ANALYSIS OF THE LABOUR MARKET IN METROPOLITAN AREAS: A SPATIAL FILTERING APPROACH

Tomasz Kossowski, Jan Hauke

Institute of Socio-Economic Geography and Spatial Management, Adam Mickiewicz University, Poznań, Poland

Manuscript received: March 29, 2011
Revised version: May 11, 2012

Kossowski T., Hauke J., Analysis of the labour market in metropolitan areas: A spatial filtering approach. *Quaestiones Geographicae* 31(2), Bogucki Wydawnictwo Naukowe, Poznań 2012, pp. 39–48. 3 tables, 5 figs. DOI 10.2478/v10117-012-0017-5, ISSN 0137-477X.

Abstract. The power of today's computers allows us to perform computation on massive quantities of data on the one hand and produces enormous amounts of analysis output on the other, as noted by Griffith in his 2003 book. Besides, visualisation and spatial filtering (the core of considerations in Griffith's book) have a chance to be widely used in research practice, especially in geosciences and, more precisely, for georeferenced data. Following the idea proposed by Patuelli et al. (2006, 2009), we analysed the labour market in Poland, focusing on metropolitan areas and their surroundings. The analysis was performed on a data set for the unemployment rate in the 2,478 Polish communes. We took into account spatial autocorrelation and used spatial filtering techniques to construct components of an orthogonal map pattern. As shown in Tiefelsdorf & Griffith (2007), the spatial filtering techniques could be employed in both, parametric and semi-parametric approaches. In this paper we adopted a parametric one.

Key words: Moran's *I* statistic, spatial autocorrelation, spatial dependence

Tomasz Kossowski, Jan Hauke, Institute of Socio-Economic Geography and Spatial Management, Adam Mickiewicz University, Dzięgielowa 27, 61-131 Poznań, Poland; e-mail: tkoss@amu.edu.pl (corresponding author), jhauke@amu.edu.pl

1. Introduction

In his significant book (2003:1), Griffith noted that:

1) „At least since the dawn of civilization data have been analyzed as numerical figures to support a decision or to understand a part of reality."
2) „One consequence of the massive quantities of data collected and analyzed today is the enormous amount of analysis output."
3) „Much of the data collected today are georeferenced, or tagged to the Earth's surface (...)".

Nowadays, commonly used computers are sufficiently powerful to perform calculations for quite large collections of data employing ever more complicated numerical methods and sophisticated algorithms. Therefore, visualisation and spatial filtering (the core of considerations in Griffith's book) have a chance to be widely used in research practice, especially in geosciences.

There are many spatial econometric procedures for a statistical analysis of georeferenced data available in the literature. One of them – very powerful – is spatial autoregression (see e.g. Anselin 1988, Griffith 1988). The method is based on spatial weights matrices measuring the spatial dependence between values of georeferenced variables. But owing to the bias of statistical efficiency and the problem of independence assumption, it is not advisable to use the ordinary least squares (OLS) method with the data.

An alternative approach is to use spatial filtering techniques (cf. Griffith 1981, Getis & Griffith 2002, Tiefelsdorf & Griffith 2007). The idea is to split variables into spatial and non-spatial components, and then OLS is allowed, after reducing the stochastic noise in the residuals. In the procedure spatial filters are computed (Griffith 1996, 2000). This technique is based on the computational formula of Moran's *I* statistic.

2. Data and methods

In the study of spatial patterns and processes, we may logically expect that close observations are more likely to be similar than those far apart (First Law of Geography). It is usual to assign a weight c_{ij} to each pair $(x_i, x_j)$to quantify it. In the simplest form, these weights will take the value of 1 for close neighbours, and 0 otherwise. We set c_{ii}=0. Moran's *I* is defined as

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

where $w_{ij} = \frac{n}{s_0} c_{ij}$, c_{ij} are elements of a matrix of spatial contiguousness C, x_i are observations, n is the number of spatial units, and s_0 is the sum of all elements of a matrix of weights C.

If the value of Moran's *I* is $I > -\frac{1}{n-1}$, we have positive spatial autocorrelation, if $I = -\frac{1}{n-1}$, we have no spatial autocorrelation, and if $I < -\frac{1}{n-1}$, we have negative spatial autocorrelation.

Let $C_s = \frac{n}{s_0}C$, and $z_i = x_i - \bar{x}$. Then, Moran's *I* presented in the matrix language takes the form

$$I = \frac{z'C_s z}{z'z}$$

or

$$I = \frac{x'MC_sMx}{x'MMx}$$

where $M = I - \frac{1}{n}11'$.

