Spatial shift-share analysis versus spatial filtering. An application to the Spanish employment

Matías Mayor Fernández (mmayorf@uniovi.es)
Ana Jesús López Menéndez (anaj@uniovi.es)
University of Oviedo, Department of Applied Economics.
Campus del Cristo, Oviedo (33006). Spain

The aim of this work is to analyze the influence of the spatial effects in the evolution of the regional employment, thus improving the explanation of the existing differences. With this aim, two non-parametric techniques are proposed: spatial shift-share analysis and spatial filtering.

Spatial shift-share models allow the identification and estimation of the spatial effects, as shown in Mayor and López (2005). On the other hand, the spatial filtering techniques can be used in order to remove the effects of the spatial correlation, thus allowing the decomposition of the employment variation into two components, respectively related to the spatial and structural effects.

The application of both techniques to the spatial analysis of the regional employment in Spain leads to some interesting findings, also showing the main advantages and limitations of each of the considered procedures and allowing the quantification of their sensibility with regard to the considered weights matrix.
1. Introduction
Shift-share analysis is a statistical tool that allows the study of regional development by means of the identification of two types of factors. The first group of factors operates in a more or less uniform way throughout the territory under review, although the magnitude of its impact on the different regions varies with its productive structure. The second type of factors has a more specific character and operates at a regional level.

Although according to Dunn (1960) the main objective of the shift-share technique is the quantification of geographical changes, the existence of spatial dependence and/or heterogeneity has barely been considered.

The classical shift-share approach analyzes the evolution of an economic magnitude between two periods by identifying three components: a national effect, a sectoral effect and a competitive effect. However, this methodology focuses on the dependence of the considered regions with respect to the national evolution but it does not take into account the interrelation among geographical units.

The need to include the spatial interaction has been acknowledged by Hewings (1976) in his revision of shift-share models. In the classical formulation, this spatial influence is gathered in a certain way, since the local predictions should converge on the national aggregate. Nevertheless, at the same time the estimation of the magnitude of sector i in region j is supposed to be independent from the growth of the same sector in another region k, an assumption which would only make sense in the case of a self-sufficient economy.

The increasing availability of data together with the development of spatial econometric techniques allows the incorporation of spatial effects into shift-share analysis. The aim is to obtain a competitive effect without spatial influence, allowing the differentiation between a common pattern in the neighbouring regions and an individual pattern of the specific considered region.

In order to achieve this objective, two different procedures are considered in this work, analyzing their suitability: the definition of a spatial weight matrix to be included into a shift-share model and the previous filtering of the considered variables.
The paper starts with a brief exposition of the classical shift-share identity, also describing the introduction of spatial dependence structures through spatial weights matrices.

In the third section some models of spatial dependence are presented, including the approach of Nazara and Hewings (2004) and some new proposals allowing the computation of spatial spillovers for each considered geographical unit and economic sector.

Section four describes the spatial filtering techniques, which can be used in order to remove the effects of the spatial correlation, thus allowing the decomposition of the employment variation into two components, respectively related to the spatial and structural effects.

An application of these models to the Spanish employment is presented in section five, and the paper ends with some concluding remarks summarized in section six.

2. Shift-share analysis and spatial dependence

The introduction of spatial dependence in a shift-share model can be carried out by two alternative methods. The first one is based on the modification of the classical identities of deterministic shift-share analysis by adding some new extensions, while the second one is based on a regression model (stochastic shift-share analysis) and the inclusion of spatial substantive and/or residual dependence.

According to Isard (1960), any spatial unit is affected by the positive and negative effects transmitted from its neighbouring regions. This idea is also expressed by Nazara and Hewings (2004), who assign great importance to spatial structure and its impact on growth. As a consequence, the effects identified in the shift-share analysis are not independent, since similarly structured regions can be considered in a sense to be “neighbouring regions” of a specified one, thus exercising some influence on the evolution of its economic magnitudes.

