GEOGRAPHICAL DISTRIBUTION OF UNEMPLOYMENT: AN ANALYSIS OF PROVINCIAL DIFFERENCES IN ITALY

M. Francesca Cracolici[ab*], Miranda Cuffaro[a] and Peter Nijkamp[b]

[a]*Dipartimento di Contabilità Nazionale e Analisi dei Processi Sociali*
Università di Palermo
V.le delle Scienze Ed. 2
90128 Palermo, Italy

[b]*Department of Spatial Economics*
Free University
De Boelelaan 1105
1081 HV Amsterdam, The Netherlands

Pn208fcmcNEW

Abstract

Unemployment rates appear to vary widely at the subregional (e.g. local or provincial) level. Using spatial econometric models for spatial autocorrelation, this paper focuses attention on the spatial structure of regional unemployment disparities of Italian provinces. On the basis of findings from the economic literature and of the available socio-economic data, various model specifications including different explanatory variables are tested to investigate the geographical distribution of unemployment in the 103 provinces of Italy for the year 2003. The results suggest that there is a significant degree of spatial dependence among labour markets at the provincial level in Italy. Provinces marked by high unemployment, as well as those characterized by low unemployment, tend to be spatially clustered, demonstrating the presence of 'spatial persistency'.

JEL classification: C21, J60, J64, R12, R23

Keywords: regional unemployment, ESDA, spatial lag model, Italian provinces

[*] Corresponding author, tel.: +39–091–6561508, fax: +39–091–6561501.
E-mail: cracolici@unipa.it (M.F. Cracolici), cuffaro@unipa.it (M. Cuffaro), pnijkamp@feweb.vu.nl (P.Nijkamp).
The authors wish to thank Roberto Patuelli for his useful comments on a previous version of the paper.

1 Setting the Scene

Geographic unemployment rates are often regarded as signposts for the socio-economic performance of regions. And, consequently, the analysis of regional unemployment differences has attracted increasing interest in the economic literature. Despite this interest, regional unemployment disparities do not represent the exclusive *core* of theories on regional economic development; most studies concentrate, principally, on growth and convergence of per capita income (see Meliciani 2006). Also the *new economic geography* – according to which multiple equilibria may exist – focuses attention on income rather than on unemployment (Fujita et al. 1999). Nevertheless, there is an abundance of empirical literature that tries to explain the differences between geographical areas in terms of unemployment rates (see, e.g., Decressin and Fatás 1995; Jimeno and Bentolila 1998; López-Bazo et al. 2000). These empirical studies have brought to light some interesting stylized facts, notably: a) regional labour markets in Europe and the US differ significantly; b) regional differences in unemployment in European regions are more persistent than in the US; c) the persistence of unemployment differences in European regions is mainly due to poor flexibility of wages and low mobility of workers. In particular, in Italy, as in several other European areas, the persistence of unemployment is due to both structural problems and the inability of Italian regions to absorb specific shocks (on the demand or on the supply side) (for details, see Dohse et al. 2002).

The functioning of regional labour markets has been the subject of intensive research in the regional economics literature (see, e.g., Fischer and Nijkamp 1987; Longhi 2005; Longhi et al. 2005). Taylor and Bradley (1997) state in a comparative empirical study that disparities between regional labour markets in Italy, Germany and the UK are more marked than unemployment disparities between each of these countries and other European areas.

In another study, Martin (1997) finds that the persistence of unemployment rates in some European countries such as the UK, France and Italy, is very different from the situation in the US, where areas with high unemployment in period t tend to have a lower rate in the next period (i.e. in period $t+1$) (see also Bertola and Ichino 1996). The persistence of unemployment differences – in other words, the stability of the regional structure of unemployment in European areas – could be due to the low mobility of the production factor labour. The spatial distribution of unemployment rates may thus be related to the different characteristics of geographical areas. In particular, "…*unemployment rates are determined by the preferences of firms and workers, given fixed endowments of land and amenities in the area*" (Marston 1985: 59). Molho (1995: 642) referred to this phenomenon as '*spatial equilibrium*'.

The principal aims of the empirical literature on regional unemployment are to examine the persistence of unemployment differentials and to define a model that investigates its determinants. The applied analyses are mainly based on time series, using standard statistical methods, both parametric and non-parametric (see Decressin and Fatàs 1995; Jimeno and Bentolila 1998; Martin 1997; López-Bazo et al. 2000). There are only a very few analyses using spatial data and spatial parametric tools (see Molho 1995; Aragon et al. 2003; Niebuhr 2003).

In our paper we will use spatial econometric methods based on spatial autocorrelation techniques to explore the geographical distribution of unemployment for the 103 Italian provinces for the year 2003. The aim of the paper is to critically review the findings in the recent literature on local unemployment, and to construct a model that is capable of both incorporating the main findings of this literature and applying it to analyses the geographical distribution of unemployment for the Italian provinces, taking into account the location of the labour market. More specifically, we will test whether the time persistency, as empirically supported by much Italian research (see Contini and Trivellato 2005), corresponds to a '*spatial persistency*' (i.e. neighbouring provinces tend to have similar unemployment rates). We will explore whether the unemployment in the provinces in Italy has a uniform spatial structure with very few 'pockets of nonstationarity' (see Anselin 1995): in other words, if the provinces with high unemployment rates are contiguous (or share a border) with those provinces with low unemployment rates. As far as we know, this is the only empirical spatial analysis on the Italian labour market.

The paper is structured as follows. Section 2 presents a review of the principal theoretical interpretations of local unemployment disparities of labour markets. Next, Section 3 presents the statistical models and the data used in our empirical application. In Section 4, the empirical findings are presented and interpreted. And, finally, some concluding remarks are made in Section 5.

2 A Literature Review on Local Unemployment Disparities

Most of the theoretical literature – like most empirical analyses on local unemployment disparities – explains and interprets unemployment differentials by starting from the hypothesis of a stable *equilibrium* of spatial labour markets. Molho (1995: 642) defined *equilibrium* as "*a situation of uniform utility across areas for (each) homogeneous labour group, such that there are no incentives for further labour migration (a further condition would be uniform profits such that capital movements are eliminated)*". The *equilibrium* interpretation of the local labour market has received empirical and theoretical support from, amongst others, Hall (1970), Marston (1985) and Rosen (1974, 1979).

When the effect on local or regional unemployment caused by short run shocks is dissipated, the persistence of differentials in unemployment rates can be interpreted in terms of *disequilibium in nature* (Marston 1985). According to Marston (1985), there is an *equilibrium* relation of unemployment rates across areas, and in each area it is a function of the amenities and the endowments of the land. Workers migrate to areas where new jobs are created until there is no further incentive to move. In other words, the spatial distribution of unemployment under an *equilibrium* interpretation is characterized by constant utility across areas: high unemployment in the i-th area is compensated for by some other positive factors (e.g. local amenities, climatic conditions, quality of life, local housing prices, etc.) which are a disincentive to migration. Similar considerations can be put forward with regard to firm migration.

In contrast to the previous interpretation, local unemployment differentials can also be explained in terms of *disequilibrium*. The *disequilibrium* view assumes that in the long run the unemployment rate will level off across areas. The adjustment process may be faster or slower and, depending on its speed, differences in unemployment across areas could persist for a long time. The speed of the adjustment may depend on different factors connected to both labour demand and supply. According to Marston (1985: 61), the labour market can be driven towards *equilibrium* by three forces: "*(1) workers migrate out of the area because of the high unemployment rate (which is uncompensated); (2) firms migrate in, attracted by a large unemployed labour force; and (3) wages tend to fall because of the excess supply of labor*". If these forces are strong, the *disequilibrium* will be brief and not significant; if they are weak, it may be long-lasting and significant.

Following the analysis conducted by Elhorst (2003), the various models in the empirical literature can be classified into four groups: single equation models, implicit models, accounting identities, and simultaneous models.