In contrast to the properties of the classical linear Pearson's and Spearman's correlation coefficients, Moran's spatial autocorrelation coefficient is not bounded in the range [–1,1]. The range of Moran's *I* values depends on matrix C_s and its eigenvalues as follows (de Jong et al. 1984):

$$\lambda_{\min}(MC_sM) \le I \le \lambda_{\max}(MC_sM)$$

and, consequently,

$$\frac{n}{s_0}\lambda_{\min}(MC_sM) \le I \le \frac{n}{s_0}\lambda_{\max}(MC_sM).$$

The global statistic *I* lets us find a spatial dependence over the studied area. Global statistics are synthetic characteristics of a spatial dependence. But they are not sensitive to local deviations from the global autocorrelation pattern. To identify such deviations, local statistics are more suitable. Their values are calculated for each spatial unit and allow us to determine the similarity of every region to its neighbours. It is also possible to check whether a region is surrounded by neighbours with high or low values of the analysed variable. Moran's *I* global and local measures of spatial autocorrelation are part of Exploratory Spatial Data Analysis (ESDA). With ESDA, we have a chance to identify patterns of global and local spatial autocorrelation, and can try to detect spatial regimes.

The use of local statistics can capture and measure local spatial dependence. Most of such tests are conducted on the basis of Local Indicators of Spatial Association (LISA). Those are indicators proposed by Anselin (1995) and they include local Moran's and local Geary's statistics. Moran's local statistic allows identifying the effects of an agglomeration and shows clusters of high and low values. Local Geary's statistic identifies spatial similarities and differences, showing the average difference between a region and its neighbours. Those statistics help us to extract so-called *hot spots*, that is, areas of high values of a test variable surrounded by areas with lower values of the variable. It is also possible find *outliers*, that is, areas with particularly low values surrounded by regions with high values of the variable, or vice versa. Boots (2003) extended the application of local measures of spatial dependence to categorised variables. In this paper local Moran's statistic is applied to identify effects of an agglomeration.

The formulation of local Moran's is:

$$I_i = z_i \sum_{j=1}^{n} w_{ij} z_j$$

where z_i, z_j are deviations from the mean.

Local Moran's statistic is approximately normally distributed. The local statistic is proportional to the global statistic. The sum of local Moran's I_i studied over a set of spatial units is equal to the global statistic.

Local Moran's I_i statistic is interpreted as an indicator of local instability. So it is possible to check whether a region is surrounded by regions with similar or different values of the variable in relation to the random distribution of those values in space. Units with statistically significant values allow the determination of clusters of low or high values of the test variable.

As a result of LISA analysis, a map of local spatial clusters is obtained. There are four types of clusters: high-high (HH), or units with high values of the characteristic surrounded by ones with high values too, low-low (LL), or units with low values surrounded by similar neighbours, and low-high (LH) and high-low (HL) units clearly standing out from their environment. These clusters are determined on the basis of essential values of local Moran's statistic.

The spatial filtering technique considered in this paper uses an eigenvector decomposition method which extracts orthogonal and uncorrelated numerical components from an $n \times n$ matrix. As Patuelli et al. (2006: 2) state, "These components can be seen as independent map patterns, and represent the latent spatial autocorrelation of a georeferenced variable concerned, according to a given geographic weights matrix. They also can be interpreted as redundant information due to spatial interdependencies, in the framework of standard regression equations".

Formally, these orthogonal components are the computed eigenvectors of the modified geographic weights matrix MC_sM. The eigenvectors of the modified matrix are computed, in sequence, to maximise the sequential residual I values. The first eigenvector, E_1, is therefore the one whose numerical values generate the largest I value among all eigenvectors of the modified matrix. Similarly, the second eigenvector, E_2, is a set of numerical values that, again, maximise the I value while being uncorrelated with E_1. The process continues until n eigenvectors have been computed. This is the complete set of all possible (mutually) orthogonal and uncorrelated map patterns (Getis 2002), and, when employed as regressors, they may function as proxies for missing explanatory variables.

A smaller set of candidate eigenvectors can then be selected from the n eigenvectors on the basis of their I values exceeding some specified threshold value. Since the eigenvectors are both orthogonal and uncorrelated, a stepwise linear regression can be used to achieve this end. In this framework, the advantage of the orthogonality of the eigenvectors is the absence of partial correlations and, therefore, of multi-collinearity issues. Also, residuals obtained with stepwise regression constitute the spatially filtered component of the georeferenced variable examined.

Following the idea proposed by Patuelli, Griffith, Tiefelsdorf & Nijkamp (2006, 2009), we analysed the labour market in Poland. We focused on metropolitan areas and their surroundings. The analysis was performed on a data set for the unemployment rate in the 2,478 Polish communes (*gmina*, a unit of the NUTS 5 level). We took into account spatial autocorrelation and used spatial filtering techniques to construct orthogonal components of the map pattern. As was shown in Tiefelsdorf & Griffith (2007), spatial filtering techniques could be used in both, parametric and semi-parametric approaches. In this paper we employed a parametric one.

The labour market analysed was characterised in detail by two variables: (a) the number of people of working age (18–64 years for men and 18–59 for women) per 100 inhabitants, and (b) the unemployment rate registered in municipalities. Variable (a) represents the supply side of the labour market, i.e., the workforce available, and variable (b), the excess of its supply over demand for it. In addition, we analysed the number of people employed in the urban/metropolitan areas of Tri-City (which consists of Gdańsk, Gdynia and Sopot), Katowice, Cracow, Łódź, Poznań, Warsaw and Wrocław. A separate study was made of Poznań poviat (a unit of the NUTS 4 level).

