, regarding each objective results and assumptions, becomes indispensable, among others, for all stakeholders, including these at regional level, to be capable of defining progress, impact and effectiveness of both projects and Operational Programmes.

The objective of the hereby study is to provide classification, using positional statistics, of the European regional space with regard to smart growth advancement degree and level in NUTS 2 regions.

2. Smart growth – vision, concept, measurement

‘Intelligence’ refers to cognitive capacity, to cleverness (in Latin *intelligentia*), while ‘smart’ is understood as intelligent, clever, reacting to changes in a given time. The most compact way to define smart growth is by referring to it as economic development based on knowledge and innovation.

Economic situation in Europe results, among others, from economic growth downturn, as compared to its most important competitors, which results from differences in labour productivity as the consequence of lower investment level in R&D and innovation, insufficient implementation of information and communication technologies, as well as inconvenient access of some social groups to innovation.

The Strategy *Europe 2020*, as the successor of the Lisbon Strategy, presents the vision of social market economy for Europe, heading towards overcoming the existing crisis and focused on preparing economy for the challenges of the decade to come. On March 3rd 2010 the European Commission issued the Communication *Europe 2020 – A strategy for smart, sustainable and inclusive growth fostering social inclusion.* The proposal announced by the

European Commission, referring to the initiation of a new strategy, was accepted by the European Council on March 26th 2010. The Strategy *Europe 2020* specifies both the position and goals of the European Union (EU) till 2020 (Europe 2010).

In the EU strategic documents smart growth is understood as obtaining better results in:

- education by encouraging to acquire knowledge, study and upgrade qualifications,
- research work and innovation by means of creating new products and services influencing economic growth and stimulating employment, as well as facilitating social problems solving,
- digital society, i.e. the implementation of IT and communication technologies.

Smart growth means increasing the role of knowledge and innovation as the driving forces in economy which requires: improving education quality, better research work results, support for innovation and knowledge transfer, full implementation of information and communication technologies, as well as undertaking efforts to transform innovation ideas into new products and services which could result in higher growth, opening new jobs and solving social problems in Europe and worldwide.

The Strategy *Europe 2020* covers three related priorities (Europe 2010):

- smart growth: development of knowledge-intensive economy and innovation;
- sustainable growth: support for economy using its resources efficiently, more environmentally friendly and a more competitive one;
- inclusive growth: support for economy featuring high employment rate, ensuring social and territorial cohesion.

Within the framework of streamlining activities and stimulating progress the European Commission accepted that the strategy should present an agreed, reduced number of measurable goals for 2020. These targets should constitute a part of the leading motive, i.e. smart and sustainable growth facilitating social inclusion. The objectives should be measurable and reflect situation diversity characteristic for the member states and, for the sake of comparisons, be based on sufficiently credible data.

However, in professional literature widespread research referring to smart growth is currently absent, both in national and regional dimension.

Strategic documents for EU countries indicate the following goals (Europe 2010; A strategy 2010):

- employment rate of population aged 20-64 should increase from the current 69% to at least 75% level, among others, as the result of the growing number of working women and elderly people, as well as better integration of migrants at the job market;
- current EU goal regarding investment in research and development activities (R&D) is 3% of GDP;
- carbon dioxide emission should be reduced to at least 20%, as compared to the level of 1990, or if the situation allows, even by 30%; the share of renewable energy sources in an overall energy consumption should increase to the level of 20% and energy efficiency should also grow by 20%;
- the goal referring to education and referring to the problem of individuals finishing school education prematurely, covers the reduction of dropout rate indicator to 10%, as compared to the current 15% level, and also increasing the percentage of people aged 30-34 who completed tertiary education from 31% to at least 40% till 2020;
- the number of Europeans living below the national poverty level should be reduced by 25% (i.e. by over 20 million people).

The implementation of priorities and meeting goals set forth by the Strategy requires numerous activities carried out at both EU level and within particular member states. Therefore the European Commission suggested seven thematic projects, including these focused on the following goals implementation:

- smart growth (Innovation 2011; Youth 2010; Digital 2011): Innovation Union, Youth on the move, A Digital Agenda for Europe,
- sustainable growth (Europe 2010): Europe taking effective advantage of resources and Industrial policy for globalization,
- social development fostering social inclusion (A strategy 2010): Programme for new skills and employment and also the European programme for fighting poverty.

3. Measures facilitating smart growth identification and assessment

Owing to the fact that professional literature, so far, has not offered any studies on smart growth, both in national and regional dimension, the identification of measures and measurement methods based on literature studies turned out, at the current stage, impossible. The approach followed by the World Bank (Knowledge 2006) was helpful in defining research areas and measures, where the methodology of difficult to measure Knowledge Based Economy was facilitated by specifying its pillars, for which lists of measures were prepared reflecting their nature in the best possible way.

In order to define areas allowing the specification of pillars and preliminary lists of measures helpful in smart growth measurement the studies of strategic objectives, flagship projects fostering smart growth as well as data base statistical resources for the European NUTS 2 level regions, were performed. Initial analysis indicates that significant elements of smart growth take the form of three components – pillars: smart specialization, creative regions and innovation. The proposals of measures for specified pillars, allowing for smart growth characteristics, are presented below.

Pillar I – smart specialization (SS), smart specialization indicators (whether a given variable is a stimulant (S) or a destimulant (D) was indicated in brackets):

(SS_1) - workforce employed in knowledge-intensive services as the share of workforce employed in services (S),
(SS_2) - average growth rate of workforce in knowledge-intensive services as the share of workforce employed in services (S),
(SS_3) - workforce in mid and high-tech industry sector (as % or workforce employed in industry) (S),
(SS_4) - average working rate of workforce in mid and high-tech industry sector (as % of workforce employed in industry) (S).

Pillar II – creative regions (CR), creativity indicators:

(CR_1) - share of tertiary education workforce in total workforce number in the region (S),
(CR_2) - share of population aged 25-64 participating in life-long learning in a region (S),
(CR_3) - human resources in science and technology as % of working population (S),
(CR_4) - people aged 15-64 born in a different country as % of population aged 15-64 (S),
(CR_5) - unemployment rate as % of active population (D),
(CR_6) - basic creative class (% of population aged 15-64) (S),
(CR_7) - share of residents in their working age who moved from different EU regions in the recent year (S),
- share of people with poor results in maths, reading and learning (D), national data,
(CR_8) - tertiary education graduates aged 30-34 (% of population aged 30-34) (S),
(CR_9) - access to broadband Internet (% of households) (S).

Pillar III – innovation (I), indicators of innovation potential, capacity and effects:

(I_1) - patents registered in the European Patent Office (EPO) per 1 million of workforce (S),

(I_2) - productivity in industry and service sectors (PPS per worker) index EU27=100 (S),
- productivity trend in industry and service sectors (annual mean productivity growth in real perspective) (S), data from a different year unavailable,

(I_3) - employment rate (% of population aged 20-64) (S),
- new foreign enterprises per 1 million of inhabitants (total number of new foreign enterprises per 1 million of inhabitants) (S), data unavailable,
- changed number of new foreign enterprises per 1 million of inhabitants (total number of new foreign enterprises per 1 million of inhabitants) (S), data unavailable,

(I_4) - investments in private sector per 1 inhabitant by purchasing power parity (S),

(I_5) - R&D expenditure in business sector (% of GDP) (S),

(I_6) - R&D expenditure (% of GDP) (S).

Indicators listed for the purposes of each smart growth pillars measurement represent the consensus between the requirements presented in EU strategic documents and database resources offering information at EU NUTS 2 level regions.

4. The proposal of research methodology

For the purposes of European regional space analysis regarding smart growth pillars, each of which is described by a set of qualities, the classification method, based on positional statistics, was suggested (Markowska, Strahl 2003, Strahl 2002).

A set of hierarchical objects is given $P = \{P_1, P_2, \ldots, P_n, \ldots, P_N\}$, within which there are placed lower order sets of objects - p, i.e.:

$$\begin{aligned} P_1 &= \{p_1^1 \cup p_2^1 \cup \ldots p_K^1\}, \\ P_2 &= \{p_1^2 \cup p_2^2 \cup \ldots p_K^2\}, \\ P_n &= \{p_1^n \cup p_2^n \cup \ldots p_K^n\}, \\ P_N &= \{p_1^N \cup p_2^N \cup \ldots p_k^N\}. \end{aligned} \tag{1}$$

In the study hierarchical objects represent EU countries, while lower order objects refer to NUTS 2 level EU regions of which each is described by a data matrix illustrating, in the first stage, the values of qualities assigned to smart growth pillars. Each pillar stands for a phenomenon characterized by a set of qualities, which were normalized using zero unitarization (considering their properties – stimulant, destimulant) (Kukuła 2000). Based on normalized values for each pillar aggregate measure was calculated as the arithmetic mean of qualities responsible for a given pillar. Following this approach aggregate measures constitute variables (m) marked by $X = \{X_1, ..., X_m\}$ symbols and observed in the studied objects – regions.

For each X_j ($j = 1, 2, \dots, m$) variable the chosen positional statistics are calculated, which in this case is the median (*Me*), i.e. the central value below and above which 50% of variable realization is present. The suggested classification procedure allows for the construction of 2^m (i.e. $G = 2^m$) classes possible to enter into combination with m variables (Markowska, Strahl 2003). Therefore:

- S_1 class covers these objects - regions from p set for which values of all m variables X_j meet condition (2), i.e.:

$$x_{kj} \geq Me\ X_j \quad \text{dla} \quad p_k^n \notin S_1. \tag{2}$$

- S_2 class includes these objects - regions from p set for which values of only (m-1) variables constructing one of $\binom{m}{m-1}$ combinations of variables meet condition (2),
- S_2 class includes these objects - regions from p set for which values of the next combination (m-1)-element variables meet condition (2),
- when no more (m-1)-element combinations are possible classes for (m-2)-element combinations are constructed and condition (2) is put forward,
- S_g ($g = 2^m$) class is constructed of regions for which x_{kj} values of all X_j variables do not meet condition (2).

5. European regional space classification regarding smart growth pillars – research results

The initial set of suggested variables was characterized by extensive data gaps, because out of 271 EU regions only 102 rows were complete (37,6%). In order to fill in data gaps extrapolation methods were applied, also including the regression method.

Years from which data originate for particular variables were different owing to their availability in Eurostat data base and in EU reports. Therefore for CR_6, I_2, I_5, I_6 qualities it was 2007, for CR_1 – CR_5, CR_8, I_1, I_3 qualities - 2008, for CR_8 – 2009, for SS_1 and SS_3 – 2010, mean value from the period of 2002 – 2006 for I_4, mean value for 2007-2008 in case of CR_7, geometric value from the period of 2000-2010 for SS_2 and SS_4 qualities.

For the classification of regions, with regard to smart growth level, aggregate measures were used in relation to the following pillars: smart specialization, creative regions and innovation which, as the result of the described procedure variants implementation, allowed for the construction of the following classes (see table 1) to which regions were assigned based on aggregate measures' values and their comparison with the median calculated for all analyzed regions.

The classification covers data for 265 regions, i.e. for 97,8% of all EU NUTS 2 level regions (Regions 2007). Due to the absence of all data the following regions were not included: French overseas Guadeloupe, Martinique, Guyane, Réunion and Spanish: Região Autónoma dos Açores, Região Autónoma da Madeira.

Table 1. Characteristics of classes

<table>
<tr><td colspan="2" rowspan="2">Class</td><td colspan="3">Measure value higher than the median</td></tr>
<tr><td>SS</td><td>CR</td><td>I</td></tr>
<tr><td colspan="2">1</td><td>+</td><td>+</td><td>+</td></tr>
<tr><td rowspan="3">2</td><td>A</td><td>+</td><td>+</td><td>–</td></tr>
<tr><td>B</td><td>+</td><td>–</td><td>+</td></tr>
<tr><td>C</td><td>–</td><td>+</td><td>+</td></tr>
<tr><td rowspan="3">3</td><td>A</td><td>+</td><td>–</td><td>–</td></tr>
<tr><td>B</td><td>–</td><td>+</td><td>–</td></tr>
<tr><td>C</td><td>–</td><td>–</td><td>+</td></tr>
<tr><td colspan="2">4</td><td>–</td><td>–</td><td>–</td></tr>
</table>

Source: Authors' compilation.

The lowest aggregate measure value was registered for innovation pillar and the highest for creative regions pillar. The largest diversification related to aggregate measure values – assessed on the basis of variability coefficient – is characteristic for EU regions from the perspective of innovation. Median values and other characteristics of aggregate measures in regions for each of the pillars are presented in table 2.

Table 2. Basic statistics of aggregate measures in pillars

Statistics	SS	CR	I
Min	0,1799	0,1600	0,0721
Max	0,6625	0,8945	0,7126
Mean value	0,3969	0,4052	0,2586
Median	0,3971	0,4087	0,2453
Standard deviation	0,0886	0,1171	0,1160
Variability coefficient	22,32	28,90	44,86

Source: Authors' compilation.

Aggregate measure value lower than 0,2 was registered for the following pillars:

- smart specialization only for the Spanish region of Comunidad Valenciana,
- creative regions for ten regions including five Greek ones (Notio Aigaio, Sterea Ellada, Peloponnisos and Ionia Nisia), four Romanian (Sud-Est, Sud-Vest Oltenia, Nord-Est and Sud – Muntenia) and the Italian region of Sicilia,
- innovation for 86 regions, however, below 0,1 in case of 16 following region: six Polish (Lubelskie, Zachodniopomorskie, Lubuskie, Opolskie, Kujawsko-Pomorskie and Warmińsko-Mazurskie), three Romanian (Nord-Vest, Centru, Sud-Est) and three Hungarian (Dél-Dunántúl, Észak-Magyarország, Észak-Alföld), two Bulgarian (Severozapaden, Severen tsentralen) and two Italian (Calabria, Sicilia).

On the other hand, aggregate measure value above 0,6 in particular pillars was registered in case of the following regions:

- smart specialization: German Bremen, Saarland and Kassel, as well as the French region of Corse,
- creative regions: British (Inner London, Outer London, Berkshire, Buckinghamshire and Oxfordshire and also Surrey, East and West Sussex), Netherlands regions (Noord-Holland, Utrecht), Swedish Stockholm, Belgian Prov. Brabant Wallon, Danish Hovedstaden and Finnish Etelä-Suomi,
- innovation – only for the German region of Stuttgart.

In the first and fourth class each of the 57 regions were classified, in the second class 78 regions (with nine in class 2A – aggregate measure value for innovation pillar was lower than median value, 26 regions in 2B class – aggregate measure value for creative regions pillar lower than the median and 44 regions in 2C class characterized by lower than median value of aggregate measure for smart specialization pillar), while the third class covered 72 regions (of which 41 were listed in 3A class – smart specialization aggregate measure value more favourable than the median, 23 in 3B class – higher than median aggregate value for the pillar including creative regions and seven in 3C class – only aggregate measure value for innovation pillar more favourable than the median value).

Among regions listed in the first class the largest representation originated from EU 15 regions with the vast majority from Germany (15), The Netherlands (9 out of 12), Great Britain (8), Belgium (7 of 8) and 5 out of 8 Swedish regions. While analyzing the number of regions in this class in the perspective of countries attention should also be paid to the fact that it includes 3 from each of 5 Danish and Finnish regions. The first class covers 26,8% from 209 EU 15 regions – see table 3.

Table 3. Regions from EU countries in the obtained classes

Class		Regions (number of regions)
1 (57)		(BE 7) Prov. Antwerpen, Prov. Oost-Vlaanderen, Prov. Vlaams-Brabant, Prov. Brabant Wallon, Prov. Liège, Prov. Luxembourg, Prov. Namur, (DK 3) Hovedstaden, Sjælland, Midtjylland, (DE 15) Stuttgart, Karlsruhe, Freiburg, Tübingen, Oberbayern, Oberpfalz, Mittelfranken, Unterfranken, Berlin, Bremen, Hamburg, Darmstadt, Gießen, Köln, Rheinhessen-Pfalz, (ES 1) Comunidad Foral de Navarra, (FR 4) Alsace, Franche-Comté, Bretagne, Midi-Pyrénées, (LU) Luxembourg, (NL 9) Groningen, Friesland, Drenthe, Overijssel, Gelderland, Utrecht, Zeeland, Noord-Brabant, Limburg, (SI 1) Zahodna Slovenija, (FI 3) Etelä-Suomi, Länsi-Suomi, Åland, (SE 5) Östra Mellansverige, Småland med öarna, Västsverige, Mellersta Norrland, Övre Norrland, (UK 8) Cheshire, Lancashire, North Yorkshire, Derbyshire and Nottinghamshire, Herefordshire, Worcestershire and Warwickshire, Berkshire, Buckinghamshire and Oxfordshire, Hampshire and Isle of Wight, Gloucestershire, Wiltshire and Bristol
2 (78)	A (9)	(FR 1) Languedoc-Roussillon, (HU 1) Közép-Magyarország, (SK 1) Bratislavský kraj, (UK 6) Tees Valley and Durham, Northumberland and Tyne and Wear, Cumbria, East Yorkshire and Northern Lincolnshire. West Midlands, Northern Ireland
	B (26)	(BE 1) Prov. Hainaut, (CZ 1) Strední Cechy, (DE 16) Niederbayern, Schwaben, Brandenburg – Südwest, Kassel, Braunschweig, Hannover, Lüneburg, Düsseldorf, Münster, Detmold, Arnsberg, Koblenz, Saarland, Dresden, Schleswig-Holstein, Thüringen, (IE 1) Border, Midland and Western, (FR 2) Haute-Normandie, Centre, (IT 2) Piemonte, Lombardia, (AT 3) Burgenland, Niederösterreich, Steiermark
		(BE 3) Région de Bruxelles-Capitale, Prov. Limburg, Prov. West-Vlaanderen (CZ 1) Praha, (DK) Syddanmark, Nordjylland, (DE 1) Trier, (IE 1) Southern and Eastern, (ES 5) País Vasco, La Rioja, Aragón, Comunidad de Madrid, Cataluña, (FR 5) Île de France, Aquitaine, Rhône-Alpes,

Class		Regions (number of regions)
	C (44)	Auvergne, Provence-Alpes-Côte d'Azur, (NL 3) Flevoland, Noord-Holland, Zuid-Holland, (AT 5) Wien, Oberösterreich, Salzburg, Tirol, Vorarlberg, (PT 1) Lisboa, (FI 1) Pohjois-Suomi, (SE 3) Stockholm, Sydsverige, Norra Mellansverige, (UK 12) Leicestershire, Rutland and Northamptonshire, East Anglia, Bedfordshire and Hertfordshire, Essex, Inner London, Outer London, Surrey, East and West Sussex, Kent, East Wales, Eastern Scotland, South Western Scotland, North Eastern Scotland
3 (71)	A (41)	(CZ 6) Jihozápad, Severozápad, Severovýchod, Jihovýchod, Strední Morava, Moravskoslezsko, (DE 5) Brandenburg – Nordost, Mecklenburg-Vorpommern, Chemnitz, Leipzig, Sachsen-Anhalt, (GR 3) Anatoliki Makedonia, Thraki, Sterea Ellada, Voreio Aigaio, (ES 2) Castilla-la Mancha, Extremadura, (FR 4) Picardie, Bourgogne, Lorraine, Corse, (IT 2) Valle d'Aosta, Provincia Autonoma Trento (LV) Latvija, (HU 6) Közép-Dunántúl, Nyugat-Dunántúl, Dél-Dunántúl, Észak-Magyarország, Észak-Alföld, Dél-Alföld, (PL 4) Podkarpackie, Lubuskie, Dolnośląskie, Warmińsko-Mazurskie, (PT 3) Alentejo, Região Autónoma dos Açores, Região Autónoma da Madeira, (RO 2) Sud – Muntenia, Vest, (SI 1) Vzhodna Slovenija, (SK 2) Západné Slovensko, Stredné Slovensko
	B (23)	(BG 1) Yugozapaden, (EE) Eesti, (ES 5) Galicia, Principado de Asturias, Cantabria, Comunidad Valenciana, Illes Balears, (FR 1) Limousin, (CY) Kypros, (LT) Lietuva, (PL 1) Mazowieckie, (RO 1) Bucuresti – Ilfov, (FI 1) Itä-Suomi, (UK 11) Greater Manchester, Merseyside, South Yorkshire, West Yorkshire, Lincolnshire, Shropshire and Staffordshire, Dorset and Somerset, Cornwall and Isles of Scilly, Devon, West Wales and The Valleys, Highlands and Islands
	C (7)	(DE 2) Oberfranken, Weser-Ems, (IT 4) Provincia Autonoma Bolzano, Veneto, Friuli-Venezia Giulia, Emilia-Romagna, (AT 1) Kärnten
4 (58)		(BG 5) Severozapaden, Severen tsentralen, Severoiztochen, Yugoiztochen, Yuzhen tsentralen (GR 10) Kentriki Makedonia, Dytiki Makedonia, Thessalia, Ipeiros, Ionia Nisia, Dytiki Ellada, Peloponnisos, Attiki, Notio Aigaio, Kriti, (ES 4) Castilla y León, Andalucía, Región de Murcia, Canarias, (FR 5) Champagne-Ardenne, Basse-Normandie, Nord - Pas-de-Calais, Pays de la Loire, Poitou-Charentes, (IT 13) Liguria, Toscana, Umbria, Marche, Lazio, Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna, (MT) Malta, (PL 11) Łódzkie, Małopolskie, Śląskie, Lubelskie, Świętokrzyskie, Podlaskie, Wielkopolskie, Zachodniopomorskie, Opolskie, Kujawsko-Pomorskie, Pomorskie, (PT 3) Norte, Algarve, Centro, (RO 5) Nord-Vest, Centru, Nord-Est, Sud-Est, Sud-Vest Oltenia, (SK 1) Východné Slovensko

Source: Authors' compilation.

