

## A SPATIAL AND TEMPORAL ANALYSIS OF LABOUR MARKET CHARACTERISTICS

---

Ewa Pośpiech, Ph.D., Eng.

*University of Economics in Katowice  
Faculty of Management  
Department of Mathematics  
1 Maja 50, 40-287 Katowice, Poland  
e-mail: ewa.pospiech@ue.katowice.pl*

Adrianna Mastalerz-Kodzis, Ph.D.

*University of Economics in Katowice  
Faculty of Management  
Department of Mathematics  
1 Maja 50, 40-287 Katowice, Poland  
e-mail: adrianna.mastalerz-kodzis@ue.katowice.pl*

Received 2 December 2015, Accepted 16 November 2016

---

### Abstract

The use of spatial methods is becoming increasingly common in social and economic research as it emphasizes the relevance of spatiality to the understanding of socio-economic facts. Once embraced, the spatial factor can substantially help explain variations in the properties being examined, thus improving the quality of their description and supporting the development of econometric models. This paper explores some of the characteristics of Poland's job market, making an inquiry into their spatial dependencies. The study looks at the country's labour market from a local perspective, examining its properties for spatial autocorrelation (both global and local). Linear econometric models are subsequently built for such variables as the number of persons in employment, the number of women and men in employment. The models are further investigated to assess the applicability of spatial modelling in their development.

**Keywords:** spatial autocorrelation, global statistics, local statistics, spatial modelling, labour market characteristics

**JEL classification:** C23, C50, E24, J10

## **Introduction**

Spatial methods are increasingly often used in studying social and economic facts, as many of these facts stem from regional or local policies and therefore tend to vary geographically. The development of spatial methods started in the 1970s (Tobler, 1970; Cliff, Ord, 1973; Paelinck, Klaassen, 1979). These methods can be used to analyse issues, the formation of which is dependent on location and which are influenced by spatial interactions (demographic processes, real estate market, ecological and socio-economic problems etc.). Therefore, many researchers deal with spatial analyses (Overmars, de Koning, Veldkamp, 2003; Sikora, Woźniak, 2007; Anselin, 2010; Pietrzykowski, 2011; Zeug-Żebro, 2014; Pośpiech, 2015; Pośpiech, Mastalerz-Kodzis, 2015). It seems that there is a growing awareness of the relevance of spatiality to the understanding of socio-economic realities and that embracing the spatial factor can be key to explaining variations in the properties being examined, thus improving the quality of their description. The incorporation and deployment of spatial methods can also help: identify spatial autocorrelation, hence determine the spatial dependence of certain variables at a given location as well as linkages to and across adjacent areas; explore spatial concentrations; conduct structural-geographic analyses; and, last but not least, it will support the development of spatial econometric models (Sucheckki, 2010).

The paper seeks to identify the spatial dependences of labour market characteristics across Poland's districts (indicative of the working population: its total as well as its gender breakdown) and to assess to the applicability of spatial methods to building econometric models for such variables.

### **1. Research methods applied**

Selected spatial econometric and statistical tools and methods were used to complete the empirical part of the research. Global and local Moran statistics were employed to identify spatial dependences at a single location and across adjacent locations. Further, the resulting models were examined using e.g. the residual Moran's *I* test, join-count statistics, and LM tests designed to capture spatial dependences under linear models.

#### **1.1. Measuring global and local spatial autocorrelation**

The research involved global and local Moran spatial statistics. The global statistic was calculated from the following formula (Cliff, Ord, 1973; Sucheckki, 2010):

$$I = \frac{1}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \times \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{n}{S_0} \times \frac{\mathbf{z}^T \mathbf{W} \mathbf{z}}{\mathbf{z}^T \mathbf{z}} \quad (1)$$

where:

$x_i, x_j$  – observations made in regions covered by the analysis,

$\bar{x}$  – the arithmetic mean (for all regions and all observations),

$n$  – the number of regions under investigation,

$\mathbf{z}$  – the column vector of values expressed as:  $z_i = x_i - \bar{x}, i = 1, \dots, n$ ,