The *single equation model* comprises three categories of explanatory approaches:
1) the *unemployment-vacancy relationship*, which defines an inverse relationship between the unemployment and vacancy rate (for details, see, e.g., Cheshire 1973; Holzer 1993);
2) the *cyclical sensitivity model* which explains the regional unemployment rate in terms of the national unemployment rate. Originally, the formulation of this model took for granted a cyclical relationship between the regional and the national unemployment rate (Thirlwall 1966; Brechling 1967). However, recent empirical studies have argued that the relationship between regional and national unemployment rate is an *equilibrium* (rather than a cyclical) relationship (Martin 1997; Baddeley et al. 1998). Clearly, the original approach has been criticized with regard to the instability of the cyclical component in relation to the chosen estimation period and the lack of a theoretical basis (Dunn 1982; Byers 1990; Chapman 1991);

3) the *amenity model,* which explains the unemployment differences between areas by the distribution of amenities: in other words, the *equilibrium* unemployment rate in each area will be a function of the amenities in the area, such that amenities (or the real wage too) balance the high unemployment (for details, see Hall 1972; Pissarides and McMaster 1990; Layard et al. 1991).

In the second group, *implicit models*, the regional unemployment rate is not explained but solved from either a theoretically postulated, or an empirically estimated, type of model. They include: the migration-based model (see Pissarides and McMaster 1990; Groenewold 1997), the NAIRU model (see Johnes and Hyclak 1989; Blackley 1989; Payne 1995), and the Blanchard and Katz model (1992).

The third group of *accounting identity models* focuses on the working age population variable and those entering and leaving it. A general accounting model, for the i-th area, can be written as: $U = P_{wa} * L + NC - E$ with $\Delta P_{wa} = G + NM$; where U is the level of unemployment; P_{wa} is the working age population; L is the labour force participation rate; NC is the net inward commuting; E is the level of employment; G is the balance between new entrants into, and departures from, the working age population; and NM is net inward migration. For more details and applications of the accounting identity model, we refer to Burridge and Gordon (1981) and Gordon (1988).

Finally, *simultaneous models* hypothesize that the regional unemployment rate both affects and is affected by one or more additional regional labour market variables (e.g. labour force participation rate, working age population, total working age population, degree of employment and earnings). Interesting applications can be found in Blanchard and Katz (1992), and Decressin and Fatás (1995).

In conclusion, the analysis of unemployment disparities at the regional level is varied and contains several complementary and sometimes contrasting frameworks. After our brief overview, we will now mention some of the most noteworthy empirical findings in the literature on the US and EU labour markets.

Marston (1985), aiming to explain the persistence of unemployment differentials between 30 US metropolitan areas for the years 1974-78, as well as to verify an *equilibrium* or *disequilibrium* interpretation for the US labour market, specified a model in which the unemployment rate U_{it} for an area *i* in period *t* was a function of three components: $U_{it} = \alpha_i + \beta_t + \varepsilon_{it}$, where α_i denotes the *equilibrium* differential, i.e. the natural unemployment rate in year *t* for each area *i*; β_t reflects changes in the national economy that affect all regional areas alike; and ε_{it} is an autocorrelated residual that reflects the fraction of the *disequilibrium* of the previous period that persists into the present (i.e. $\varepsilon_{it} = \rho\ \varepsilon_{i,\ t-1} + \eta_{it}$, with η_{it} a white noise error). He found the residuals of this regression were serially uncorrelated, arguing that any shocks which disturb the local steady-state value tend to be eliminated by geographical mobility within a year. Moreover, he affirmed that the variation of unemployment rates amongst US metropolitan areas has an equilibrium explanation.

This interpretation of the US labour market seems to be confirmed by other empirical studies, but these – in contrast to Marston (1985) – argued that the adjustment process towards the *equilibrium*-state takes longer than one year (Blanchard and Katz 1992; Holzer 1991; Treyz et al. 1993).

Blanchard and Katz (1992) showed that temporary local demand shocks have short-term effects on US local unemployment, so regional persistence in the US is low and depends mainly on worker migration rather than on firm migration.

In contrast to the empirical evidence from the US market, empirical analyses on the European labour market have shown a strong stability and persistence of high regional unemployment. The main reason for this may be connected to the low flexibility of the labour market, i.e. wage rigidity and employment inertia, and to labour force dynamics (Jimeno and Bentolila 1998). But, actually, a detailed comparative analysis between the EU and the US shows that the first impression of high and low regional unemployment persistence in the EU and the US, respectively, is wrong (see Eichengreen 1992; Decressin and Fatás 1995; Jimeno and Bentolila 1998). Eichengreen (1992) observed, with regard to the comparison between the EU and the US labour markets, that the persistence of unemployment in European regions is not higher than in the US, but the main difference concerns the different way the market responds. With regard to Britain and Italy, Eichengreen (1992) argued that the responsiveness of migration to a regional labour market *disequilibrium* is greater in the US than in either of the other two countries. And hence, it is reasonable to hypothesize that the adjustment process towards the *equilibrium*-state is due to other mechanisms (e.g. wage adjustment, labour-leisure choice, interregional capital mobility, etc.) which balance the limited mobility of workers.

Agreeing with Eichengreen's findings (1992), Decressin and Fatás (1995: 1646), focussing on Germany, Italy and the UK, showed that "*in the short run participation is the main adjustment mechanism, while, except for Italy, the role played by the employment rate over both the short and long run is negligible*". They found that the persistence of regional unemployment rates is actually lower in Europe than in the US, and that it is connected to large movements in and out of the labour force.

Jimeno and Bentolila (1998), following the findings from the previous literature, classified the areas of unemployment into three groups, on the basis of the degree of persistence of the aggregate and regional relative unemployment: 1) low persistence of aggregate and regional relative unemployment (this is the case for the US); 2) high and low persistence of, respectively, aggregate and regional relative unemployment (this is the case for most of the EU); 3) high persistence of aggregate and regional relative unemployment (this is the case of some European countries like Italy or Spain).

The previous literature stressed the different behaviour of the Italian labour market with respect to both other European countries and the US. This different performance of the Italian labour market leads us to investigate it at a more detailed territorial level. A *stylized* fact of the Italian labour market is the North-South dichotomy (see, e.g., Faini et al. 1997; Prasad and Utili 1998; Brunello et al. 2001). However, this dichotomy actually hides a patchwork of local facts that could be better explained by a provincial analysis (see Gambarotto and Maggioni 2002).

On the basis of the above considerations we propose to analyse the geographical distribution of Italian unemployment by using a single equation model. In the model proposed, the unemployment is explained by a proxy for labour demand and by some control variables; moreover, to take into account the territorial aspect of labour markets, spatial economic tools have been used. In the next section the empirical models and data will be presented.

3 Models and Data on Unemployment in Italian Provinces

The aim of the present analysis is to investigate whether there is a spatial relationship among provincial unemployment disparities in Italy. Using observations on a provincial level, the present analysis emphasises the territorial aspect of labour markets and explores in more detail stylized facts on the dual unemployment structure of the Italian market (the North-South dichotomy), in order to highlight the differences between the complex and varied structure of local labour markets.

According to the literature, regional differentials of unemployment and its spatial patterns may be explained by differences at the local level due to structural and non-coincidental factors. Usually, the variables used involve the following aspects: natural change, participation, migration, commuting, wages, unionization, employment, gross regional product, market potential, size and density, industry-mix explanation, economic and social barriers, and educational attainment of the population (for details, see Elhorst 2003).

Because of lack of data, the present analysis considers only some proxies of the previous features which characterize the local labour market. Using spatial econometric models we investigate the significance of spatial interaction of unemployment disparities in the 103 Italian provinces for the year 2003.

