

# **The evolution of regional productivity disparities in the European Union, 1975-2000\***

**Yiannis Kamarianakis**

Regional Analysis Division,  
Institute of Applied and Computational  
Mathematics,  
Foundation for Research and Technology,  
Vasilika Vouton, P.O. Box 1527,  
GR 711 10, Heraklion Crete, Greece

Department of Economics,  
University of Crete,  
University Campus,  
Rethymno 74 100,  
Greece  
[kamarian@iacm.forth.gr](mailto:kamarian@iacm.forth.gr)

and

**Julie Le Gallo**

IERSO (IFReDE-GRES),  
Université Montesquieu-Bordeaux 4,  
Avenue Léon Duguit,  
33608 Pessac Cedex, France  
[legallo@u-bordeaux4.fr](mailto:legallo@u-bordeaux4.fr)

**December 2003**

---

\* Part of this work was done when both authors were visiting scholars at the Regional Economics Applications Laboratory, University of Illinois at Urbana-Champaign. A previous version of this paper has been presented at the 50<sup>th</sup> North American Meetings of the Regional Science Association International (RSAI), Philadelphia, November 20-22, 2003. We would like to thank S. Nazara, S. Dall'erba and J. Paelinck for their valuable comments. The usual disclaimer applies.

# **The evolution of regional productivity disparities in the European Union, 1975-2000**

## **Abstract**

*The aim of this paper is to assess the evolution of regional productivity disparities in the European Union. Using a sample of 205 regions and 8 sectors on the 1975-2000 period, we use Esteban's shift-share analysis to investigate the extent to which the existing interregional inequalities in productivity can be attributed to differences in sectoral composition between regions and/or to uniform productivity gaps across sectors. After a specification search on the bivariate functional forms that relate productivity differentials to their shift-share decomposition, the difference between regional and EU average productivity is regressed on the three shift-share components: industry-mix, differential and allocative (i.e. the covariance between the first two components). In that purpose, spatial seemingly unrelated (SUR) regressions are carried out to study the evolution of the impact of the components on the productivity gap over time, while allowing for intertemporal covariance. Moreover, spatial autocorrelation is also included in the SUR regressions, and its evolution over the period is analyzed. Results indicate that both spatio-temporal dependencies are essential in model specification.*

**Keywords:** European regions, productivity disparities, spatial autocorrelation, SUR

**JEL Classification:** C31, C33, O52, R11

## 1. Introduction

European integration has stimulated numerous studies of regional economic convergence within the European Union.<sup>2</sup> One approach dealing with the dynamics of regional inequality in Europe is presented by Esteban (1994) who examines to what extent disparities can be attributed to regional differences in various factors, beginning by breaking down per capita income into production per worker, employment rate and participation rate. His findings suggest that regional differences in productivity are the main reason for regional inequality in per capita income in the European Union.<sup>3</sup>

In order to gain a deeper insight into regional inequality in income per capita, Esteban (2000) analyzes the causes that generate regional productivity disparities in Europe. He uses shift-share analysis to additively decompose regional productivity differentials with respect to the European mean into three components: structural, regional and allocative factors and demonstrates that productivity differentials in the E.U. are uniformly distributed across sectors, e.g. each region's industry mix contributes relatively little to regional dispersion in average productivity.<sup>4</sup>

As is still usual in the inequality and convergence literature, the empirical methods used by Esteban (2000) at a regional level do not take into account spatial effects, particularly spatial autocorrelation, defined as the coincidence of value similarity with locational similarity (Anselin, 2001). However, there are a number of factors - trade between regions, technology and knowledge diffusion and more generally regional spillovers - that lead to geographically dependent regions. Because of spatial interactions between regions, geographical location is important in accounting for their economic performance. The role of

---

<sup>2</sup> A summary of the main findings in this area can be found in Cuadrado-Roura (2001), Armstrong (2002) or Terrasi (2002).

<sup>3</sup> In contrast to the situation in Europe, Browne (1989) and Carlino (1992) report the main cause of regional inequality in per capita income in the United States to be regional variability in unemployment rates.

<sup>4</sup> A detailed description of these factors lies in the next section.

spatial effects in economic processes needs to be examined using the appropriate spatial statistics and econometric methods. Such studies appeared in the literature after the mid-nineties (see Rey and Montouri (1999), Le Gallo and Ertur (2003) and Rey and Janikas (2003) and references therein for a literature review).

This paper aims at investigating regional productivity disparities and their relation with the three aforementioned shift-share components in space and time. It extends Esteban's approach by performing shift-share for several years, by allowing for intertemporal covariance between the different years and by explicitly taking into account spatial autocorrelation. Using a dataset that corresponds to 205 NUTS 2 European regions from 1975 to 2000, we find that spatial autocorrelation is an unavoidable feature. Indeed, we use recently developed tools of exploratory spatial data analysis to identify both global and local spatial autocorrelation and thus characterize the way economic activities are located in the E.U. and the way this pattern of location has changed over time. Moreover, we employ spatial seemingly unrelated regressions (SUR) to model the temporal evolution of the relation of productivity with each one of its shift-share components while at the same time accounting for intertemporal covariances and spatial dependencies.

In section 2, we set out Esteban's (2000) shift-share decomposition where regional productivity growth is modeled as the sum of three components: structural, differential and allocative. Section 3 presents the sample of 205 European regions over the 1975–2000 period as well as the spatial weight matrices used in this paper. In section 4, we perform exploratory spatial data analysis (ESDA) methods on productivity and the three shift-share components. The fifth section starts with a specification search on the functional relations between productivity and the shift-share components; these results are used in the specification of SUR and spatial SUR models that are presented next in order to assess the evolution of the impact

of the components on the productivity gap over time, while allowing for intertemporal covariance and spatial autocorrelation. The sixth and final section concludes the paper.

## 2. The Shift-Share Approach

In this section, regional labor productivities are decomposed via traditional shift-share analysis as depicted in Esteban (1972, 2000). A number of studies have focused on analyzing changes in employment and productivity as determinants of income growth using shift-share analysis or a related methodology (Wadley and Smith, 2003). First used by Dunn (1960) as a forecasting technique for regional growth employment, the shift-share approach has been applied more recently by Esteban (1972, 2000) to analyze productivity changes among the European regions.

Esteban's approach can be formulated as follows: let  $p_i^j$  be sector  $j$ 's employment share in region  $i$  so that  $\sum_j p_i^j = 1$  for all regions  $i$ . We denote by  $p_{EU}^j$  sector  $j$ 's employment share at the European level. Thus, we shall also have  $\sum_j p_{EU}^j = 1$ . Similarly, we denote by  $x_i^j$  the productivity per worker in sector  $j$  and region  $i$ , respectively  $x_{EU}^j$  at the European level. In our case eight sectors are concerned: agriculture, construction, total energy and manufacturing, distribution, transport and communications, banking and insurance, other market services and non-market services.

Based on the above, the following equalities hold:

$$x_i = \sum_j p_i^j x_i^j \quad \text{and} \quad (1a)$$

$$x_{EU} = \sum_j p_{EU}^j x_{EU}^j . \quad (1b)$$

The regional differential in productivity per worker between region  $i$  and the European average is therefore:  $x_i - x_{EU}$ .