The second class includes 18 British, 17 German, 8 Austrian and 8 French regions with regard to aggregate measures values of smart growth pillars, which gives the total result of 35,4% of all EU 15 regions.

In the third class 44 EU 15 regions and 28 EU 12 regions were classified including 11 British, 7 German, 7 Spanish, 6 Italian, 6 Hungarian, 6 Czech and also 5 French and 5 Polish regions.

In the fourth class, which covered regions for which each of the three aggregate measures illustrating smart growth pillars was lower than the median defined for all analyzed EU regions, the most numerous were Italian regions (13), Polish (11), Greek (10), as well as French, Bulgarian and Romanian ones (5 regions each).

Isolated regions from EU 12 were registered in the first class (Zahodna Slovenija) and the second class (Bratislavský kraj, Praha and Strední Cechy, as well as Közép-Magyarország). 50% of the remaining ones were classified in the third class, while 41,1% in the fourth class. It means that as many as 91,1% of regions from EU 12 are listed in classes where two or none of aggregate measures, for the distinguished smart growth pillars, did not present better values than the median value.

While analyzing positions of capital regions or including the country capital in the overall classification it has to be emphasized that out of 28 first class regions (aggregate measures values for each pillar higher than the median) five regions were listed including only one from countries of the recent accessions - Zahodna Slovenija – see table 4.

Table 4. Capital regions or including the country capital in the obtained classes

Class (number of regions)		SS	CR	I	Capital regions
1 (5)		+	+	+	Hovedstaden, Berlin, Luxembourg, Zahodna Slovenija, Etelä-Suomi
2 (13)	A (2)	+	+	–	Közép-Magyarország, Bratislavský kraj
	B	+	–	+	
	C (11)	–	+	+	Région de Bruxelles-Capitale, Praha, Southern and Eastern, Comunidad de Madrid, Île de France, Noord-Holland, Wien, Lisboa, Stockholm, Inner London, Outer London
3 (7)	A (1)	+	–	–	Latvija
	B (6)	–	+	–	Yugozapaden, Eesti, Kypros, Lietuva, Mazowieckie, Bucuresti – Ilfov
	C	–	–	+	
4 (3)		–	–	–	Attiki, Lazio, Malta

Source: Authors' compilation.

The following two EU 12 regions (Közép-Magyarország, Bratislavský kraj) are currently included in 2A class (aggregate measure value for innovation pillar is lower than the median). The most of regions, because as many as 11, were listed in 2C class which contains regions characterized by aggregate measure value for smart specialization pillar lower while for innovation and creative regions pillars higher than the median value).

Polish regions, in positional classification, with regard to aggregate measures values for smart growth pillars, were listed in 3A class (Podkarpackie, Lubuskie, Dolnośląskie, Warmińsko-Mazurskie – aggregate measure for smart specialization pillar better than the median) and 3B class (Mazowieckie – aggregate measure value for creative regions pillar better than the median) see table 5.

Table 5. Polish regions in the obtained classes

<table>
<tr><th colspan="2">Class (number of regions)</th><th>SS</th><th>CR</th><th>I</th><th>Polish regions</th></tr>
<tr><td colspan="2">1</td><td>+</td><td>+</td><td>+</td><td></td></tr>
<tr><td rowspan="3">2</td><td>A</td><td>+</td><td>+</td><td>–</td><td></td></tr>
<tr><td>B</td><td>+</td><td>–</td><td>+</td><td></td></tr>
<tr><td>C</td><td>–</td><td>+</td><td>+</td><td></td></tr>
<tr><td rowspan="3">3 (5)</td><td>A (4)</td><td>+</td><td>–</td><td>–</td><td>Podkarpackie, Lubuskie, Dolnośląskie, Warmińsko-Mazurskie</td></tr>
<tr><td>B (1)</td><td>–</td><td>+</td><td>–</td><td>Mazowieckie</td></tr>
<tr><td>C</td><td>–</td><td>–</td><td>+</td><td></td></tr>
<tr><td colspan="2">4 (11)</td><td>–</td><td>–</td><td>–</td><td>Łódzkie, Małopolskie, Śląskie, Lubelskie, Świętokrzyskie, Podlaskie, Wielkopolskie, Zachodniopomorskie, Opolskie, Kujawsko-Pomorskie, Pomorskie</td></tr>
</table>

Source: Authors' compilation.

The majority of Polish regions – 11 out of 16 were classified in the fourth class in which each of aggregate measures was lower than the EU median value).

6. Conclusions

The presented analysis constitutes a part of broader research stream evaluating the results of "Europe 2020" strategic goals such as smart growth, social cohesion and social exclusion prevention.

It illustrated the occurring extensive differences between EU 12 and EU 15 regions with regard to aggregate measures. The largest number of capital regions was listed in class 2 (13/27), in particular in class 2A (Közép-Magyarország, Bratislavský kraj) and 2C (Région de Bruxelles-Capitale, Praha, Southern and Eastern, Comunidad de Madrid, Île de France, Noord-Holland, Wien, Lisboa, Stockholm, Inner London, Outer London).

Mazowieckie was placed in class 3B – one aggregate measure higher than the median. The majority of Polish regions (11/16) are included in the class where all measures are smaller than the median.

The need occurs for conducting more dynamic analyses in order to assess changes of positions in the obtained classifications as well as the distance to be covered in accordance with Europe 2020 strategy.

References

A strategy for smart, sustainable and inclusive growth (2010), COMMUNICATION FROM THE COMMISSION, EUROPE 2020, European Commission, COM(2010) 2020 final, Brussels

Digital Agenda Scoreboard (2011), Commission Staff Working Paper, SEC(2011) 708, European Commission, Brussels

EUROPA 2020. Strategia na rzecz inteligentnego i zrównoważonego rozwoju sprzyjającego włączeniu społecznemu (2010), Komisja Europejska, Komunikat Komisji, KOM(2010)

Innovation Union Competitiveness report (2011), Directorate-General for Research and Innovation, Directorate-General for Research and Innovation, Research and Innovation, European Commission, Luxembourg: Publications Office of the European Union

Knowledge Assessment Methodology (KAM) (2006), World Bank Institute, World Bank, Washington

Kukuła K. (2000), Metoda unitaryzacji zerowanej, Biblioteka Ekonometryczna, Wydawnictwo Naukowe PWN, Warszawa

Markowska M., Strahl D. (2003), *Statystyki pozycyjne w klasyfikacji porównawczej*, [w:] K. Jajuga, M. Walesiak (red.), *Klasyfikacja i analiza danych – teoria i zastosowania*. *SKAD*, Prace Naukowe Akademii Ekonomicznej nr 988, Taksonomia 10, AE, Wrocław

Regions in the European Union. Nomenclature of territorial units for statistics NUTS 2006/EU-27 (2007), Series: Methodologies and Working Papers, European Commission, Luxemburg

Strahl D. (2002), *Klasyfikacja regionów z medianą*, [w:] J. Dziechciarz (red.), *Zastosowania metod ilościowych*, Prace Naukowe Akademii Ekonomicznej nr 950, Ekonometria 10, AE, Wrocław

Youth on the Move (2010), Publications Office of the European Union, European Union, Luxembourg

Streszczenie

KLASYFIKACJA EUROPEJSKIEJ PRZESTRZENI REGIONALNEJ ZE WZGLĘDU NA POZIOM INTELIGENTNEGO ROZWOJU

Strategia Europa 2020, jako sukcesor Strategii Lizbońskiej, jest wizją społecznej gospodarki rynkowej dla Europy XXI wieku, obejmującą trzy powiązane priorytety: 1/ rozwój inteligentny: rozwój gospodarki opartej na wiedzy i innowacji; 2/ rozwój zrównoważony: wspieranie gospodarki efektywniej korzystającej z zasobów, bardziej przyjaznej środowisku i bardziej konkurencyjnej; 3/ rozwój sprzyjający włączeniu społecznemu: wspieranie gospodarki o wysokim poziomie zatrudnienia, zapewniającej spójność społeczną i terytorialną.

W pracy przedstawiono wyniki klasyfikacji ze statystykami pozycyjnymi europejskiej przestrzeni regionalnej ze względu na stopień zaawansowania i poziom inteligentnego rozwoju w regionach szczebla NUTS 2. Rozwój inteligentny regionów uwzględnia trzy filary: innowacyjność, kreatywność (Gospodarka Oparta na Wiedzy) i inteligentną specjalizację ilustrowane charakterystykami znajdującymi się w zasobach informacyjnych Eurostatu. Cechy te stanowiły podstawę do konstrukcji miar agregatowych dla określonych filarów, a także do klasyfikacji regionów. Wyniki uzyskanych podziałów oceniono w całej europejskiej przestrzeni regionalnej, w grupach regionów wydzielonych zgodnie z chronologią procesów integracji, a także w regionach stołecznych i zawierających stolicę oraz w układzie regionów polskich.

10.2478/v10103-012-0039-1

PIOTR STROŻEK*

Comparative Analysis of the Level of Knowledge-based Part of Economies in European Union Countries with KAM Methodology

Abstract

The article presents regional classification of EU countries according to the knowledge development of economy, which in these days is treated as fundamental factor of international competitiveness. This differentiation created with Knowledge Economy Index (KEI) and Knowledge Index (KI), which are use by The World Bank in Knowledge Assessment Methodology (KAM).

In the analysis used four main pillars (i.e. The Economic Incentive and Institutional Regime, The Innovation System, Education and Human Resources, Information and Communication Technology), which showed relation between individual components.

The purpose of this article is to identify disparities in the use of knowledge in socio-economic life in the EU countries. This research was conducted with use of the cluster analysis (tools belonging to multidimensional comparative analysis).

1. Introduction

Creating an economy based on knowledge is a multidimensional process which occurs on many planes of social and economic life. The process involves not only technology of production but also the basis of society and the peoples' ability to absorb knowledge. Effective social relationships and economical partnership ensure social capital forming. If modern infrastructure is equally

* University of Łódź

available to everyone that means that innovative activity among manufactures and consequently general competition increase.

Measuring such a complicated phenomenon like a knowledge based economy is very difficult. The World Bank suggests the use of KAM (Knowledge Assessment Methodology) which consists of two basic indexes:

1. Knowledge Index (KI) which determines the whole knowledge potential of a countries while combining the creation, applying and flow of knowledge.
2. Knowledge Economy Index (KEI) in a more complicated tool, which is used to create global statistics of certain economies according to economic aspects.

Both of them are created on the basis of the 4 GOW pillars: The Economic Incentive and Institutional Regime, The Innovation System, Education and Human Resources, Information and Communication Technology (ICT).

In different countries the meaning of the above pillars on overall concept can be diversified. Clarity of KAM methodology allows to analyse certain indexes thanks to which that we can observe which country has some arrears. However, knowledge index and knowledge based economy index ensure the overall comparison. In the article there is also used analysis of concentration is also used. Thanks to that method we can indicate similarities among some European countries and show them on the map of the knowledge based economies.

2. The essence of knowledge based economy

The phenomenon of knowledge based economy (KBE) was first observed at the beginning of the 1990s. It soon spread on other well developed economies all over the world. People noticed that knowledge is the main index of production and it determines its growth. It guarantees more effective use of manufacturing potential mainly through the use of qualified human resources. (Malara 2006, p. 126). Knowledge must be perceived more extensively, taking into consideration at least four factors:

- whole society together with its traditions, culture and social behaviour patterns,
- state, its units and its policy,

- manufacturing plants and their potential, growth strategies, ways of using knowledge, demand for knowledge and the ability of its absorption,
- system of education, system of research and system of popularization of knowledge resources (Świtalski 2005, pp. 139–140).

Together with the growth of knowledge significance for economy appeared new terms, such as: “net economy”, “digital economy” or “new economy”. For many the most suitable definition describing growing trend of the meaning of knowledge for efficiency of social–economic system is “knowledge based economy”. It is a type of economy which uses knowledge and information both in production process and in the distribution of the products (*The Knowledge…* 1996, p. 7). Subjects (i.e. people, institutions, companies) building the economy gain knowledge to spread it and consequently be able to use it in a more effective way. The above subjects grow their competence (Kukliński 2003, p. 195).

The solid basis of KBE functioning is technical factor, i.e. temporarily very fast development of ICT Sector. This crucial development delivered effective information flows technology and data transfer which have a very positive influence on the total factor productivity (TFP) and consequently on all parts of a country’s economy (Porwit 2001, p. 115). Some people ever claim that we can say about the third or even fourth technological revolution, because the conditions of KBE enable achieving faster pace of long–term growth without the threat of inflation. However, sceptics say that the Internet cannot be compared to such inventions as steam engine or electricity. They claim that the influence of the Internet and the expansion of ICT on the efficiency of an economy will be not so effective as the influence of the inventions of the XIX and XX centuries. That fact can be proved by Solow’s paradox which shows that the era of computers can be seen in all aspects of an economy except the statistics showing the efficiency of work. Moreover, we have to take into account the growth of risk which can unsettle the fluidity of decision making. Sceptics also doubt the theory of lack of inflation threat, because even if fast economic growth, faster than the pace of efficiency and employment, is noticed, that means the decrease of unemployment rate. If the index falls below stated nonaccelerating inflation rate of unemployment (NAIRU) the inflation increases automatically (Wojtyna 2001, p. 6).

Nobody doubts that we can talk about spectacular changes in economy. The priorities of functioning of the whole system change. Service sector, investments into non–material assets, popularization of new technologies and creating information technology society are gaining importance (Platform, Sysko-Romańczuk, Moszoro 2004, p. 87). All these aspects influence innovation implementation. This term was first defined by Joseph Alois Schumpeter while presenting his theory in 1920 which indicated 5 rules describing definition of innovation:

1. Innovative good is unconventional good so far unknown by the consumers because they could not purchase it earlier.
2. Innovative process of production means applying technics not used so far, which enable most of all reduction of cost, influence productivity and effectiveness in a positive way and they also ensure that interference into natural environment is smaller than with the use of traditional production methods.
3. Innovation is opening a new market, i.e. the market on which a certain branch of industry was not introduced earlier.
4. Innovation is the use of natural resources or semi–finished products which come from an unknown so far source.
5. Innovation is co-ordination of new industry, e.g. through creation or breaking monopolistic position (Schumpeter 1960, p. 60).

We have to pay attention to the fact, that today innovation is perceived wider, because it involves changes which occur in intellectual, economical, organizational and administration spheres. In a wide perspective innovation is an original idea which influences the changes in social system, economy structure, technology and environment. That is why we can say that innovation means number of actions, which meet the expectations of consumers both in material and non-material spheres. The term “innovation” is often identified with the idea, method of acting or a thing which was unknown so far. Another meaning of this issue in the ability to discover the new. Here innovation is perceived as the opposite of traditional and routine action (Janasz 2003, pp. 47–51).

3. Theoretical aspects of the use of KAM methodology

The whole process of measuring the knowledge of economies is very complex and complicated. In the area of the programme of The Knowledge for Development (K4D) the World Bank suggests the use of KAM methodology, which has been constantly upgraded since 1998. Now the methodology consists of 148 variables (both quantitative and qualitative) which are collected for 146 countries. Periodical and precise analysis is to ensure the possibility to collect more and more accurate data. KAM methodology is created by 4 important pillars:

1. The Economic Incentive and Institutional Regime which is responsible for the perfection of economic policy and the activity of different institutions. Deepening, spreading and using the knowledge in the units can ensure effective actions through sharing resources and stimulation of creativity.
2. The Innovation System which involves acting of economic subjects, research centres, universities, advisory committees and other organizations which adopt their activity to the preferences of more and more demanding customers.
3. Education and Human Resources which mean personnel who can adopt to constantly developed technological solutions thanks to developing their skills.
4. Information and Communication Technology (ICT), which ensures effective communication and faster transfer of data. All these aspects influence transfer and processing information and knowledge (Chen, Dahlman 2006, pp. 5–9).

Theory of the World Bank says that the knowledge factor can only be involve in the national production when a certain financial structure of the above pillars in kept. Such actions are to ensure the economic success of country, through raising the added value of goods, services and social and economic growth which raises competitiveness of the country on global market.

The process of collecting certain data in some countries can differentiate a little. Such situation leads to problems when we want to compare the countries being surveyed. To ensure the clear researchers use 12 sub-indexes – 3 connected with one of the pillars and additionally 2 connected with general functioning of an economy (see table 1).

Table 1. Sub-indexes used in individual pillars

Overall functioning of economy	GNP – Gross National Product (in %), HDI – Human Development Index,
The Economic Incentive and Institutional Regime	RB – rates barriers and non-rates barriers(based on trade policy), RQ – regulation quality (applies to frequency of unfriendly policy which hinders international trade and business development), LR – law regulations (which apply to effectiveness of crime detection and the efficiency of judiciary),
The Innovation System	EBR – employment in the sector B+R per million of citizens, P – patents granted by USPTO (Us Patent and Trademark Office) per million of citizens, SRA – scientific and research articles published per million of citizens,

Education and Human Resources	RLA – rate of literacy among adults' (people above 15 years of age in relation to the number of citizens), PPS – percentage of people who attend secondary schools in relation to the population who should attend suck school, PPA – percentage of people attending colleges and universities in relation to people who can attend them,
Information and Communication Technology	PL – number of phone lines per one thousand of citizens, C – number of computers per one thousand people, I – Internet users per one thousand people.

Source: own studies based on www.worldbank.org/kam, states on 10.02.2012.

The four pillars described above two key indexes:

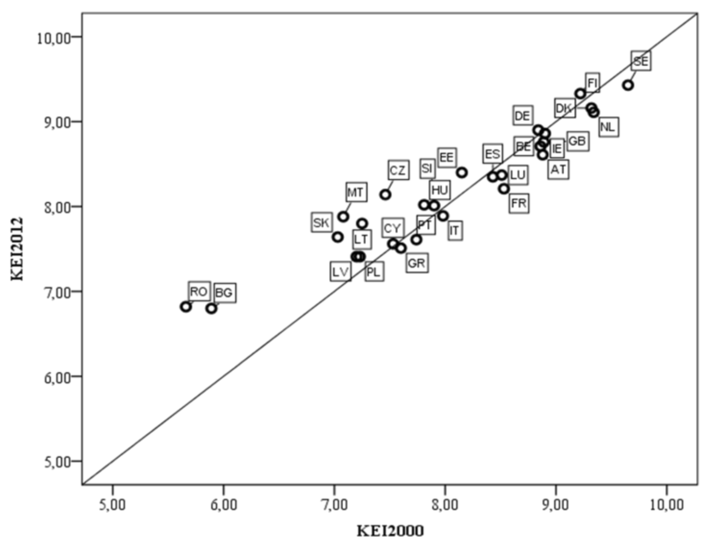
- Knowledge Index which determines the whole knowledge potential of a state while combining the creation, applying and flow of knowledge. The index is represented by the average of standardized findings for a certain country with the use of three most important pillars (without the economic incentive and institutional regime).
- Knowledge Economy Index in a more complicated tool, which is used to create global statistics of certain economies according to economic aspects. The most important characteristics of this group of variables assigned data is subjected to a normalization process through assigning certain. Later, partial indexes are calculated and finally they are valued again until the final index is assigned (www.worldbank.org/kam, states on 10.02.2012).

The effectiveness of this method is provided by its simplicity, clearness and versatility. Cross-sectional analysis of its certain aspects allows to create overall vision of a knowledge based economy. Another advantage is the fact that comparisons are done periodically and in international dimension, both from the synthetic and detailed perspective. Moreover, KAM methodology ensures clear graphic presentation of the analyses economic changes. There are some disadvantages of it as well. First of all, it is the aspiration to comprehensive picture of GOW. Furthermore, it is duplicating of information through taking into consideration highly correlated variables and of data in some countries which leads to difficulties in comparative analysis (Piech 2005, pp. 17–31).

4. KAM methodology in the light of empirical research

Constant analysis of a knowledge based economy conducted by the World Bank allows to determinate the growth and decreases of KEI and KI indexes in stated periods of time. The visual way of presenting the changes are scattering graphs (see picture 1 and 2), which are layouts of points (each point means one country) according to real index values- which are presented by standardised measurements ranging 0–10. The higher the index rate, the more developed the economy is from the point of view the measured phenomenon. On the horizontal axis there are index values from the year 2000 whereas on the vertical axis current values from 2012. The graphs were prepared for 27 European Union countries. Through the centre of the graph runs a straight line expressing the equation y=x, which divides the area into equal parts. Points which are on the line show countries whose index level does not change in two research periods. Points which are above the line represent countries in which index growth was observed in 2012 in comparison with 2000. However, points under the line represent the countries where there was index value decrease in 2012 in comparison with the year 2000. The further from the beginning of the system of coordinates, the better because it indicates bigger potential of knowledge and its better allocation in a country[1].