3. Results

Fig. 1 presents the spatial distribution of the population of working age per 100 inhabitants and the unemployment rate in 2007 in Poland. Visible on the right-hand map is a spatial imbal-

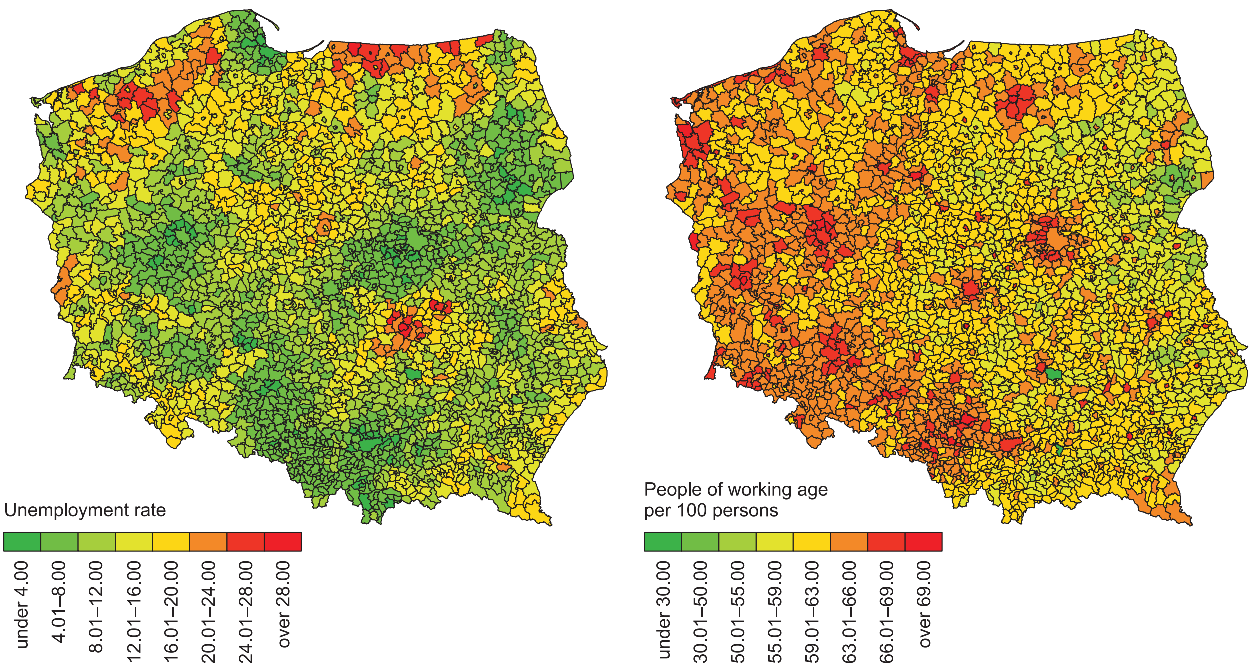

Fig. 1. Unemployment and people of working age.

ance in labour supply on the market as measured by the number of people of working age. The supply of labour increases in the western and northern parts of the country. The highest supply figures can be found in the cities of Poznań with part of Poznań poviat, Wrocław, Cracow, part of the Silesian agglomeration, and Szczecin (not included in the detailed study). All the examined cities are in the two highest classes. From the labour market perspective, the demographic structure is more advantageous in western Poland, which is much younger than in eastern Poland (exceptions being Warsaw, Białystok and some larger cities).

The index of the spatial autocorrelation of the number of people of working age per 100 inhabitants is relatively high and Moran's coefficient is $I = 0.4958$, statistically significant at $p = 0.0000$.

The map on the left in Fig. 1 shows the unemployment rate in Polish municipalities in 2007. The spatial distribution of unemployment refers to the supply of workforce. Areas with a low level of unemployment concentrated along two main axes: Wrocław-Warsaw-Łódź-Białystok and Poznań-Upper Silesia-Cracow, which intersect in the south of Wielkopolska. The Tri-City agglomeration is a separate island. Along the two axes connecting the analysed metropolitan areas, low levels of unemployment coincide with high labour supply. Those are areas where labour supply is almost entirely balanced by demand. Areas in the north of Poland are places with a large imbalance of labour supply and demand for it. This applies especially to Central Pomerania, Warmia and Mazuria, and also eastern Poland. The high supply of labour is not equilibrated here by strong demand for it at present. The result is soaring unemployment, even up to 30%. The spatial autocorrelation of unemployment is even higher than the distribution of people of working age. Moran's coefficient calculated for the 2007 data is $I = 0.7426$ and is statistically significant (at $p = 0.0000$).