Esteban (2000) shows that the regional differential in productivity per worker can be attributed to three possible causes. The first one is due to the specialization of a region in the more productive sectors, which would result in a regional aggregate productivity above the mean, even if the productivity of each single sector was the same at any location. It may result from local advantages that have been growing with history. The second cause comes from each region's sector-by-sector productivity differential to the average, assuming that the sectoral composition of the regional industry is the same to the one at the European level. It may come from previous investments in technology, human capital and public infrastructures. The third cause of differential in productivity per worker is due to a combination of both.

In order to assess the extent to which each of these component impacts on the different levels of regional productivity per worker compared to the EU average, the three components of the regional deviation in productivity are defined as follows:

a) *The industry-mix component*  $\mu_i$  of region  $i$  measures the differential in productivity per worker between region  $i$  and the EU average due to the specific sectoral composition of its industry. Here we assume that the productivity per worker in each sector is the same across all the regions and the European average. We thus write:

$$\mu_i = \sum_j (p_i^j - p_{EU}^j) x_{EU}^j \quad (2)$$

$\mu_i$  takes positive values if the region is specialized (i.e.  $p_i^j > p_{EU}^j$ ) in sectors with high productivity compared to the European level or de-specialized (i.e.  $p_i^j < p_{EU}^j$ ) in sectors of low productivity.  $\mu_i$  is at a maximum if the region is specialized in the most productive sector.

Note that (2) can be rewritten as:

$$\sum_j p_i^j x_{EU}^j = x_{EU} + \mu_i . \quad (3)$$

The left hand side of (3) is the average productivity per worker in region  $i$  if European and regional productivities coincide sector by sector. According to (3), region  $i$ 's average productivity is equal to the European average plus the regional industry-mix component.

b) *The productivity differential component*  $\pi_i$  focuses on productivity differentials due to region  $i$ 's sector by sector productivity differential to the EU average, assuming that the region's industry mix coincides with the European one. We then define  $\pi_i$  as:

$$\pi_i = \sum_j p_{EU}^j (x_i^j - x_{EU}^j) \quad (4)$$

$\pi_i$  takes on positive values if the region has sectoral productivities above the European average. Equation (4) can also be written as follows:

$$\sum_j p_{EU}^j x_i^j = x_{EU} + \pi_i . \quad (5)$$

The left hand side of (5) stands for the average productivity of region  $i$  when its industry mix equals the European one and hence any differential in average productivity must be caused by sectoral productivity differences. Region  $i$ 's average productivity could thus be expressed as the sum of the European average plus the regional productivity differential component.

c) *The allocative component*  $\alpha_i$  is a combination of the two previous components and is defined as follows:

$$\alpha_i = \sum_j (p_i^j - p_{EU}^j)(x_i^j - x_{EU}^j). \quad (6)$$

This component is positive if the region is specialized, relative to the European average, in sectors whose productivity is above the European average, and negative if below it.  $\alpha_i$  is at its maximum if the region is completely specialized in the sector with the largest productivity differential with respect to the European average. This component is an indicator of the efficiency of each region in allocating its resources over the different industrial sectors. The allocative component can also be viewed as measuring the covariance between the two previous components. The gap between regional and European average productivities decomposed into the three components can be formulated as follows:

$$y = x_i - x_{EU} = \mu_i + \pi_i + \alpha_i. \quad (7)$$

In order to measure the role played by each component in explaining regional differences in aggregate productivity per worker, Esteban computes the relative weight of the variance of each component in the overall observed variance. From (7) we have:

$$\text{var}(y) = \text{var}(\mu) + \text{var}(\pi) + \text{var}(\alpha) + 2[\text{cov}(\mu, \pi) + \text{cov}(\mu, \alpha) + \text{cov}(\pi, \alpha)]. \quad (8)$$

Finally, he tests whether interregional differences in aggregate productivity per worker can be explained by a model including one single component of the shift-share decomposition (7).

In that purpose, the following models are estimated:

$$x_i - x_{EU} = a_\mu + b_\mu \mu_i + \varepsilon_{\mu,i} \quad i = 1, \dots, N \quad (9a)$$

$$x_i - x_{EU} = a_\pi + b_\pi \pi_i + \varepsilon_{\pi,i} \quad i = 1, \dots, N \quad (9b)$$

$$x_i - x_{EU} = a_\alpha + b_\alpha \alpha_i + \varepsilon_{\alpha,i} \quad i = 1, \dots, N \quad (9c)$$

where  $N$  is the total number of regions, and  $\varepsilon_\mu$ ,  $\varepsilon_\pi$  and  $\varepsilon_\alpha$  are error terms with the usual properties (i.e., they are  $Nid(0, \sigma^2)$ ). Using 4 datasets –three of them corresponding to 1986 and one corresponding to 1989- with different regional/sectoral combinations, he finds that most of the observed interregional variance in aggregate productivity per worker is attributable to pure productivity differentials.

It should be noted though, that variance is not a typical measure of inequality since it does not satisfy the requirement of scale independence. This could give rise to a serious restriction if, as in the case at hand, the aim is to make comparisons over time. Moreover, spatial autocorrelation is not taken into account in these regression models, and, when ignored, it can result in major model misspecification (Anselin, 1988a). Recent developments in spatial econometrics offer procedures for testing for the potential presence of these misspecifications and suggest the proper estimator for models that treats spatial dependence explicitly. Based on these two aspects, we extend Esteban's analysis in two respects. First, we estimate equations (9a), (9b) and (9c) for different years in order to capture the evolution of the role played by each component in the explanation of the labor productivity gaps. Since there is no reason to assume that the different years are uncorrelated, we allow for intertemporal covariance using Seemingly Unrelated Regressions (SUR). Second, spatial autocorrelation is explicitly taken into account, so that spatial SUR are estimated (Anselin, 1988b). In that purpose, a spatial weight matrix has to be defined for our sample, which we present in the next section.

### 3. Data and spatial weight matrix

The computation of shift share components presented in the previous section is based on European regional data on gross value added and employment for eight economic sectors: agriculture, energy and manufacturing, construction, distribution, transport and communications, banking and insurance, other market services and non market services. The data come from the Cambridge Econometrics database. Our sample includes 205 regions in 15 countries (NUTS2 level) over the 1975-2000 period: Luxemburg, Belgium (10), Denmark, Germany (31), Greece (12), Spain (16), France (22), Ireland (2), Italy (20), Netherlands (12), Austria (9), Portugal (5), Finland (6), Sweden (21), United Kingdom (37).

Spatial data analysis needs modeling the spatial interdependence between the observations using a spatial weight matrix  $W$ . In this matrix, each region is connected to a set of neighboring regions by means of a purely spatial pattern introduced exogenously. The elements  $w_{ii}$  on the diagonal are set to zero whereas the elements  $w_{ij}$  indicate the way the region  $i$  is spatially connected to the region  $j$ . These elements are non-stochastic, non-negative and finite.

Various matrices have been considered in the spatial statistic and econometric literature: a simple binary contiguity matrix; a binary spatial weight matrix with a distance-based critical cut-off, above which spatial interactions are assumed negligible; more sophisticated generalized distance-based spatial weight matrices with or without a critical cut-off; nearest-neighbor matrices, etc. Note that in the European context, existence of islands doesn't allow considering simple contiguity matrices since they would imply a weight matrix that includes rows and columns with only zeros for the islands. Since unconnected observations are eliminated from the results of the global statistics, this would change the sample size and the interpretation of the statistical inference.