2.1. Classical shift-share analysis

If we denote by $X_{ij}$ the initial value of the considered economic magnitude corresponding to the i sector in the spatial unit j, $X'_{ij}$ being the final value of the same magnitude, then the change undergone by this variable can be expressed as follows:

$$\Delta X_{ij} = X'_{ij} - X_{ij} = X_{ij} \left( r_i - r_j \right) + X_{ij} \left( r_j - r_i \right)$$

where

$$r_i = \frac{\sum_{j=1}^{R} (X'_{ij} - X_{ij})}{\sum_{j=1}^{R} \sum_{i=1}^{S} X_{ij}}$$

$$r_j = \frac{\sum_{i=1}^{S} X_{ij}}{\sum_{j=1}^{R} \sum_{i=1}^{S} X_{ij}}$$

$$r' = \frac{X_{ij} - X_{ij}}{X_{ij}}$$
The three terms of this identity correspond to the shift-share effects:

- **National Effect** \( \text{NE}_{ij} = X_{ij}r \)
- **Sectoral or structural Effect** \( \text{SE}_{ij} = X_{ij}(r_i - r) \)
- **Regional or competitive Effect** \( \text{CE}_{ij} = X_{ij}(r_j - r_i) \)

As it can be appreciated, besides the national growth we should consider the positive or negative contributions derived from each spatial environment, known as the net effect. Thus the sectoral effect collects the positive or negative influence on the growth of the specialization of the productive activity in sectors with growth rates over or under the average, respectively. In its turn, the competitive effect collects the special dynamism of a sector in a region in comparison to the dynamism of the same sector at national level.

Once the regional and sectoral effects are calculated for each industry, their sum provides a null result, a property which Loveridge and Selting (1998) call “zero national deviation”.

Is spite of its limitations\(^1\), the shift-share technique is widely used in the analysis of spatial dynamics. In order to solve one of the main drawbacks of this method, related to the fact that the sectoral and regional effects depend on the industrial structure, Esteban-Marquillas (1972) introduced the idea of “homothetic change”, defined as the value that would take on the magnitude of sector \( i \) in region \( j \), if the sectoral structure of that region were assumed to be coincident with the national one. In this way, the homothetic change of sector \( i \) in region \( j \) is given by the expression:

\[
X^*_{ij} = \sum_{i=1}^{S} \frac{X_{ij}}{\sum_{j=1}^{R} \sum_{j=1}^{R} X_{ij}} = \frac{\sum_{i=1}^{S} X_{ij}}{\sum_{j=1}^{R} \sum_{j=1}^{R} X_{ij}} \sum_{j=1}^{R} X_{ij}
\]

leading to the following shift-share identity:

\[
\Delta X_{ij} = X_{ij}r + X_{ij}(r_i - r) + X^*_j (r_j - r_i) + (X_{ij} - X^*_j)(r_j - r_i)
\]

The third element of the right hand side of the equation is known as the “net competitive effect”, which measures the advantage or disadvantage of each sector in the region with respect to the total. The part of growth not included in this effect when \( X_{ij} \neq X^*_j \) is called the

---

\(^1\) Some limitations have been detected in shift-share analysis, derived, in the first place, from an arbitrary choice of the weights, which are not updated with the changes of the productive structure. Secondly, the obtained results are sensitive to the degree of sectoral aggregation and, furthermore, the growth attributable to secondary multipliers is assigned to the competitive effect when it should be collected by the sectoral effect, resulting in the interdependence of both components. Besides these problems, some authors as Dinc et al. (1998) emphasize the complexity related to the increasing of the spatial dependences between the sectors and the regions, which should be reflected in the model by means of the incorporation of some term of spatial interaction.
“locational effect”, corresponding to the last term of identity (1.3) and measuring the specialization degree.

2.2. The structure of spatial dependence: Spatial Weights

Since each region should not be considered as an independent reality, it would be advisable to develop a more complete version of the shift-share identity, keeping in mind that the economic structure of each spatial unit will depend on others, which are considered “neighbouring regions” in some sense. A suitable approach is the definition of a spatial weights matrix, thus solving the problems of multi-directionality of spatial dependence.