Fig. 2 shows the change in the number of employees in the surveyed cities. In the years 1995–2008 there was a general decline in their numbers, except in Warsaw and Poznań poviat. The downward trend is broken by small increases in the years 2004–2007 and generally does not exceed the initial 1995 level. It is only in the case of Wrocław that the 2008 figure exceeds the 1995 one. The leader of the metropolitan labour market is Warsaw with more than 800,000 jobs, and its position is unthreatened. Classified on the second position is Cracow, which exceeded the Katowice figure in 2005. However, only in Katowice did employment surpass 300,000 in the years 1995–1999. The other cities had employment of

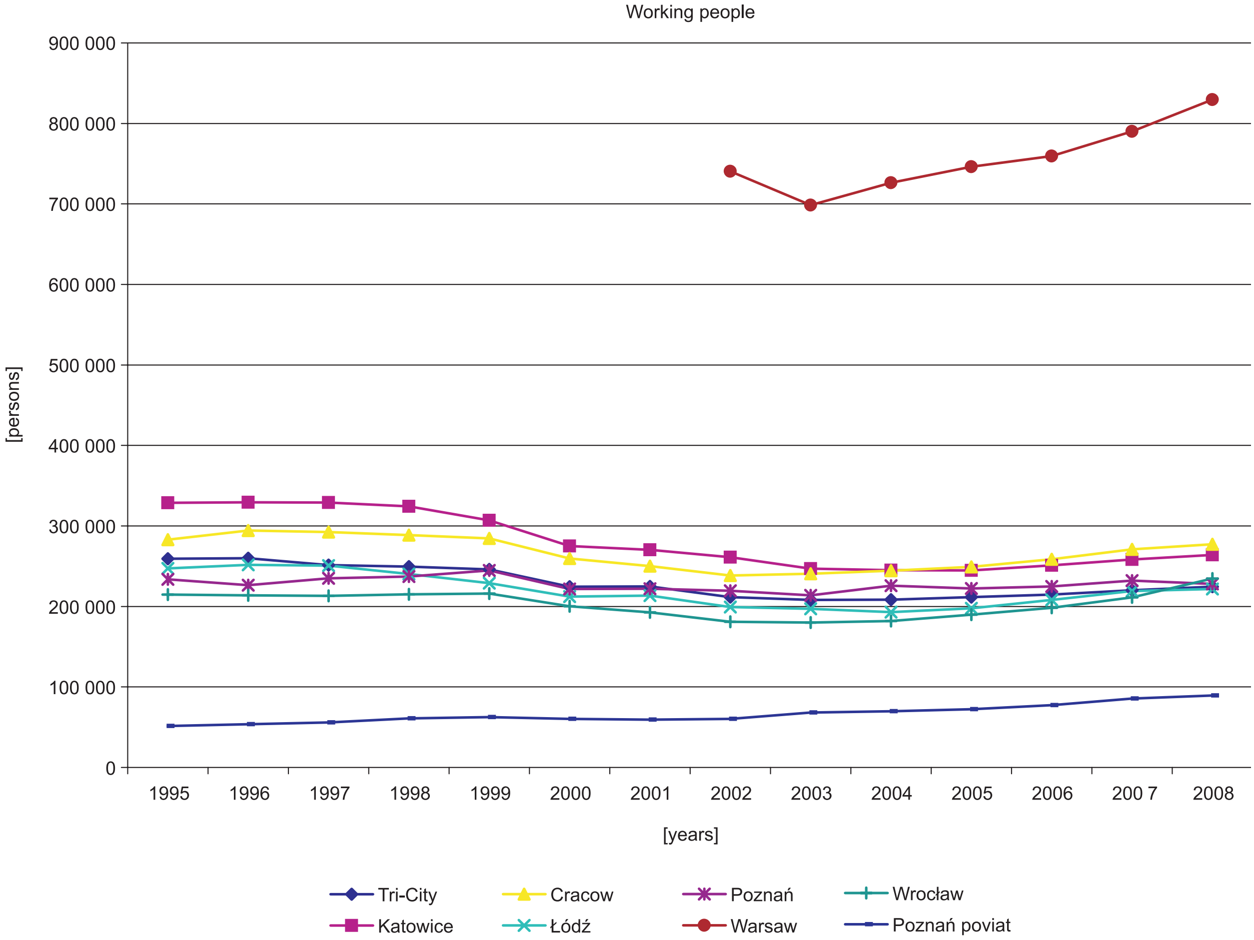

Fig. 2. Working people in metropolises.

200,000–270,000 (apart from Wrocław, where employment dropped periodically to a lower level), and their 2008 figures are almost the same.

In the years 2003–2009 the unemployment rate in the Polish cities varied considerably (see Table 1). Starting from 2003, unemployment kept decreasing until 2008 and then in the last year of the analysed period increased again due to the global financial crisis. The leaders of the metropolitan labour market were Poznań, Warsaw and Poznań poviat, with the level of unemployment rate below 3% in 2009, and even below 2% in 2008. The highest rates were always recorded in Łódź and Katowice, with an unemployment rate below 5% happening only periodically.