The weights should be exogenous to the model in order to avoid the identification problems raised by Manski (1993) in social sciences. Therefore, we consider pure geographical distance, more precisely great circle distance between regional centroids, which is indeed strictly exogenous.

The inverse of squared distance between regions is used as a functional form and the general form of the distance weight matrix  $W$  is defined as following:

$$\begin{cases} w_{ij}^* = 0 & \text{if } i = j \\ w_{ij}^* = 1/d_{ij}^2 & \text{if } d_{ij} \leq D(1) \\ w_{ij}^* = 0 & \text{if } d_{ij} \geq D(1) \end{cases} \quad \text{and} \quad w_{ij} = w_{ij}^* / \sum_j w_{ij}^* \quad (10)$$

where  $w_{ij}^*$  is an element of the unstandardized weight matrix;  $w_{ij}$  is an element of the standardized weight matrix such that the elements of a row sum up to one;  $d_{ij}$  is the great circle distance between centroids of regions  $i$  and  $j$ ;  $D(1)$  is the first quartile of the great circle distance distribution. It is the cutoff parameter above which interactions are assumed negligible. Since we use a row standardized matrix  $W$ , it is relative and not absolute distance that matters. For a given variable  $x$ , this transformation also means that the expression  $Wx$ , called the spatial lag variable, is the weighted average of the neighboring observations. Note that an additional advantage of this weight matrix is that it guarantees connections between United Kingdom and continental Europe and between Greek and Italian regions so that a bloc-diagonal structure of the weight matrix can be avoided.

Since all analyses are conditional upon the choice of the spatial weight matrix, several alternatives have been considered to check for robustness of our results: distance-based weight matrices with different cut-offs and nearest-neighbor matrices.

#### 4. Exploratory spatial data analysis of productivity and its shift-share decomposition

Using the dataset presented in the previous section, we compute the productivity of each region in deviation from the EU average and the three shift-share components for every year of the sample, 1975-2000. This section aims at showing that spatial autocorrelation characterizes the distributions of regional productivity and its shift-share decomposition.

Spatial autocorrelation can be defined as the coincidence of value similarity with locational similarity (Anselin 2001). There is positive spatial autocorrelation when high or low values of a random variable tend to cluster in space and there is negative spatial autocorrelation when geographical areas tend to be surrounded by neighbors with very dissimilar values. This effect is highly relevant in Europe since spatial concentration of economic activities in European regions has already been documented (Lopez-Bazo *et al.*, 1999, Le Gallo and Ertur, 2003; Dall'erba, 2003). Here we are interested in both global and local spatial autocorrelation.

The measurement of global spatial autocorrelation is usually based on Moran's  $I$  statistic (Cliff and Ord, 1981). For each year of the period 1975-2000, this statistic is written in the following matrix form:

$$I_t = \frac{n}{S_0} \cdot \frac{z_t' W z_t}{z_t' z_t} \quad t = 1, \dots, 25 \quad (11)$$

where  $z_t$  is the vector of the  $n$  observations for year  $t$  in deviation from the mean. Moran's  $I$  statistic gives a formal indication of the degree of linear association between the vector  $z_t$  of observed values and the vector  $Wz_t$  of spatially weighted averages of neighboring values,

called the spatially lagged vector. Values of  $I$  larger (resp. smaller) than the expected value  $E(I) = -1/(n-1)$  indicate positive (resp. negative) spatial autocorrelation.

Table 1 displays Moran's  $I$  statistic for regional productivity in deviation from the EU average and the three shift-share components for 1975 and 2000 period for the 205 European regions of our sample. Inference is based on the permutation approach with 9999 permutations (Anselin, 1995). It appears that all four variables are positively spatially autocorrelated since the statistics are significant with  $p = 0.0001$  for 1975 and 2000.<sup>5</sup> This result suggests that the distributions of regional productivity and its three shift-share components are by nature clustered over the whole period. Comparing the results for 1975 and 2000 shows that the standardized values of the statistic slightly decrease over the period, especially for the allocative component. These results therefore indicate a very small decrease of the geographical clustering of similar regions.

[Table 1 about here]

Moran's  $I$  statistic is a global statistic and does not allow assess the regional structure of spatial autocorrelation. In order to gain more insight into the way regions with high or low labor productivity are located in the European Union, we now analyze local spatial autocorrelation using Moran scatterplots (Anselin, 1996), and Local Indicators of Spatial Association "LISA" (Anselin, 1995).

First, Moran scatterplots plot the spatial lag  $Wz_i$  against the original values  $z_i$ . The four different quadrants of the scatterplot correspond to the four types of local spatial association between a region and its neighbors: HH a region with a high<sup>6</sup> value surrounded by regions with high values, LH a region with a low value surrounded by regions with high values, etc.

---

<sup>5</sup> All computations are carried out using SpaceStat 1.90 (Anselin 1999) and Arcview 3.2 (Esri).

<sup>6</sup> High (resp. low) means above (resp. below) the mean.

Quadrants HH and LL (resp. LH and HL) refer to positive (resp. negative) spatial autocorrelation indicating spatial clustering of *similar* (resp. *dissimilar*) values. The Moran scatterplot may thus be used to visualize atypical localizations, i.e. regions in quadrant LH or HL. Note that the use of standardized variables makes the Moran scatterplots comparable across time.

Second, Anselin (1995) defines a Local Indicator of Spatial Association (LISA) as any statistics satisfying two criteria: (i) the LISA for each observation gives an indication of significant spatial clustering of similar values around that observation; (ii) the sum of the LISA for all observations is proportional to a global indicator of spatial association. The local version of Moran's  $I$  statistic for each region  $i$  and year  $t$  is written as:

$$I_{i,t} = \frac{(x_{i,t} - \mu_t)}{m_0} \sum_j w_{ij} (x_{j,t} - \mu_t) \quad \text{with } m_0 = \sum_i (x_{i,t} - \mu_t)^2 / n \quad (12)$$

where  $x_{it}$  is the observation in region  $i$  and year  $t$ ,  $\mu_t$  is the mean of the observations across regions in year  $t$  and where the summation over  $j$  is such that only neighboring values of  $j$  are included. A positive value for  $I_{i,t}$  indicates spatial clustering of *similar* values (high or low) whereas a negative value indicates spatial clustering of *dissimilar* values between a region and its neighbors. Due to the presence of global spatial autocorrelation, inference must be based on the conditional permutation approach with 9999 permutations (Anselin, 1995). It should be stressed that  $p$ -values obtained for local Moran's statistics are actually pseudo-significance levels.

Combining the information in a Moran scatterplot and the significance of LISA yields to "Moran significance maps", showing the regions with significant LISA and indicating by a color code the quadrants in the Moran scatterplot to which these regions belong. Figures 1a, 1b, 2a, 2b, 3a, 3c, 4a and 4b display the Moran scatterplot maps using a 5% pseudo-

significance level for regional productivity in deviation from the EU average, its three shift-share components for the initial and final years of our sample.