The spatial interaction between economic phenomena introduces the concept of spatial autocorrelation, which is linked to the territorial shape of the observed phenomena and to the connections between observations. Measures of spatial autocorrelation take into account the dependence between observations by a spatial weights matrix **W**. For a set of N observations the spatial matrix **W** is an N x N matrix with the diagonal elements equal to 0; the other elements w_{ij} represent the intensity of the effect of territorial area i on territorial area j (see Anselin and Bera 1998). The matrix defines the structure and the intensity of spatial effects, and it may be either a contiguity matrix or a matrix based on a distance decay function. In our application, we use a contiguity matrix, i.e. a binary spatial weight such that $w_{ij} = 1$ if the provinces i and j are contiguous (i.e. share a border), and $w_{ij} = 0$ otherwise. In order to explore the significance of spatial clusters of high or low unemployment, our starting point is a cross-sectional regression model on regional unemployment without spatial effects:

$$U = \beta_o + \beta_1 E + \sum_{k=2}^{R} \beta_k C_k + \varepsilon, \quad (1)$$

where **U** is the log of the provincial unemployment rate; **E** is the log of provincial employment over working age population as a proxy for labour demand; $\mathbf{C_k}$ are control variables; and ε is a vector of residuals. The control variables are: log of employment in the service sector over total provincial employment (E_{serv}); log of employment in the manufacturing sector over total provincial employment (E_{manif}); log of the size of the younger population (age from 15 to 29 years) over the total population ($P_{15\text{-}29}$) and log of the number of occupied houses over the total number of available houses (H_{occ}).

The variables E_{serv} and E_{manif} are proxies for the industry-mix explanation, though it is not always clear which sign these control variables should have; intuitively, provinces specializing in a declining economic sector such as manufacturing might show higher structural unemployment rates than provinces specializing in modern sectors such as services.

The size of the younger population with respect to the total population ($P_{15\text{-}29}$) is a proxy for natural change. Many studies have investigated whether the age structure of the population affects the local unemployment rate. In the main, these studies have shown that areas with a relatively young population have a more stubborn unemployment problem, and that areas with a relatively old population experience a less persistent problem (Elhorst 1995; Molho 1995).

The H_{occ} is a proxy for economic and social barriers. The housing market where there is a lower proportion of occupied housing should have cheaper housing prices and more chance of finding housing compared with provinces where there is a high proportion of occupied housing. We will expect a negative sign of the variable, the reason being that workers are not available to move from area i with a high number of vacancies to province j with a low number of vacancies.

If the spatial effects are substantive but they are ignored, the OLS regression of equation (1) will provide a biased estimation of the parameters. Therefore, in order to explore the spatial interaction of the geographical distribution of unemployment, we follow Florax and Folmer's procedure (1992) (see also Florax et al. 2003). The most general model is the following:

$$U = \rho WU + \beta X - \delta WX + (I - \lambda W)^{-1}\xi, \tag{2}$$

where X is an ($n \times k$) matrix of observations on the k independent variables (in our application the **E** and $\mathbf{C_k}$ variables). ρ is the spatial autocorrelation coefficient and measures the *spillover* effects: in other words, $\rho \neq 0$ implies that unemployment in province *i* depends directly on unemployment in other neighbouring provinces. Moreover, in order to capture spillover effects connected to the explanatory variables, we include their spatial lags (i.e. C_k and E) with the coefficient δ.

Model 2 cannot be estimated directly because the parameters cannot be fully identified. So, in brief, Florax and Folmer proposed that first it is necessary to test whether it is appropriate to include autoregressive disturbances (Is $\boldsymbol{\lambda} = 0$?) or a spatially lagged dependent variable (Is $\boldsymbol{\rho} = 0$?). Secondly, if neither are included, it is possible to test whether spatially-lagged independent variables should be included (Is $\boldsymbol{\delta} = 0$?). Therefore, departing from a model without spatial effects (Model 1), using a separate Lagrange multiplier test – proposed by Burridge (1980) (see also Anselin, 1988, 1992) – we test whether $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ are equal to 0; <u>if neither are equal to 0</u>, we could choose between a spatial error or a spatial lag model, on the basis of which a LM statistic is larger[1]. Moreover, it is worth asking whether a more general model would be preferable. In this case, an LR test on a common factor hypothesis should be done[2].

In addition to this procedure, we follow a general empirical strategy according to Hendry's methodology (see, e.g., Spanos 1988). We distinguish between the theoretical model (i.e. the mathematical formulation of the theory, in our application Model 1) and the statistical models written in terms of observable random variables. If the assumptions of the statistical model are tested and not rejected, this indicates that the postulated probabilistic structure is appropriate for the data. If not, an alternative model, which has a more appropriate informative structure, must be chosen. In other words, we will try to maximize the 'statistical adequacy' of the theoretical model.

---

[1] The LM statistic is distributed as a χ^2 with one degree of freedom (see Anselin 1992, p. 180).

[2] The autocorrelated error model is equivalent to a special form of spatial lag model by the following transformation of dependent and independent variables: (Y - λWY) e (X - λWX); so the spatial lag model can be written as Y = λWY + Xβ - λWXβ + ε. This is a subset, known as the common factor hypothesis model, of the more general model Y = λWY + Xβ + WXδ + ε. The LR test of the common factor hypothesis tests the hypothesis δ = λβ: if the null hypothesis is rejected a more general model with lagged independent variables must be estimated.

The empirical findings, discussed in the next section, were obtained in the light of this empirical strategy.

4 Empirical Findings for Provincial Italian Unemployment

4.1 Preliminary findings

Although the previous section identified, on the basis of theory and the availability of the data, some relevant variables that explain the regional disparities of unemployment rates in Italy, it is not expected that all these variables (i.e the variables included in the estimatable model) would be required in an adequate statistical model.

We first estimate a cross-sectional model without spatial effects. The estimations obtained are shown in column 1 of Table 1. All the coefficients of independent variables – except E_{serv} and H_{occ} – are statistically significant. The reason why the coefficients of E_{serv} and H_{occ} are not statistically significant could be connected to the spatial correlation between observations, as is highlighted by Burridge Lagrange Multiplier (LM test). The LM spatial test gives a significant value of 9.58 and this indicates that $\rho \neq 0$; so that – according to the procedure of Folmer and Florax – a spatial lag model has to be estimated.

The estimations of a spatial lag regression model are shown in column 2 of Table 1. The coefficients of the variables – except the variable E – are not statistically significant, but the coefficient of the variable WU is statistically significant and equal to $\rho = 0.29$. The positive value of ρ implies that unemployment in province *i* depends directly on unemployment in other neighbouring provinces. Moreover, the significant value of the coefficient of the variable E implies that unemployment in one area depends only on the employment in the same area. Finally, the value of an LM error test of 1.22 with a p-value of 0.27 ($\lambda = 0$) implies that the residuals from the unemployment regression are not spatially autocorrelated.

<< Insert Table 1 about here>>

The comparison between OLS and the spatial lag model (i.e. columns 1 and 2 of Table 1) highlights that the coefficients of the variables are very different. According to the diagnostic of spatial correlation, we should stop the analysis at the second step with Model 2 (column 2 of Table 1). This model eliminates the spatial correlation, but because nearly all the coefficients are not statistically significant, this leads us to investigate the spatial structure of each variable.

Could this lack of statistical significance of the coefficients be connected to the different spatial structure of the variables? In order to answer this question, we performed a more detailed analysis using Exploratory Spatial Data Analysis (ESDA) and other models. These estimates are reported in columns 3, 4, 5 and 6. The results are discussed in the next subsection.