$S_0$  – the sum-total of weights matrix elements,  $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$ ,

$\mathbf{W}$  – the weights matrix  $w_{ij}, i = 1, \dots, n, j = 1, \dots, n$ , comprising elements

$$w_{ij} = \begin{cases} 1, & \text{when objects } i \text{ and } j \text{ share a border} \\ 0, & \text{when objects } i \text{ and } j \text{ do not share a border} \\ 0, & \text{for diagonal elements} \end{cases} \quad (2)$$

Hypotheses on the absence ( $H_0$ ) or presence ( $H_1$ ) of spatial autocorrelation can be validated through testing, where the test statistic  $Z_I \sim N(0, 1)$  is expressed by the following formula:

$$Z_I = \frac{I - E(I)}{\sqrt{\text{Var}(I)}} \quad (3)$$

in which  $E(I)$  and  $\text{Var}(I)$  are defined as follows:

$$E(I) = -\frac{1}{n-1} \quad (4)$$

$$\text{Var}(I) = \frac{n[(n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2] - k[(n^2 - n)S_1 - 2nS_2 + 6S_0^2]}{(n-1)(n-2)(n-3)S_0^2} - \frac{1}{(n-1)^2} \quad (5)$$

and

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}, \quad S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2, \quad S_2 = \sum_{i=1}^n \left( \sum_{j=1}^n w_{ij} + \sum_{j=1}^n w_{ji} \right)^2 \quad (6)$$

$$k = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} \quad (7)$$

Values adopted by the statistic  $Z_I$  (equal, greater or lower than zero) stand for, respectively:

- no spatial autocorrelation (random distribution of the variable across the locations in question),
- positive autocorrelation (similar values tend to cluster together at a certain location),
- negative autocorrelation (high and low values are contiguous).

The local Moran  $I$  statistic can be utilized to identify clusters, i.e. areas where high or low values of the variable of interest are concentrated, as well as to detect atypical locations known as outliers. The statistic is computed using the following formula:

$$I_i = \frac{(x_i - \bar{x}) \sum_{j=1}^n w_{ij}^* (x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (8)$$

where:

$w_{ij}^*$  – elements of the weights matrix, given from the formula  $w_{ij}^* = \frac{w_{ij}}{\sum_{j=1}^n w_{ij}}$ , other symbols as in the above.

Assuming an asymptotic normal distribution of the statistic  $Z_{I_i}$  expressed as:

$$Z_{I_i} = \frac{I_i - E(I_i)}{\sqrt{\text{Var}(I_i)}} \quad (9)$$

whose estimated expected value and variance (Anselin, 1995) are given from the formulas shown below:

$$E(I_i) = -\frac{\sum_{j=1}^n w_{ij}}{n-1}, \quad \text{Var}(I_i) = \frac{(n-k) \sum_{j \neq i} w_{ij}^2}{n-1} + \frac{2(2k-n) \sum_{k \neq i} \sum_{l \neq i} w_{ik} w_{il}}{(n-1)(n-2)} - \frac{\left( \sum_{j=1}^n w_{ij} \right)^2}{(n-1)^2} \quad (10)$$

A positive statistically significant value of the statistic  $I_i$  substantiates the conclusion that the region  $i$  is surrounded by regions where a given property takes similar values (i.e. similar values are clustered), while a negative statistically significant value of the statistic indicates that a particular location is surrounded by areas where the property has clearly different values (outlier).

## 1.2. Selected tools for spatial model identification

To ascertain and to find out whether spatial estimation methods need to be used, the model should be examined for spatial properties. The first step in identifying the need for spatial modelling is to investigate the spatial autocorrelation of a given model's residuals, which is estimated via the values of selected variables as well as their logarithms. This step can be accomplished owing to the residual Moran's  $I$  test that is defined as follows:

$$I = \frac{n}{S_0} \times \frac{\mathbf{u}^T \mathbf{W} \mathbf{u}}{\mathbf{u}^T \mathbf{u}} \quad (11)$$

where  $\mathbf{u}$  is an  $n$ -element column vector of the model's residuals (the model is estimated with the use of the OLS method). The null hypothesis proposes that there is no spatial autocorrelation in the model residuals, whereas the alternative hypothesis presupposes that there is. The analysis is supported by a graphical representation showing the spatial distribution of the residuals. A non-random distribution of the residuals suggests that there would be other factors affecting the explained variable. Further, positive residuals for a specific region indicate that the model underestimates the values taken by a specific characteristic, while negative residuals signal that the model overestimates them. Positive and negative residuals can be explored by running the join-count test that is used to check the randomness of residuals from a regression model.<sup>1</sup>