The concept of spatial autocorrelation attributed to Cliff and Ord (1973) has been the object of different definitions and, in a generic sense, it implies the absence of independence among the observations, showing the existence of a functional relation between what happens at a spatial point and in the population as a whole. The existence of spatial autocorrelation can be expressed as follows:

\[
\text{Cov}(X_j, X_k) = E(X_jX_k) - E(X_j)E(X_k) \neq 0
\]  

(1.4)

\(X_j, X_k\) being observations of the considered variables in units \(j\) and \(k\), which could be measured in latitude and length, surface or any other spatial dimension. In the empirical application included in this paper these spatial units are the European territorial units NUTS-III at the Spanish level.

The spatial weights are collected in a squared, non-stochastic matrix whose elements \(w_{jk}\) show the intensity of interdependence between the spatial units \(j\) and \(k\).

\[
W = \begin{bmatrix}
0 & w_{12} & \cdots & w_{1N} \\
w_{21} & 0 & \cdots & w_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
w_{N1} & w_{N2} & \cdots & 0
\end{bmatrix}
\]  

(1.5)

According to Anselin (1988), these effects should be finite and non-negative and they could be collected according to diverse options. A well-known alternative is the Boolean matrix, based on the criterion of physical contiguity and initially proposed by Moran (1948) and Geary (1954). These authors assume \(w_{jk}=1\) if \(j\) and \(k\) are neighbouring units and \(w_{jk}=0\) in another case, the elements of the main diagonal of this matrix being null.
In order to allow an easy interpretation, the weights are standardised by rows, so that they satisfy $0 \leq w_{jk} \leq 1$ and $\sum_k w_{jk} = 1$ for each row $j$. According to this fact, the values of a spatial lag variable in a certain location are obtained as an average of the values in its neighbouring units.

The consideration of different criteria for the development of the spatial weights matrix can deeply affect the empirical results. Thus, the contiguity can be defined according to a specific distance: $w_{jk} = 1$ if $d_{jk} \leq \delta$, being $d_{jk}$ the distance between two spatial units and $\delta$ the maximum distance allowed so that both can be considered neighbouring units.

In a similar way the weights proposed by Cliff and Ord depend on the length of the common border adjusted by the inverse distance between both locations:

$$w_{jk} = \frac{b_{jk}^6}{d_{jk}^{12}}$$

$b_{jk}$ being the proportion that the common border of $j$ and $k$ represents with respect to the total $j$ perimeter. From a more general perspective, weights should consider the potential interaction between the units $j$ and $k$ and could be computed as: $w_{jk} = d_{jk}^{-\alpha}$ and $w_{jk} = e^{-\beta d_{jk}}$.

In some cases the definition of weights is carried out according to the concept of “economic distance” as defined by Case et al. (1993) with $w_{jk} = \frac{1}{\sum_j |X_j - X_k|}$, $X_j$ and $X_k$, being the per capita income or some related magnitude. Some other authors as López-Bazo et al. (1999) suggest the use of weights based on commercial relations. Some alternative definitions have been developed by Fingleton (2001), with $w_{ij} = \sum_{t=0}^{J} d_{ij}^{-t}$ and Boarnet (1998), whose weights increase with the similarity between the investigated regions.

$$w_{ij} = \frac{1}{\sum_j \frac{1}{|X_i - X_j|}}$$

Together with the advantages of simplicity and easy use, the considered matrix shows some limitations, such as the non-inclusion of asymmetric relations, which is a requirement included in the five principles established by Paelink and Klaasen (1979).

The consideration of a binary matrix with weights based only on distance measures guarantees exogeneity but it can also affect the empirical results as indicated by López-Bazo, Vayá and Artís (2004). In this sense, it would be interesting to compare these results with those related to some alternative weights defined as a function of the economic variables of interest.