Figure 1. Scattering graph of KEI index in European Union countries

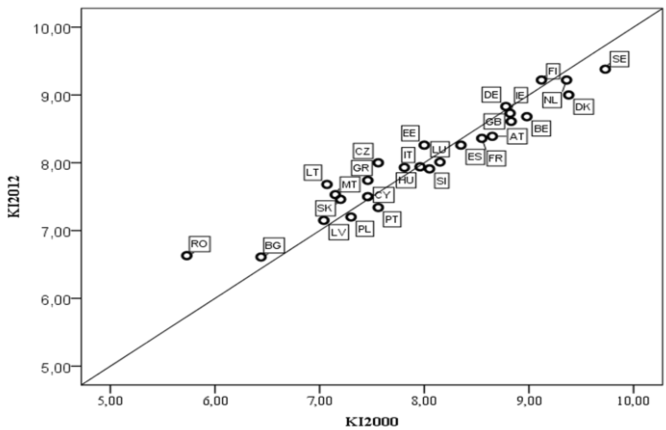


Source: own studies based on www.worldbank.org/kam, states on 10.02.2012 with SPSS Statistics.

[1] To make the graphs more clear a set of values ranging from 5 to 10 was set and the abbreviations of the names of countries were used according to the ISO 3166 system: AT – Austria, BE – Belgium, BG – Bulgaria, CY – Cyprus, CZ – The Czech Republic, DE – Germany, DK – Denmark, EE – Estonia, ES – Spain, FI – Finland, FR – France, GB – Great Britain, GR – Greece, HU – Hungary, IE – Ireland, IT – Italy, LT – Latvia, LU – Luxemburg, LV – Lithuania, MT – Malta, NL – Holland, PL – Poland, PT – Portugal, RO – Romania, SE – Sweden, SI – Slovenia, SK – Slovakia.

Looking at picture 1 we can assume that in the majority of top countries there was an index value decrease in the knowledge based economies in 2012 in comparison with the year 2000. The decrease also affected the leading country – Sweden. The only top countries, which reported index growth are Finland and Germany. Bottom and middle countries registered index, growth, which may indicate slow but constant process of catching up the arrears to the countries which can make a better use of knowledge in the economy. It is also worth mentioning that Bulgaria and Romania despite the biggest KEI changes, diverge from the rest of the countries.

Figure 2. Scattering graph of the KI index in European Union countries



Source: own studies based on www.worldbank.org/kam, states on 10.02.2012 with SPSS Statistics.

KI is very similar to KEI in most top countries (except Germany and Finland which noticed a slight growth of knowledge potential) are characterized by the index value decrease. Some middle and bottom countries (Romania, Bulgaria, Latvia, Lithuania, Slovakia, Malta, Cyprus, the Czech Republic, Hungary, Slovenia, Estonia and Greece) noticed the growth of the index. Situation is similar to the first graph where Bulgaria and Romania diverge from other countries, whereas Poland and Latvia are on the third position from the end in the year 2012 (after a decrease of a few positions comparing with the year 2000).

The graphs show two main indexes used in KAM methodology. But to be able to define the elements describing the indexes we must analyse the four main pillars (see table 2). Clarity of this method rely on possibility to construct four separate rankings which are assigned to all key pillars. Thanks to this it is easily to notice in which field certain state have development backlog. Only high values in all pillars guarantee this high values in main pillars (KI and KEI) and high place in general ranking.

Table 2. Ranking of the European Union countries by KAM (2012)

EU rank	Global rank		Country	KEI	KI	The Economic Incentive and Institutional Regime	The Innovation System	Education and Human Resources	Information and Communication Technology
1	1	0	Sweden	9.43	9.38	9.58	9.74	8.92	9.49
2	2	6	Finland	9.33	9.22	9.65	9.66	8.77	9.22
3	3	0	Denmark	9.16	9.00	9.63	9.49	8.63	8.88
4	4	-2	Holland	9.11	9.22	8.79	9.46	8.75	9.45
5	8	7	Germany	8.90	8.83	9.10	9.11	8.20	9.17
6	11	0	Ireland	8.86	8.73	9.26	9.11	8.87	8.21
7	14	-2	Great Britain	8.76	8.61	9.20	9.12	7.27	9.45
8	15	-1	Belgium	8.71	8.68	8.79	9.06	8.57	8.42
9	17	-4	Austria	8.61	8.39	9.26	8.87	7.33	8.97
10	19	7	Estonia	8.40	8.26	8.81	7.75	8.60	8.44
11	20	2	Luxemburg	8.37	8.01	9.45	8.94	5.61	9.47
12	21	2	Spain	8.35	8.26	8.63	8.23	8.82	7.73
13	24	-3	France	8.21	8.36	7.76	8.66	8.26	8.16
14	26	7	The Czech Republic	8.14	8.00	8.53	7.90	8.15	7.96
15	27	2	Hungary	8.02	7.93	8.28	8.15	8.42	7.23
16	28	0	Slovenia	8.01	7.91	8.31	8.50	7.42	7.80
17	30	-3	Italy	7.89	7.94	7.76	8.01	7.58	8.21
18	31	8	Malta	7.88	7.53	8.94	7.94	6.86	7.80
19	32	2	Lithuania	7.80	7.68	8.15	6.82	8.64	7.59
20	33	7	Slovakia	7.64	7.46	8.17	7.30	7.42	7.68
21	34	-4	Portugal	7.61	7.34	8.42	7.62	6.99	7.41
22	35	-3	Cyprus	7.56	7.50	7.71	7.71	7.23	7.57
23	36	-5	Greece	7.51	7.74	6.80	7.83	8.96	6.43
24	37	0	Latvia	7.41	7.15	8.21	6.56	7.73	7.16
25	38	-3	Poland	7.41	7.20	8.01	7.16	7.76	6.70
26	44	9	Romania	6,82	6,63	7,38	6,14	7,55	6,19
27	45	6	Bulgaria	6.80	6.61	7.35	6.94	6.25	6.66

Source: own studies based on www.worldbank.org/kam, states on 10.02.2012.

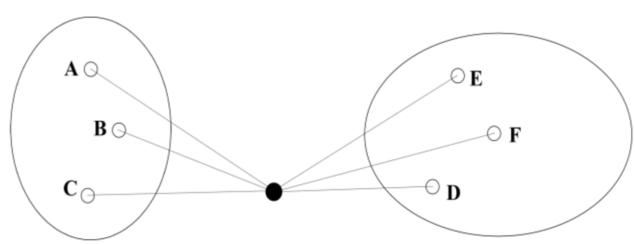
The first column in table 2 shows European ranking, the second world ranking while the third one represents growth/decrease in the world ranking in comparison with the year 2010. Both in the world and Europe Scandinavian countries dominate. Finland with its spectacular growth (6 points up within the last 2 years) in the ranking is also worth paying attention to. We must also mention that leading Sweden is constantly on the same position. Poland has a bottom position in the European Union. Unfortunately, Poland has registered a decrease of 3 positions in the world ranking since 2010.

Analysing the key pillars we can notice that the most work must be done in the education and human resource pillars (the level of these indexes in the majority of the countries is relatively lower than the level of indexes in other pillars). Spain, Hungary and Romania are the only exceptions. There indexes connected with education are higher than in other pillars. In Luxemburg (placed 11 in Europe) education index only 5,61. But high index in other pillars allow the country a rather high position in the general ranking.

5. Territorial classification of the EU based on KAM methodology

Similarities of education potential and its implementing in the economy between the European Union countries can be seen with the use of cluster analysis. The analysis comes from the multidimensional statistics which combines methods of data classification. The technique of cluster analysis guaranties the division of a researched area into consistent class objects. Object in one area are similar according to a stated measure of similarity and can be identified with some distance between them and differ from the objects in other areas. In the article hierarchic agglomeration method of Ward was used. To estimate the distance between the clusters the method uses variation analysis.

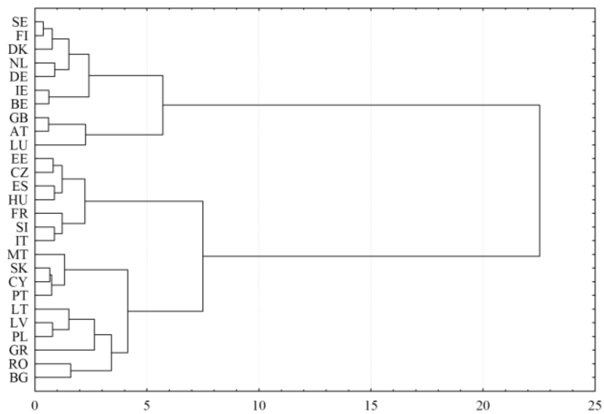
Figure 3. Semi-group distances in Ward's minimal variations method



Source: own studies based on Suchecki 2010, p. 62.

The distance between clusters is defined as a module of the difference between sums of squares of distances of points from the small created subsets join into numerous groups until the biggest subset contains all the objects. In the end the whole research process can be expressed as a hierarchic tree, thanks to which we can observe the consecutive analysis stages. This method is most frequently used in the economic analysis. By many researchers it is said to be the best, because if ensures the highest effectiveness of structure recognition in data matrix which describes analysed areas (Suchecki 2010, p. 62).

Figure 4. The diagram of Ward's agglomeration method



Source: own studies based on www.worldbank.org/kam, states on 10.02.2012 with Statistica 8.

Analysing picture 4 we can notice that two most similar countries are Sweden and Finland, which join into clusters on the level of consistent bond at 0,5. The most different is Greece which joins the group containing Poland, Latvia and Lithuania on the level slightly lower than 3.

On the basis of the diagram a map was constructed showing special relation of researched countries in different parts of the continent. The number of groups was decided on the level of consistent bond 3, which allowed to determine 6 groups. Two groups were distinguished. There are 7 countries which belong to each of them. In the first group there is Sweden, Finland, Denmark, Holland, Germany, Ireland and Belgium. The second Estonia, the Czech Republic, Spain, Hungary, France, Slovenia and Italy. Other groups contain 4 countries. In the first one there are: Poland, Lithuania, Latvia and Greece. The second 4–part group consists of Malta, Slovakia, Cyprus and Portugal. The least numerous is 3–part group, formed of Great Britain, Austria and Luxemburg. Bulgaria and Romania form the smallest, 2–element group.

Figure 5. Map of GOW potential in the European Union Countries (in 2012)

Source: own studies.

Looking at the map (drawing 5) we can see, that knowledge is the most effectively used in the economies of Scandinavian countries and in central Europe. The measures are slightly worse in South-West European countries. The worst situation is in East Europe. It is also worth paying attention on the fact, that countries with the highest grades are in immediate proximity (except Ireland). The situation is similar as far at the countries with the lowest grades are concerned (two weakest groups), which are on the east border (from Latvia to Bulgaria).

6. Conclusions

The use of knowledge potential in the European Union countries is very diversified. Between the top countries in the ranking and those which are at the bottom there is a very big developmental difference. But in the last few years the index growth of lowest classified countries and the slight decrease in the leading countries have been observed. It can indicate a slow make up in the arrear worse developed areas. But it will take a lot of time and effort to balance the levels of KI an1d KEI. There is also a threat that this can be impossible.

Poland, in terms of the use of knowledge in the economy presents average. It has the third position from the end in Europe and very distant position in the world ranking. Moreover, in the last few years the position of Poland has been weakening (which may indicate too slow pace of knowledge allocation in social and economic areas). What makes the situation worse, countries which joined European Union together with Poland (1 May, 2004), i.e. Cyprus, Lithuania, Latvia, Malta, Slovakia and Slovenia have got higher indexes. We cannot forget Estonia, the Czech Republic and Hungary where the indexes are much higher. It must be also mentioned that the KI and KEI are constantly increasing in Bulgaria and Romania (the only two countries which are lower in the ranking comparing with Poland).

Therefore, we can ask a question if Poland fully used the development chances after joining EU? There is no clear answer to this question. But to make Polish manufactures more competitive it is necessary to unite education and business. The government has a very important role to play in this area. Its policy should be for innovation and should support the use of knowledge in economic life through raising expenditure on research and developmental sectors. Changes should also be introduced in education which should not deal with perfecting people's abilities but first of all it should be able to respond to the needs of human resources on the job market. More frequent use of computers and the Internet by entrepreneurs is crucial. According to the results in KAM, ITC pillar has the lower index in comparison with the other three.

There are no doubts that in Poland there are a lot of areas which demand bigger involvement of knowledge to increase its effectiveness. But we must look into future with hope that it will happen and that the indexes of knowledge and economy based on knowledge will be higher and that it will influence not only the higher position in the ranking but also social–economic reality and real effects taking place visible in the economy.

References

Chen D. H. C., Dahlman C. J. (2006), *The Knowledge Economy, the KAM Methodology and World Bank Operations*, World Bank Institute, Washington, D.C. 20433

Janasz W. (ed.) (2003), *Innowacje w modelach działalności przedsiębiorstw*, Uniwersytet Szczeciński, Szczecin

Kukliński A. (ed.) (2003), *Gospodarka oparta na wiedzy. Perspektywy Banku Światowego*, Biuro Banku Światowego w Polsce, Komitet Badań Naukowych, Warsaw

Malara Z. (2006), *Przedsiębiorstwo w globalnej gospodarce. Wyzwania współczesności*, PWN, Warsaw

Migut G. (2009), *Zastosowanie technik analizy skupień i drzew decyzyjnych do segmentacji rynku*, StatSoft Polska, Cracow

Piech K. (2005), *Mierzenie rozwoju edukacji i kapitału ludzkiego w krajach transformacji systemowej z punktu widzenia gospodarki opartej na wiedzy*, [in:] Kopycińska D. (ed.), *Konkurencyjność rynku pracy i jego podmiotów*, Katedra Mikroekonomii, Uniwersytet Szczeciński, Szczecin

Platonoff A. L., Sysko-Romańczuk S., Moszoro B. (2004), *Innowacyjność polskich firm w gospodarce opartej na wiedzy* [in:] Grudzewski W. (ed.), *Ekonomika i Organizacja Przedsiębiorstwa nr.1*, ORGMASZ, Warsaw

Porwit K. (2001), *Cechy gospodarki opartej na wiedzy (G.O.W.)* [in:] Kukliński A. (ed.), *Gospodarka oparta na wiedzy. Wyzwanie dla Polski XXI wieku*, Warsaw

Schumpeter A. J. (1960), *Teoria rozwoju gospodarczego*, PWN, Warsaw

Świtalski W. (2005), *Innowacje i konkurencyjność*, Wydawnictwa Uniwersytetu Warszawskiego, Warsaw

Suchecki B. (ed.) (2010), *Ekonometria Przestrzenna. Metody i modele analizy danych przestrzennych*, C.H.BECK, Warsaw

The Knowledge-Based Economy (1996), General Distribution OECD/GD(96)102, Organisation for Economic Co-operation and Development, Paris

Wojtyna A. (2001), *Czy tradycyjna ekonomia pozwala zrozumieć nową gospodarkę?*, VII Kongres Ekonomistów Polskich, Warsaw

Websites:

Organisation for Economic Co-operation and Development: www.oece.org

The World Bank: www.worldbank.org

Streszczenie

ANALIZA PORÓWNAWCZA POZIOMU GOSPODAREK OPARTYCH NA WIEDZY W PAŃSTWACH UNII EUROPEJSKIEJ Z WYKORZYSTANIEM METODOLOGII KAM

W artykule przedstawiono klasyfikację terytorialną państw Unii Europejskiej według rozwoju wiedzy gospodarek, która w dzisiejszym świecie traktowana jest jako determinanta międzynarodowej konkurencyjności. Zróżnicowanie to zostało skonstruowane na podstawie indeksów KEI (Knowledge Economy Index) oraz KI (Knowledge Index) wykorzystywanych przez Bank Światowy w metodologii KAM (Knowledge Assessment Methodology).

Uwzględnienie czterech głównych filarów (tj. system bodźców ekonomicznych, system innowacyjny, edukacja i jakość zasobów ludzkich oraz nowoczesna infrastruktura informacyjna) umożliwiło wskazanie relacji pomiędzy poszczególnymi składowymi.

Celem artykułu jest wskazanie dysproporcji wykorzystania wiedzy w życiu społeczno-gospodarczym w państwach UE. Badanie zostało przeprowadzone przy użyciu analizy skupień (narzędzia zaliczanego do wielowymiarowej analizy porównawczej).

10.2478/v10103-012-0040-8

GRAŻYNA TRZPIOT*

Spatial Quantile Regression

Abstract

In a number of applications, a crucial problem consists in describing and analyzing the influence of a vector X_i of covariates on some real-valued response variable Y_i. In the present context, where the observations are made over a collection of sites, this study is more difficult, due to the complexity of the possible spatial dependence among the various sites. In this paper, instead of spatial mean regression, we thus consider the spatial quantile regression functions. Quantile regression has been considered in a spatial context. The main aim of this paper is to incorporate quantile regression and spatial econometric modeling. Substantial variation exists across quantiles, suggesting that ordinary regression is insufficient on its own. Quantile estimates of a spatial-lag model show considerable spatial dependence in the different parts of the distribution.

1. Introduction

1.1 Linear Regression - introduction

Linear regression is the standard tool for many empirical studies. When the relationship between a dependent variable, y, and a set of explanatory variables, $\boldsymbol{X}$, can be written as $y = X\beta + u$, a simple ordinary least squares

* Ph.D., Professor at the University of Economics in Katowice

(OLS) regression of y on X can provide unbiased estimates of the parameters, β, and a predicted value, $\hat{y} = X\hat{\beta}$.

This heavy reliance on linear regression models has been carried over to the analysis of spatial data. The most commonly used spatial model adds a weighted average of nearby values for the dependent variable to the list of explanatory variables:

$$y = \rho WY + X\beta + u\,. \tag{1}$$

In this model, ***W*** is a "spatial weight" matrix that specifies the relationships between observations. W is a "spatial weight matrix" with rows that sum to one and zeros on the diagonals, and ρ is a parameter measuring the strength of the relationship.

The model can be useful when X does not fully account for the tendency for the dependent variable to be highly correlated over space, so that nearby values of Y provide significant explanatory power. The endogeneity of WY poses challenges for estimation. Most empirical applications are based on maximum likelihood estimation of the model under the assumption of normally distributed errors. Other approaches are based on instrumental variables (IV) estimation, usually with spatially lagged values of X (such as WX and WWX) as instruments for WY. Several researchers have used the spatial AR model as the basis for quantile regressions in which both ρ and β are allowed to vary across quantiles.

Typical specifications of the spatial weight matrix are based on first-order contiguity when the data are drawn from geographic units such as counties or census tracts. Though the approach is used less commonly for point data, typical specifications are similar in that the spatial weights are assumed to decline rapidly with distance between observations.

Predicted values are then based on

(1) the structural model

$$\hat{Y}(\tau) = \hat{\rho}WY + X\hat{\beta}$$

(2) the reduced form

$$\hat{Y}(\tau) = (I - \hat{\rho}W)^{-1} X\hat{\beta}$$

(3) or a decomposition into "trend" and"signal" components

$$\hat{Y}(\tau) = X\hat{\beta} + \hat{\rho}W(I - \hat{\rho}W)^{-1} X\hat{\beta}$$

Spatial effects generally appear as noise around a spatial trend that looks much like the predicted values from an OLS regression of Y on X. The objective of a regression analysis is to estimate the coefficients, ρ and β, and to obtain predictions of y at given values of $\boldsymbol{X}$.

Regression analysis is not well suited to explaining the *distribution* of a variable. When the predicted values from a regression are $\hat{y} = X\hat{\beta}$, then the distribution of the predicted values simply mimics the distribution of the variables in X. The implied effect of a change in one of the explanatory variables is to cause a *parallel* shift of $\hat{y}$ by an amount determined by the variable's estimated coefficient. Though a parallel shift may be reasonable in some cases, it is limitation that a research may not want to impose beforehand.

1.2. Quantile Regression

Quantile regression is a method for estimating functional relations between variables for all portions of probability distribution. Typically a response variable Y is some function of predictor variable X. Regression application focus in estimating rates of changes in the mean of the response variable distribution as some function of a set of predictor variables. In the other words the function is defined for the expected value of Y conditional X, $E(\mathrm{Y}|\mathrm{X})$. Regression analysis gave incomplete picture of the relationships between variables especially for regression models with heterogeneous variances.

Quantile regression was developed as an extension of the linear model for estimating rate of change in all parts of the distribution of response variables. The estimates are semi parametric in the sense that no parametric distributional form (eg. normal, Poisson, negative binominal, etc.) is assumed for the random error part of the model ε, although a parametric form is assumed for the deterministic portion of the model (eg. $\beta_0 X_0 + \beta_1 X_1$). The conditional quantiles denoted by $Q_y(\tau|\mathrm{X})$ are the inverse of the conditional cumulative distribution function of the response variable $\mathrm{F}_{\mathrm{y}}^{-1}(\tau|\mathrm{X})$, where $\tau \in [0, 1]$ denotes quantile rank.