4.2 Final results

In order to detect patterns of spatial association, spatial outliers or forms of spatial heterogeneity, we use some of the tools of Exploratory Spatial Data Analysis (ESDA). ESDA is a set of techniques aimed at: describing and visualizing spatial distribution; identifying atypical localizations or spatial outliers; detecting patterns or spatial association, clusters or hot spots; and suggesting spatial regimes or other forms of spatial heterogeneity (for details, see Haining 1990; Anselin 1998a,b). To know the spatial structure of variables and whether or not it is similar among variables will enable us – as with the time-structure in time series analyses – to identify the correct variables to include in a regression model. It could be that the lack of statistical significance of the coefficients in Model 2 (column 2 Table 1) might be connected with the different spatial structure of the variables considered.

According to Spanos (1988: 117), the problem "*arises as to how to coalesce the relevant theoretical and sample information in the specification of statistical models*". In other words, we need to identify an estimatable model – with a theoretical basis – that is bound up with an adequate statistical model.

Table 2 displays the Moran's I statistic of all the variables. Inference is based on a standardized z-value that follows a normal distribution.

All the variables are positively spatially autocorrelated since the statistics are significant, with p = 0.00001 for every variable. Because, Moran's *I* is similar (but not equal) to a correlation coefficient, we could say that the variables show a different intensity of spatial association. This is higher for the variables U, E, E_{manif} and P_{15-29} than for the variables E_{serv} and H_{occ}. The high intensity of the global association index (Moran's *I*) indicates a tendency towards geographical clustering of similar provinces with a high (or low) value of the variable (e.g. provinces with high or low value of unemployment are geographically clustered). Conversely, the low positive value of Moran's *I* with regard to the variables E_{serv} and H_{occ} could indicate a non-geographical clustering of similar provinces; i.e. the low value of Moran's I indicates lack of similarity among provinces with respect to E_{serv} and H_{occ}.

<< Insert Table 2 about here >>

The Moran's *I* statistic is a global statistic and does not allow us to investigate the provincial structure of spatial autocorrelation of each variable. It does not enable us to discover aspects such as: Which provinces contribute more to the global spatial autocorrelation? Are there local spatial clusters of high or low values? If so, do these clusters identify a dual structure (North-South)? Do variables have the same or similar spatial heterogeneity?

Therefore, a closer investigation of the spatial distribution of variables which explain the disparities of unemployment could be useful to identify the correct variables we need to include in the regression model. Maybe, in order to obtain an adequate statistical model, we have to include in the model variables with the same order of spatial association and spatial structure. In order to explore the spatial distribution of our variables, the Moran scatterplott was used (see Anselin 1995).

The Moran scatterplott allows us to study the local spatial instability by plotting the spatially-lagged variable (e.g. WY) against the unlagged variable (e.g. Y). It may be subdivided into four quadrants corresponding to four spatial associations. The first, of these associations on the top-right, are provinces with a large Y surrounded by large WY (quadrant HH). The second, on the top-left, are provinces with a small Y surrounded by large WY (quadrant LH). The third, on the bottom-left, are provinces with a small Y surrounded by small WY(quadrant LL). And the fourth, on the bottom-right, are provinces with a large Y surrounded by small WY (quadrant HL)[3]. The first and third quadrants (HH and LL) contain provinces with a positive spatial association, i.e. they indicate clusters of provinces with similar values. The second and fourth quadrants, however, display clusters of provinces with dissimilar values or a negative spatial association.

In Figures 1-6, the Moran scatterplot for the variables U, E, E_{serv}, E_{manif}, $P_{15\text{-}29}$ and H_{occ} is displayed. It can be seen (Figure 1) that, with respect to the variable U, positive spatial association characterizes most Italian provinces: 91.3% of Italian provinces have a positive association or similar values (35.0% in quadrant HH, and 56.3% in quadrant LL). With respect to the variable E, it has an equal but opposite spatial structure in comparison with the unemployment variable. The percentage of provinces with a positive association is again equal to 91.3% (56.3% in quadrant HH, and 35.0% in quadrant LL: see Figure 2).

<< Insert Figures 1 and 2 about here >>

[3] A value is large or small with respect to its average value.

The same could be done with respect to the variables E_{manif} and P_{15-29} (Figures 4 and 5); they present a spatial structure very similar to that of U and E. In particular, with respect to E_{manif}, 80.1% of provinces present a positive spatial association, whereas with respect to P_{15-29}, 91.3% have a positive spatial association. In contrast to the previous variables, the variables E_{serv} and H_{occ} do not show a high positive spatial association. In fact the percentage of provinces with positive association is equal to 65.0% for E_{serv} and 69.9% for H_{occ} (Figures 3 and 6).

<< Insert Figures 3, 4, 5 and 6 about here >>

Moreover, the Moran scatterplots help us to identify atypical provinces, i.e. those provinces characterized by an association of dissimilar values. In particular, the Moran scatterplot of the dependent variable U shows 2 and 7 provinces in the second (LH) and fourth (HL) quadrant, respectively (i.e. 9 in total). Similarly, also in the Moran scatterplot of variables E, E_{manif} and P_{15-29}, there are, respectively, 9, 9 and 20 provinces in the HL and LH quadrants taken together. Conversely, there are more provinces deviating from the global pattern of positive correlation for the variables E_{serv} and H_{occ} than for the others, viz. there are 26 provinces with an association of dissimilar values in the case of E_{serv}, and 31 provinces in the case of H_{occ}.

All these results lead us to the conclusion that there is a different spatial structure among variables. This is similar or almost identical for the variables U, E, E_{manif} and P_{15-29}, but not for the variables E_{serv} and H_{occ}. The Moran scatterplot of variables such as E, E_{manif} and P_{15-29} indicates the presence of spatial heterogeneity that may be synthesized in four distinct spatial regimes. The first corresponds to the HH scheme and includes mainly the Central-northern provinces. The second corresponds to the LL scheme and includes mainly the Southern provinces. Both regimes have a positive spatial association. The third and the fourth regimes correspond, respectively, to the LH and the HL scheme which both exhibit an atypical negative spatial association and include mainly Central and Southern provinces.

The Moran scatterplot of variable U indicates the presence of spatial heterogeneity that may be subdivided into three clusters. The first of these corresponds to the HH scheme, including mainly Southern provinces. The second corresponds to the LL scheme and includes mostly Central-northern provinces. Both regimes exhibit a positive spatial association. The third corresponds to the HL scheme which has only seven provinces with an atypical negative spatial association.

The exploratory spatial data analysis of the distribution of the variables U, E, E_{manif} and P_{15-29} highlights a Central-northern and Southern spatial polarization scheme.

In contrast to the above, the other independent variables, E_{serv} and H_{occ} show a spatial structure that cannot be encapsulated in specific spatial regimes. The large number of provinces contained in the LH and the HL scheme is an expression of 'pockets of nonstationarity'.

In the light of the above implications, the different spatial structure between both the dependent variable and the independent variables E_{serv} and H_{occ}, and between the latter two variables and the other independent variables could explain why most of the coefficients of Model 2 of Table 1 are not statistical significant. Therefore, we have estimated a new OLS model leaving out the variables E_{serv} and H_{occ}.

Column 3 of Table 3 shows the results of a cross-sectional model that neglects spatial effects. The coefficients have the expected sign and they are statistically significant. The statistically significant value of 7.76 of the LM spatial lag test indicates a value of $\rho \neq 0$, so a spatial lag model has to be estimated.

In column 4 of Table 1 the estimations of the spatial lag model are shown. All the coefficients are statistically significant and have the expected sign. The value of the LM error test of 0.39 with a p-value of 0.53 ($\lambda = 0$) implies that the residuals from the unemployment regression are not spatially autocorrelated.

The estimations highlight that the differentials of unemployment are mainly connected to those of employment. A marginal increase in employment produces a more proportional decrease in unemployment, while a marginal increase in employment in the manufacturing sector is translated into a less proportional decrease in unemployment. The demographic variable also has the expected sign but, contrary to expectations, a marginal increase of younger people has very small effect on unemployment, i.e. unemployment decreases less proportionally.