Boarnet (1998) defines a spatial weights matrix based on population density, per-capita income and the sectoral structure of the employment in each region. The considered matrix is also standardised by rows, since its expression guarantees that the aggregation of the weights for each region leads to a unitary result.

---

2 Together with the advantages of simplicity and easy use, the considered matrix shows some limitations, such as the non-inclusion of asymmetric relations, which is a requirement included in the five principles established by Paelink and Klaasen (1979).

3 The consideration of a binary matrix with weights based only on distance measures guarantees exogeneity but it can also affect the empirical results as indicated by López-Bazo, Vayá and Artís (2004). In this sense, it would be interesting to compare these results with those related to some alternative weights defined as a function of the economic variables of interest.

4 Boarnet (1998) defines a spatial weights matrix based on population density, per-capita income and the sectoral structure of the employment in each region. The considered matrix is also standardised by rows, since its expression guarantees that the aggregation of the weights for each region leads to a unitary result.
The choice of the spatial weight matrix is a key step in the spatial econometric modelling and nowadays there is not a unique method to select an appropriate specification of this matrix. In fact, this problem is suggested for future research by Anselin et al. (2004), and Paelink et al. (2005) among others.

3. Models of spatial dependence

The extension of the shift-share model proposed by Nazara and Hewings (2004) introduces the spatially modified growth rates according to the previously assigned spatial weights:

\[ r_{ij} = r + (r_{ij}^y - r) + (r_{ij} - r_{ij}^y) \]  \hspace{1cm} (1.8)

where \( r_{ij}^y \) is the rate of growth of the i sector in the neighbouring regions of a given spatial unit j which can be obtained as follows:

\[ r_{ij}^y = \frac{\sum_{k \in V} w_{jk} X_{ik}^{t+1} - \sum_{k \in V} w_{jk} X_{ik}^t}{\sum_{k \in V} w_{jk} X_{ik}^t} \]  \hspace{1cm} (1.9)

It must be noted that the \( w_{jk} \) elements correspond to the previously defined matrix of rows-standardized weights. In any case, regional interactions are supposed to be constant between the considered periods of time, as it is usually assumed in spatial econometrics.

Three components are considered in expression (1.8), the first one corresponding to the national effect, which is equivalent to the first effect of the classical (non-spatial) shift-share analysis. The second one, the sectoral effect or industry mix neighbouring regions-nation effect, shows a positive value when the evolution of the considered sector in the neighbouring regions of j is higher than the average. Finally, the third term is the competitive neighbouring regions effect and compares the rate of growth in region j of a given sector i with the evolution of the spatially modified sector. Thus, a negative value of this effect shows a regional evolution that is worse than the one registered in the neighbouring regions, meaning that region j fails to take advantage of the positive influence of its neighbouring regions.

A weakness can be found in the previously defined model, since a single spatial weights matrix is considered for the computation of the different spatially modified rates of sectoral and global growth. This assumption would not be so problematic if we used the binary matrix, instead of endogenous matrices which would vary sensitively depending on the sectoral or global adopted perspective. On the other hand, the use of the same structure of weights in the
initial and final periods could be considered too simplistic, suggesting the need of developing a dynamic version.

Mayor and López (2005) develop an alternative approach in order to compute to what extent a spatial unit is being affected by the neighbouring territories. This procedure consists on introducing homothetic effects analogous to those defined by Esteban-Marquillas (1972) but referring to a regional environment. In this way, we would be able to define the value that the magnitude of sector i in region j would have taken if the sectoral structure of j were similar to its neighbouring regions. More specifically, the homothetic change with respect to the neighbouring regions would be given by the expression:

\[
X_{ij}^v = \sum_{i=1}^{S} X_{ik} \frac{\sum_{k \in V} X_{ik}}{\sum_{i=1}^{S} \sum_{k \in V} X_{ik}}
\]

(1.10)