The quantile model posits the τ^{th} quantile of Y conditional on x to be,

$$\mathrm{Q}(\tau|\mathrm{x}) = \alpha(\tau) + \mathrm{x}\beta(\tau),\ 0 < \tau < 1. \tag{2}$$

If $\beta(\tau)$ is a constant β, the model reduces to the standard conditional expectation model, $\mathrm{E}(\mathrm{Y}|\,\mathrm{x}) = \alpha + \mathrm{x}\beta$, with constant variance errors. When $\beta(\tau)$ depends on τ, the model allows the distribution of Y to depend on x in different ways at different parts of the distribution. The traditional linear model can be viewed as a summary of all the quantile effects; that is, $\int Q(\tau|x)d\tau = E(Y|x)$.

Under this interpretation, traditional analysis loses information due to its aggregation of possibly disparate quantile effects. Many different quantile paths, for example, can lead to $\beta_k = 0$. On the one hand, $\beta_k = 0$ can mean x_k does not matter – does not affect the distribution of Y. But it can also mean there are important but compensating quantile effects relating Y and x.

Quantile regression is much better suited to analyzing questions involving changes in the distribution of a dependent variable. Quantile regressions allow for separate effects of an explanatory variable on different points of the dependent variable distribution. The coefficient estimates are then frequently interpreted as being analogous to standard linear regression estimates, albeit for different points in the distribution of the dependent variable (Trzpiot 2008, 2009 a, b, c, 2010, 2011 a, b).

It is less commonly recognized that quantile regression can produce estimates of changes in the full distribution of the dependent variable when the values of the explanatory variables change. The set of coefficients produced for independent variables imply a change in the full distribution dependent variables.

Special issues do not necessarily arise when estimating quantile regressions using spatial data. Several researchers have proposed variants of the spatial autoregressive (AR) model, $y = \rho WY + X\beta + u$, for quantile analysis.

These procedures treat ***WY*** as just another endogenous explanatory variable. The spatial AR model may not necessarily be the best choice for spatial modeling, particularly for large data sets comprising individual geographic points rather than large zones or tracts. In situations where the distribution of the dependent variable changes smoothly over space, a nonparametric procedure may be a much better approach.

2. Distribution of the Dependent Variable

In general, the conditional quantile function for y given a set of variables X can be written:

$$Q_y(\tau | X) = X\beta(\tau | X), \quad 0 < \tau < 1. \tag{3}$$

Usually, we have limited our attention to a small number of values for the quantile, τ. Focusing on that values provides useful information about the distribution of the dependent variable given values of *X*, but it certainly does not provide a complete picture of the full distribution of *y*.

One way to use quantile regression estimates to simulate the distribution of the dependent variable is to draw randomly from possible values of and then estimate a separate quantile regression for each value of τ. For example, we might draw 1000 values of τ from a uniform distribution ranging from 0 to 1, i.e., $\tau \sim U(0, 1)$. If we let J represent the number of draws from the U(0,1) distribution, then we have:

$$\hat{Q}_y(\tau_j \mid X) = X\hat{\beta}(\tau_j \mid X) \quad j = 1,\ldots, J \tag{4}$$

With J estimates of the conditional quantile in hand, a standard kernel density function can be applied to $X\hat{\beta}(\tau_j \mid X)$ to estimate the density function for the dependent variable. Since quantile estimates are generally fairly smooth across, drawing multiple values of τ from a U(0,1) distribution is a very inefficient way of constructing the density function. Using a limited range of value for τ is more efficient. For example, we might restrict the estimates to τ, , or a still more limited set of values for that provides good coverage of the set of permissible values for τ. Since quantile estimates are likely to have very high variances at extreme values of τ such as 0,01 or 0,99, it generally is a good idea to trim the extreme observations if a grid of values is used for τ.

3. The Effect of a Discrete Change in an Explanatory Variable

Quantile regression estimates can have interesting implications for the distribution of y values even in a model with a single explanatory variable. Consider a model with k explanatory variable in addition to the intercept. After estimating quantile regressions for J quantiles, the predicted values for quantile are simply τ_j:

$$\hat{Q}_y(\tau_j \mid X) = \hat{\beta}_0(\tau_j) + \hat{\beta}_1(\tau_j)x_1 + \ldots + \hat{\beta}_k(\tau_j)x_k, \quad j = 1,\ldots,J \tag{5}$$

When simplified the notation by replacing $\beta(\tau_j \mid X)$ with $\beta(\tau_j)$, but it should be clear that the estimates depend on the observed values of X. Even in the single-explanatory case where k = 1, the implied effect of changing from to produces J separate values for

$$\hat{Q}_y(\tau_j \mid X, x_1 = \delta_0) = \hat{\beta}_0(\tau_j) + \hat{\beta}_1(\tau_j)\delta_0 + \ldots + \hat{\beta}_k(\tau_j)x_k, \; j = 1,\ldots,J \tag{6}$$

$$\hat{Q}_y(\tau_j \middle| X, x_1 = \delta_0) = \hat{\beta}_0(\tau_j) + \hat{\beta}_1(\tau_j)\delta_1 + ... + \hat{\beta}_k(\tau_j)x_k\,,\; j = 1,\ldots,J \tag{7}$$

With J quantiles and n observations, equation (6) and (7) imply nJ values for the conditional quantile functions. Since $\hat{\beta}_1(\tau_j)$ is not constant, the conditional quantile functions imply a full distribution of values for y even when x_1 is the only variable in the model.

4. Quantile version of the Spatial AR Model

The analysis up to this point has not been explicitly spatial. Although the explanatory variables might include measures of access to various amenities such as a city's central business districts, parks, or lakes, nothing yet is unique to the analysis of spatial data. Several attempts have been made to adapt the standard spatial autoregressive (AR) model for quantile regression.

The spatial AR model adds a weighted average of nearby values of the dependent variable to the list of explanatory variables. The model is written $y = \rho WY + X\beta + u$, where $\boldsymbol{X}$ is the nxk matrix of explanatory variables, $\boldsymbol{Y}$ is the dependent variable, and $\boldsymbol{W}$ is an nxn matrix specifying the spatial relationship between each value of $\boldsymbol{Y}$ and its neighbors.

Suppose the observations represent census tracts. If each tract is contiguous to four other tracts, then W_{ij} = ¼ for each of the four tracts that is contiguous to observation i, and $W_{ij} = 0$ for all other values of j. In this example, each of the n elements of $\boldsymbol{WY}$ is simply the average, for each observation, of the four neighboring values of $\boldsymbol{Y}$. More generally, if observation i is contiguous to other tracts, then $W_{ij} = 1/n_i$ for the tracts that are contiguous to observation i, and $W_{ij} = 0$ otherwise. For point data, $\boldsymbol{WY}$ might form a weighted average of the nearest K neighbors, or the weights might decline with distance. $\boldsymbol{WY}$ is clearly an endogenous variable. Indeed, one interpretation of $\boldsymbol{WY}$ is that it is the set of predicted values from kernel regressions of $\boldsymbol{Y}$ on the set of geographic coordinates. For example, suppose we were to write $y_i = f(lo_i, la_i) + u_i$. If we use a rectangular kernel with a very small window size e.g., the four closest observations – then the cross-validation version of the kernel regression estimator is

$$\hat{y}_i = \frac{1}{n_i}\sum_{j=1}^{n} I_j y_j\,, \tag{8}$$

where I_j indicates that observation j is one of the nearest neighbors to observation i, and indicates the number of observations that are being given weight when constructing the estimate for observation i. Not surprising, adding the predicted value of $\boldsymbol{Y}$ as an explanatory variable for $\boldsymbol{Y}$ often produces highly significant results.

Although the spatial lag variable, $\boldsymbol{WY}$, is formally equivalent to a kernel regression, the approaches could hardly be more different in spirit.

The spatial AR model is based on an assumption that the researcher can truly specify the full spatial relationship between all of the observations. After specifying the entire path by which each of n observations can influence all of the other observations, all that is left is to determine the strength of the relationship by estimating ρ.

In contrast, nonparametric and semi-parametric regressions involve far less difficult. We could easily write the model in semiparametric form as

$$y_i = f(lo_i, la_i) + X_i\beta + u_i \tag{9}$$

or in the conditionally parametric form

$$y_i = X_i\beta(lo_i, la_i) + u_i\,. \tag{10}$$

The spatial AR model is based on the assumption that the researcher can specify a simple parametric function that accounts for both the relationship between $\boldsymbol{X}$ and $\boldsymbol{Y}$ and the entire spatial relationship between all observations. Nonparametric approaches are based on an assumption that the researcher can correctly specify the variables that influence $\boldsymbol{Y}$, but they allow for local variation in the marginal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$. The spatial AR model and its variants may be useful in situations where the objective is to estimate a causal relationship between $\boldsymbol{Y}$ and neighboring values of the dependent variable.

4.1 Quantile Regression with an Endogenous Explanatory Variable

The spatial AR model is most commonly estimated by maximizing the log-likelihood function that is implied under the assumption of normally distributed errors. An alternative approach based on the generalized method of moments method allows the model to be estimated using a variant of two-stage least squares (2SLS). In the first stage, the endogenous variable, $\boldsymbol{WY}$, is regressed on a set of instruments.

The predicted value of ***WY*** is then used as an explanatory variable in the second stage regression. Most researchers use ***X***, ***WX***, and, sometimes, additional orders of the spatial lags such as ***WWX*** as instruments for ***WY***.

Though this method can work well when the goal is to estimate a standard regression, quantile regression may be more complex because instrumental variables are needed for ***WY*** when estimating a regression for each quantile, τ.

Two methods have been used to form the instrumental variables needed these quantile regressions. The simpler version was proposed by Kim and Muller (2004). Their approach is a straightforward extension of 2SLS.

For each value of τ, they first estimate a quantile regression for ***WY*** using the set of instruments (e.g., ***X*** and ***WX***) as explanatory variables. The predicted values from the quantile regression are $\widehat{WY}(\tau)$. In the second stage, they estimate another quantile regression for the same value of τ, this time with ***Y*** as the dependent variable and ***X*** and $\widehat{WY}(\tau)$ as the explanatory variables.

Only 10 quantile regressions are needed to estimate the model for 5 quantiles (e.g., τ = 0,10, 0,25, 0,50, 0,75, 0,90). Zietz et al. (2008) and Liao and Wang (2012) use this approach to estimate quantile versions of the spatial AR model. They use bootstrap procedures to construct standard error estimates.

Though somewhat more complicated, the Chernozhukov and Hansen (2006) approach may be more robust than the Kim and Muller (2004) approach because it does not require that the same quantile be used in both stages of the procedure.

An additional advantage is that Chernozhukov and Hansen present a covariance matrix estimate that is easy to construct.

In the version describe here, the predicted values, $\widehat{WY}$, from an OLS regression of ***WY*** on the instruments are used as the instrumental variable for ***WY***. This instrumental variable is then used as an explanatory variable for a series of quantile regressions $\boldsymbol{Y} - \rho \boldsymbol{WY}$ of on ***X*** and $\widehat{WY}$ The same quantile, τ, is used for each of the regressions, while a grid of alternative values is used for ρ. The estimated value of ρ is the value that produces the coefficient on $\widehat{WY}$ that is closest to zero. After finding $\hat{\rho}$ the estimated values of β are calculated by a quantile regression of $\boldsymbol{Y} - \hat{\rho} \boldsymbol{WY}$ on X. The motivation behind this estimator is a property of two-stage least squares: when instruments are chosen optimally, the coefficient on $\widehat{WY}$ will be zero when both the actual variable, ***WY***, and the instrumental variable are included in a regression.

Standard error estimates are easy to construct for the Chernozhukov and Hansen method. Let e represent the residuals from the quantile regression of $\boldsymbol{Y} - \hat{\rho} \boldsymbol{WY}$ on ***X***, and define $I(|e_i| < h)/(2h)$, where h is a constant bandwidth.

Define $\Phi_i = f_i \overset{\wedge}{WY}_i$ and $Z_i = f_i X_i$. Then the covariance matrix for $\hat{\theta} = (\hat{\rho}, \hat{\beta})$ is:

$$V(\hat{\theta}) = J(\tau)^{-1} S(\tau) J(\tau)^{-1'} \tag{11}$$

where

$$J(\tau) = \begin{bmatrix} \Phi' WY & \Phi' X \\ Z' WY & Z' X \end{bmatrix} \quad \text{and} \quad S(\tau) = \tau(1-\tau)\begin{bmatrix} \overset{\wedge}{WY}' WY & \overset{\wedge}{WY}' X \\ X' WY & X' X \end{bmatrix}$$

As is the case for any instrumental variables (IV) estimator, the estimates from either approach can be sensitive to the choice of instruments. However, an important advantage of the IV approach over maximum likelihood estimation, which is commonly used for the non-quantile version of the spatial AR model, is that that there is no need to invert the nxn matrix $(I - \rho W)^{-1}$ when estimating the model.

It may still prove necessary to invert large matrices when constructing predicted values for Y. Let $\hat{Y}(\tau)$ denote the set of predicted values of the dependent variable for quantile τ.

Three procedures are often used to construct $\hat{Y}(\tau)$:

quantile version of the structural model

$$\hat{Y}(\tau) = \hat{\rho}(\tau) WY + X\hat{\beta}(\tau) \tag{12}$$

quantile version of the reduced form

$$\hat{Y}(\tau) = (I - \hat{\rho}(\tau) W)^{-1} X\hat{\beta}(\tau) \tag{13}$$

quantile version of a decomposition into “trend” and“signal” components

$$\hat{Y}(\tau) = X\hat{\beta}(\tau) + \hat{\rho}(\tau) W (I - \hat{\rho}(\tau) W)^{-1} X\hat{\beta}(\tau) \tag{14}$$

Though the first procedure may be viewed as cheating because it uses actual values of ***WY*** to predict ***Y***, it is commonly used for standard linear simultaneous equations models.

The second version follows directly from the original model specification: the equation implies $Y = (I - \rho W)^{-1} (X\beta + u)$, from which equation (13) follows by setting u = 0. Finally, equation (14) is derived by noting that equation (13) also provides a way to estimate ***WY*** for the expression given in equation (12):

$$\overset{\wedge}{WY}(\tau) = W(I - \hat{\rho}(\tau) W)^{-1} X\hat{\beta}(\tau) \tag{15}$$

5 Conditionally Parametric Quantile Regression

Nonparametric approaches can be adapted to quantile regression models. In the case of a single explanatory variable, x, all that is necessary to make the model nonparametric is to add a kernel weight function $K((x\text{-}x_t)/h)$ when estimating a quantile regression for a target point x_t. After estimating the function for a series of target points, the estimates can then be interpolated to all values of x. The nonparametric approach is a flexible way to add nonlinearity to the estimated quantile regressions. Although nonparametric approaches can potentially be applied to variables with many explanatory variables, the variance of the estimated can become very high when there are more than two or three explanatory variables.

The problem can be simplified significantly in the case of spatial data sets. It is usually reasonable to assume that a simple linear model fits the data well in small geographic areas. The difficulty is in specifying a global parametric model that fits the data well across a large region. The spatial AR model allows for local variation around a global parametric trend. In contrast, a conditionally parametric approach allows for flexible trends by allowing the coefficients to vary smoothly over space. Nonparametric models can be hard to summarize because they produce separate estimates for every observation. As we have seen, this is not a problem for quantile regression models because the easiest way to interpret the results is to present sets of density estimates showing how the distribution of the dependent variable changes as the values of individual explanatory variable change. As a result, it is just as easy to estimate nonparametric quantile regressions as it is to use a linear approach.

5.1 CPAR Quantile Regression for Spatial Data

Consider the conditional quantile function $Q_y(\tau|X) = X\beta(\tau|X)$, in which the dependent variable is a linear function of a set of explanatory variables, X. Now suppose that we want to allow the coefficients to vary over space. Using *lo* and *la* to represent the geographic coordinates, we can write the conditional quantile function as:

$$Q_y(\tau|X, lo, la) = X\beta(\tau|X, lo, la) \tag{16}$$

Although it is possible to include *lo* and *la* as explanatory variables – in which case they are part of the X matrix – the more common approach is to keep them separate. Also, note that *lo* and *la* can represent any geographic coordinate system rather than just longitude and latitude.

All that is necessary to estimate a nonparametric version of equation (16) is to specify a kernel weight function that indicates the weight given an observation with coordinates (*lo*, *la*) when estimate the function at a target point (lo_t, la_t). One approach is to use a simple product kernel:

$$K\left(\left(lo - lo_t\right)/ h_1, \left(la - la_t\right)/ h_2\right) \tag{17}$$

A more commonly used alternative is to make the weights depend simply on the straight-line distance between each observation and the target point, d_t:

$$K(d_t / h) \tag{18}$$

The kernel weight function in equation (19) draws a circle around the target point to form the weights. Although, equation 18 is slightly more general, there is little difference between the two in practice. With *J* different quantiles, the set of estimated coefficients for explanatory variable *k*, $\hat{\beta}_k$, is an nxJ matrix. With K explanatory variables in addition to the intercept, the *n*x*J* matrix of quantile predictions is

$$\hat{y} = \hat{\beta}_0 + \sum_{k=1}^{K} x_k{}' \hat{\beta}_k \tag{20}$$

These predictions can be used to calculate density functions for predicted values of the dependent variable for arbitrary values of the explanatory variables. Suppose we want to evaluate the model at $x_1 = \delta$. Then the *n*x*J* set of predicted values is simply

$$\hat{y} = \hat{\beta}_0 + \delta x_1 + \sum_{k=2}^{K} x_k{}' \hat{\beta}_k \ . \tag{21}$$

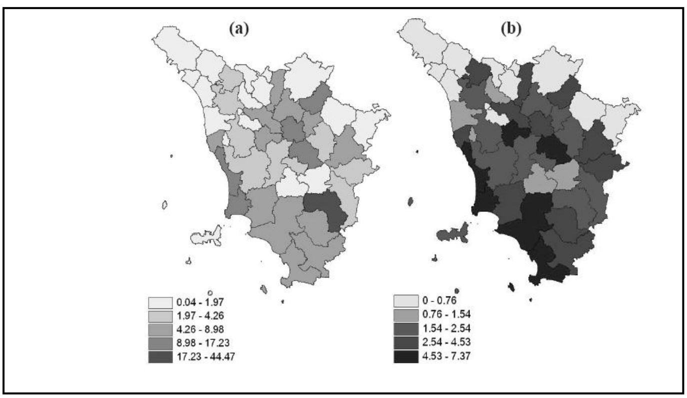
The calculations can be repeated for other values of δ and for other explanatory variables. The results can then be summarized using estimated kernel density functions.

Model is analogous to conditionally parametric (CPAR) local linear regression. The estimation procedure involves estimating separate quantile regressions for various target points, with more weight placed on observations that are close to the target. Unlike a fully nonparametric approach, the CPAR approach produces coefficient estimates for the explanatory variables. But unlike the spatial AR version of quantile regression, the estimated coefficients vary over space. The CPAR approach is less sensitive to model misspecification than the fully parametric spatial AR approach, and it accounts for local variation in an overall spatial trend. The approach is well suited for quantile analysis in situations where the distribution of the dependent variable is, for example, highly skewed in some locations, tightly clustered in others, while all the time varying smoothly over space. Moreover, the CPAR approach does not require the specification of a large (n *x* n) spatial weight matrix, making it amenable to large data sets.

We present results (Chambers, Pratesi, Salvati, Tzavidis 2005) obtained for the estimation spatial distribution of the mean and median production of olives per farm LES. The data are from Farm Structure Survey (2003). ***Z*** the incidence matrix of dimensions 2508 farm per 42 LESs. The neighborhood structure **W is** defined as follows: spatial weight w_{ij}, is 1 if area shares an edge with *j* and 0 otherwise.

The median map is intensive to the presence a few big farms that raise the medium level of production as a consequence the spatial distribution of the median is more homogenous.

Figure 1 a) Mean b) median production of olives



Source: Chambers R. & all (2005).

6. Conclusions

In summary we can say that the classic paper for quantile regression is Koenker and Bassett (1978). Koenker (2005) presents an extensive examination of the econometric theory related to a wide variety of quantile models. Buschinsky (1998) helped popularize the use of quantile regression analysis on the distribution of wages. The spatial AR version of the quantile model relies on approaches developed by Chernozhukov and Hansen (2006) and Kim and Muller (2004). The approaches have been applied to studies of house prices by Kostov (2009), Liao and Wang (2012) and Zeitz et al. (2008). The studies rely on the IV approach for estimating the spatial AR model. Nonparametric versions of quantile models relies heavily on Koenker work. Splines are also a potential alternative to kernel smoothing; it was done in Koenker and Mizera (2004). The use of nonparametric methods for spatial models has been forced by the invention of new terms by geographers for procedures that have already been used extensively in statistics and economics.