Concerning first the Moran significance maps for regional productivity, a relative stability of the spatial patterns can be observed between 1975 and 2000. It appears that most European regions are characterized by positive local spatial association, i.e. they are significantly located in the HH or the LL quadrant. The significant HH regions are mostly to be found in Germany, Sweden and Austria. The regions in these countries therefore perform well in terms of productivity compared to the EU average. On the contrary, the significant LL regions are located in the South of France, Spain, Greece, South of Italy and most UK regions. The examination of these maps also allows detecting atypical regions characterized by negative local spatial autocorrelation. For example, some French, UK and Spanish regions perform well compared to their neighbors since they are significantly HL.

[Figures 1a and 1b about here]

The Moran significance maps for the three shift-share components in 1975 and 2000 are analyzed next. It appears that the spatial patterns for the first two components are relatively similar to that of labor productivity while the spatial pattern for the third component seems reversed. Therefore, we can expect a positive relationship between regional productivity, the industry-mix and the productivity differential components and a negative relationship between regional productivity and the allocative component.

[Figures 2a, 2b, 3a, 3b, 4a and 4b about here]

All the results presented in this section reveal the presence of a significant and positive spatial autocorrelation for all variables that is persistent over the period. This feature should be taken into account in our econometric estimations that are presented next.

## 5. Spatial SUR modeling

In this section, we first perform a specification search on the functional relations between productivity and the shift-share components. Second, these results are used in the specification of the pooled SUR and spatial SUR models.

### *Search of the appropriate functional form*

In order to have a sharper interpretation of the role played by each shift-share component, Esteban (2000) tests whether a model including a single component can explain interregional differences in aggregate productivity per worker. To this effect, he estimates models (9) on 4 datasets and for all of them he reports an almost perfect fit from the second component, with  $R^2$  statistics ranging from 0.9 to 0.975. The first component (industry mix) explains slightly more than 50% of the sample variability whereas the third one (allocative) does not seem to have a linear relation with productivity since its  $R^2$  statistics range from 0.06 to 0.2.

The aforementioned modeling procedure presupposes a linear relationship between productivity and the three components of its shift-share decomposition. This assumption needs to be tested; the first or third component may be very strongly related to productivity in a nonlinear fashion. For that purpose, we performed a model specification search on the pooled data. As can be seen from figures 5a, 5b and 5c, the relationship between productivity and the second shift share component is clearly linear, with increasing variability at increasing levels of the component. For the first and third components though, one should definitely look for some optimal transformations of the variables that strengthen linear relations and homogenize variance.

[Figures 5a, 5b and 5b about here]

Thus, we applied the Box-Cox method that seeks (via maximum likelihood) an optimal power transformation for the response. Despite the fact that we allowed polynomial forms of the explanatory (up to third order) the optimum power ranged between 0.9 and 1.2 in every case, indicating very little changes in our final relations. We continued by applying two nonparametric transformation procedures; the first proposed by Young *et al.* (1976)<sup>7</sup> and the second by Tibshirani (1986)<sup>8</sup>. In both cases, we neither observed clear functional relationships nor a significant improvement in the relationships between the transformed variables, as shown in figures 6a, 6b and 6c that represent the optimal functional forms for the relation between productivity and the allocative component, estimated by Tibshirani's method.

[Figures 6a, 6b and 6b about here]

#### *Estimation results for pooled and SUR regressions*

Since specification search indicates that no dramatic strengthening of linear relations occurs by a parametric or non-parametric transformation, we now fit a regression model on the pooled data. The main results of this analysis are displayed in table 2. It appears that the coefficients associated to the first and second shift-share component are significantly positive while the coefficient associated to the third component is significantly negative. These results are consistent with those previously obtained for ESDA. Compared to Esteban (2000), we observe a significantly worse fit for the first and a much better fit for the third component. The second model performs the best according to the information criteria.

[Table 2 about here]

Pooling the data implies that we cannot capture its temporal dimension. In particular, it is interesting to estimate how the relation of productivity with each shift-share component

---

<sup>7</sup> The SAS PROC TRANSREG procedure was used in that purpose.

<sup>8</sup> The R software-acepack package was used in that purpose.

evolves through time. For that purpose, we perform seemingly unrelated regressions (SUR) that allow the coefficients to be different in each time period and intertemporal dependence through the covariance matrix of the system of the regression equations.

In other words, we estimate for each component, the following relationships:

$$x_{it} - x_{EU,t} = a_{\mu,t} + b_{\mu,t}\mu_{it} + \varepsilon_{\mu,it} \quad i = 1, \dots, 205 \quad \text{and} \quad t = 1, \dots, 26 \quad (13a)$$

$$x_{it} - x_{EU,t} = a_{\pi,t} + b_{\pi,t}\pi_{it} + \varepsilon_{\pi,it} \quad i = 1, \dots, 205 \quad \text{and} \quad t = 1, \dots, 26 \quad (13b)$$

$$x_{it} - x_{EU,t} = a_{\alpha,t} + b_{\alpha,t}\alpha_{it} + \varepsilon_{\alpha,it} \quad i = 1, \dots, 205 \quad \text{and} \quad t = 1, \dots, 26 \quad (13c)$$

In this framework, the regression coefficients are assumed to be constant over space, but vary for each year. Wald statistics can therefore be used to test for the temporal stability of the coefficients. Moreover, the error terms are allowed to be correlated between years, such as:

$$E[\varepsilon_{it} \varepsilon_{is}] = \sigma_{ts}, \text{ or in matrix form:}$$

$$E[\varepsilon_t \varepsilon_t'] = \sigma_{ts} I_N \quad i = 1, \dots, 205 \quad \text{and} \quad t = 1, \dots, 26 \quad (14)$$

This assumption of dependence between equations can also be tested by means of a Lagrange multiplier test or a likelihood ratio test of the diagonality of the error covariance matrix.

Equations (13) with the error structure as in (14) are estimated via maximum likelihood<sup>9</sup> (ML) and via minimization of  $\frac{r'(S_{ols}^{-1} \otimes I)r}{n}$  where  $r$  stands for the pooled residuals vector and the  $S_{ols}$  matrix estimates the covariances of the errors across OLS equations. The basic advantage of the latter approach is that it does not require normality for the residuals.<sup>10</sup> The two SUR estimation approaches lead to practically the same results. The evolution of the magnitude of the coefficients when estimated by OLS regressions and by SUR (ML), are displayed in the first two rows of figure 7.

---

<sup>9</sup> The estimation results of ML were obtained using programs written in Python 2.2. They are available upon request from the authors.

[Figure 7 about here]

The reader may observe in figure 7 that there are large differences between coefficients that correspond to the same year/relation when estimated by OLS, compared to SUR. These differences are due to omitted temporal dependence, which in the SUR case is captured from the variance covariance matrix of the system of regression equations. As expected, the SUR covariance matrix (not presented due to space constraints) indicates a declining pattern for covariances for increasing temporal distance between equations.