Where diagnostic tests for linear models detect spatial autocorrelation in residuals, it can be assumed that spatial estimation should be applied. Two types of spatial models are mainly distinguished: spatial lag and spatial error models. Selection between them is based on the findings of LM tests (LMlag, LMerr can be performed alongside their robust variants: RLMlag, RLMerr) and an appropriate decision rule.<sup>2</sup>

## 2. Empirical analysis

The empirical analysis was conducted for the period of 2005–2013 and involved variables relating to job market characteristics as well as demographic ones (Marciniak, 2013). The spatial autocorrelation analysis was geared to find out about the following:

- the proportion of working population to total population (designated as UP),
- the percentage of women in the working population (UK),
- the percentage of men in the working population (UM).

<sup>1</sup> A discussion of the method and the statistics used can be found e.g. in Suchecki (2010).

<sup>2</sup> The method is described at length in Kopczewska (2011) and Suchecki (2010).

The follow-up analyses aiming to verify the need to build a spatial model looked into the following variables:

- the number of persons in employment (PO),
- the number of women in employment (PK),
- the number of men in employment (PM),
- the average gross monthly pay (PW),
- the age dependency ratio (OD),
- the population density – inhabitants per square kilometer (G),
- the population growth rate (N),
- the population below working age (Lpr),
- the population at working age (L),
- the population above working age (Lpo).

Input data were sourced from GUS<sup>3</sup> Local Data Bank, while the graphics and computations presented in the paper were developed using the MS Excel and the R CRAN package (Kopczewska, 2011; Kopczewska, Kopczewski, Wójcik, 2013).

## 2.1. Spatial autocorrelation analysis results

Table 1 shows the values of the global Moran statistic produced for specific poviats.<sup>4</sup>

Table 1. Global Moran's *I* statistics for poviats

Year	Global Moran's <i>I</i> statistics (UP)		Global Moran's <i>I</i> statistics (UK)		Global Moran's <i>I</i> statistics (UM)	
	I	p-value	I	p-value	I	p-value
2005	0.0190	0.2627	0.1191	0.0002	0.1191	0.0002
2006	0.0151	0.3004	0.1148	0.0003	0.1148	0.0003
2007	0.0151	0.3012	0.1307	4.8E-05	0.1307	4.8E-05
2008	0.0407	0.1011	0.1581	1.3E-06	0.1581	1.3E-06
2009	0.0500	0.0609	0.1536	2.4E-06	0.1536	2.4E-06
2010	0.0477	0.0698	0.1612	8.2E-07	0.1612	8.2E-07
2011	0.0502	0.0599	0.1342	3.1E-05	0.1342	3.1E-05
2012	0.0597	0.0335	0.1163	0.0003	0.1163	0.0003
2013	0.0554	0.0440	0.1197	0.0002	0.1197	0.0002

Source: own calculations.

<sup>3</sup> GUS (Główny Urząd Statystyczny – Central Statistical Office) is Poland's chief government agency for national statistics.

<sup>4</sup> In Poland, powiat is the second-level unit of local government, consisting of several municipalities (communes) and equivalent to a county or district in other countries.

The global Moran's  $I$  statistic for the UP ratio (percentage of working population in total population) in 2005–2011 takes positive but low and statistically insignificant values. The statistics only appeared statistically significant in the final two years of the period under examination, revealing a very weak positive spatial autocorrelation (a generally random distribution of the values between locations). The global Moran's  $I$  statistic for the UK and the UM ratios reflecting the proportion of women and men in the working population are positive and statistically significant throughout the period, although low values of the  $I$  statistic denote a weak spatial autocorrelation (not much clustering of poviats with similar variable values at a single location).