References

Buchinsky M. (1998), Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research, Journal of Human Resources, 33

Chambers R., Pratesi M., Salvati N., Tzavidis N. (2005), Small Area Estimation: spatial information and M- quantile regression to estimate the average production of olive per farm, Wp 279, Dipatimento di Statistica e Matematica Applicata all'Economia, Universita di Pisa

Chernozhukov V., Hansen Ch.. (2006), Instrumental Quantile Regression Inference for Structural and Treatment Effect Models, Journal of Econometrics, 132

Cleveland, W. S. (1994), Coplots, Nonparametric Regression, and Conditionally Parametric Fits. In T.W. Anderson, K.T. Fant, and I. Olkin (Eds.), Multivariate Analysis and its Applications, Hayward: Institute of Mathematical Statistics.

Kim T.-H., Muller Ch.. (2004), Two-Stage Quantile Regression when the First Stage is Based on Quantile Regression, Econometrics Journal, 7

Koenker R., Basset B., (1978), Regression Quantiles, Econometrica, Vol 46

Koenker R., Ng P., (2005), Inequality Constrained Quantile Regression, The Indian Journal of Statistics, Vol. 67

Koenker R., Hallock K. F. (2001), Quantile Regression, Journal of Economic Perspectives, 15

Koenker R., Mizera I. (2004). Penalized Triograms: Total Variation Regularization for Bivariate Smoothing, Journal of the Royal Statistical Society: Series B, 66

Kostov, P. (2009), A Spatial Quantile Regression Hedonic Model of Agricultural Land Prices, Spatial Economic Analysis, 4

Liao, Wen-Chi and Xizhu Wang (2012), "Hedonic House Prices and Spatial Quantile Regression," Journal of Housing Economics, 21, 16-27

Mcmillan. D. P. (2010), Issues in Spatial Data Analysis, Journal of Regional Science, 119-141.

Trzpiot G., (2008), The Implementation of Quantile Regression Methodology in VaR Estimation "Studies and Researches of Faculty of Economics and Management University of Szczecin",

Trzpiot G., (2009 a), Quantile Regression Model versus Factor Model Estimation, in: Financial Investments and Insurances", University of Economics in Wrocław, Vol 60

Trzpiot G., (2009 b), Application weighted VaR in capital allocation, Polish Journal of Environmental Studies, Olsztyn, Vol 18, 5B

Trzpiot G., (2009 c), Estimation methods for quantile regression, Economics Studies 53, Karol Adamiecki University of Economics in Katowice

Trzpiot G., (2010), Quantile Regression Model of Return Rate Relation – Volatility for Some Warsaw Stock Exchange Indexes, (in Polish), Finances, Financial Markets and Insurances. Capital Market, University of Szczecin, Vol 28, 61-76

Trzpiot G. (2011a). Bayesian Quantile Regression, Studia Ekonomiczne, Zeszyty Naukowe nr 65, Uniwersytet Ekonomiczny w Katowicach, 33-44

Trzpiot G. (2011b). Some tests for quantile regression models, Acta Universitatis Lodziensis, Łódź, Folia Economica, 255, 125-135

Zeitz J., Zietz E. N., Sirmans G. S. (2008) Determinants of House Prices: A Quantile Regression Approach, Journal of Real Estate Finance and Economics, 37

Streszczenie

PRZESTRZENNA REGRESJA KWANTYLOWA

W wielu zastosowaniach, podstawowym problemem jest opis i analiza wpływu wektora skorelowanych zmiennych objaśniających X na zmienna objaśnianą Y. W przypadku, gdy obserwacje badanych zmiennych są dodatkowo rozmieszczone przestrzennie, zadanie jest jeszcze trudniejsze, ponieważ mamy dodatkowe zależności, wynikające ze zmienności przestrzennej.

W tej pracy, w miejsce przestrzennej regresji wykorzystującej średnią, rozpatrzymy przestrzenna regresję kwantylową. Regresja kwantylowa zostanie omówiona w przestrzennym kontekście. Głównym celem pracy jest wskazanie na

możliwości powiązania metodologii regresji kwantylowej i ekonometrycznego modelowania przestrzennego. *Dodatkowe zasoby informacji o zmienności otrzymujemy badając kwantyle, wychodząc poza tradycyjny opis klasycznej regresji. Estymacja kwantylowa w modelu przestrzennym uwydatnia zależności przestrzenne dla różnych fragmentów rozważanych rozkładów.*

10.2478/v10103-012-0041-7

WIESŁAW WAŃKOWICZ*

Land Management of the Areas of High Landscape Values: An Economic Model

Abstract

This paper presents selected results of the research entitled Planning the Space of High Landscape Values, Using Digital Land Analysis, with Economic Appraisal, supervised by Dr. Paweł Ozimek, Cracow Technical University, conducted since 2009. Usually, we do not pay attention to surrounding landscapes in our everyday life. However, for the persons who deal with spatial planning, geography, natural environment, or cultural heritage, the validity and value of landscape are the terms which do not have to be defined. The first part of the paper is dedicated to the landscape features that decide about its value. The author discusses whether those features are the same as those we want to protect and how we can appraise landscape values. The next part contains an analysis of the economic bases of development. In reference to space, the analysis and opinion on land use in the context of the development of usable functions are essential. Consequently, the identification of the limitations connected with the protection of landscape and delimitation of the areas on which such limitations exist are required. Another component consists in the determination of the land requirements associated with existential and economic needs of the local population. Such a general balance of needs and requirements is the starting point of the adoption of development policies and action programmes. The programmes should include the location of individual projects and capital investments on land, as well as their proper timing co-ordination. Owing to the complexity of the tasks, the option analysis is the preferred method of search for the best possible solution. The reconciliation of individual land use (title to land), public and business land uses, with the protection of environmental and cultural values, can be difficult or next to impossible to

* Ph.D., Institute of Urban Development

attain. Therefore, we need some mechanisms to compensate the losses occurring in individual interests and in local, regional, or national development. The choice of options for local or regional development is based on balancing the costs and benefits that depend on the sizes of both protected and non-protected areas. In conclusion, the author attempts at answering the questions whether the landscape and landscape values can be saved owing to their economic assets, and what instruments should be implemented to utilize economic mechanisms of protection.

1. Introduction

We do not pay much attention to local landscape in our everyday life, especially when the need to satisfy our daily needs and our aspirations associated with the consumer life style are in opposition to the rules of sustainable spatial management. People are active in the contexts of their living conditions. They are aware of such conditions depending on people's needs, interests or objectives. People create their projections not only of their social environment, but also of the shape of that environment. Landscape is simply defined as the appearance of land surface in a specific location, or the view of our surroundings. Although landscape is a common commodity, it may not be "good" for somebody, and people may have various opinions about landscape. For those who are involved in spatial planning, geography, natural environment or cultural heritage, landscape importance and value are the premises that do not have to be defined.

2. Landscape and its Characteristics

Landscape may have different values for various people or groups: property owners, land users, or public authorities. Space is appraised differently on a local, national or international scale. Space becomes landscape in the observer's eyes, and it can become a source of aesthetic experiences associated with the attractiveness of a given place or view opening to distant areas. We should, therefore, turn to man (Jałowiecki and Szczepański 2006, pp. 333-335) to try to discover how one perceives, values or adapts space. Based on displayed human imagination and verbalized needs, people can shape space. In the research dedicated to the perception and evaluation, it is important not only to obtain answers to the question of how a given person (persons or communities) perceives and appraises space, but also why the method of perception and evaluation is so particular. Thus, we intend to discover the variables that determine such processes. The perception of space is equivalent to informed reception of stimuli. The patterns of space image shaping depend on the following:

- perception of space in whole or in part,
- separate perception of objects and people,
- observer's experience, his or her skills, and the rules of appraisal (which are different for land users, land owners, or developers: designers, urban planners, architects, or decision-makers).

Various perceptions of space makes it either easy or difficult to communicate on the matters of appraisal of both existing status or vision of the future. Landscape perceived in a particular manner indicates that we evaluate space when appraising landscape. When shaping landscape, we actually shape space.

How can we practically resolve the problem of the space appraisal during spatial planning and project design or implementation (in case on new developments, replacements, remodelling etc.)? To appraise space, we can apply the indicators which allow us to measure space for the following purposes:

- problem description (appraisal of landscape values in our case),
- project need evaluation (an intervention when landscape is negatively appraised),
- definition of the purpose of action (indication of success measures) and
- evaluation of the degree of success (have the landscape values been improved owing to intervention?).

Referring to our assumption that narrowly understood landscape is the object of our analysis, we have applied an approach based on aesthetic and appearance values for the purpose of landscape evaluation (including space evaluation as a result) (a broader description can be found in: Bajerowski, 2007, pp. 7-29). The features that allow us for such landscape evaluation include beauty, sublimity, and curiosity (aesthetically-oriented landscape in: Böhm, 2006, pp. 291-293). As to the primary (original) landscape value, the proposed method is subjective in nature: the first step in managing space of high landscape values belongs to experts. The appraisal will be an average evaluation made by a competent group. The competent group, selected for the purpose of analysing specific land, will indicate all attractiveness factors, such as volumetric objects (buildings, structures, and landscape units which decide about the landscape value). Those objects (o_i) should be assigned to two separate groups, with negative and positive impact on the analysed landscape, respectively. Each object should be evaluated on the basis of selected criteria (see above, c_{ji} – value of object *i* for the criterion *j*, w_j – weight of the criterion *j*) and assigned aggregated total value (LVo_i). The issue of the evaluation scale, standardization of variables or aggregation rules are omitted here as they are not associated with the main topic of the paper.

$$LVo_i = f(c_{ji}, w_j) \tag{1}$$

Nominal landscape value (LVA) of the analysed area (A) will be the following (the objects' values referring to negative influences on the landscape will reduce the total value):

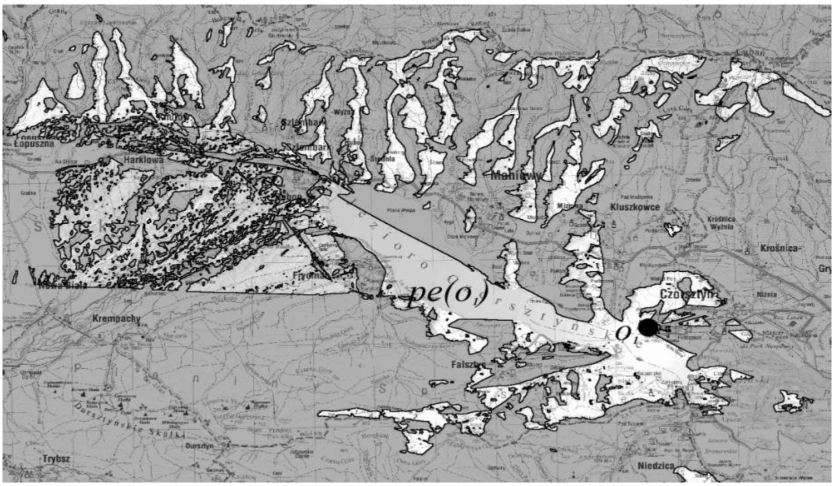
$$LVA = \sum_i LVo_i \tag{2}$$

3. Space and Landscape as the Elements of Development Processes

The value of the preserved (restored and conserved) natural and cultural environment, although associated with a possibility of useful and effective utilization of selected space cab be a measure of the business value of landscape (Wańkowicz 2010, pp. 352-359). The general balance of the needs and possibilities should be a starting point for the determination of the development policy and later for drafting of spatial management plans. In particular, we should aim at balancing the land needs in respect of the indication of land use (specific manner of development, e.g. land for house building, or businesses) of the selected areas, taking into account the protection of those components whose use and development should be subjected to special needs, owing to the features of the natural and cultural environment. That would allow for the implementation of the principles of sustainable development: on the one hand, preservation of selected resource, and, on the other hand, a possibility of the development of local communities.

What is an essential element in reference to landscape is the analysis and evaluation of the usefulness of particular plots of land for the development of various useful functions which will allow us to maintain and utilize landscape values. Assuming that the landscape use is associated with a possibility of providing aesthetic experiences, the landscape value increases with the increase of the possibility of providing such experiences to observers, just like in case of masterpieces (landscape has real value only when one can see it). The areas which are affected by the objects mentioned before are the lands on which the objects are passively exposed (or the places where we can see such specific objects). To put it simply, the larger the area from which an object can be seen the higher landscape value. Omitting at this stage the dependence of the object exposition on other factors (such as the size of the object which can limit the distance of exposition), or its form, colour etc., we can determine the passive exposition area for a given object, using a digital land model and GIS type of software. An example is shown in Fig. 1. The exposition area was determined by the team of Dr. Paweł Ozimek (Cracow Technical University).

Figure 1. Passive exposition area ($pe(o_1)$) of the Czorsztyn Castle (o_1)



Source: P. Ozimek (Cracow Technical University).

If we take into account the exposition, the modified value (ELV) of the object (o_i) will reflect the size (surface area) of object exposition ($pe(o_i)$), e.g. in respect of the analysed surface area (A).

$$ELVo_i = f\left(c_{ji}, w_{jj}\right) \times \frac{pe(o_i)}{A} \tag{3}$$

And the landscape value ($ELVA$) of the analysed area (A) will be as follows (the negative object values affecting the landscape will reduce the total value):

$$ELVA = \sum_i ELVo_i \tag{4}$$

Depending on the nominal value (LVo_i) of the objects recognized to be important for the landscape, we can determine their scope of protection on the exposition area. Consequently, the management and use of both object and its exposition area will be subjected to restrictions. Using the terms from the field of landmark preservation, the actions intended to protect and preserve the landmark substance and stopping of destruction processes (conservation), as well as the actions intended to display the artistic and aesthetic values of the object, including supplementing or recreation of parts of the object (restoration), if necessary, seem to deny the possibility of business use of the objects and protected space at first sight. If, however, landscape care consists e.g. in the use of space in the manner that ensures durable preservation of landscape's values, the business use of objects and protected space seems to be possible provided that we can properly determine which features decide about the object's value and which ones we want to preserve or recreate. The essence of planning of the space representing high landscape values consists in finding balance between protection and availability.

4. Balance of Needs and Possibilities

Another stage of analysis of a selected area includes development programming, taking into account protection measures and the use of landscape potential. This analysis requires taking into account actual land needs that are associated with the operation of the local community, both at the level of standard solutions and with inclusion of community aspirations. That concerns primarily the areas designated for housing projects, the services meant for the local community and for the business activities which are the basic sources of income. Consequently, it is necessary to identify, within the analysed area, the land designed for house building (M), services (U), and various types of businesses (G). However:

$$A \geq M + U + G \tag{5}$$

Based on the above considerations, the analysed area is divided into three subdivisions as follows:

- The area subjected to complete strict protection (CP), excluding a possibility of land use, except for the purposes of protection (quoting the Polish forms of nature and heritage protection measures, we can create: strict reservation, archaeological protection area, landmark, or the highest-class landmark).
- The area covered by partial protection (PP) allowing for various although limited types of uses (e.g. landscape park, or cultural park).
- Unprotected area (NP) subjected only to the principles of land development and building which determine not only possible land uses (designation for

farming or housing purposes), but also the types of structures (size, floor area etc.).

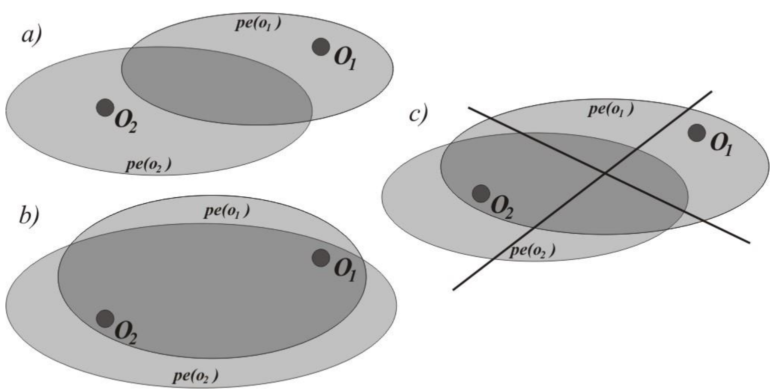
The areas determined in that way should fulfil the following condition:

$$A = CP + PP + NP \tag{6}$$

$$PP + NP \geq M + U + G \tag{7}$$

To determine fully protected areas, partially protected areas, and unprotected areas, we can use the previously indicated objects, which may either positively or negatively affect the landscape, together with the designated passive exposition areas. At this point, we should point out the dependence between passive exposition (*pe*) and active exposition (*ae*), as well as the exposition area of several objects. Those issues are illustrated in Fig. 2; the case (c) does not occur.

Figure 2. Examples of the relationships between exposition areas (both passive and active exposition areas) of two objects; case (c) does not occur.



Source: P. Ozimek (Cracow Technical University).

Without going to the depth of our considerations owing to symmetric and non-transition aspects, the following relationships occur:

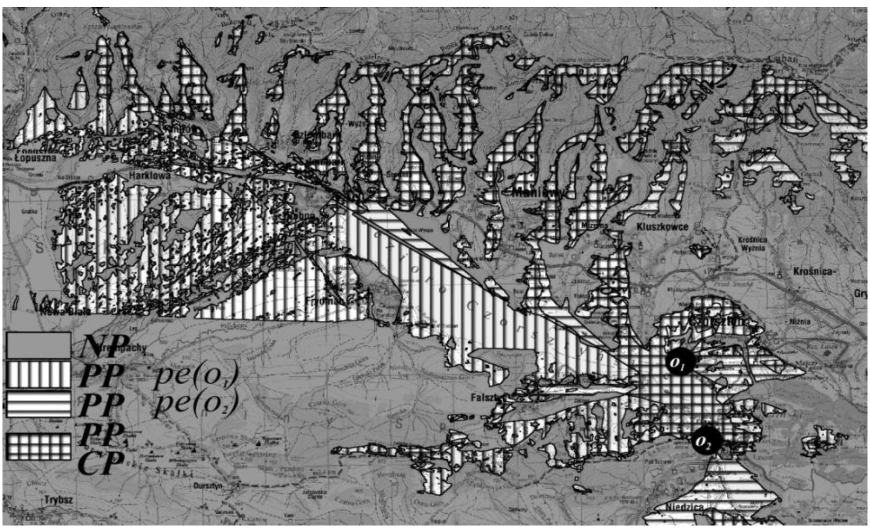
$$pe(o_i) = ae(o_i) \tag{8}$$

$$pe(o_1 i o_2) = pe(o_1) \cap pe(o_2) \tag{9}$$

$$pe(o_1 \,\mathrm{lub}\, o_2) = pe(o_1) \cup pe(o_2) \tag{10}$$

An analysis leading to the determination of fully protected, partially protected, and unprotected areas can be carried out with the use of a digital land model and GIS type of software. An example is shown in Fig. 3. The exposition area was determined by the team of Dr. Paweł Ozimek (Cracow Technical University). That example covers only the objects that positively affect the landscape. The lands which are not exposition areas for any object that positively affects landscape or object areas are determined to be unprotected lands. The lands subjected to partial protection are the passive exposition areas of all the objects which positively affect the landscape. The lands subjected to full protection are the areas of the objects which positively affect the landscape and the selected portions of the areas of concurrent passive exposition of several objects (two in our example).

Figure 3. Determination of full protection (*CP*), partial protection (*PP*), and unprotected areas (*NP*)



Source: P. Ozimek (Cracow Technical University).

This selection requires deeper analysis associated with precise determination of the scopes and objects of protection. Accounting of the objects which negatively influence the landscape will require a similar approach. However, an analysis is intended to limit active exposition to the objects, especially when such negative exposition exists in the background or foreground of the objects which positively affect the landscape.

5. Development Programming

The general balance of land needs, possibilities, and restrictions becomes a starting point for the determination of development policy and action programmes. The spatial aspect of development is an element of development policy. It entails the distribution of particular components of programmes (projects and capital investment) in space, with proper timing coordination (including drafting of Local Physical Plans). Owing to the complexity of the tasks, option analysis is the preferred method of seeking the best possible solutions.

The value of space can be identified with its ability to satisfy specific needs, including primary needs (place of residence, or production of food for one's needs), and higher-order needs (appearance). The capability of space to satisfy so determined specific needs refers straight to the space owner or user. Space can also be understood as the capability of the generation of goods that can be exchanged for other goods, or the capability of generation of income and of participation in the cash and commodity exchange. That is associated on the one hand with bearing costs, and, on the other hand, with possible benefits. Drafting of a Local Physical Plan, a development policy, and action programmes for the analysed area will provide model predictions of costs and benefits. That will lead directly to the formulation of a balance equation. The following are the variables that can be calculated in that way: surface areas covered by full protection (CP), partially protected surface areas (PP), and unprotected surface areas (NP). The balance equation has the following general form:

$$C_{CP} + C_{PP,p} = B_{CP} + B_{PP} + B_{NP} - C_{PP,np} - C_{NP} \qquad (11)$$

where:

Costs of protection of fully protected areas: $C_{CP} = f(CP)$

Costs of protection of partially protected areas: $C_{CP,p} = f(PP)$

Costs of operation of partially protected areas: $C_{CP,np} = f(PP)$

Costs of operation of unprotected areas: $C_{NP} = f(NP)$

Benefits of fully protected areas: $B_{CP} = f(CP)$

Benefits of partially protected areas: $B_{PP} = f(PP)$

Benefits of unprotected areas: $B_{NP} = f(NP)$

When the left-hand part of the equation is smaller or equal to the right-hand part, we are dealing with an economically beneficial situation: we can afford the planned arrangements and protective actions. Otherwise, it will be necessary to think over the assumed spatial solutions: change the balance of the protected surface area within the total area subjected to analysis (in accordance with the principle that when we want to protect everything, we will protect nothing). From the viewpoint of the landscape values, space, landscape value protection, and space value constitute public objectives. Of course, in this study, we consider only the space with high landscape values, not all space. Consequently, the balance equation may assume a different form:

$$C_{CP} + C_{PP,p} = B_{CP} + B_{PP} + B_{NP} - C_{PP,np} - C_{NP} + S_p \quad (12)$$

where: S_p - Expected subsidy for selected value protection.