Table 3 displays in columns 3, 4 and 5 the diagnostics and specification tests for the SUR models for each component. It appears that the choice of SUR is justified in every case by Lagrange multiplier and likelihood ratio statistics on the diagonality of the covariance matrix, which always reject the null hypothesis of intertemporal covariances equal to 0. Moreover, the Wald test on coefficient homogeneity across equations also rejects the null hypothesis in every case so that the coefficients associated to the three components are significantly different over time. However, these models seem to be misspecified since spatial autocorrelation is not taken into account. Indeed, the Lagrange Multiplier tests for spatial autocorrelation in the form of a spatial error (LMERR) and in the form of spatial lag (LMLAG) are all significant. The models should therefore be reestimated by allowing for spatial autocorrelation in each time period.

[Table 3 about here]

To determine the form taken by spatial autocorrelation, we compare the significance levels of the two tests, as in a cross-sectional setting (Anselin and Rey, 1991; Anselin and

---

<sup>10</sup> The SAS PROC SYSLIN procedure was used in that purpose.

Florax, 1995). For the first and third component, it appears that LMLAG is more significant, therefore a SUR model with a spatial lag should be estimated for those components:

$$x_{it} - x_{EU,t} = \rho_{\mu,t}(x_{it} - x_{EU,t}) + a_{\mu,t} + b_{\mu,t}\mu_{it} + \varepsilon_{\mu,it} \quad i = 1, \dots, 205 \text{ and } t = 1, \dots, 26 \quad (15a)$$

$$x_{it} - x_{EU,t} = \rho_{\alpha,t}(x_{it} - x_{EU,t}) + a_{\alpha,t} + b_{\alpha,t}\alpha_{it} + \varepsilon_{\alpha,it} \quad i = 1, \dots, 205 \text{ and } t = 1, \dots, 26 \quad (15b)$$

where  $\rho_t$  indicating the extent of spatial correlation in the dependent variable in each equation. On the contrary, for the second component, LMERR is more significant than LMLAG, so a SUR model with spatial autocorrelation error terms in each equation is the most appropriate specification:

$$\varepsilon_{\pi,t} = \lambda_{\pi,t} W \varepsilon_{\pi,t} + u_{\pi,t} \quad \text{with} \quad E[u_{\pi,t} u'_{\pi,s}] = \sigma_{\pi} I_N \quad (16)$$

where  $\lambda_{\pi,t}$  is a coefficient indicating the extent of spatial correlation between the residuals. Models (15a), (15b) and (15c) are estimated by ML. The evolutions of coefficients over time are displayed in the last row of figure 7 and the diagnostics and specification tests are displayed in columns 6, 7 and 8 of table 3.

Again, the choice of SUR is justified in every case by Wald tests on coefficient homogeneity across equations that reject the null hypothesis so that the coefficients associated to the three components are significantly different over time. The spatial coefficients however, are not significantly different over time.

What can be observed in figure 7 is that for the second shift share component, the evolution of intercepts and slopes remains stable regardless of the type of analysis used. For the first and third components though, results change dramatically with respect to the modeling framework. Error covariance among equations corresponding to different years proves to have much more explanatory power in this case. The three time series plots at the left column of figure 7 indicate that one gets misleading results if he/she performs separate

OLS regressions for each year separately. In that case, one overestimates slopes and intercepts, i.e. the effect of a unit change in industry mix (regional specification in more productive industries) to productivity changes and the productivity levels for regions corresponding to average EU industry mix. We should also underline the clear negative trend in intercepts through time, i.e. the decline of productivities for regions corresponding to average EU industry mix. A similar observation holds for the differential component (second shift share component); productivities tend to decline as time goes by for regions corresponding to EU average as far as technological or locational advantages are concerned. For the industry mix component on the other hand, slopes tend to slightly increase so that a unit change in industry mix affects productivities more as time goes by. For the allocative component, one has the same conjectures as with the industry mix concerning the evolution of coefficients, a result that is consistent with the similarity of their scatterplots in the descriptive analysis. However, separate OLS regressions for each year of the study tend to underestimate (instead of overestimate) intercepts and slopes. The coefficients that correspond to spatial effects are practically constant through time. One has to notice how intercepts change for the first and third components if spatial effects are not included in the SUR model.

## **6. Conclusion**

This study used Esteban's (2000) article as a benchmark and extended it both in terms of methods used and in terms of conjectures derived after using these techniques. Using a dataset of much larger temporal dimension than the one used in Esteban (2000), we observed significant spatial autocorrelation for productivity differentials and the components that define their decomposition according to industry mix, locational/technological advances or their covariance. Spatial autocorrelation appears to be a significant factor in correct model

specification. The same holds for temporal dependencies, which in fact lead to dramatic model misspecifications if omitted as displayed in the last part of the application.

Instead of making conjectures based on a pooled regression model, we use SUR models to capture temporal dependencies. Thus, we estimate sets of regressions equations, each one corresponding to a cross section of observations for a specific year. Temporal dependencies are captured implicitly through the common variance-covariance matrix of this system of regression equations. To account for significant spatial associations we add a spatial lag as an extra explanatory variable, or a spatial error autocorrelation coefficient for each equation. Our results provide a view of the evolution of European economy from 1975 to 2000: regional productivities tend to decline with respect to the EU average as time goes by, for regions that correspond to EU average in terms of technological/locational advantages or in terms of their industry mix proportions. Moreover, the industry mix tends to be a more significant factor as time goes by for productivity differentials.

## References

- Anselin L (1988a) Spatial econometrics: methods and models. Kluwer, Dordrecht
- Anselin L (1988b) A test for spatial autocorrelation in seemingly unrelated regressions. *Economics Letters* 28:335-341
- Anselin L (1995) Local indicators of spatial association-LISA. *Geographical Analysis* 27:93-115
- Anselin L (1996) The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In: Fisher M, Scholten HJ, Unwin D (eds) *Spatial analytical perspectives on GIS*. Taylor & Francis, London
- Anselin L (2001) Spatial econometrics. In: Baltagi B (ed) *Companion to econometrics*. Basil Blackwell, Oxford
- Anselin L, Florax R (1995) Small sample properties of tests for spatial dependence in regression models. In: Anselin L, Florax RJGM (eds) *New directions in spatial econometrics*. Springer-Verlag, Berlin
- Anselin L, Rey S (1991) Properties of tests for spatial dependence in linear regression models. *Geographical Analysis* 23:112-131
- Armstrong HW (2002) European Union regional policy: reconciling the convergence and evaluation evidence. In: Cuadrado-Roura JR, Parellada M (eds) *Regional convergence in the European Union: facts, prospects and policies*. Springer-Verlag, Berlin
- Browne LE (1989) Shifting regional fortunes: the wheel turns. *New England Economic Review*, Federal Reserve Bank of Boston
- Carlino GA (1992) Are regional per capita earnings diverging? *Business Review*, Federal Reserve Bank of Philadelphia, 3-12.
- Cliff AD, Ord JK (1981) *Spatial processes: models and applications*. Pion, London
- Cuadrado-Roura JR (2001) Regional convergence in the European Union: from hypothesis to the actual trends. *Annals of Regional Science* 35:333-356
- Dall'erba S (2003) Distribution of regional income and regional funds in Europe 1989-1999: an exploratory spatial data analysis. *Annals of Regional Science*, forthcoming
- Dunn ES (1960) A statistical and analytical technique for regional analysis. *Papers and Proceedings of the Regional Science Association* 6:97-112