Fig. 3 presents the results of an analysis of the local spatial association (LISA). The map on the left shows statistically significant cluster values of Moran's local statistics calculated for the 2007 unemployment rate in communes. The spatial

Table 1. Unemployment rate in Polish metropolises (in %)

City	2003	2004	2005	2006	2007	2008	2009
Tri-City	7.7	6.9	5.7	3.8	2.1	1.5	3.2
Katowice	10.4	9.6	8.7	6.8	4.5	3.0	4.4
Cracow	6.0	5.3	5.0	4.0	2.9	2.1	3.2
Łódź	12.8	12.1	10.8	7.7	5.8	4.6	6.6
Poznań	5.7	5.6	5.1	4.2	2.5	1.5	2.8
Warsaw	5.7	5.8	5.2	4.4	2.9	1.9	2.8
Wrocław	8.7	8.3	7.4	5.5	3.2	2.5	3.9
Poznań poviat	6.4	6.2	5.6	4.2	2.1	1.1	2.3

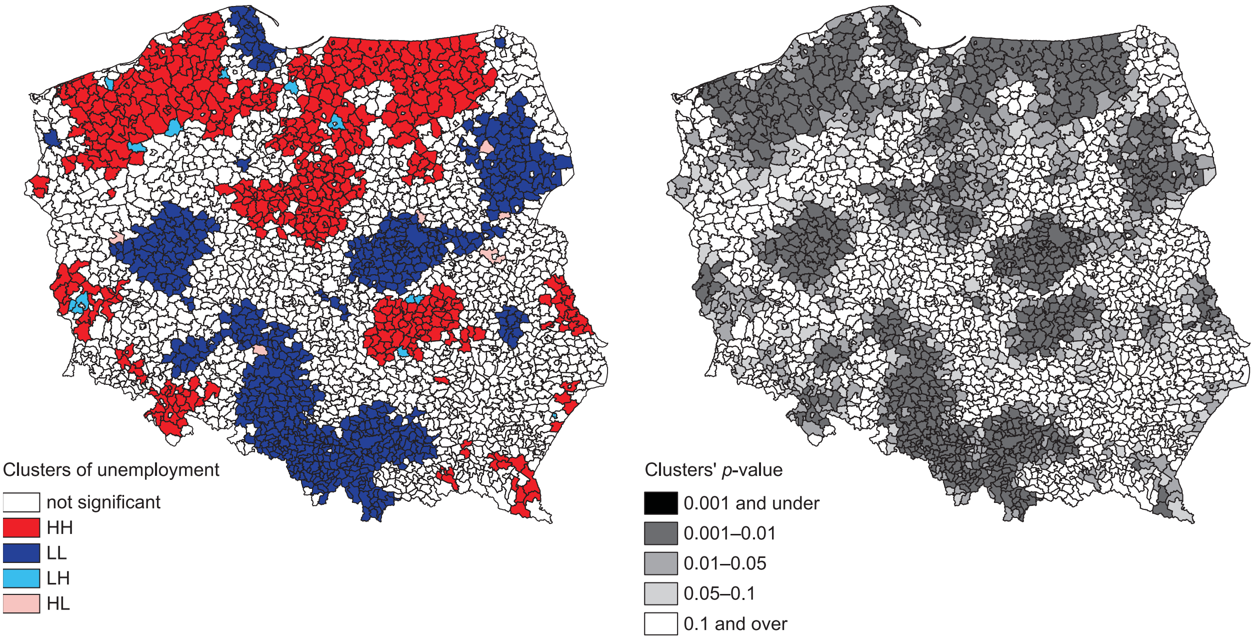

Fig. 3. Clusters of unemployment and their significance.

distribution of clusters and their ranges indicate the strength of agglomeration effects for communes with high and low unemployment rates.

On the basis of the left-hand map in Fig. 3, it can be stated that there is an imbalance of spatial agglomeration effects of the unemployment rate. Clusters of areas of low unemployment rates include southern and western Poland, Warsaw, and Białystok, as well as all the analysed metropolitan cities except Łódź. In turn, northern Poland (except Tri-City) and the Kielce area show clusters with a high rate of unemployment (being often of a structural nature, connected with the fall of state-owned agricultural or the industrial sector).

The map of the statistical significance of Moran's *I* shows the range of metropolitan labour markets. Areas with the smallest *p*-value can be interpreted as the cores of the metropolitan areas. The arrangement of low-unemployment clusters is polarised into east-west by two cores: Poznań with Poznań poviat and some neighbouring communes and poviats, and Warsaw with its adjacent communes. The characteristic of clusters built around Poznań and Warsaw is their extension in a westerly, and to a lesser extent, in a southerly direction. The Białystok cluster extends north-south in the system, and so does the Tri-City cluster. In the latter case, the impact is primarily due to the territorial system of the Polish borders.

The southern cluster covers the largest part of Poland's territory, stretching from the surroundings of Wrocław via southern Wielkopolska and Upper Silesia to Małopolska and Tarnów. There are a few cores in this cluster: Wrocław, the Upper Silesia agglomeration with Bielsko-Biała, and Cracow. An important characteristic of the distribution of this cluster is that its western part is connected with the main area by the shortest line via Opole, but also via the border between Wielkopolska, Małopolska and Silesia, in the immediate vicinity of Wieluń and Bełchatów. Other clusters of low unemployment rates are located in the Lublin area and south of Łódź. Their sizes do not make any change in the cluster system of low unemployment rates.