The findings of an analysis of local Moran's  $I_i$  statistics are shown in Figures 1 and 2. The graphics depict clusters and outliers (atypical areas) for the initial and the final year in the period being investigated.

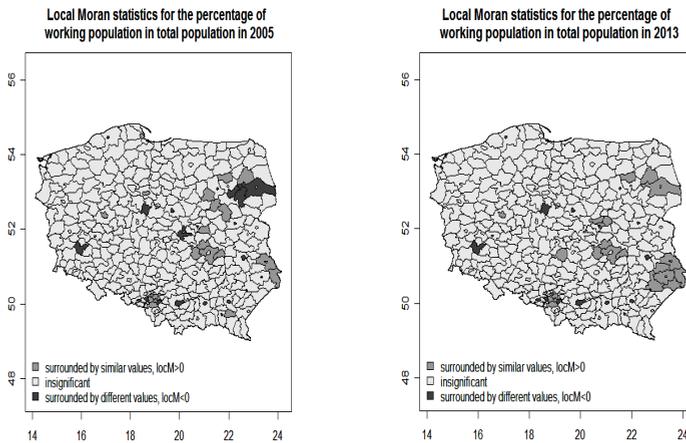


Figure 1. Local statistics for the percentage of the working population in total population

Source: own.

For the proportion of working population to total population, there are few clusters and only a small number of outliers to be seen. Among poviats that are atypical with regard to their neighbouring districts in both the years there are e.g. the city of Kraków (greater proportions of working population compared to its neighbors – 32.9 and 39.3%, respectively), the poviats of Radziejów (much lower values than in adjacent poviats – 9.8% in 2005 and 11.7% in 2013), and the poviats of Polkowice (considerably greater values than in adjoining poviats – 40.5 and 45.9%, respectively). More outliers can be observed in 2005, while more conspicuous clusters

of areas surrounded by ones with similar values can be seen in 2013. Attention should be given to the clusters of poviats situated in the country's southern (with the proportion ranging between 25 and 30%) and central-eastern part (where the proportion of the working population is mostly below 10%).

The remaining two properties are closely inter-related – the clusters and the outliers roughly overlap, yet are to be differently interpreted: an outlier with high values of one of the properties is an outlier with low values for the other, and vice versa (the same is true for clusters). Figure 2 illustrates the locations of clusters and outliers in the first and final year of the period being analysed. A smaller number of clusters and outlying poviats in 2013 implies a more random distribution of the percentages of working women and men in the working population. Instances of areas that stand out from their neighbouring districts include the poviat of Dąbrowa Górnicza (63 and 61.6% of working men and, respectively, 37 and 38.4% of working women) and the poviat of Łęczna (62.5 and 67.2% for men, 37.5 and 32.8% for women, respectively). A cluster example, corresponding to areas with a high percentage of working men (oscillating around 60%), can be found in the Śląskie voivodeship (region).

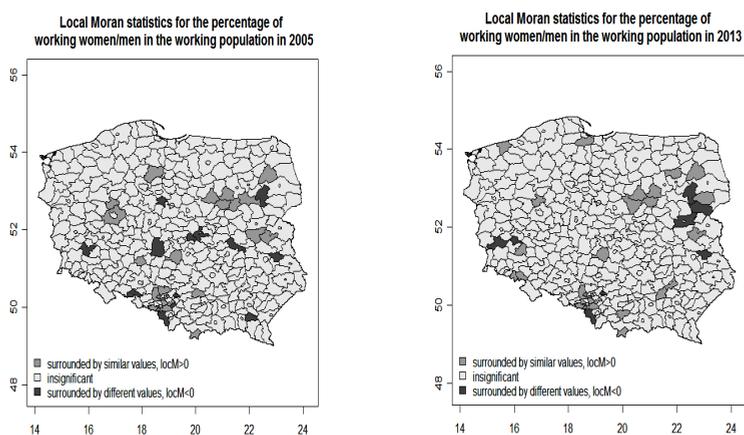


Figure 2. Local statistics for the percentage of working women/men in the working population  
Source: own.