The same formula can be applied on a local, regional or national scale. It is important, however, that the reconciliation of individual land use (title to land), public and business land uses, with the protection of environmental and cultural values, can be difficult or next to impossible to attain. Therefore, we need some mechanisms to compensate the losses, e.g. by subsidies granted to individual interests (local scale), municipalities (regional policy), or regions under a national policy. Equation (11) will be proper for the national scale.

6. Conclusions

In reality, the border on which valuable space, with the value of a public wealth, ends and the space without such a feature starts is rather fuzzy (here, public space is different than in case of freely accessible space). Landscape protection is associated on the one hand with the limitation of the title to land, and in particular with the limitation of land uses, and, on the other hand, with the landscape having significant influence on the property value. The influence can be either positive or negative when we take into account satisfaction of individual and collective needs, or when we consider the use of a property only by its owner or with the purpose of income generation and participation in commodity and cash exchange. It is essential to make a calculation which will demonstrate the interdependence of benefits and losses (lack of benefits) associated with landscape protection. However, it is not only an economic issue, but also a social and political one, because it is associated with such matters as legal order or social justice.

A correctly designed space management system, also involving the space representing high landscape values, should take into account economic calculations on micro and macro scales. On a macro scale, the calculation should produce a positive or break-even balance (it can also be temporarily negative), both in public and private sectors. On a micro scale, the calculation should produce a positive or break-even balance (it can also be temporarily negative) in the private sector, or it can be negative in the public sector. That is associated with the possibility of participation in costs of the entities which do not receive direct benefits (redistribution of resources from other areas). The balance will allow to indicate which portion of space (landscape) can and should be protected and to what extent. On a macro scale, the balance can help to prepare development policies understood not only as action plans, but also as tools for the creation, protection, and revision of general principles and standards of life, including the principles of redistribution of costs and benefits (taxes and finances available from the public budget, based on: Heywood A., 2008, pp. 3-27).

Still, the questions: Will landscape survive owing to its business value? and What instruments should be implemented to use the economic mechanisms for landscape protection? remain opened.

References

Bajerowski T. (red.), Biłozor A., Cieślak I., Senetra A., Szczepańska A., 2007, *Ocena i wycena krajobrazu*, Educaterra, Olsztyn

Böhm A., 2006, *Planowanie przestrzenne dla architektów krajobrazu*, Politechnika Krakowska, Kraków

Fiedor B. (red.), Czaja S., Graczyk Z., Jakubczyk Z., 2002, *Podstawy ekonomii środowiska i zasobów naturalnych*, Wydawnictwo C.H. Beck, Warszawa.

Fujita M., Krugman P., Venables A.J., *The Spatial Economy*, The MIT Press, Cambridge, Massachusetts

Heywood A., 2008, *Politologia*, PWN, Warszawa

Jałowiecki B., Szczepański M.S., 2006, *Miasto i przestrzeń w perspektywie socjologicznej*, Scholar, Warszawa

Ponsard C. (red.), 1992 (wydanie polskie), *Ekonomiczna analiza przestrzenna*, Wydawnictwo Akademii Ekonomicznej, Poznań

Stiglitz J. E., 2004, *Ekonomia sektora publicznego*, PWN, Warszawa

Ustawa z dnia 16 kwietnia 2004 r. *o ochronie przyrody*, Dz. U. nr 92, poz. 880 z późn. zm

Ustawa z dnia 23 lipca 2003 r. *o ochronie zabytków i opiece nad zabytkami,* Dz. U. nr 162, poz. 1568 z późn. zm

Ustawa z dnia 27 marca 2003 r. *o planowaniu i zagospodarowaniu przestrzennym*, Dz. U. nr 80, poz. 717 z późn. zm

Wańkowicz W., 2010, *Planowanie przestrzeni o wysokich walorach krajobrazowych, problemy ekonomiczne*, *Krajobraz a turystyka*, Prace Komisji Krajobrazu Kulturowego Polskiego Towarzystwa Geograficznego nr 14, Sosnowiec

Streszczenie

GOSPODAROWANIE PRZESTRZENIĄ O WYSOKICH WALORACH KRAJOBRAZOWYCH. MODEL EKONOMICZNY

Artykuł przedstawia wybrane rezultaty pracy Planowanie przestrzeni o wysokich walorach krajobrazowych przy użyciu cyfrowych analiz terenu wraz z ocena ekonomiczną (kierownik dr Paweł Ozimek, Politechnika Krakowska). W codziennym życiu zwykle na krajobraz nie zwracamy uwagi. Ale dla osób zajmujących się planowaniem przestrzennym, geografią, środowiskiem naturalnym i dobrami kultury ważność i wartość krajobrazu jest tezą nie wymagającą dowodu. Pierwszą część artykułu poświęcono zagadnieniu cech krajobrazu, które decydują o jego wartości. Także – czy są to te cechy krajobrazu, które chcemy chronić oraz jak możemy ocenić wartość krajobrazu. Kolejna część to analiza gospodarczych podstaw rozwoju. W odniesieniu do przestrzeni istotnym elementem jest ocena przydatności poszczególnych terenów dla rozwoju funkcji użytkowych. W konsekwencji określenie ograniczeń wynikających z ochrony krajobrazu oraz wskazanie obszarów, na których ograniczenia te występują. Drugim elementem jest określenie potrzeb terenowych związanych z bytowymi i ekonomicznymi potrzebami ludności. Tak opracowany ogólny bilans potrzeb i możliwości służy za punkt wyjścia do określenia polityki rozwoju i programów działania. Programy winny zawierać lokalizację przedsięwzięć i inwestycji w przestrzeni oraz odpowiednią ich koordynację w czasie. Ze względu na złożoność preferowaną metodą poszukiwania możliwie najlepszego rozwiązania jest analiza wariantów. Pogodzenie indywidualnego (prawo własności), społecznego i ekonomicznego użytkowania przestrzeni z ochroną jej walorów środowiskowych i kulturowych może być trudne lub wręcz niemożliwe. Konieczne są więc mechanizmy rekompensujące straty zarówno w sprawach indywidualnych, jak i w kontekście rozwoju lokalnego, regionalnego czy krajowego. Wybór wariantu rozwoju opiera się na bilansowaniu kosztów i korzyści zależnych od wielkości chronionego i niechronionego obszaru.

Podsumowaniem artykułu jest próba odpowiedzi na pytanie, czy krajobraz i walory krajobrazu ocaleją dzięki swojej ekonomicznej wartości oraz jakie instrumenty winno się wdrożyć w celu wykorzystania mechanizmów ekonomicznych do jego ochrony.

10.2478/v10103-012-0042-6

WŁADYSŁAW WELFE*

Multicountry and Regional Macroeconometric Models

Abstract

Multicountry models were developed in the previous century to serve the analyses and projections of the world economy and/or its regions (for instance Latin America). They distinguish the largest countries and the rest of the world (ROW) composed of particular countries. Hence, their structure is based on the specifications of equations for individual countries using full statistical information available at the countries level.

The regional macroeconomic models are built for either administrative or geographical units distinguished within large countries (USA, China, Russia). Their structure should be in principle similar to those of the national economy. However, the statistical information of the regional economies is typically uncomplete: no sufficient information is available on exports and imports of the region, migrations and financial flows. Appropriate approximations are necessary. As in majority of countries the prices and wages movements are in general unified over the country, the variables representing the national level excerpt an impact on the regional variables. The regional impact on the national variables is rather unusual.

The paper shows the skeleton model applied in the multicountry models and the skeleton of the macroeconomic regional model. The specific properties of the regional model are discussed and the possibilities of its extensions analysed.

* Ph.D.,Full Professor at the University of Łódź

1. Introduction

The development of macro econometric national models and regional models has a long tradition. The use of national models in testing the economic theories and in policy simulations and analyses and economic forecasting was over the last 50 years successful. It lead in the 70-ties to a tendency the leading modelling centers to model the surrounding world i.e. to construct multicountry models. It became the domain of international organisations and also public research centers and of national banks. The multicountry models are composed of a sample of national models mostly linked via commodity flows. Their structure developed along the national models being their components. We shall briefly comment on this process in the next paragraph[1].

On the other extreme, the tendency evolved to analyse the development of particular regions, especially of large countries as the USA both to support the regional science, the local policies and also local forecasting. Here, we constrain the discussion to the building of regional models in a narrow sense, as of models explaining the relationships within the first subdivision of the country i.e. region like California, Lyon (France) Poznań voirodship (Poland). The generalization to lower units like metropolitan areas will be skipped. If the models cover all regions within a country this system is called a multiregional model.

The structure of the regional model should in principle be similar to the structure of national models, following L.R. Klein (1969) suggestions. It shows up specific properties, however, because of special "regional" characteristics and mainly because the regional data bases are not sufficient enough to construct many equations in a similar way as in the national models. This will be the subject of discussions presented in paragraph 3.

2. The multicountry econometric models.

The origins are to be found in the Project LINK, that was developed in Philadelphia in the early 70-ties by the LINK center, led by L.R. Klein. The Project LINK grew enormously from 7 to above 100 countries from 1987. The national models had different structure being elaborated by national units

---

[1] The multicountry models contain mainly national models of the major countries. The models of remaining countries representing either particular regions (groups of countries, like LATIN America) or the rest of the world (ROW), built according to similar principles for the regions as a whole or using the bottom-up principle. Notice, that "region" is understood here as a summary characteristics for countries (nations).

(annual and quarterly). All were interlinked by commodity flows (4 groups). To achieve it a matrix of export shares in international trade was estimated and systematically updated (Hickman 1991). From 1971-72 the model LINK was systematically used in forecasting world economy and in policy scenario analyses. In the last decade it came over to the Toronto University and to the United Nations in New York.

This development was followed by a construction of multicountry models in international organisations and central banks. These models were composed of national models constructed at the particular institutions using the same specifications. They were elaborated for large countries and simplified for small countries. Let us mention the OECD INTERLINK model (Richardson 1988), World Bank models and IMF MULTIMOD models, mainly used is simulation excercisens (Laxton et al. 2004), European Commission QUEST (Roeger, in't Veld 1997).

The development of multicountry models in large countries led to the process of combining the national models with the world models. Let us mentions the FRB/WORLD model, that included the FRB/MCM model for the USA economy (Levin et al. 1999). At the Bundesbank the MEMOD multicounty model, was constructed 2000. At the National Institute of Economic and Social Research (NIESR London) a quarterly world model GEM was constructed in the 80-ties. Its new version NIGEM contained the quarterly model for the UK as well as for the OECD countries and 15 other countries and groups of countries (2005). French model MIMOSA contained also France and world economy and was fairly disaggregated (Delessy et al. 1996).

We also have to mention the enormous FUGI annual model by A. Onisbi that included over 150 000 equations (180 countries), specifying both demand and supply of products (1993). On the other extreme, we have to indicate the model MC of the world economy by R.C. Fair 2004), that included 39 countries. It was used in many interesting policy simulations.

The structure of the above models is not unique and it changed in time following the developments in national model building, like incorporation of rational expectations in several models, emphasis on the microeconometric underpinning of equations and in the last years relying more intensively on economic theory, developed within the Dynamic Stochastic General Equilibrium (DSGE) models (cf. W. Welfe 2010).

Below a skeleton structure of a country model is shown

Accounting identity

$$X_t = C_t + J_t + G_t + \Delta R_t + E_t - M_t \tag{1}$$

Equations explaining final demand

Consumer demand

$$C_t = c(H, V_t r_t), \tag{2}$$

Investment demand

$$J_t = j(X_t, r_t w_t / p_t^s), \tag{3}$$

Inventory increase

$$\Delta R_t = r(X_t wx_t), \tag{4}$$

Imports

$$M_t = m(X_t, wx_t, p_t / e_t p_t^w) \tag{5}$$

Exports

$$E_t = e(WT_t, wx_t, p_t / e_t p_t^w). \tag{6}$$

Consumption depends on expected income H_t, personal assets, V_t and interest rate r_t. Investment depends on GDP (X_t), user costs represented by interest rate r_t and relation of wages (w_t) and capital prices (p_t^s). The inventory increase depends on GDP (X_t) and indicator of market tensions (wx_t). In the foreign trade equations the relative prices and the rate of capacity utilization (wx_t) are present except for the demand variables like GDP and world trade (WT_t) .

Equations explaining supply sector:

Employment

$$N_t = n(X_t, N_{t-1}, A_t), \quad (7)$$

Real wages

$$w_t / p_t = w(u_t, \pi_t, tw_t), \quad (8)$$

Producer prices

$$p_t = p(wx_t, w_t / \pi_t, ep^w), \quad (9)$$

Exchange rate

$$e_t = e(e_t^e, r_t / r_{t,}^w \gamma) \quad (10)$$

Employment is obtained from inverting the production function, hence, it depends on GDP (X_t) and total factor productivity (A_t). Real wages change as a result of negotiations and depend on the unemployment rate (u_t), labour productivity (π_t) and income tax (tw_t). Producer prices depend on labour costs, (w_t / π_t) world prices (p_t^w) and market tensions (wx_t). The exchange rate depends on its expected value, ratio of domestic and foreign interest rates (UIP) and risk premium (γ).

Equations explaining financial sector:

Budget income

$$BP_t = b(tw_t YP_t, ta_t AP_t, tv(CP_t + GP_t)), \quad (11)$$

Budget deficit

$$BD_t = BP_t - BC_t, \quad (12)$$

Money demand

$$MD_t = m(Y_t, \Delta Y_t, P_t, r_t), \quad (13)$$

Interest rate

$$r_t = r(P_t - P_{t-1}) / P_{t-1} \quad (14)$$

The budget revenues are mainly determined by inflows from taxes – direct taxes on personal income (YP_t), on corporate incomes (AP_t) and indirect taxes – the tax base being sales of consumer goods $(CP_t + GP_t)$. The budget deficit is commonly obtained as a residual by subtracting budget expenditures (BC_t) from budget income. Money supply equals money demand being dependent on personal incomes and their increase, prices and interest rates. Interest rates if they are endogenous are determined following the rate of inflation.

The above stylised equations represent long-term relations. In the short-run the adjustments are introduced, either in the from of lags and leads. The expectations are in many models introduced – there are either rational expectations or adaptive expectations.

3. The regional models

The econometric regional model developed mainly in the USA in the 60's and 70's under the influence of national modellers (cf. Bolton 1991) . Later their construction served the needs of regional science and specific tasks of regional governments. Their development was constraint by the lack of regional data in many areas, like commodity and money flows from and to particular regions, migrations etc.

These development were followed by similar research in France, where R. Courbis has built the REGINA model for France (Courbis 1979), for Europe and Japan by B. Issaev et al. (1982) for Canada by R. Bolton (1982), and for Spain: Castilla La Mancha (Montero et al 2008).

The majority of models are econometric models for single region. Multiregional models are seldom built. The single regions models are linked to the national models such that several national variables are introduced into the regional models (being thus exogenous). This is so called "top-down" approach. The reverse links ("bottom-up") are rather-seldom-for instance they may express the role of fuel output exported to the whole country. The links between single regions were poorly emphasised.

Following the research and administrative goals the single region models were substantially disaggregated; many sectors were distinguished. The national models to be linked with them had to show a similar level of sectoral disaggregation – they were large macroeconometric models like Wharton or DRI for the US economy – in contrast to the multicountry models being manly one sectoral.

The structure of the single regional models could not fully follow the Klein's suggestions, mainly because of the scarcity of data as regards the regional accounts. The Gross Regional Product (GRP) and its decomposition similar to (1) was available in annual terms and only rarely presented a time series. No quarterly data was typically available.

Several proxies were used instead. Retail sales (in constant prices) were applied instead of private consumption. Private investment was not always available. The residential building was representing the residential investment expenditures. The regional exports and imports were in general not accessible. The demand for output was represented by either gross output $\left(Q_t^R\right)$ or value added $\left(X_t^R\right)$ decomposed by industries (mainly manufacturing). Similar decomposition was available for employment $\left(N_t^R\right)$, rather than working hours $\left(H_t^R\right)$. The capital stock $\left(K_t^R\right)$ was hardly available and its increase was approximated by investments. That in exceptional cases lead to the construction of potential output using either Cobb-Douglas or CES production functions. Estimates of labour supply became available after allowing for unemployment rates and migration. The producer prices were rather national, whereas the CPI regionally distinguished. The nominal and real wages were in principle observable and hence the calculation of nominal and real personal income $\left(Y^R\right)$feasible. The financial flows were in principle limited to regional budgets.

This concise characteristic of the regional data base makes possible to demonstrate a skeleton structure of a single regional model. It will be partly illustrated by the structure applied for an econometric model of the Łódź – region in Poland[2].

The skeleton econometric for a single region

Final demand:

Real retail sales

$$S_{tj}^R = s\left(Y_t^R, P_{tj}^R / P_t^R\right), \tag{15}$$

Business investment

$$J_{ti}^R = j\left(Q_{ti}^R, J_{ti}, r_t\right), \tag{16}$$

[2] See Florczak et. al. 2008. A model for Mazowsze region was recently announced. The paper by Florczak 2011 is, however, very general and does not contain any information on the models structure.

Residential investment

$$JR_t^R = j\left(Y_t^R, r_t\right), \tag{17}$$

The consumer demand being represented by retail sales depends on regional real disposable income $\left(Y_t^R\right)$ and relative prices of distinguished commodity and services groups. Similar factors affect residential investment. The business investment by industries if available, depends on their output, the national investment and the national interest rate.

Supply sector:

Regional gross output

$$Q_{ti}^R = q\left(Q_{ti}, S_t^R, J_{ti}^R\right), \tag{18}$$

Regional value added

$$X_{ti}^R = x\left(Q_{ti}^R, X_{ti}\right), \tag{19}$$

Potential output

$$\Delta X_{ti}^{RO} = x\left(J_{ti}^R, \Delta N_{ti}^R, \Delta TFP_{ti}^R\right), \tag{20}$$

Employment

$$N_{ti}^R = n\left(X_{ti}^R N_{t-1,i}^R, N_{ti}\right), \tag{21}$$

Real wages

$$w_{ti}^R / pc_t^R = w\left(\mu_t^R, \pi_{ti}^R, w_{ti} / p_t\right), \tag{22}$$

Producer prices

$$p_{ti}^R = p\left(p_{ti}\right), \tag{23}$$

Consumer prices

$$pc_{ti}^R = p\left(pc_{ti}, p_{ti}^R\right). \tag{24}$$

The regional gross output by industries depends on local consumer and investment demand and also on national output. Gross output is transferred into value added (by industries), sometimes using I-O coefficients. The output determines employment with some lags. Potential output can be obtained using production function and allowing for TFP increase, being dependent on R&D

local expenditures. Real wages are affected by local factors: the unemployment rate $\left(\mu_t^R\right)$ and productivity $\left(\pi_{ti}^R\right)$ and are influenced by the nation- wide dynamics of wages. Producer prices do not differ from national-wide prices, whereas consumer prices are affected by national-wide prices and regional producer prices.

The financial flows description is mainly constrained to the local budget receipts and expenditures. The equations do not differ much from the nation-wide shown above.

4. Conclusions

The above review is by no means very simplified. It however, allows to draw the conclusion that as the multicounty models are forrunners in developing the structure of macromodels following the most recent contributions of economic theory, the regional models follow these development with certain lags. Main reason remains the scarcity of data, especially dealing with inter regional commodity flows and also migrations.

Nevertheless a serious progress has been made in construction of econometric regional models. It has to be stressed that there were more or less clear lines drown between relations that have specific, “regional” character and relations that are national wide (wages, prices, interest rates), that can be implemented into regional models.

The development of data bases in the last years is so quick, that it may be expected a further decline of existing gaps in the structures of nation wide and sectoral models, that might bring more common construction of multiregional models.

References

Adams F.G., Ganges B., Shishido S., Movchauk, (1996), *Disaggregate Modeling of the US-Japan-World Economy,* Paper prepared for the meeting of Project Link, Lausanne, mimeo

Barrell R., Becker B., Byrne J., Gottschalk S., Hurst I., van Welsum D., (2004), *Macroeconomic Policy in Europe: Experiments with Monetary Responses and Fiscal Impulses*, “ Economic Modelling”, vol. 21

Bayoumi T., with assistance from Laxton D. et al. (2004), *GEM*: *A New International Macroeconomic Model*, “ Occasional Paper”, No. 239, IMF, Washington D.C.