Although, the LM error test does not indicate the presence of spatial correlation in the residuals, before choosing the best statistical model, we estimate an autocorrelated error model.

The estimations of the autocorrelated error model show that all coefficients are statistically significant and with the expected signs. The LM spatial lag test with the significant value of 7.27 indicates a value of $\rho \neq 0$ that points us to a spatially-lagged dependent variable model. Moreover, the LR test of common factor hypothesis rejects the null hypothesis, so a model with lagged independent variables must be estimated.

Column 6 of Table 1 shows the estimation of the model with lagged independent variables: on the left of column 6 the coefficients of the independent variables are shown, and on the right the coefficients of the lagged independent variables. The diagnostic of spatial dependence shows that no spatial autocorrelation remains in the errors.

Our adequate statistical model is, therefore, the spatial lag model (column 4 of Table 1) because it is the model that best puts together the relevant theoretical insight with the sample information. It has one of the highest values of AIC (Akaike Information Criteria) and is the best model according to the spatial diagnostic. Moreover, even though the autocorrelated error model has the same value of AIC, we still preferred the spatial lag model over the autocorrelated error model because the evaluation of the spatial dependence of the latter model is the expression of the joint effect of omitted variables, the model misspecification and spatial autocorrelation.

5 Conclusions

The most important aim of this paper was to find the most adequate statistical model to explain the provincial unemployment differences of the labour market in Italy. We followed a general empirical strategy according to both a single equation model and the probabilistic structure of spatial data. We first estimated two regression models, one without and one with spatial effects. Then, using the Exploratory Spatial Data Analysis (ESDA) approach, we identified the variables with a similar spatial structure to be inserted into the statistical model.

The most adequate statistical model appears to be the spatial lag model. The regional differences in unemployment are strictly related to employment rather than to demographic variables. This result, already highlighted by previously undertaken Italian research (see, e.g., Amendola et al. 1999; Amendola et al. 2004) leads us to the conclusion that the differentials of unemployment which have characterized the Italian labour market for a long time are mainly due to the labour demand.

The other relevant result is that, in 2003, local labour markets were characterized by both positive spatial autocorrelation and spatial heterogeneity. In other words, local labour markets with high or low values of unemployment tend to cluster in space (i.e. positive spatial autocorrelation). At the same time, the economic behaviour of the actors is not stable across space but generates specific spatial regimes: the cluster of Central-northern provinces is clearly distinguished from the cluster of Southern provinces (i.e. a case of spatial heterogeneity). We call the presence of these joint characteristics '*spatial persistency*'.

The results obtained lead us to a new and hitherto unexplored research challenge: the link between spatial and time series analysis. In other words, if the phenomenon is characterized for a long period by a constant spatial persistency, this could explain the tendency for a high/low unemployment rate to be maintained for years (i.e. *time persistency*). This last issue will clearly be the subject of future research.

References

Amendola, A., F.E. Caroleo and G. Coppola. 1999. Differenziali territoriali nel mercato del lavoro e sviluppo in Italia. In *Struttura della contrattazione. Differenziali salariali e occupazione in ambiti regionali* edited by M. Biagioli, F.E. Caroleo and S. Destefanis. Napoli: ESI.

Amendola, A., F.E. Caroleo and G. Coppola. 2004. Divari regionali in Europa, *Discussion Paper* 78 CELPE Salerno.

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

______. 1992. *SpaceStat tutorial. A workbook for using SpaceStat in the analysis of spatial data*. University of Illinois, Urbana-Champaign.

______. 1995. Local indicators of spatial association-LISA, *Geographical Analysis* 27:93–115.

______. 1998a. Interactive techniques and exploratory spatial data analysis. In *Geographical information systems: Principles, techniques, management and applications* edited by P.A. Longley, M.F. Goodchild, D.J. Maguire and D.W. Wind. New York: Wiley,

______. 1998b. Exploratory spatial data analysis in a geocomputational environment. In *Geocomputation, a primer* edited by P.A. Longley, S.M. Brooks, R. McDonnell and B. Macmillan. New York: Wiley.

Anselin, L. and A. Bera. 1998. Spatial dependence in linear regression models with an application to spatial econometrics. In *Handbook of applied economic statistics* edited by A. Ullah and D.E.A. Giles. Berlin/Heidelberg/New York: Springer.

Aragon, Y., D. Haughton, J. Haughton, E. Leconte, E. Malin, A. Ruiz-Gazen and C. Thomas-Agnan. 2003. Explaining the pattern of regional unemployment: The case of the Midi-Pyrénées region, *Papers in Regional Science* 82:155–74.

Baddeley, M., R. Martin and P. Tyler. 1998. Transitory shock or structural shift? The impact of the early 1980s recession on British regional unemployment, *Applied Economics* 30: 19–30.

Bertola, G. and A. Ichino. 1996. Wage inequality and unemployment: U.S versus Europe, *Discussion Paper* 1186 CEPR London.

Blackley, P.R. 1989. The measurement and determination of state equilibrium unemployment rates, *Southern Economic Journal* 56:440–456.

Blanchard, O. and L. Katz. 1992. Regional evolutions. *Brookings Papers on Economic Activity* 1:1–75.

Brechling, F. 1967. Trends and cycles in British regional unemployment, *Oxford Economics Papers* 19:1–21.

Brunello, G., C. Lupi and P. Ordine. 2001. Widening differences in Italian regional employment, *Labour Economics* 8:103–129.

Burridge, P. 1980. On the Cliff-Ord test for spatial correlation, *Journal of the Royal Statistical Society* B 42: 107–108.

Burridge, P. and I. Gordon. 1981. Unemployment in the British metropolitan labour areas, *Oxford Economic Papers* 33:274–297.

Byers, J.D. 1990. The cyclical sensitivity of regional unemployment: An assessment, *Regional Studies* 24:447–453.

Chapman, P.G. 1991. The dynamics of regional unemployment in the UK, 1974-1989, *Applied Economics* 23:1059–1064

Cheshire, P.C. 1973. *Regional unemployment differentials in Great Britain*. Cambridge University Press, National Institute of Economic and Social Research.

Contini, B. and U. Trivellato. 2005. Eppure si muove. Dinamiche e persistenze nel mercato del lavoro Italiano. Bologna: Il Mulino.

Decressin, J. and A. Fatás. 1995. Regional labour dynamics in Europe, *European Economic Review*, 39:1627–1695.

Dohse, D. C. Krieger-Boden and R. Soltwedel. 2002. EMU and regional labor market disparities. In *Regional convergence in the European union* edited by J. Cuadrado-Roura and M. Parellada. New York: Springer.

Dunn, R. 1982. Parameter instability in models of local unemployment responses, *Environment and Planning A* 14:75–94.