Clusters of high unemployment covering north-eastern Poland practically form one continuous area broken by the Tri-City metropolitan area with a low level of unemployment. This cluster extends far to the south, covering the region of Kujavia and forming a wide wedge between the clusters of Warsaw and Poznań.

The second large cluster of high unemployment is Świętokrzyska Land, adjacent from the north to the Warsaw cluster. Other high unemployment clusters are located near the borders (Hrubieszów in the east, Kłodzko, the Sudeten foreland in the south, and the Lubuska-Lusatia region in the west). The cores of those clusters are

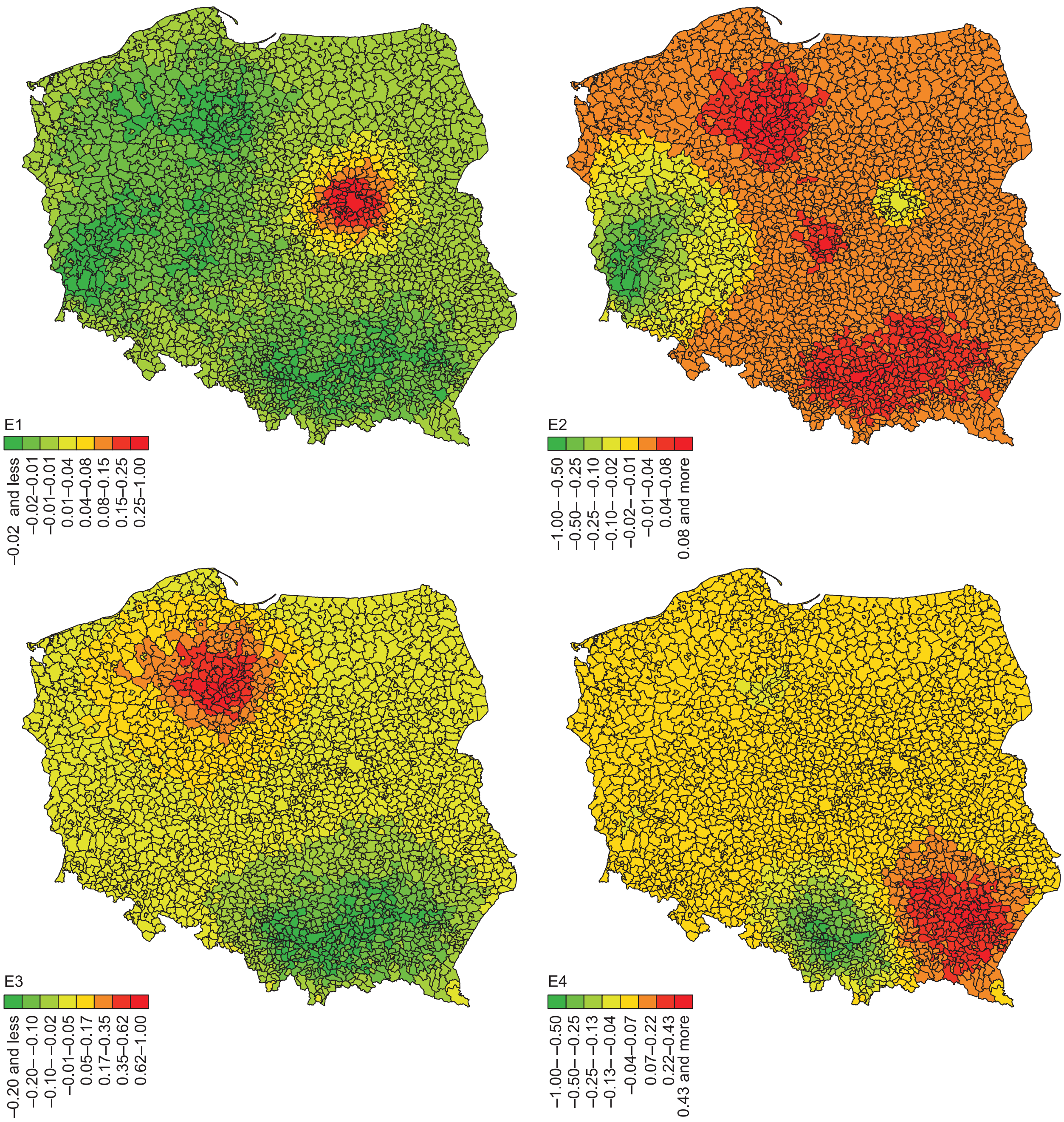

Fig. 4. Eigenvectors E_1-E_4.

mostly communes located in rural areas far away from major cities.

The next step of the analysis involved the procedure of the spatial filtering of variables in order to characterise the unemployment rate. First, the spatial weights matrix MC_sM was modified. Then its 2,478 eigenvectors were calculated. For the members of the set of those vectors $E(MC_sM)$ spatial autocorrelations of their values $I(E_1)$, $I(E_2)$, ..., $I(E_{2478})$ were calculated. From the set we selected those elements that satisfied the condition:

$$\frac{I(E_i)}{\max\limits_{i=1,\ldots,2478} I(E_i)} > 0.25$$

In this way, the set of eigenvectors was reduced to 576 elements.