- Esteban J (1972) A reinterpretation of shift-share analysis. *Regional and Urban Economics* 2:249-261
- Esteban J. (1994) La desigualdad interregional en Europa y en España: description y analysis. In: *Crecimiento y convergencia regional en España y Europa*, vol. 2, Instituto de Analisis Economico-CSIC y Fundacion de Economia Analytica, Barcelona
- Esteban J (2000) Regional convergence in Europe and the industry mix: A shift-share analysis. *Regional Science and Urban Economics* 30:353-364
- Le Gallo J, Ertur C (2003) Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995. *Papers in Regional Science*. 82:175-201
- Lopez-Bazo E, Vayà E, Mora AJ, Suriñach J (1999) Regional economic dynamics and convergence in the European Union. *Annals of Regional Science*. 33:343-370
- Manski CF (1993) Identification of endogenous social effects: the reflection problem. *Review of Economic Studies* 60:531-542
- Rey SJ, Montouri BD (1999) U.S. regional income convergence: a spatial econometric perspective. *Regional Studies* 33:145-156
- Rey SJ, Janikas MV (2003) Regional convergence, inequality and space. Department of Geography, San Diego State University.
- Terrasi M (2002) National and spatial factors in EU regional convergence. In: Cuadrado-Roura JR, Parellada M (eds) *Regional convergence in the European Union: facts, prospects and policies*. Springer-Verlag, Berlin
- Tibshirani R (1987) Estimating optimal transformations for regression. *Journal of the American Statistical Association* 83:394-405
- Wadley D, Smith P (2003) Straightening up shift-share analysis. *Annals of Regional Science*. 37:259-261
- Young FW, de Leeuw J, Takane Y (1976) Regression with qualitative and quantitative variables: an alternating least squares approach with optimal scaling features. *Psychometrika*. 41:505-529.

## Tables and maps

**Table 1: Moran's  $I$  statistics for regional productivity and the three shift-share components for 1975 and 2000**

Variable	1975			2000		
	<i>Moran's I</i>	<i>St. dev.</i>	<i>St. value</i>	<i>Moran's I</i>	<i>St. dev.</i>	<i>St. value</i>
<b>Productivity differential</b>	0.720	0.032	22.754	0.690	0.032	21.899
<b>Industry-mix component</b>	0.658	0.032	21.146	0.573	0.032	18.237
<b>Productivity differential component</b>	0.704	0.032	22.404	0.681	0.032	21.674
<b>Allocative component</b>	0.458	0.032	15.006	0.390	0.032	12.918

*Notes:* the expected value for Moran's  $I$  statistic is -0.005 for all variables. All statistics are significant at 1% level.

**Table 2: Ordinary least squares regression results on the pooled data**

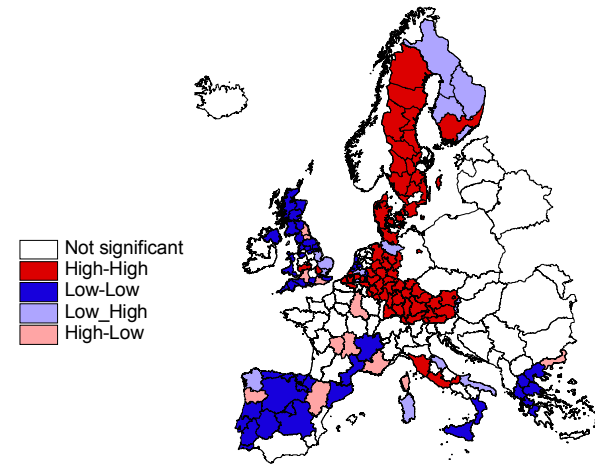
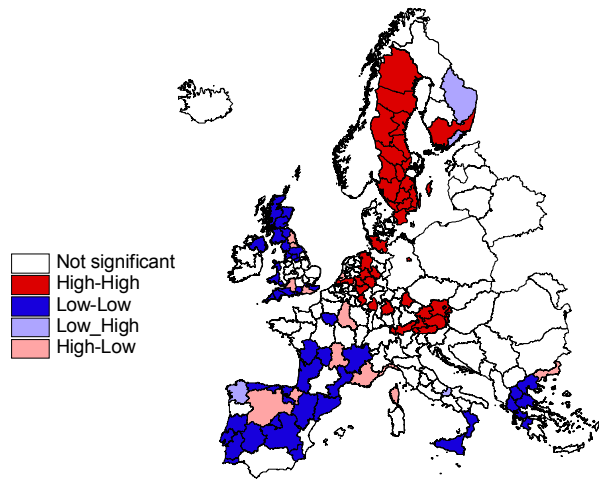
	Industry mix component	Regional component	Allocative component
$\hat{a}$	-0.291 (0.012)	-1.671 (0.000)	-3.439 (0.012)
$\hat{b}$	1.985 (0.000)	0.891 (0.000)	-1.821 (0.000)
R2	0.2273	0.9065	0.2394
R2 adjusted	0.2271	0.9065	0.2393
LIK	-17491.227	-111429.732	-17445.712
AIC	34986.453	22863.464	34895.424
SC	34999.764	22876.774	34908.734
$\hat{\sigma}^2$	70.595	8.540	69.484

*Notes:*  $p$ -values are in brackets. *LIK* is value of the maximum likelihood function. *AIC* and *SC* stand respectively for the Akaike and the Schwartz information criteria.

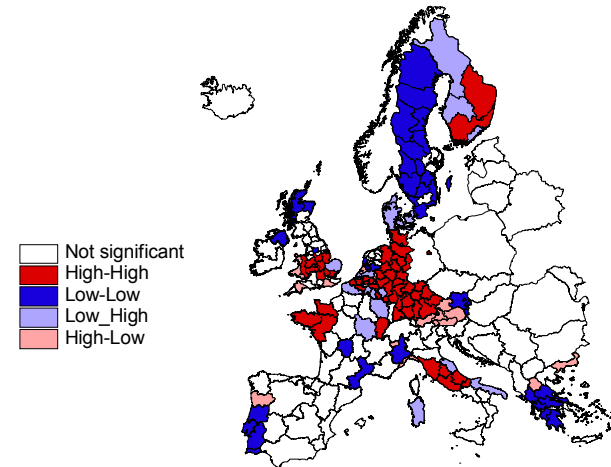
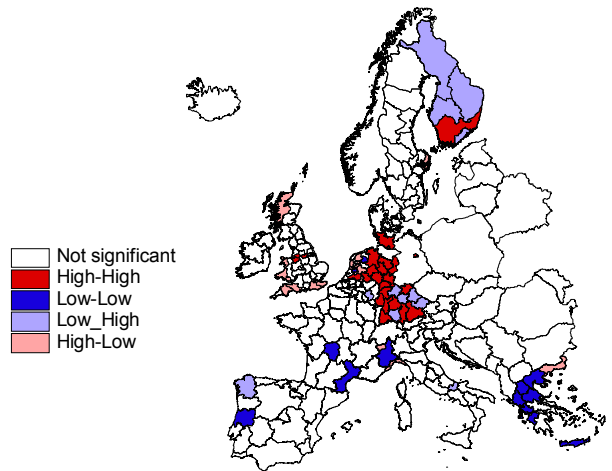
**Table 3: Tests results for SUR and spatial SUR**