Eichengreen, B. 1992. *Labor markets and european monetary unification*, Mimeo, University of California, Berkeley, CA.
Elhorst, J.P. 1995. Unemployment disparities between regions in the European union. In *Convergence and divergence among European regions* edited by H.W. Armstrong and R.W. Vickerman. London: Pion.
______. 2003. The mystery of regional unemployment differentials: Theoretical and empirical explanations, *Journal of Economic Surveys* 17:709–748.
Faini, R., G. Galli, P. Gennari and F. Ross. 1997. An empirical puzzle: Falling migration and growing unemployment differential among Italian regions, *European Economic Review* 41:571–579.
Fischer, M and P. Nijkamp. 1987. *Regional Labour Markets*. Amsterdam: Elsevier.
Florax, R.J.G.M. and H. Folmer. 1992. Specification and estimation of spatial linear regression models: Monte Carlo evaluation and pre-estimators, *Regional Science and Urban Economics* 22:405–432.
Florax, R.J.G.M., H. Folmer and S.J. Rey. 2003. specification searchers in spatial econometrics: The relevance of Hendry's methodology, *Regional Science and Urban Economics* 33:557-579.
Fujita, M., P. Krugman and A. Venables. 1999. *Spatial economy*. MIT Press Cambridge MA.
Gambarotta, F. and M.A. Maggioni. 2002. Localizzazione e agglomerazione produttiva: L'analisi territoriale dei mercati del lavoro, *Proceedings of the Italian Association of Labour Economics, XVII Scientific Meeting*.
Gordon, I.R. 1988. Evaluating the effects of employment changes on local unemployment, *Regional Studies* 22:135–147.
Groenewold, N. 1997. Does migration equalise regional unemployment rates? Evidence from Australia, *Papers in Regional Science* 76:1–20.
Haining, R. 1990. *Spatial data analysis in the social and environment sciences*. Cambridge: Cambridge University Press.
Hall, R. 1970. Why is the unemployment rate so high at full employment?, *Brooking Papers on Economic Activity* 3:369–402.
______. 1972. Turnover in the Labour Force, *Brookings Papers on Economic Activity* 3:709–764.
Holzer, H.J. 1991. Employment, unemployment and demand shifts in local labor markets, *Review of Economics and Statistics* 73:25–32.
______. 1993. Structural/frictional and demand-deficient unemployment in local labor markets, *Industrial Relations* 32:307–328.
Jimeno, J.F. and S. Bentolila 1998. Regional unemployment persistence (Spain, 1976-1994), *Labour Economics* 5:25–51.
Johnes, G. and T.J. Hyclak. 1989. Wage inflation and unemployment in Europe: the regional dimension, *Regional Science* 23:19–26.
Layard, R., S. Nickell and R. Jackman. 1991. *Unemployment, macroeconomic performance and the labour market*. Oxford: University Press.
Longhi, S. 2005. *Open Regional Labour Markets and Socio-Economic Developments*, PhD Thesis, Vrije Univerisiteit Amsterdam.
Longhi, S., P. Nijkamp, A. Reggiani and E. Maierhofer. 2005. Neural network modelling as a tool for forecasting regional employment patterns, *International Regional Science Review* 28(3):330-346.
López-Bazo, E. T. del Barro and M. Artis. 2000. *The geographical distribution of unemployment*, Mimeo, University of Barcelona.
Marston, S. 1985. Two views of the geographic distribution of unemployment, *Quarterly Journal of Economics* 100:57–69.
Meliciani, V. 2006. Income and employment disparities across European regions, *Regional Studies* 40(1):75-92.
Martin, R. 1997. Regional unemployment disparities and dynamics, *Regional Studies* 31:237–252.
Molho, I. 1995. Spatial autocorrelation in British unemployment, *Journal of Regional Science* 35: 641– 658.
Niebuhr, A. 2003. Spatial interaction and regional unemployment in Europe, *European Journal of Spatial Development* 5:2–24.
Payne, J.E. 1995. A note on real wage rigidity and state unemployment rates, *Journal of Regional Science* 35:319–332.
Pissarides, C.A. and I. McMaster. 1990. Regional migration, wages and unemployment: Empirical evidence and implication for policy, *Oxford Economic Papers* 42:812–831.

Prasad, E.S. and F. Utili. 1998. The Italian labor market: stylized facts, institutions, and directions for reform, *IMF Working Paper*.
Rosen, S. 1974. Hedonic prices and implicit prices, *Journal of Political Economy* 86:34–53.
______. 1979. Wage-Based Indexes of Urban Quality of Life. In *Current issues in urban economics* edited by P. Mieszkowki and M. Straszheim. Baltimore: Johns Hopkins.
Spanos, A. 1988. Towards a unifying methodological framework for econometric modelling, *Economic Notes* 1:107–141.
Taylor, J. and S. Bradley. 1997. Unemployment in Europe: A comparative analysis of regional disparities in Germany, Italy and the UK, *Kyklos* 50:221–245
Thirlwall, A.P. 1966. Regional unemployment as a cyclical phenomenon, *Scottish Journal of Political Economy* 13:205–219.
Treyz, G.I., D.S. Rickman, G.L. Hunt and M.J. Greenwood. 1993. The dynamics in US internal migration, *Review of Economic Statistics* 60: 209–214.