## 2.2. Outputs of spatial model identification

The next step was to develop a model and tentatively assess the applicability of spatial modelling to linear models (such as those built for the year 2013) in which the explained

(output) variables were as follows: the number of persons in employment (PO), the number of women in employment (PK), and the number of men in employment (PM). The explanatory (input) variables were those listed at the outset of subchapter 2. The variable G coefficient (population density) proved statistically insignificant under all models; hence it is omitted in further discussions. Table 2 brings together the estimates produced by the econometric models<sup>5</sup> along with the coefficient of determination  $R^2$ .

Table 2. Estimates produced by the econometric models

Variables	Coefficients values		
	(PO) model	(PK) model	(PM) model
Intercept	1.649E+04	1.336E+04	3.131E+03
PW	3.576	0.772	2.804
OD	-1,024	-575.6	-448.4
N	37.59	19.72	17.88
Lpr	-1.703	-0.719	-0.983
L	-0.525	-0.387	-0.138
Lpo	4.629	2.656	1.973
$R^2$	0.983	0.986	0.976

Source: own calculations.

The resulting models are characterized by a very high coefficient of determination – a condition that implies their positive assessment. Model coefficients show the influence of the explanatory variables on the explained variable. In each model the strongest positive influence on the explained variable had the population growth rate (N) whilst the strongest negative influence had the age dependency ratio (OD).

Next, the spatial autocorrelation of errors was examined. The outcomes are presented in Table 3.

Table 3. Moran statistics for model residuals

Year	(PO) model		(PK) model		(PM) model	
	I	p-value	I	p-value	I	p-value
2013	0.0399	0.0965	0.0429	0.0820	0.0540	0.0429

Source: own calculations.

<sup>5</sup> Models based on logarithms of the variables were not tested, since one of the variables, namely the population growth rate (N), has negative values.

The (PM) model is the only one whose Moran statistic is statistically significant, which might suggest that some (albeit weak) spatial autocorrelation exists. Figure 3 provides a graphical representation of the residuals distribution.

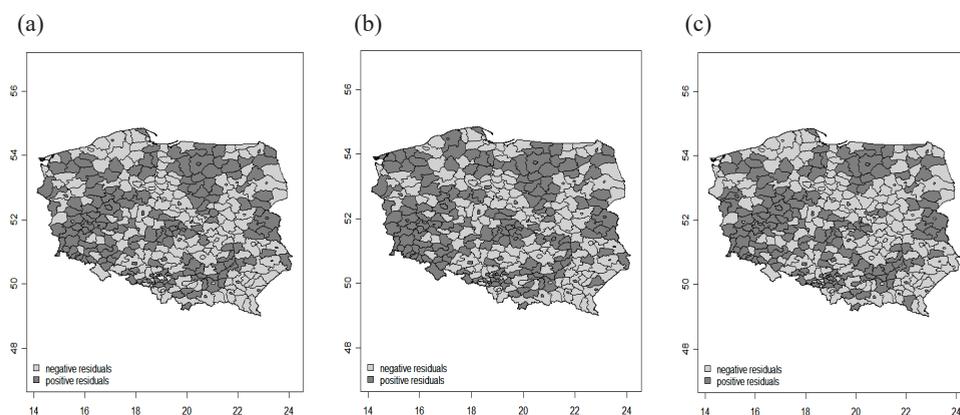


Figure 3. Positive and negative residuals for models: PO – (a), PK – (b), PM – (c)

Source: own.

It can be thus presumed that in the first two models the residuals are distributed randomly, that is, they do not depend on location. The weak spatial autocorrelation for the PM model indicates that its residuals are somewhat dependent on location.

Eventually, the residuals were sorted into positive-valued and negative-valued ones and subjected to the join-count test. The test demonstrated that the positive residuals of the first two models are distributed randomly, while their negative counterparts and both the positive and negative residuals of the third (PM) model are characterized by a degree of spatial autocorrelation. The LM tests that were subsequently performed showed that the most adequate spatial model for each of the variables being investigated would be that of spatial lag.