Examples of the spatial distributions of values of eigenvectors E_1-E_4 are shown in Fig. 4.

The interpretation of the eigenvectors is as follows: vector E_1 attains a maximum value in the Warsaw metropolis and at the same time has the largest spatial autocorrelation. It can be called the Warsaw component of the labour market, with a dominant role in moulding the spatial structure of unemployment in Poland. Eigenvectors E_2-E_3 show the polarising components of the labour market in Poland in the east-west and the north-south system, while eigenvector E_4 shows the hidden structure of spatial dependence in the core-periphery system. As has already been mentioned, "These components can be seen as independent map patterns, and represent the latent spatial autocorrelation of a georeferenced variable concerned" (Patuelli et al. 2006: 2). In the next step, an analysis of the linear regression of variable U_{2007} (unemployment) was performed on the reduced $E^*(MC_sM)$ set of eigenvectors of the modified weights matrix, and parameters of the model were estimated. The model was:

$$U_{2007} = \sum_{j=1}^{576} \alpha_j E_j^* + \varepsilon$$

The estimation of its parameters was based on OLS estimators. A stepwise regression method, implemented under the Matlab program in the stepwisefit procedure, was used for calculations. The critical p-value for the variable added to the model was 0.05, and for the deleted variable, 0.1. The finally obtained model consisted of 212 statistically significant explanatory variables (eigenvectors E*). Its simplified diagnostics is presented in Table 2; the test results for α_i parameters are not shown because of their large number (212).

Table 2 shows that the resulting model is relatively good. The coefficients of determination – both standard and adjusted – exceed 0.7. The specification of the model is sufficiently correct, as evidenced by the highly significant value of F statistics. Next, we calculated the values of the spatial filter $SF(E^*)$, which is a combination of the 212 statistically significant linear eigenvectors, and the values of spatial filter residuals e from the formula:

Table 2. Model diagnosis

Statistics	Value
SS_R	13,654
SD_R	2.45525
R^2	0.74942
Adjusted R^2	0.72597
$F(212, 2265)$	31.9529 ($p < 0.00001$)
LogLik	–5,630.58
Akaike	11,687.2
Schwarz	12,925.8
Hannan-Quinn	12,137.1