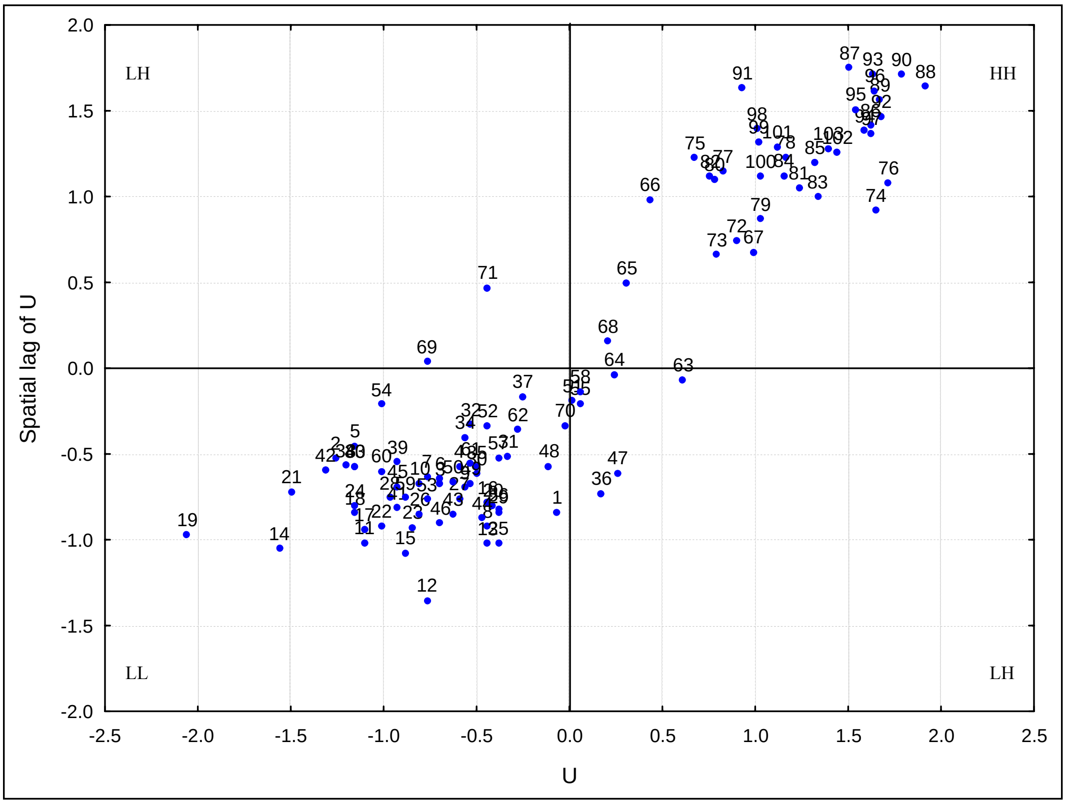

Figure 1. Moran scatterplot of U

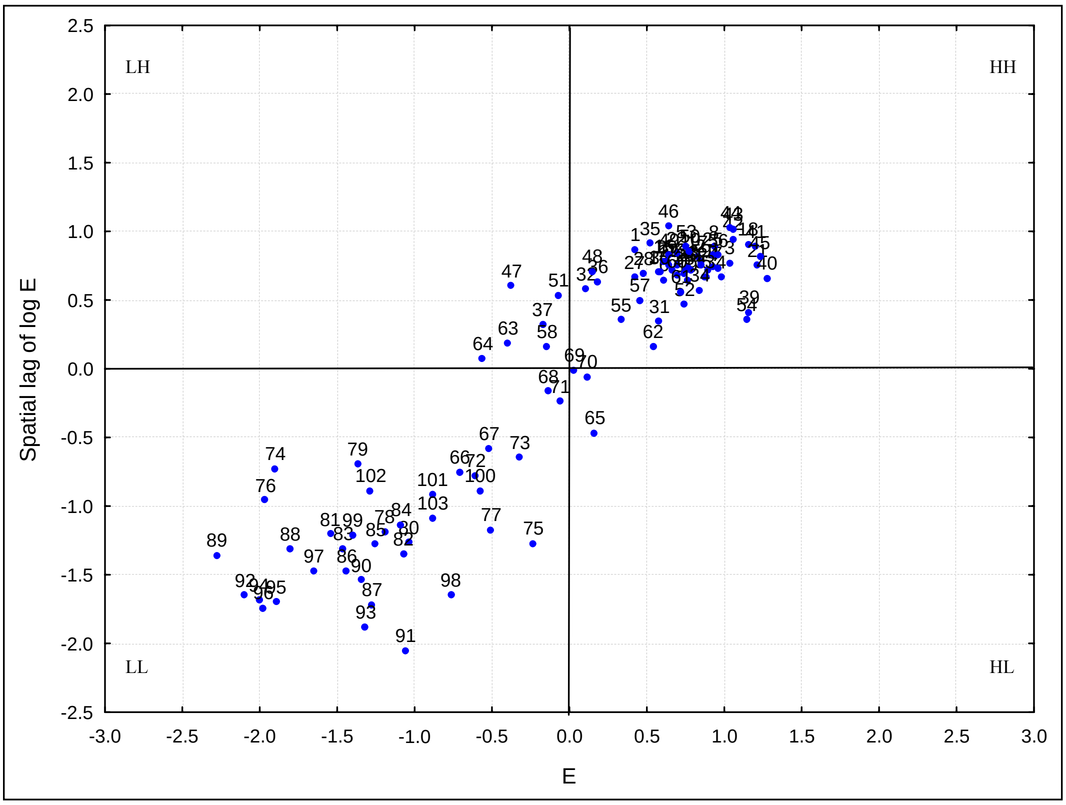

Figure 2. Moran scatterplot of E

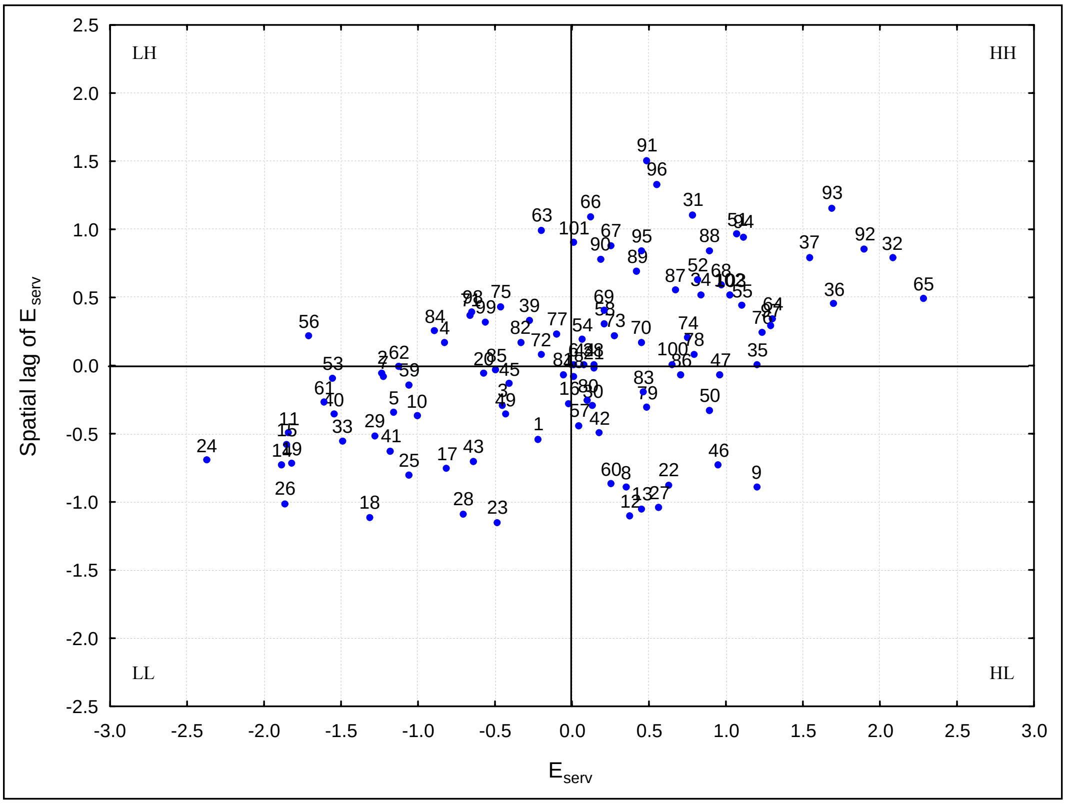

Figure 3. Moran scatterplot of E_{serv}

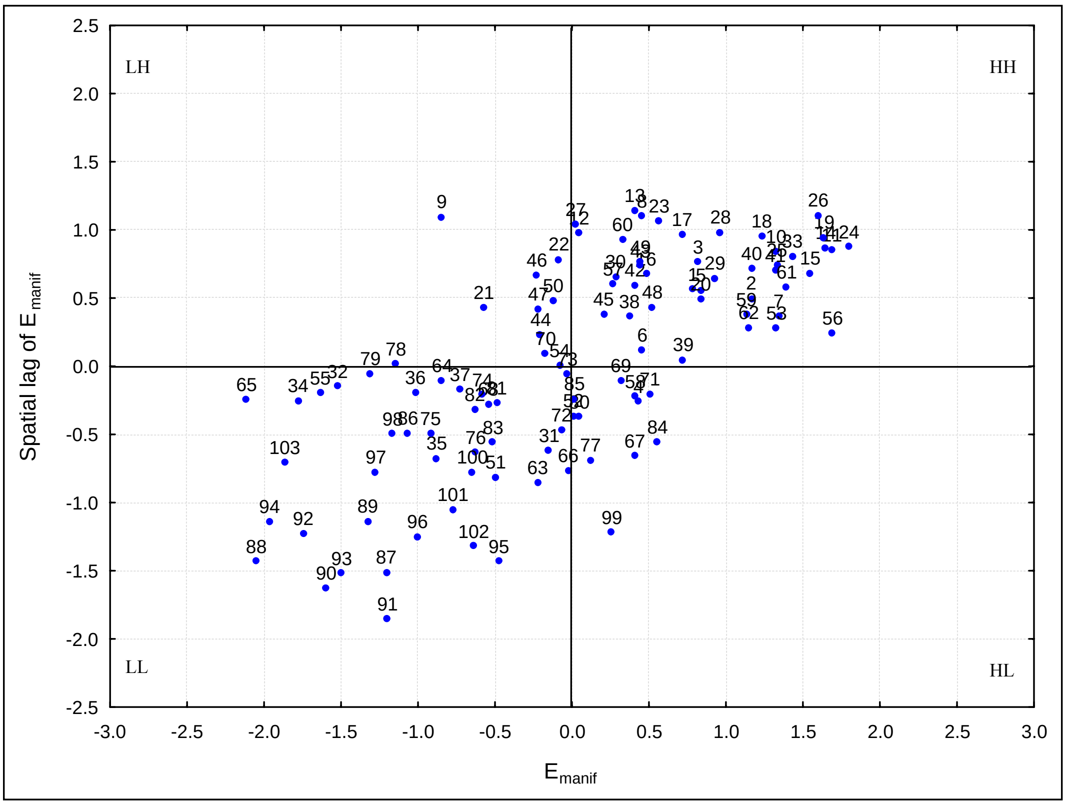

Figure 4. Moran scatterplot of E_{manif}

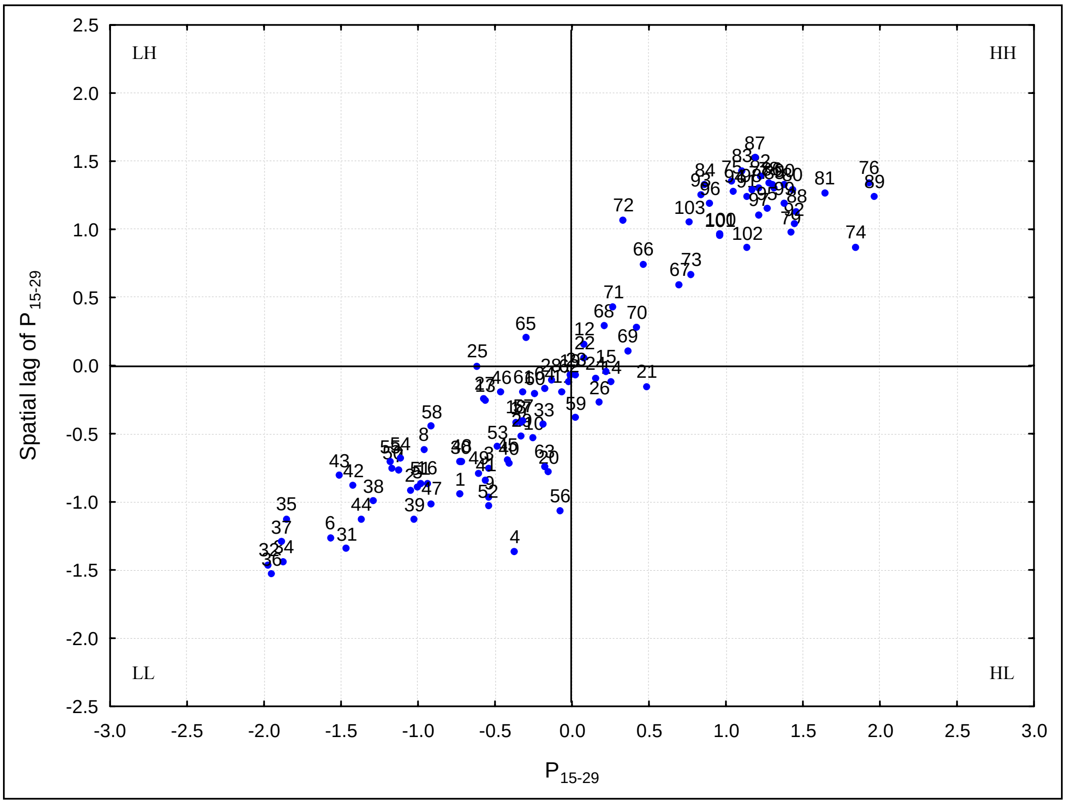

Figure 5. Moran scatterplot of $P_{15\text{-}29}$

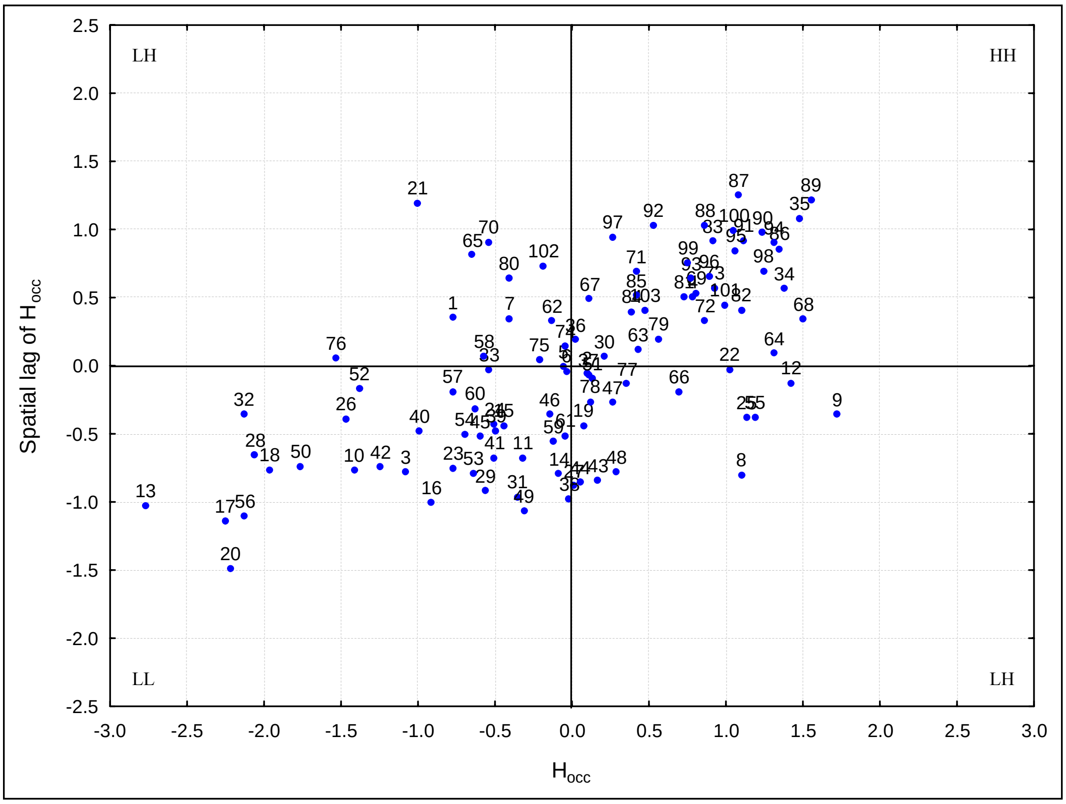

Figure 6. Moran scatterplot of H_{occ}

Table 1. Regression Results

Variable	Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 5	
		Lagged dep. var.		Lagged dep. var.	Autocorrelated Error	Lagged indep. var.	
						unlagged terms	lagged terms
	OLS	MLE	OLS	MLE	MLE	MLE	
E	-3.301	-2.604	-3.289	-2.937	-3.298	-2.536	0.122
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.83)
E_{serv}	0.012	0.396	-	-		-	-
	(0.96)	(0.14)					
E_{manif}	-0.423	-0.131	-0.436	-0.413	-0.399	-0.284	-0.384