## Conclusions

The research presented in the paper was primarily focused on identifying spatial dependences for selected variables relating to Poland's job market: the percentage of the working population in total population and the proportions of women and men in the working population. The analysis spanned local government districts of Poland (known as poviats) and revealed a weak spatial autocorrelation throughout the period being examined as well as the

presence of a small number of clusters and atypical regions in the initial and final years. Next, linear models for three variables – those of total working population, number of women in employment, and number of men in employment – were estimated, mainly using demographic data as explanatory variables. The models were then pre-examined for the extent to which the use of spatial estimation methods would be necessary to describe a given variable. Weak spatial dependences were discovered, and the LM tests for spatial dependences that were subsequently performed pointed to the spatial lag model as an adequate solution to be applied in interpreting the data. It must be, however, realized that the findings are of a tentative nature and that further research, based on a more complete specification of the econometric model, is required to assess the applicability of spatial modelling more accurately. Nevertheless, it can be stated even at this point that the phenomena reflected by the job market variables do exhibit certain spatial relationships and that therefore using spatial methods to address these issues is very likely to be a sound approach.

## References

---

- Anselin, L. (1995). Local Indicators of Spatial Association – LISA. *Geographical Analysis*, 27 (2), 93–115. DOI: 10.1111/j.1538-4632.1995.tb00338.x.
- Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, 89 (1), 3–25. DOI: 10.1111/j.1435-5957.2010.00279.x.
- Cliff, A.D., Ord, J.K. (1973). *Spatial Autocorrelation*. London: Pion.
- Kopczewska, K. (2011). *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R Cran*. Warszawa: CeDeWu.
- Kopczewska, K., Kopczewski, T., Wójcik, P. (2013). *Metody ilościowe w R. Aplikacje ekonomiczne i finansowe*. Warszawa: CeDeWu.pl.
- Marciniak, S. (ed.) (2013). *Makro- i mikroekonomia. Podstawowe problemy współczesności*. Warszawa: Wydawnictwo Naukowe PWN.
- Overmars, K.P., de Koning, G.H.J., Veldkamp, A. (2003). Spatial autocorrelation in multi-scale land use models. *Ecological Modelling*, 164, 257–270. DOI: 10.1016/S0304-3800(03)00070-X.
- Paelinck, J.H.P., Klaassen, L.H. (1979). *Spatial Econometrics*. Farnborough: Salon House.

- Pietrzykowski, R. (2011). Wykorzystanie metod statystycznej analizy przestrzennej w badaniach ekonomicznych. *Roczniki Ekonomiczne Kujawsko-Pomorskiej Szkoły Wyższej w Bydgoszczy*, 4, 97–112.
- Pośpiech, E. (2015). Analiza przestrzenna bezrobocia w Polsce. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, 227, 59–74.
- Pośpiech, E., Mastalerz-Kodzis, A. (2015). Autokorelacja przestrzenna wybranych charakterystyk społeczno-ekonomicznych. *Metody ilościowe w badaniach ekonomicznych*, 16 (4), 85–94.
- Sikora, J., Woźniak, A. (2007). Autokorelacja przestrzenna wskaźników infrastruktury wodno-ściekowej woj. małopolskiego. *Infrastruktura i ekologia terenów wiejskich*, 4 (2), 315–329.
- Suchecki, B. (ed.) (2010). *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*. Warszawa: C.H. Beck.
- Tobler, W. (1970). A Computer Model Simulating Urban Growth in Detroit Region. *Economic Geography*, 46 (2), 234–240.
- Zeug-Żebro, K. (2014). Analiza przestrzenna procesu starzenia się polskiego społeczeństwa. *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania*, 36 (2), 441–456.
- Data base: Główny Urząd Statystyczny, Bank Danych Lokalnych [Central Statistical Office, Local Data Bank]. Available at: <http://www.stat.gov.pl> (accessed Aug 25, 2015).