A more complete option is based on the use of a spatial weights matrix. In this case the economic magnitude is defined as a function of the neighbouring values, and, therefore, the concept of homothetic employment would be substituted by the spatially influenced employment, which would be computed according to a certain structure of spatial weights (W) and the effectively computed employment for each combination region-sector. The following identity would then hold:

\[
\Delta X_{ij} = X_{ij}^r + X_{ij}^v \left( r_i - r \right) + X_{ij}^{r^*} \left( r_j - r \right) + (X_{ij}^r - X_{ij}^{r^*}) \left( r_i - r \right)
\]

(1.11)

where the value of the magnitude is obtained from its neighbouring regions as:

\[
X_{ij}^{r^*} = \sum_{k \in V} w_{jk} X_{ik}
\]

(1.12)

V being the set of neighbouring regions of j. One of the drawbacks of this spatially influenced employment is related to the fact that, as a consequence of the considered expression, it can be observed that: \( \sum_{i,j} X_{ij}^{r^*} \neq \sum_{i,j} X_{ij} \). This fact leads to two considerations with respect to the usefulness of the proposed definition: on the one hand, the amounts of the effects for each sector-region are going to be in some cases sensitively different to those obtained in the equivalent model of Esteban-Marquillas (1972), leading to a more difficult interpretation and comparison of the obtained results. On the other hand, as a result of the structure of the spatial weights, the expected level of employment would be different to the effective one.
In order to solve both problems, an alternative concept is proposed using new spatially modified sectoral weights based on the spatially influenced employment (1.18):
\[
\frac{\sum_{j=1}^{R} X_{ij}^{*}}{\sum_{j=1}^{R} \sum_{i=1}^{R} X_{ij}^{*}} = \frac{X_{ij}^{*}}{X^{*}},
\]
leading to the so-called homothetic spatially influenced employment:
\[
X_{ij}^{**} = X_{ij} \frac{X_{ij}^{*}}{X^{*}} \tag{1.13}
\]

It must be pointed out that this new concept satisfies the identity \( \sum_{i,j} X_{ij}^{**} = \sum_{i,j} X_{ij} \), although substantial differences are found in the distribution of the variable for each combination sector-spatial unit. The substitution of the expression (1.19) in (1.17) leads to the identity:
\[
X_{ij} r + X_{ij} (r_i - r) + X_{ij}^{**} (r_j - r_j) + (X_{ij} - X_{ij}^{**}) (r_j - r_i) \tag{1.14}
\]
where the third expression is the spatial competitive net effect (SCNE**) and the fourth is the spatial locational effect (SLE**).

4. Spatial filtering

An alternative approach in order to deal with spatial autocorrelation in regression analysis involves the filtering of variables allowing the elimination of the spatial effects. The most well-known filtering procedures are those proposed by Getis (1990, 1995) based on the statistic of local association \( G_i \) (Getis and Ord, 1992) and the Griffith’s (1996, 2000) alternative procedure based on the eigenfunction decomposition associated with the Moran statistic.

Since one of the main problems in spatial regressions is related to the presence of stochastic regressors, leading to biased Ordinary Least Squares (OLS) estimations, Getis (1990) develops a new procedure, based on the decomposition of a variable into two components (spatial and non-spatial) through the use of a filter or screen which removes the spatial component of each of the considered variables.

In this work we consider this screening procedure as a decomposition technique previous to further analyses. The spatial filtering developed by Getis (1990) is based on the consideration of a spatial vector \( S \):
\[
S \approx \rho W \tag{1.15}
\]
which takes the place of both the spatial weights matrix and the auto-regressive coefficient $\rho$.

We must point out that the $S$ vector must be designed in order to capture the spatial dependence in the considered data. Its construction is based on data points, but in dealing with surface partitions, points could be considered as the reference of different spatial areas and this vector allows the conversion of the dependent variable in its non-spatial equivalence:

$$y^* = y - S$$  \hspace{1cm} (1.16)

Once the model includes all the non-spatial variables, it can be specified and estimated through the OLS method, leading to an unbiased estimation.