Bolton R., (1982), The *Development of Multiregional Economic Modeling in North America: Multiregional Models it Transition for Economies in Transotion*, [in:]. B. Issaev et al. (eds). *Multiregional Economic Modeling: Practice and Prospect*, North-Holland, Amsterdam

Bolton R., *Regional Econometric Models* (1991), [in:] R.G. Bodkin, L.R. Klein, K., Marwah, *A History of Macroeconometric Model Building*, E. Elgar, Aldeshot

Brayton F., Levin A., Tryon R., Williams J.C. (1997), *The Evolution of Macro Models at the Federal Reserve Board*, "Finance and Economics Discussion Series", No 1997-29, FRB, mimeo

Commission of the European Communities (CEC), (1991), *Quest: A Macroeconomic Model for the Countries of the European Community as Part of the World Economy*, "European Economy", No.47, EEC, Brussels

Courbis R., (1979), *The REGINA Model: A Regional-national Model for French Planning* , "Regional Science and Urban Economics", vol. 9

D'Adda, Poli G., Vincenzi L., (1997), *PRIAMO: Prometea International Annual Model*, [in:] Welfe (ed.), *Economies in Transition and the World Economy. Models Forecasts and Scenarios*, Peter Lang , Frankfurt/M.

Delessy H., Sterdyniak H., et al., (1996), *The New Version of MIMOSA. A Macroeconomic Model* of *the World Economy*, (version 1995). Mimosa Team, no. 96-01, CEPII/OFCE

Dramais H., Roeger W., in't Veld J. (1997), *Quest II. A Multi Country Business Cycle and Growth Model*, paper presented at the Project Link Meeting, Malaysia, European Commission, mimeo

Edison H., Marquez J., Tryon R., (1989), *The Structure and Properties of the Federal Reserve Board Multicountry Model*, "Economic Modelling" vol, 4

Fair R.C., (2004), *Estimating how the Macroeconomy Works*, Harvard University Press, Cambridge MA

Florczak W., Karp P., Świeczewska I., Welfe W., (2008), *Rozwój regionalny ujęcie ekonometryczne. Raport w ramach realizacji projektu LORIS WIZJA. Regionalny Foresight Technologiczny*, Wyd. Wyższa Szkoła Przedsiębiorczości i Zarządzania, Łódź

Florczak W., (2011), *O możliwościach modelowania zrównoważonego rozwoju regionalnego. Uwagi metodyczne na temat implementacji projektu Trendy Rozwoju Mazowsza*, Studia Prawno-Ekonomiczne, vol. 84

Freyz G., (1993), *Regional Economic Modeling*, Kluwer Academic Publisher, Boston

Helliwell J., Sturm P., Garrett P., Salou G., (1986), *The Supply Side in OECD's, Macroeconomic Model*, "OECD Economic Studies", No. 6, OECD Publishers

Hickman B.G., (1991), *Project LINK and Multi-country Modelling*, [in:], R.G. Bodkin, L.R. Klein, K. Marwah, *A History of Macroeconometric Model-Building*, E, Elgar, Aldershot

Hickman B.G., Ruffing K.G., (1995), *Project LINK: Past, Present and Future*, [in:] L. Klein, F. Lo (eds), UNU Press, Tokyo

Issaev B., (ed.) (1982), *Multiregional Economic Modeling Practice and Prospect*, North-Holland, Amsterdam

Klein L.R., (1969), *The Specification of Regional Econometric Models*, "Papers Regional Science Assoctation, vol. 23

Klein L.R., (1999), *Project LINK*, [in:] L.R. Klein, A. Welfe, W. Welfe, *Principles of Macroeconometric Modeling*, North Holland, Amsterdam

Laxton D., Isard P., Faruqee H., Prasad E., Turlelboom B., (1998), *MULTIMOD MARK III*: *The Core Dynamic and Steady State Models*, "IMF Occasional Papers", No 164,

Levin A.T., Rogres J.H., Tryon R.W., (1999), *A Guide to FRB/GLOBAL*, "International Finance Discussion Papers", No. 588, FRS, Washington D.C. ,mimeo

Macroeconometric Multi Country Model: MEMMOD, (2000), Deutsche Bank, Frankfurt.Main, mimeo

Montero L.,J.,M., et al. (2008), *Alternative Clasification of EU27 Eligible for European Cohesion Policy*, paper presented to the 65, IAE Conference, Warsaw 9-12.04.2008, mimeo

Onishi A., (1993), *FUGI Global Model 7.0. A New Frontier Science of Global Economic Modelling*, "Economic and Financial Computing", vol. 3

Overview of NIGEM, (2005), NIESR, London

Petersen C.E., Pedersen K.N., Riordan E.J., Lynn R.A., Bradley T., (1991), *BANK-GEM: A World Bank Global Economic Model*, paper presented at the Project LINK Conference, Moscow, SECAP, The World Bank, mimeo

Richardson P., (1988), *The Structure and Simulation Properties of OECD's INTERLINK Model*, "OECD Economic Studies", No. 10, OECD Publishing

Roeger W., in't Veld J., (1997), *QUEST II, A Multicountry Business Cycle and Growth Model*, "Economic Papers", No 123, European Commission

Socha M., Sztanderska, (2000), *Strukturalne podstawy bezrobocia*, PWN, Warszawa

Waelbroeck J., Dramais A., (1974), *DESMOS: A Model for the Cordination of Economic Policies in the EEC Countries*, [in:] A. Ando, R. Herring, R. Marston (eds.), *International Aspects of Stabilization Policies*, FRB of Boston, Boston

Welfe W., (2010), *Zarys historii ekonometrycznego modelowania gospodarki narodowej*, Wydawnictwo UŁ, Łódź

Streszczenie

MAKROEKONOMETRYCZNE MODELE GOSPODARKI ŚWIATOWEJ I MODELE REGIONALNE

Modele gospodarki światowej lub jej części (regionów) zostały zbudowane w ubiegłym stuleciu, służąc celom analizy rozwoju i prognozowaniu gospodarki całego świata lub wyróżnionych regionów (np. Ameryki Łacińskiej). Modele te zwykle wyróżniały największe, uprzemysłowione kraje świata oraz sumarycznie pozostałe kraje (ROW). Stąd ich struktura została oparta na specyfikacjach równań typowych dla poszczególnych krajów, wykorzystując pełnię statystycznych informacji gromadzonych dla każdego kraju.

Makroekonometryczne modele regionalne są budowane albo dla wyróżnionych jednostek administracyjnych albo geograficznych w obrębie na ogół dużych krajów (USA, Chiny, Rosja, etc.). Ich struktura winna być zbliżona do struktury modeli danych krajów. Jednakże, dane statystyczne dotyczące regionów są na ogół niekompletne: brak jest danych o eksporcie i imporcie regionu, migracjach ludności, przepływach finansowych etc.. Wymaga to odpowiednich aproksymacji. Ponieważ w wielu krajach dynamika cen i plac przebiega w sposób podobny w skali kraju – wprowadza się do modeli regionalnych zmienne dotyczące całego kraju. Z drugiej strony oddziaływanie zwrotne występuje na ogół rzadko w tych modelach.

W referacie przedstawiono typowe struktury makroekonometrycznych modeli występujących w modelach gospodarki światowej oraz struktury modeli charakterystyczne dla modeli regionalnych. Specyficzne własności modeli regionalnych są przedmiotem odrębnej dyskusji a także możliwości rozbudowy tych modeli są analizowane.

10.2478/v10103-012-0043-5

TOMASZ ŻĄDŁO*

On Misspecification of Spatial Weight Matrix for Small Area Estimation in Longitudinal Analysis

Abstract

The problem of prediction of subpopulation (domain) total is studied as in Rao (2003). Considerations are based on spatially correlated longitudinal data. The domain of interest can be defined after sample selection what implies its random sample size. The special case of the General Linear Mixed Model is proposed where two random components obey assumptions of spatial and temporal moving average process respectively. Moreover, it is assumed that the population may change in time and elements' affiliations to subpopulation may change in time as well. The proposed model is a generalization of longitudinal models studied by e.g. Verbeke, Molenberghs (2000) and Hedeker, Gibbons (2006). The best linear unbiased predictor (BLUP) is derived. It may be used even if the sample size in the subpopulation of interest in the period of interest is zero. In the Monte Carlo simulation study the accuracy of the empirical version of the BLUP will be studied in the case of correct and incorrect specification of the spatial weight matrix. Two cases of model misspecification are studied. In the first case the misspecified spatial weight is used. In the second case independence of random components is assumed but the variable which is used to compute elements of spatial weight matrix in the correct case will be used as auxiliary variable in the model.

* Ph.D., University of Economics in Katowice

1. Introduction

We start from some general description of the problem. In the paper considerations are based on the model approach in survey sampling. In the survey sampling the main purpose is the estimation or prediction of characteristics of the whole population e.g. the population mean or the population total (i.e. sum of values of the variable of interest). In the practice of survey sampling typically it is not the only purpose of the survey – the estimation or prediction of subpopulation (domain) characteristics may be of interest of survey statistician as well. For example, from some population of people a sample is drawn. The key issue is to estimate the total amount of money spent for some type of goods in the whole population. Additional purpose of the survey is to estimate the total amount of money spent for the considered type of goods but not in the whole population but for inhabitants of some geographical region which (additionally) belong to the group of households consisting from 3 persons. If the division of the population due to geographical regions and household size is not taken into account in the sampling plan, the subpopulation size in the sample will be random. It means that it may be very small or even zero. What is more, if the problem will be considered in the case of longitudinal survey, we should take into account that the population may change in time, population elements may change their subpopulation's affiliation in time (that the household size, to which some person belongs to, may change in time) and that the temporal and spatial autocorrelation is observed.

2. Basic notations

Let us introduce some notation presented earlier by Żądło (2009). Longitudinal data for periods t=1,...,M are considered. In the period t the population of size N_t is denoted by Ω_t. The population in the period t is divided into D disjoint subpopulations (domains) Ω_{dt} each of size N_{dt}, where d=1,...,D. Let the set of population elements for which observations are available in the period t be denoted by s_t and its size by n_t. The set of subpopulation elements for which observations are available in the period t is denoted by s_{dt} and its size by n_{dt}. Let: $\Omega_{rdt} = \Omega_{dt} - s_{dt}$, $N_{rdt} = N_{dt} - n_{dt}$.

Let M_{id} denotes the number of periods when the i-th population element belongs to the d-th domain. Let us denote the number of periods when the i-th population element (which belongs to the d-th domain) is observed by m_{id}. Let

$m_{rid} = M_{id} - m_{id}$. It is assumed that the population may change in time and that one population element may change its domain affiliation in time (from technical point of view observations of some population element which change its domain affiliation are treated as observations of new population element). It means that *i* and *t* completely identify domain affiliation but additional subscript *d* will be needed as well. More about this assumptions will be written at the end of the next section.

The set of elements which belong at least in one of periods *t*=1,...,*M* to sets Ω_t is denoted by Ω and its size by *N*. Similarly, sets Ω_d, s, s_d, Ω_{rd} of sizes N_d, *n*, n_d, N_{rd} respectively are defined as sets of elements which belong at least in one of periods *t*=1,...,*M* to sets Ω_{dt}, s_t, s_{dt}, Ω_{rdt} respectively. The *d**-th domain of interest in the period of interest t^* will be denoted by Ω_{d*t*}, and the set of elements which belong at least in one of periods *t*=1,...,*M* to sets Ω_{d*t*} will be denoted by Ω_{d*}.

Values of the variable of interest are realizations of random variables Y_{idj} for the *i*-th population element which belongs to the *d*-th domain in the period t_{ij}, where *i*=1,...,*N*, *j*=1,...,M_{id}, *d*=1,...,*D*. The vector of size $M_{id} \times 1$ of random variables Y_{idj} for the *i*-th population element which belongs to the *d*-th domain will be denoted by $\mathbf{Y_{id}} = \left[Y_{idj} \right]$, where $j = 1,...,M_{id}$. Let us consider values of the variables of interest $Y_{i'd'j'}$ for the *i'*-th population element which belongs to the *d'*-th domain observed in periods $t_{i'j'}$, where *i'*=1,...,*n*, *j'*=1,...,$m_{i'd'}$, *d'*=1,...,*D*. The vector of random variables $Y_{i'd'j'}$ (where *i'*=1,...,*n*, *j'*=1,...,$m_{i'd'}$, *d'*=1,...,*D*) of size $m_{i'd'} \times 1$ will be denoted by $\mathbf{Y_{s\,i'd'}} = \left[Y_{i'd'j'} \right]$, where *j'*=1,...,$m_{i'd'}$. The vector of random variables $Y_{i''d''j''}$ of size $m_{ri''d''} \times 1$ for the *i''*-th population element which belongs to the *d''*-th domain for observations which are not available in the sample is denoted by $\mathbf{Y_{r\,i''d''}} = \left[Y_{i''d''j''} \right]$, where *j''*=1,...,$m_{ri''d''}$.

The proposed approach may be used to predict the domain total for any (past, current and future) periods but under assumption that values of the auxiliary variables and the division of the population into subpopulations in the period of interest are known.

3. Superpopulation model

We consider superpopulation models used for longitudinal data (compare Verbeke, Molenberghs, 2000; Hedeker, Gibbons, 2006) which are – what is important for further considerations – special cases of the General Linear Model (GLM) and the General Linear Mixed Model (GLMM). The following model is assumed:

$$\mathbf{Y_d} = \mathbf{X_d}\boldsymbol{\beta}_\mathbf{d} + \mathbf{Z_d}\mathbf{v_d} + \mathbf{e_d}, \tag{1}$$

where

$\mathbf{Y_d} = col_{1\le i\le N_d}(\mathbf{Y_{id}})$, where $\mathbf{Y_{id}}$ is a random vector of size $M_{id}\times 1$, $\mathbf{X_d} = col_{1\le i\le N_d}(\mathbf{X_{id}})$, where $\mathbf{X_{id}}$ is known matrix of size $M_{id}\times p$, $\mathbf{Z_d} = diag_{1\le i\le N_d}(\mathbf{Z_{id}})$, where $\mathbf{Z_{id}}$ is known vector of size $M_{id}\times 1$, $\mathbf{v_d} = col_{1\le i\le N_d}(v_{id})$, where v_{id} is a random component and $\mathbf{v_d}$ (d=1,2...,D) are assumed to be independent, $\mathbf{e_d} = col_{1\le i\le N_d}(\mathbf{e_{id}})$, where $\mathbf{e_{id}}$ is a random component vector of size $M_{id}\times 1$ and $\mathbf{e_{id}}$ (i=1,...,N; d=1,...,D) are assumed to be independent, $\mathbf{v_d}$ and $\mathbf{e_d}$ are assumed to be independent.

What is more, that vector of random components $\mathbf{v_d}$ obey assumptions of spatial moving average process, i.e.

$$\mathbf{v_d} = \lambda_{(sp)}\mathbf{W}_d\mathbf{u_d} + \mathbf{u_d}\ , \tag{2}$$

where $\mathbf{W}_d$ is the spatial weight matrix for profiles $\mathbf{Y_{id}}$, $\mathbf{u_d} \sim (\mathbf{0}, \sigma_u^2\mathbf{I}_{\mathbf{N_d}})$. Hence,

$$\mathbf{v_d} \sim \left(\mathbf{0}, \mathbf{R_d}\right), \tag{3}$$

where $\mathbf{R_d} = \sigma_u^2\mathbf{H_d}$ and $\mathbf{H_d} = \mathbf{I}_{\mathbf{N_d}} + \lambda_{(sp)}(\mathbf{W_d} + \mathbf{W_d^T}) + \lambda_{(sp)}^2\mathbf{W_d}\mathbf{W_d^T}$. Moreover, elements of $\mathbf{e_{id}}$ obey assumptions of MA(1) temporal process, i.e.

$$e_{idj} = \varepsilon_{idj} - \lambda_{(t)}\varepsilon_{idj-1}\,. \tag{4}$$

Hence,

$$\mathbf{e_{id}} \sim (\mathbf{0}, \mathbf{\Gamma_{id}}), \tag{5}$$

where elements of $\mathbf{\Gamma_{id}} = \sigma_\varepsilon^2 \begin{bmatrix} 1+\lambda_{(t)}^2 & -\lambda_{(t)} & 0 & \dots & 0 \\ -\lambda_{(t)} & 1+\lambda_{(t)}^2 & -\lambda_{(t)} & \dots & 0 \\ 0 & -\lambda_{(t)} & 1+\lambda_{(t)}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1+\lambda_{(t)}^2 \end{bmatrix}$.

Let $\mathbf{Y} = col_{1\le d\le D}(\mathbf{Y_d})$, $\mathbf{V} = D_\xi^2(\mathbf{Y}) = diag_{1\le d\le D}(D_\xi^2(\mathbf{Y_d}))$ and $\mathbf{V_d} = D_\xi^2(\mathbf{Y_d})$.
Hence,

$$\mathbf{V} = diag_{1\le d\le D}(\mathbf{V_d}) = diag_{1\le d\le D}\left(\sigma_u^2\mathbf{Z_d}\mathbf{H_d}\mathbf{Z_d}^T + diag_{1\le i\le N_d}(\mathbf{\Gamma_{id}})\right). \tag{6}$$

Let $\mathbf{Y_s} = col_{1\le d\le D}(\mathbf{Y_{sd}}) = col_{1\le d\le D}(col_{1\le i\le N_d}(\mathbf{Y_{sid}}))$,
$\mathbf{V_{ss}} = D_\xi^2(\mathbf{Y_s}) = diag_{1\le d\le D}(D_\xi^2(\mathbf{Y_{sd}}))$, $\mathbf{V_{ss\,d}} = D_\xi^2(\mathbf{Y_{sd}})$

Hence,

$$\mathbf{V_{ss}} = diag_{1\le d\le D}(\mathbf{V_{ss\,d}}) = diag_{1\le d\le D}\left(\sigma_u^2\mathbf{Z_{sd}}\mathbf{H_d}\mathbf{Z_{sd}}^T + diag_{1\le i\le n_d}(\mathbf{\Gamma_{ss\,id}})\right), \tag{7}$$

where

$\mathbf{Z_{sd}} = diag_{1\le i\le n_d}(\mathbf{Z_{sid}})$, where $\mathbf{Z_{sid}}$ is known vector of size $m_{id}\times 1$, $\mathbf{\Sigma_{ss\,id}}$ is a submatrix obtained from $\mathbf{\Sigma_{id}}$ by deleting rows and columns for unsampled observations,

What is more,

$$\mathbf{V_{ss}^{-1}} = diag_{1\le d\le D}(\mathbf{V_{ss\,d}^{-1}}) = diag_{1\le d\le D}\left(\left(\sigma_u^2\mathbf{Z_{sd}}\mathbf{H_d}\mathbf{Z_{sd}}^T + diag_{1\le i\le n_d}(\mathbf{\Gamma_{ss\,id}})\right)^{-1}\right). \tag{8}$$

Let $\mathbf{Y_r} = col_{1\le d\le D}(\mathbf{Y_{rd}}) = col_{1\le d\le D}(col_{1\le i\le N_{rd}}(\mathbf{Y_{rid}}))$,
$\mathbf{V_{rr}} = D_\xi^2(\mathbf{Y_r}) = diag_{1\le d\le D}(D_\xi^2(\mathbf{Y_{sd}}))$, $\mathbf{V_{rr\,d}} = D_\xi^2(\mathbf{Y_{rd}})$,
Hence,

$$\mathbf{V_{rr}} = diag_{1\le d\le D}(D^2_{\xi}(\mathbf{V_{rr\,d}})) = diag_{1\le d\le D}\left(\sigma_u^2 \mathbf{Z_{rd}} \mathbf{H_d} \mathbf{Z_{rd}^T} + diag_{1\le i\le N_{rd}}(\mathbf{\Gamma_{rr\,id}})\right) \quad (9)$$

where $\mathbf{Z_{rd}} = diag_{1\le i\le N_{rd}}(\mathbf{Z_{rid}})$, where $\mathbf{Z_{rid}}$ is known vector of size $M_{rid} \times 1$, $\mathbf{\Gamma_{rr\,id}}$ is a submatrix obtained from $\mathbf{\Gamma_{id}}$ by deleting rows and columns for sampled observations. Let $\mathbf{V_{sr}} = Cov_{\xi}(\mathbf{Y_s}, \mathbf{Y_r}) = diag_{1\le d\le D}(Cov_{\xi}(\mathbf{Y_{sd}}, \mathbf{Y_{rd}}))$, $\mathbf{V_{sr\,d}} = Cov_{\xi}(\mathbf{Y_{sd}}, \mathbf{Y_{rd}})$ **. Hence,**

$$\mathbf{V_{sr}} = diag_{1\le d\le D}(\mathbf{V_{sr\,d}}) = diag_{1\le d\le D}\left(\sigma_u^2 \mathbf{Z_{sd}} \mathbf{H_d} \mathbf{Z_{rd}^T} + diag_{1\le i\le N_{rd}}(\mathbf{\Gamma_{sr\,id}})\right)$$

where $\mathbf{\Gamma_{sr\,id}}$ is a submatrix obtained from $\mathbf{\Gamma_{id}}$ by deleting rows for unsampled observations and column for sampled observations.