	SUR			Spatial SUR		
	Industry mix component	Regional component	Allocative component	Industry mix component	Regional component	Allocative component
<b>R2</b>	0.0072	0.7401	0.0085	-	-	-
<b>R2 adjusted</b>	0.0558	0.9041	0.0627	-	-	-
<b>LIK</b>	-6197.148	-1632.722	-6166.429	-5863.912	-1405.201	-5824.364
<b>LM test on diagonality</b>	58303.612 (0.000)	50883.297 (0.000)	55954.890 (0.000)	-	-	-
<b>LR test on diagonality</b>	24975.751 (0.000)	22885.808 (0.000)	25019.654 (0.000)	-	-	-
<b>Test on coefficient homogeneity of slope</b>	224.466 (0.000)	272.330 (0.000)	186.667 (0.000)	74.676 (0.000)	222.470 (0.000)	133.515 (0.000)
<b>Test on coefficient homogeneity of spatial coefficient</b>	-	-	-	21.370 (0.672)	27.917 (0.312)	22.544 (0.604)
<b>LMERR</b>	636.579 (0.000)	482.218 (0.000)	675.381 (0.000)	-	-	-
<b>LMLAG</b>	655.239 (0.000)	188.957 (0.000)	692.918 (0.000)	-	-	-
<b>Wald test on spatial dependence</b>	-	-	-	1240.018 (0.000)	762.221 (0.000)	1242.793 (0.000)

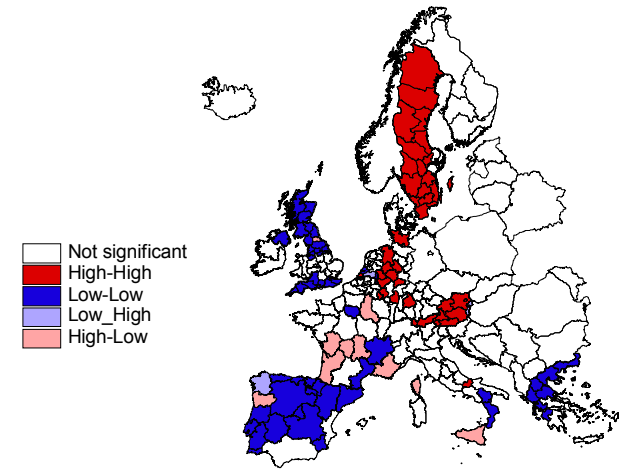
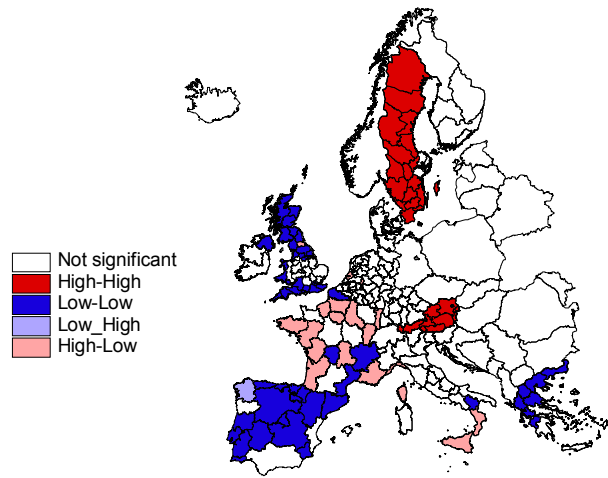
*Notes:* *p*-values are in brackets. LIK is value of the maximum likelihood function. LMERR and LMLAG stand for the Lagrange Multiplier tests respectively for residual spatial autocorrelation and spatially lagged endogenous variable in a SUR model (Anselin, 1988b). The LM and LR tests of diagonality stand respectively for the Lagrange Multiplier and the Likelihood Ratio test of diagonality of the error variance-covariance matrix.



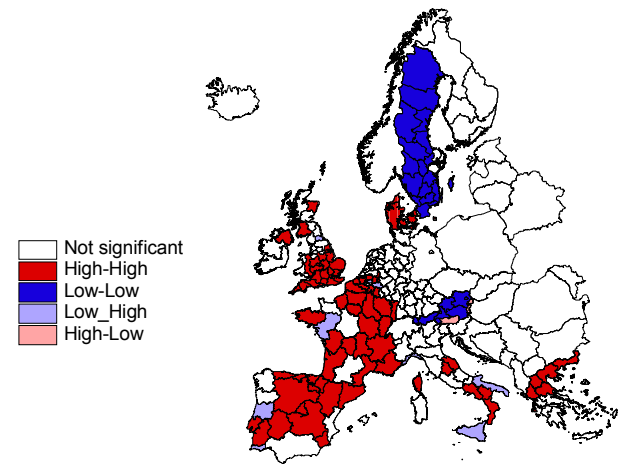
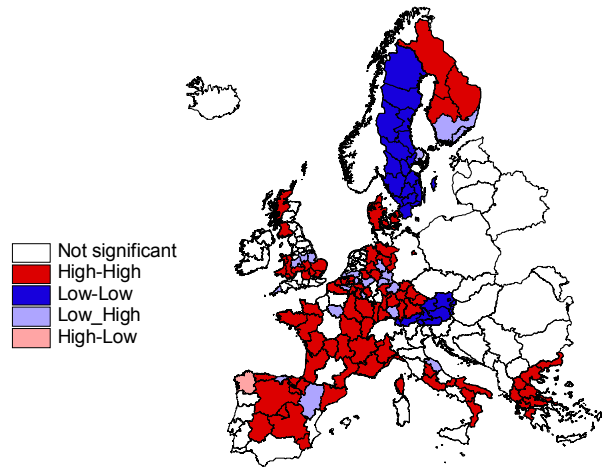
**Figures 1a and 1b: Moran's significance map for regional productivity in deviation from the EU average in 1975 and 2000**



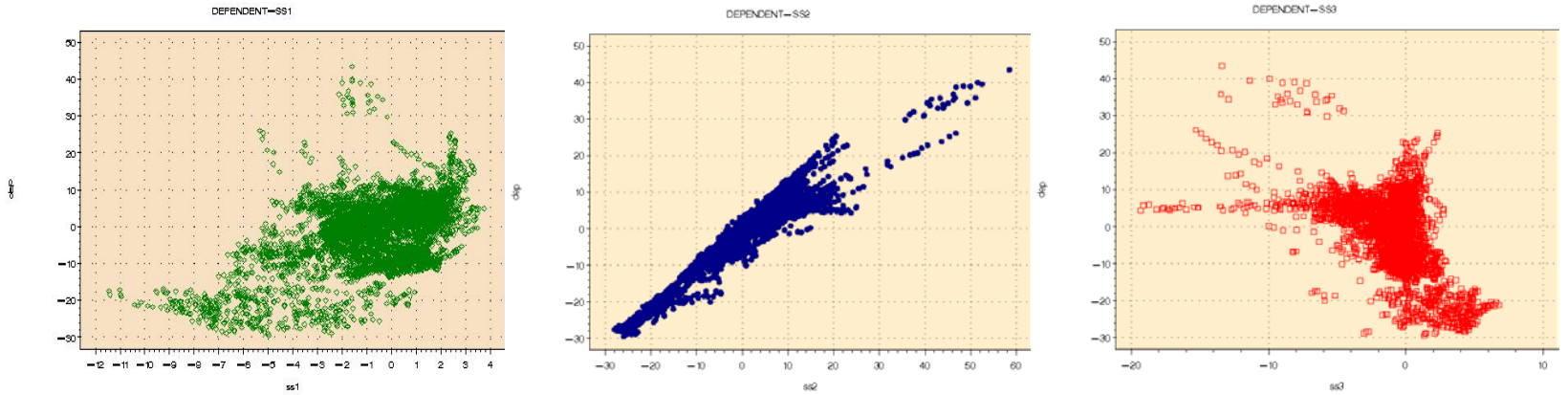
**Figures 2a and 2b: Moran's significance map for the industry-mix component in 1975 and 2000**