In Getis (1990), $S$ is found by means of the multistep second-order method developed by Ripley (1981). Haining (1982) asserts the importance of the second-order moment properties since one specific location on a map can not be considered independent from other locations$^5$.

Getis (1990) applied the local $K$-function $L_i(d) = A \sum_{j \neq i, j \neq i} k_{ij}(d) / \pi(n-1)$ as an association ratio and compares the expected and observed values for each individual observation being $A$ the region size and $\sum_{j \neq i, j \neq i} k_{ij}(d)$ the aggregation over all points located within distance $d$ of point $i$.$^6$

Although global tests as Moran I and Geary $c$ are generally used in a global context, a more detailed (local) detection of the spatial association is often required. Therefore, a modified version of the filtering procedure is developed by Getis (1995), based on the local statistic $G_i(d)$ by Getis and Ord (1992), which computes the degree of association due to the concentration of points within a distance $d$.

Given a region divided into $n$ subregions (which are considered as points with known values) $G_i(d)$ is the ratio between the sum of the $x_j$ values included in a $d$ distance from the $i$ point and the sum of the values in all the regions excluding $i$:

$$G_i(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j}{\sum_{j=1}^{n} x_j}; i \neq j$$  \hspace{1cm} (1.17)

The matrix of spatial weights is binary, being $w_{ij}(d) = 1$ if $d_{ij} \leq d$ and $w_{ij}(d) = 0$ if $d_{ij} > d$.

Getis and Ord (1992) deduce the expressions of the expected value and the variance under the spatial independence hypothesis:

---

$^5$ The idea is based on the consideration of the set of distances between all pairs of points because of the great information provided by the $N(N-1)/2$.

$^6$ Getis (1990) proposes as choice criteria the maximization of the expression $\sum[(L_i(d) - L_i(d))^2]$.
Expression \( G_i(d) \) measures the concentration of the sum of values in the considered area, and would increase their result when high values of \( X \) are found within a \( d \) distance from \( i \). In general terms, the null hypothesis is that the values within a \( d \) distance from \( i \) are a random sample drawn without replacement from the set of all possible values. Then, assuming that the statistic is normally distributed, the existence of spatial dependence can be tested from the following expression:

\[
Z_i = \frac{G_i(d) - E[G_i(d)]}{\sqrt{\text{Var}(G_i(d))}} \tag{1.20}
\]

Getis (1995) proposes the computation of the filtering vector from the values of \( G_i(d) \) statistic. Since the expected value of the Getis statistic, \( E[G_i(d)] \), represents the value in location \( i \) when the spatial autocorrelation is not present, then the ratio \( G_i(d)/E[G_i(d)] \) is used in order to remove the spatial dependence included in the variable. If the considered statistic is higher than its expected value then the spatial dependence results to be positive. In order to remove this spatial dependence from the considered variable we obtain the filtered series:

\[
\tilde{x}_i = \frac{x_i \left( \frac{W_i}{n-1} \right)}{G_i(d)} \tag{1.21}
\]

leading the difference between the original and the filtered series to a new variable which shows the spatial dependence \( L = X - \tilde{X} \).

According to Getis and Griffith (2002), two main ideas can be identified in the filtering procedure: firstly, it is necessary to identify a correct distance \( d \) to include the spatial

---

As expected, the variance of this statistic would be null when no neighbouring regions exist \((W_i=0)\), when all the \( n-1 \) regions result to be neighbouring regions of \( i \) \((W_i=n-1)\) and also when values assigned to the \( n-1 \) observations are coincident \((Y_{i1}=0)\).
dependence among the regions and secondly, the contribution to the spatial dependence of each individual observation should be computed.