	(0.017)	(0.45)	(0.00)	(0.00)	(0.00)	(0.02)	(0.08)
$P_{15\text{-}29}$	0.328	0.085	0.349	0.333	0.326	0.168	0.232
	(0.03)	(0.57)	(0.00)	(0.00)	(0.00)	(0.69)	(0.58)
H_{occ}	0.011	0.006	-	-		-	-
	(0.85)	(0.91)					
λ	-	-	-	-	0.161	-	-
					(0.21)		
ρ	-	0.290	-	0.113	-	-	0.164
		(0.00)		(0.06)			(0.20)
Log Likeklihood	-2.77	2.77	-2.79	-1.18	-2.18	3.17	
LM (λ)	1.13	1.22	1.08	0.39		1.26	
	(0.28)	(0.27)	(0.29)	(0.53)		(0.26)	
LM (ρ)	9.58	-	7.76		7.27		
	(0.00)		(0.00)		(0.00)		
LR Common Factor	-	-	-	-	10.7	-	-
					(0.01)		
AIC	-15.5	-6.5	-11.6	-10.4	-10.4	-7.7	

Table 2. Moran's *I* statistics

Variable	Moran's *I*	Standard deviation	Standardized value
U	0.856	0.069	12.595
E	0.865	0.069	12.726
E_{serv}	0.300	0.069	4.497
E_{manif}	0.530	0.069	7.848
$P_{15\text{-}29}$	0.852	0.069	12.533
H_{occ}	0.402	0.069	5.985

Note: The expected and the standard deviation value for the Moran's I statistic are -0.01 and 0.07, respectively. All statistics are significant at p = 0.00001