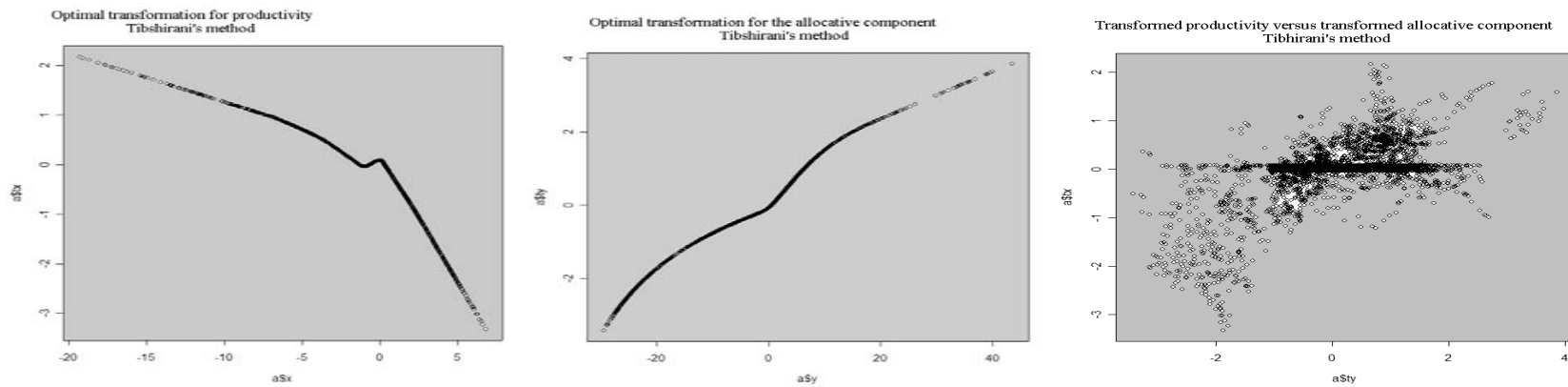
**Figures 3a and 3b: Moran's significance map for the productivity differential component in 1975 and 2000**



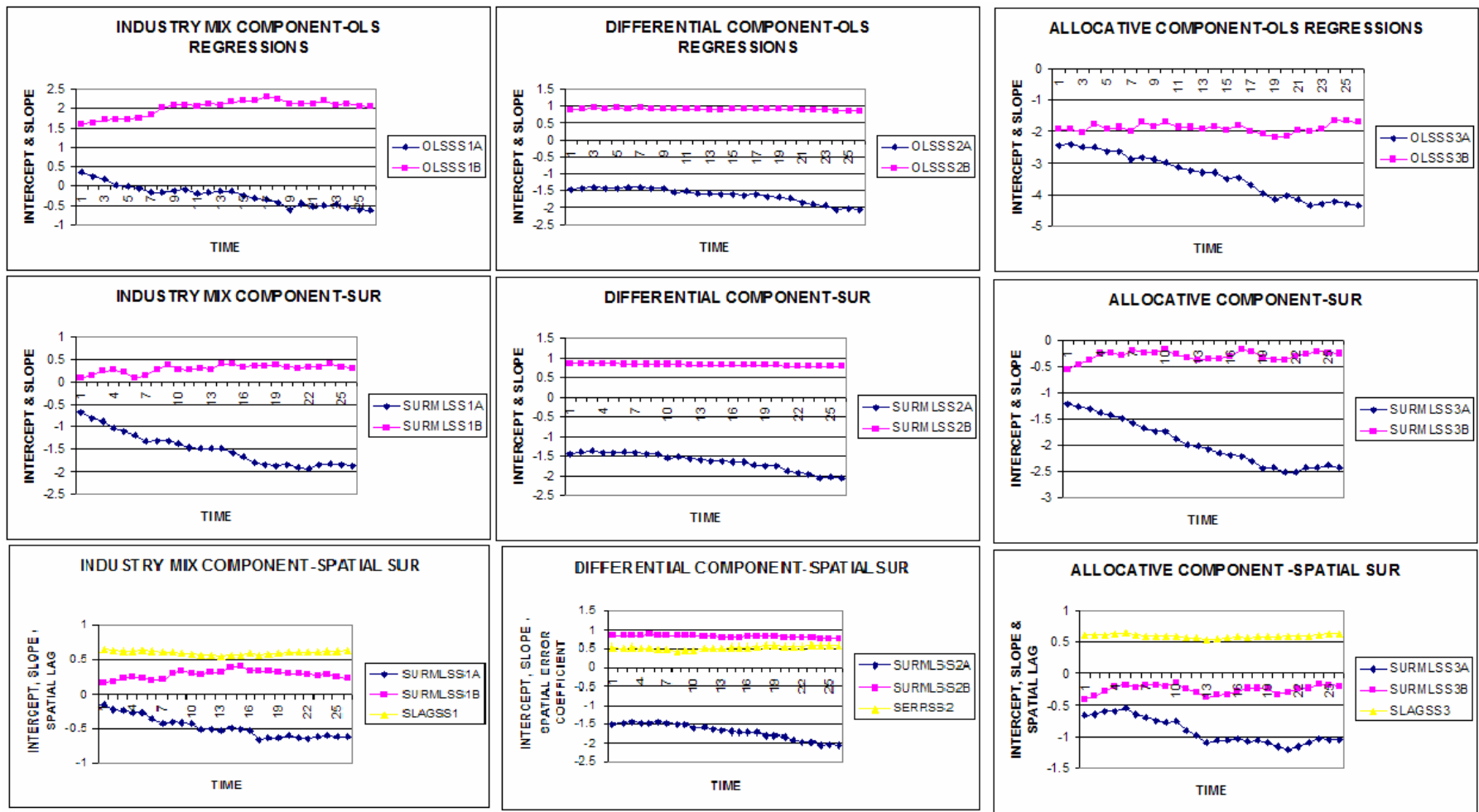
**Figures 4a and 4b: Moran's significance map for the allocative components in 1975 and 2000**



**Figures 5a 5b and 5c: Scatterplots of productivity with each shift share component**



**Figures 6a 6b and 6c: Optimal transformations (Tibshirani's method) for the productivity-allocative component relation.**



**Figure 7: The evolution of each model's coefficients through time. From left to right: Industry mix, Differential and Allocative component. From top to bottom: Separate OLS Regressions for each year, SUR and Spatial SUR**