Similar model is considered by Żądło (2011) but instead of spatial and temporal MA models for vectors of random components considered in this paper (see assumptions (2) and (5)) he studied simultaneously spatial autoregressive (SAR) process and temporal AR(1) model. Model (1) (with assumptions (2) and (5)) similarly to the model proposed by Żądło (2011) may be used when the population changes in time and the domain affiliation of population elements changes in time. In this case observations of new element of the population or observations of the population element after the change of its domain affiliation form new profile $\mathbf{Y_{id}}$. It means that observations of new population element will be temporally correlated and spatially correlated with other population elements within the subpopulation. If the population element changes its domain affiliation its new observations will be temporally correlated (but temporally uncorrelated with old observations) and spatially correlated with other population elements within new subpopulation (but spatially uncorrelated with elements of the previous subpopulation).

In next sections three predictors of the total (of the sum of random variables) $\theta_{d^*t^*} = \sum_{i\in\Omega_{d^*t^*}} Y_{id^*t^*}$ in the domain of interest in the period of interest will be proposed.

4. First predictor – Spatial EBLUP

In the section, based on the Royall (1976) theorem, we derive the formula of the best linear unbiased predictor (BLUP) of the population total under the model (1). Let us introduce following notations:

$$\hat{\boldsymbol{\beta}}_{d*} = \left(\mathbf{X}^{\mathbf{T}}_{\mathbf{sd}*}\mathbf{V}^{\mathbf{-1}}_{\mathbf{ss\ d}*}\mathbf{X}_{\mathbf{sd}*}\right)^{-1}\mathbf{X}^{\mathbf{T}}_{\mathbf{sd}*}\mathbf{V}^{\mathbf{-1}}_{\mathbf{ss\ d}*}\mathbf{Y}_{\mathbf{sd}*} \tag{10}$$

where

$$\mathbf{V}_{\mathbf{ss\ d}*} = \sigma_u^2\mathbf{Z}_{\mathbf{sd}*}\mathbf{H}_{\mathbf{d}*}\mathbf{Z}^T_{\mathbf{sd}*} + diag_{1\le i\le n_{d*}}(\boldsymbol{\Gamma}_{\mathbf{ss\ id}*}), \tag{11}$$

$\mathbf{X}_{\mathbf{sd}*}$ is known $\sum_{i=1}^{n_{d*}} m_{id*}\times p$ matrix of auxiliary variables, $\mathbf{Y}_{\mathbf{sd}*}$ is a $\sum_{i=1}^{n_{d*}} m_{id*}\times 1$ vector of random variables Y_{idj}.

The BLUP of the total in the domain of interest in the period of interest is given by:

$$\hat{\theta}_{(1)} = \sum_{i\in s_{d*t*}} Y_{id*t*} + \tilde{\mathbf{x}}_{\mathbf{rd*t*}}\hat{\boldsymbol{\beta}}_{\mathbf{d*}} +$$

$$+\boldsymbol{\gamma}^{\mathbf{T}}_{\mathbf{rd*}}\left(\sigma_u^2\mathbf{Z}_{\mathbf{rd}*}\mathbf{H}_{\mathbf{d}*}\mathbf{Z}^{\mathbf{T}}_{\mathbf{sd}*} + diag_{1\le i\le N_{rd*}}(\boldsymbol{\Gamma}_{\mathbf{rs\ id}*})\right)\mathbf{V}^{-1}_{\mathbf{ss\ d}*}\left(\mathbf{Y}_{\mathbf{sd}*} - \mathbf{X}_{\mathbf{sd}*}\hat{\boldsymbol{\beta}}_{\mathbf{d*}}\right) \tag{12}$$

where

$\tilde{\mathbf{x}}_{\mathbf{rd*t*}}$ is a $1\times p$ vector of totals of auxiliary variables in Ω_{rd*t*},

$\boldsymbol{\gamma}_{\mathbf{rd*}}$ is a $\sum_{i=1}^{n_{d*}} M_{rid*}\times 1$ vector of one's for observations in period $t*$ (i.e. in Ω_{rd*t*}) and zero otherwise.

If the unknown parameters σ_u^2, σ_ε^2, $\lambda_{(sp)}$, $\lambda_{(t)}$ in (12) will be replaced by some estimators we obtain the empirical best linear unbiased predictor (EBLUP) which remains unbiased under some weak assumptions (see Żądło (2004)). Because the spatial correlation is included in the assumed model, the EBLUP may be called spatial EBLUP (SEBLUP).

5. Second predictor – misspecified spatial weight matrix

In the previous section the BLUP and its empirical version were derived under the model (1) assuming that the structure of the spatial weight matrix $\mathbf{W}_{\mathbf{d}}$ (where d=1,...,D) is correct. In this case the new predictor is derived under the model given by formula (1) but the assumed structure of the spatial weight matrix is not correct. The misspecified spatial weight matrix will be denoted by

$\mathbf{W}_{\mathbf{d}\,(mis)}$ (where d=1,...,D). In this case we obtain some empirical predictor which is not EBLUP under (1) due to the misspecification of the spatial weight matrix – in the simulation study it will be denoted by SEBLUPmis.

6. Third predictor – independent random components

In this section we assume that population data obey assumptions of the model (1) but with $\lambda_{(sp)} = 0$ and $\lambda_{(t)} = 0$ what means that random components are assumed to be uncorrelated. For this model (i.e. under assumption that $\lambda_{(sp)} = 0$ and $\lambda_{(t)} = 0$) BLUP is given by:

$$\hat{\theta}_{(3)} = \sum_{i \in s_{d^*t^*}} Y_{id^*t^*} + \tilde{\mathbf{x}}_{\mathbf{rd^*t^*}} \hat{\boldsymbol{\beta}}_{\mathbf{d^*}} + \sigma_u^2 \boldsymbol{\gamma}_{\mathbf{rd^*}}^{\mathbf{T}} \mathbf{Z}_{\mathbf{rd^*}} \mathbf{Z}_{\mathbf{sd^*}}^{\mathbf{T}} \mathbf{V}_{\mathbf{ss\,d^*}}^{-1} \left(\mathbf{Y}_{\mathbf{sd^*}} - \mathbf{X}_{\mathbf{sd^*}} \hat{\boldsymbol{\beta}}_{\mathbf{d^*}} \right) \tag{13}$$

where

$$\mathbf{V}_{\mathbf{ss\,d^*}} = \sigma_u^2 \mathbf{Z}_{\mathbf{sd^*}} \mathbf{Z}_{\mathbf{sd^*}}^T + \sigma_\varepsilon^2 \mathbf{I}_{\sum_{i=1}^{n_{d^*}} m_{id^*} \times \sum_{i=1}^{n_{d^*}} m_{id^*}} \tag{14}$$

where $\mathbf{I}_{\sum_{i=1}^{n_{d^*}} m_{id^*} \times \sum_{i=1}^{n_{d^*}} m_{id^*}}$ is identity matrix of size $\sum_{i=1}^{n_{d^*}} m_{id^*} \times \sum_{i=1}^{n_{d^*}} m_{id^*}$, $\hat{\boldsymbol{\beta}}_{\mathbf{d^*}}$ given by formula (10) where $\mathbf{V}_{\mathbf{ss\,d^*}}$ given by (11) is repl aced by (14).

The predictor given by (13) is not BLUP under the model (1) due to the misspecification of the assumed model (because it is derived under assumption that $\lambda_{(sp)} = 0$ and $\lambda_{(t)} = 0$). In the simulation study the empirical version of the predictor (13) will be denoted by SEBLUPmis2.

7. Monte Carlo simulation study

The simulation study was conducted using R package (R Development Core Team (2012)). It is based on artificial longitudinal data from M=3 periods. The population size in each period equals N=200 elements which consists of D=20 domains (subpopulations) each of size 10 elements. The balanced panel sample is considered – in each period the same 40 elements are observed. The sample sizes in D=20 domains are as follows {1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3}. Relatively small population size is due to very time-consuming computations. Relatively large sample size (comparing to the population size) is assumed to increase accuracy of SEBLUP due to the existence of spatial effect. The number of iterations is 2000.

In the simulation data are generated based on the model (1) assuming arbitrary chosen parameters $\sigma_u^2 = 1$, $\sigma_\varepsilon^2 = 1$, $\forall_d \beta_d = \beta = 100$, $\mathbf{X_{id}} = [1]_{M_{id} \times p}$ and different values of $\lambda_{(sp)}$ and $\lambda_{(t)}$. The spatial weight matrix is based not on geographical distances between profiles (in this case between elements) but based on the values of the auxiliary variable. In the simulation study the elements of spatial weight matrix (denoted by $\mathbf{W_d}$) were row-standardized inverses of absolute differences between sorted values of auxiliary variable generated from Beta(1,5) distribution. The Beta(1,5) distribution is distribution with positive asymmetry as many economic variables what means that considered distance between elements may be treated as a distance in some economic sense. Values of the variable were sorted to obtain the biggest values of the spatial weight matrix close to the diagonal. For this type of assumed as correct spatial weight matrix, assumption of row-standardized neighborhood matrix (where one element has two neighbours – one before and one after) as a misspecified spatial weight matrix (denoted by $\mathbf{W_{d\,(\mathit{mis})}}$) is reasonable solution.

In the simulation three predictors are considered:

- The first predictor (denoted in the simulation by SEBLUP) is spatial EBLUP which is empirical version of (12), where parameters are estimated using restricted maximum likelihood method. The spatial weight matrix is given by $\mathbf{W_d}$ (described above) and $\mathbf{Z_{id}} = [1]_{M_{id} \times 1}$.
- The second predictor (denoted in the simulation by SEBLUPmis) is empirical version of (12) but where $\mathbf{Z_{id}} = [1]_{M_{id} \times 1}$ and $\mathbf{W_d}$ is replaced by row-standardized neighborhood matrix (denoted by $\mathbf{W_{d\,(\mathit{mis})}}$) and parameters are estimated using restricted maximum likelihood method based on misspecified likelihood function (where $\mathbf{W_d}$ is replaced by $\mathbf{W_{d\,(\mathit{mis})}}$)
- The third predictor (denoted in the simulation by SEBLUPmis2) is given by (13) where parameters are estimated using restricted maximum likelihood method based on misspecified likelihood function (assuming that $\lambda_{(sp)} = 0$ and $\lambda_{(t)} = 0$ and $\mathbf{Z_{id}}$ is a vector of auxiliary variable generated based on Beta(1,5) distribution which was used earlier (in the case of the first and the

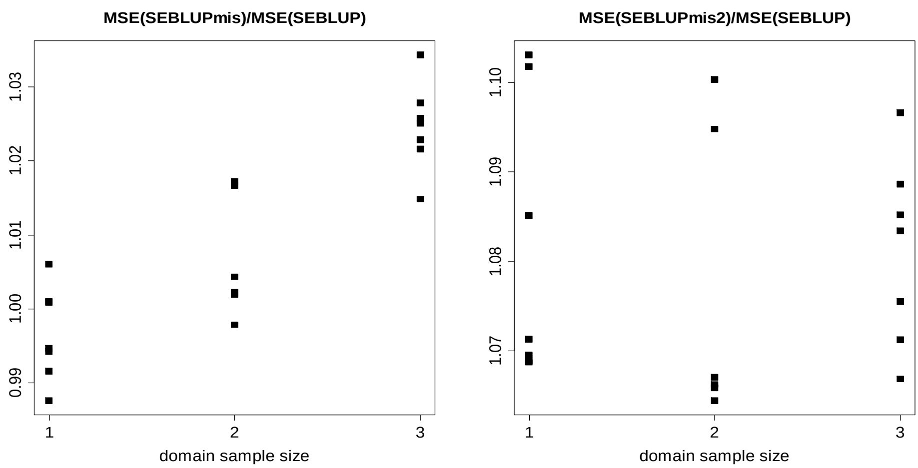
second predictor) to compute distances and then elements of spatial weight matrix.

In the simulation special cases of the predictors' equations are used for balanced panel sample and under assumption that $\forall_d \beta_d = \beta$.

Because we are mainly interested in the spatial effect, in 5 simulations we study different values of $\lambda_{(sp)}$ ($\lambda_{(sp)} = \{0,6; 0,7; 0,8; 0,9; 1\}$) and one value of $\lambda_{(t)} = 0,5$. To study maximum effect of both spatial and time effect in the last simulation it is assumed that $\lambda_{(sp)} = \lambda_{(t)} = 1$. Cases for $\lambda_{(sp)} = \{0,6; 0,7\}$ are not presented at graphs.

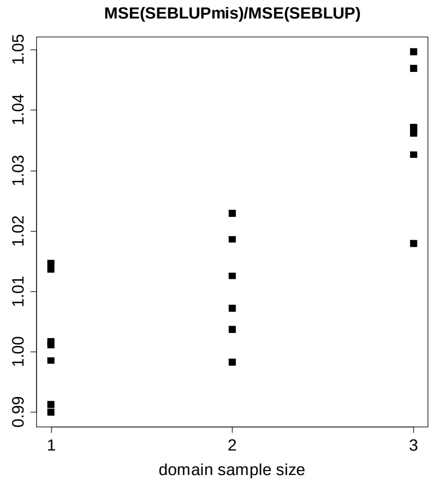
In the following graphs it is shown (on the right side) that the predictor SEBLUP may be more accurate comparing to BLUPind maximum for the considered cases from c.a. 9% to c.a. 17% for $\lambda_{(sp)} = 1$ and $\lambda_{(t)} = 0,5$.- see the right side of graph 3.

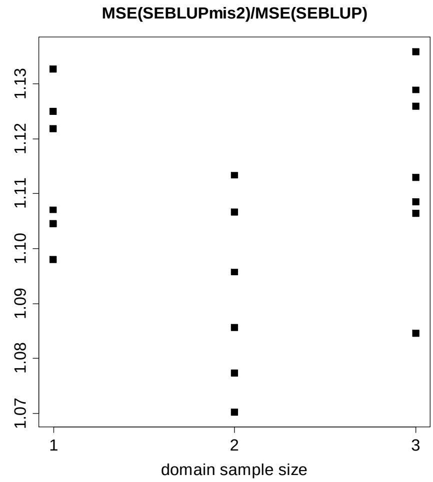
Graph 1. Simulation results for $\lambda_{(sp)} = 0,8$ and $\lambda_{(t)} = 0,5$



Source: Results based on own computations.

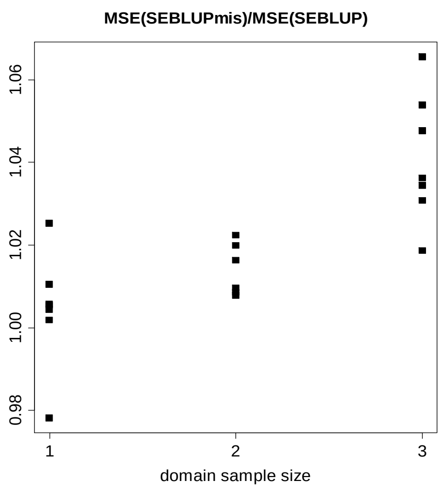
Graph 2. Simulation results for $\lambda_{(sp)} = 0{,}9$ **and** $\lambda_{(t)} = 0{,}5$

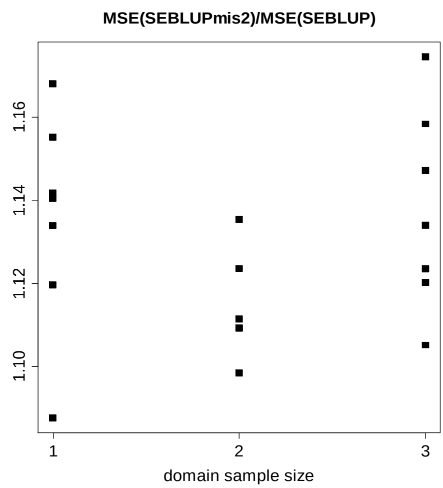




Source: results based on own computations.

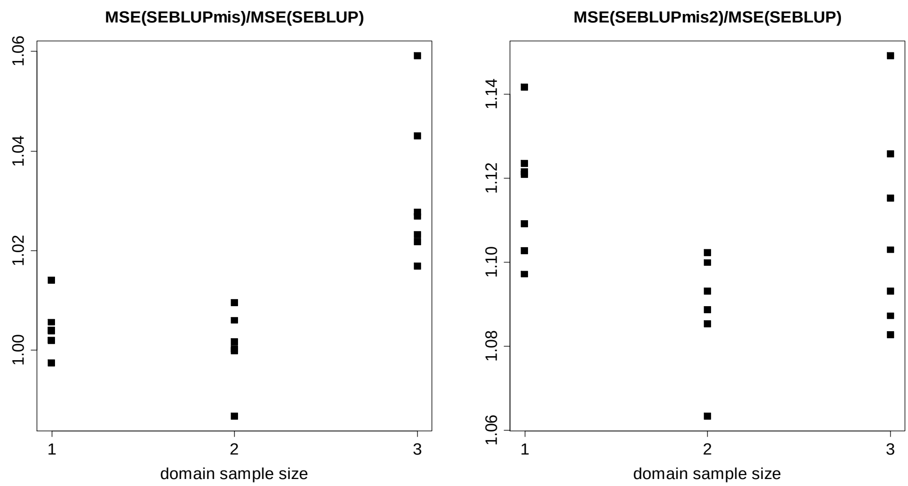
Graph 3. Simulation results for $\lambda_{(sp)} = 1$ **and** $\lambda_{(t)} = 0{,}5$





Source: results based on own computations.

Graph 4. Simulation results for $\lambda_{(sp)} = 1$ and $\lambda_{(t)} = 1$



Source: results based on own computations.

The predictor SEBLUP may be more accurate comparing to SEBLUPmis maximum for the considered cases by c.a. 6% - see the left side of graph 3 and graph 4. What is very interesting, for all of the considered pairs of $\lambda_{(sp)}$ and $\lambda_{(t)}$ there are some cases when SEBLUPmis is better than SEBLUP. It results from the decrease of the accuracy of the first predictor due to the estimation of the model parameters (the decrease of the accuracy of spatial EBLUP comparing with spatial BLUP) – in some cases studied in the simulation study the maximum decrease was even greater than 5%.

Summarizing, the proposed predictor (the first predictor - SEBLUP) which takes the spatial and temporal autocorrelation into account may be better than the predictor derived under assumption of the lack of spatial and temporal autocorrelation (the third predictor – SEBLUPmis2) in the studied cases from c.a. 1% to c.a. 17%. The misspecification of the spatial weight matrix has small influence on the accuracy – the maximum decrease of the accuracy observed in the simulation was c.a. 6% but in many cases the predictor under assumption of the misspecified spatial weight matrix is better than the correct predictor what results from the decrease of the accuracy of SEBLUP due to the estimation of model parameters.

8. Conclusions

In the paper three predictors of the subpopulation total are proposed for longitudinal data and their accuracy is studied in the simulation study. It is shown that the considered misspecification of spatial weight matrix decreases the accuracy of the predictor only slightly.

References

Hedeker D., Gibbons R.D. (2006), Longitudinal Data Analysis, John Wiley, New Jersey

R Development Core Team (2011), *A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna

Rao J.N.K (2003), *Small area estimation*, John Wiley and Sons, New Jersey

Royall R.M. (1976), The linear least squares prediction approach to two-stage Sampling, *Journal of the American Statistical Association*, 71, 657-473

Verbeke G., Molenberghs G. (2000*), Linear Mixed Models for Longitudinal Data*, Springer-Verlag, New York

Żądło T. (2004), *On unbiasedness of some EBLU predictor*, [in:] J. Antoch (ed.), *Proceedings in Computational Statistics*, Physica-Verlag, Heidelberg-New York, 2019-2026

Żądło T. (2009), On prediction of domain totals based on unbalanced longitudinal data, [in:] Wywiał J., Żądło T. (eds.) *Survey Sampling in Economic and Social Research*, University of Economic in Katowice, Katowice

Żądło T. (2011), *On accuracy of two predictors for spatially and temporally correlated longitudinal data*, submitted to publication

Streszczenie

O BŁĘDNEJ SPECYFIKACJI MACIERZY WAG PRZESTRZENNYCH W STATYSTYCE MAŁYCH OBSZARÓW W BADANIACH WIELOOKRESOWYCH

Rozważany jest problem predykcji wartości globalnej w podpopulacji (domenie) podobnie jak w Rao (2003). Zaproponowano przypadek szczególny Ogólnego Mieszanego Modelu Liniowego, gdzie dwa składniki losowe spełniają założenia

odpowiednio przestrzennego i czasowego procesu średniej ruchomej. Proponowany model jest uogólnieniem modeli wielookresowych rozważanych przez Verbeke, Molenberghs (2000) oraz Hedeker, Gibbons (2006). Wyprowadzona zostanie postać najlepszego liniowego nieobciążonego predyktora wartości globalnej w domenie. W badaniu symulacyjnym dokładność empirycznej wersji najlepszego liniowego nieobciążonego predyktora była analizowana zarówno w przypadkach prawidłowej jak i nieprawidłowej specyfikacji macierzy wag.

Contents