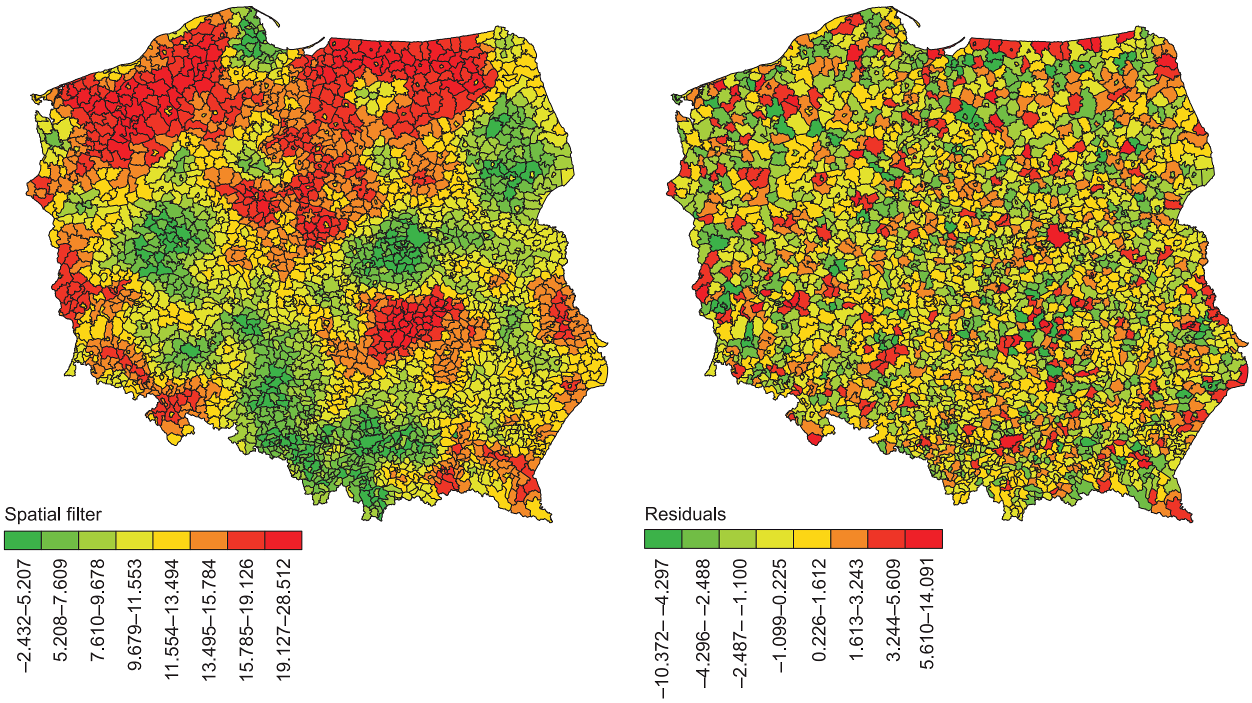

Fig. 5. Spatial filter and residuals

$$U_{2007} = SF(E^*) + e$$

The spatial distribution of the values of the spatial filter and the residual components are shown in Fig. 5. The distribution of the spatial filter is related to the decomposition of the unemployment rate, but it is more contrasted. The values of the spatial filter in communes range from about –2.5 to just over 28.5, while the unemployment rate cannot take any negative values. The filter assumes negative values for metropolitan areas, underestimating the true level of unemployment.

We compared the values of Moran's *I* for unemployment in 2007 (raw data), for the spatial filter, and for the spatial residuals. The results are presented in Table 3.

The autocorrelation analysis of the unemployment rate, the filter zoning and the residual component allows us to state that the level of spatial dependence of the unemployment rate can be explained almost entirely by the eigenvectors of the modified weights matrix acting as the spatial filter. It is worth mentioning that the spatial autocorrelation of values of the spatial filter is close to unity, but the autocorrelation of residuals is not statistically significant

The map of the spatial distribution of the residual components (Fig. 5) confirms their random distribution, but with one exception. In the metropolitan areas studied, as already mentioned, the spatial filter assumed negative values. For those spatial units, the value of the residuals are positive on the map.

Table 3. Comparison of Moran's *I* coefficients

Data	Moran's *I*
2007 unemployment (raw data)	0.7426
2007 spatial filter residuals	–0.0075
2007 spatial filter	0.9715

4. Conclusions

In this article we presented an analysis of the labour market in Poland, with particular emphasis on selected metropolitan areas. There are significant disparities between the labour market resources of communes (the number of people of working age) and their unemployment rates. It is only in metropolitan areas that labour finds sufficient employment. On the other hand, it revealed spatial imbalance in the unemployment rate when examined by geographical directions, metropolitan areas, and their surroundings. The analysis of spatial autocorrelation allows the conclusion that there is a strong relationship among selected variables characterising the Polish labour market. Particularly strong are spatial relations in the unemployment rate, revealed by LISA analysis. This approach has led to questions about the identification of spatial structures depending on the unemployment rate. The spatial maps of modified eigenvectors of the spatial weights matrix obtained through the use of filtering techniques revealed the spatial structure of unemployment to have hidden dimensions. As shown in Fig. 4, such structures exist: one is centred on Warsaw and the others extend along the east-west, north-south, and core-periphery lines. The remaining eigenvectors show the spatial structure of the local pattern of dependence.

On the basis of the eigenvectors of the modified weights matrix, a spatial filter was constructed and spatial filter residuals were calculated. The spatial filter showed spatial autocorrelation to be close to unity, while residual autocorrelation was absent. It follows that the eigenvectors representing the hidden structure of the autocorrelation give a good explanation and allow a better identification of the spatial dimension of the unemployment rate in Poland.

References

Anselin L., 1988. *Spatial econometrics: Methods and models*. Kluwer, Dordrecht.

Anselin L., 1995. Local Indicators of Spatial Association – LISA. *Geographical Analysis*, 27(2): 93–115.

Boots B., 2003. Developing local measures of spatial association for categorical data. *Journal of Geographical Systems*, 5: 139–160.

Getis A., Griffith D.A., 2002. Comparative spatial filtering in regression analysis. *Geographical Analysis*, 34: 130–140.

Griffith D.A., 1981. Towards a theory of spatial statistics: A rejoinder. *Geographical Analysis*, 13: 91–93.

Griffith D.A., 1988. *Advanced spatial statistics*. Kluwer, Dordrecht.

Griffith D.A., 1996. Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying georeferenced data. *The Canadian Geographer*, 40: 351–367.

Griffith D.A., 2000. A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems*, 2: 141–156.

GRIFFITH D.A., 2003. *Spatial autocorrelation and spatial filtering*. Springer, Berlin.

JONG P. DE, SPRENGER C., VEEN F. VAN, 1984. On extreme values of Moran's *I* and Geary's *c*. *Geographical Analysis*, 16(1): 17–24.

PATUELLI R., GRIFFITH D.A., TIEFELSDORF M., NIJKAMP P., 2006. The use of spatial filtering techniques: The spatial and space-time structure of German unemployment data. *Tinbergen Institute Discussion Papers*, 06–049/3.

PATUELLI R., GRIFFITH D.A., TIEFELSDORF M., NIJKAMP P., 2009. Spatial filtering and eigenvector stability: Space-time models for German unemployment data. *Quaderni della facolta di Scienze economiche dell'Universita di Lugano*, 0902.

TIEFELSDORF M., GRIFFITH D.A., 2007. Semiparametric filtering of spatial autocorrelation: The eigenvector approach. *Environment and Planning*, A, 39(5): 1193–1221.