The main point is to find an optimal value \( d \) which maximizes the existing spatial dependence. With this aim, Getis (1995) proposes to maximize the absolute value of the sum of the standard variation of statistic \( G_k(d) \) for all the observations of \( X \).

\[
\max \sum_{k=1}^{g} |Z_k| = \max \sum_{k=1}^{g} \left| G_k(d) - E\left(G_k(d)\right) \right| \sqrt{\text{Var}\left(G_k(d)\right)} \quad (1.22)
\]

### 4.1 Spatial filtering models

Once we have described the filtering process we propose some useful models, analyzing their main characteristics, advantages and limitations.

**Model 1:** Once the filtering process has finished, a traditional shift-share analysis can be carried out considering both the spatial and non-spatial (filtered) component of the variables. The obtained results are not strictly comparable to those related to the original data, due to two different reasons: first, we must take into account that different filters are applied to the original and final periods and second, the considered rates of growth are different in each case. Thus, the rates of growth for the filtered variable \( \tilde{X} \) are:

\[
\tilde{r} = \frac{\tilde{X}^t - \tilde{X}^{t-k}}{X^{t-k}} \quad \tilde{r}_i = \frac{\tilde{X}^t_i - \tilde{X}^{t-k}_i}{X^{t-k}_i} \quad \tilde{r}_{ij} = \frac{\tilde{X}^t_{ij} - \tilde{X}^{t-k}_{ij}}{X^{t-k}_{ij}} \quad (1.23)
\]

leading to the following shift-share decomposition:

\[
\Delta \tilde{X}_{ij} = \tilde{X}_{ij}^{t} \tilde{r} + \tilde{X}_{ij}^{t-k} \tilde{r}_i + \tilde{X}_{ij}^{t-k} \tilde{r}_{ij} \quad (1.24)
\]

In a similar way, we can define the rates of growth for the spatial component \( X - \tilde{X} \) = \( L \), leading to the following decomposition:

\[
\Delta L_{ij} = L_{ij}^{t} r^t + L_{ij}^{t-k} (r_i^{t-k} - r^t) + L_{ij}^{t-k} (r_{ij}^{t-k} - r_{ij}^{t}) \quad (1.25)
\]

where \( r^t, r_i^t, r_{ij}^{t} \) are respectively the global, sectoral and regional-sectoral rates of growth.

The described model leads to some interesting results although, as explained above, it is non comparable with the traditional shift-share and, therefore, the sum of spatial and non spatial effects is not expected to coincide with that obtained in the classical identity applied to the original data. In fact, the coincidence is only verified by the national effect.
**Model 2:** In this option, two new effects can be defined: the spatial competitive effect (SCE) and the non-spatial or filtered competitive effect (FCE). The proposal is similar, in some sense, to the Esteban-Marquillas decomposition. In this case, the homothetic employment is substituted by the expected level of the variable without spatial influences and the deviation between the expected and real values is due to the spatial spillover effects. The deviation, the spillover effect is measured in terms of a variable\(^8\).

Thus, the Spatial competitive effect and the Filtered competitive effect are given by the following expressions:

\[
SCE_{ij} = L_{ij} (r_j - r_i) = (X_{ij} - \bar{X}_j)(r_j - r_i) \tag{1.26}
\]

\[
FCE_{ij} = \bar{X}_{ij} (r_j - r_i) \tag{1.27}
\]

It can be proved that this decomposition verifies the additivity property, in a similar way to the original Esteban-Marquillas model. Thus, the filtered competitive effect is strictly comparable to that of the traditional shift-share, since the identity \(CE=FCE+SCE\) is held.

**Model 3:** A new option could be the comparison between the results obtained with filtered values and those obtained with the spatial shift-share developed by Mayor and Lopez (2005). We are trying to define an alternative concept to the homothetic change by Esteban-Marquillas (1.10) and the homothetic spatially influenced variable by Mayor and López (2005) (1.13) by using a modified sectoral weight (without spatial spillovers) based on the values of the filtered variable. The spatially influenced variable (1.12) is substituted by the filtered value (\(\bar{X}\)):

\[
\sum_{j=1}^R \bar{X}_{ij} \sum_{i=1}^R \bar{X}_{ij} = \bar{X}_i \bar{X} \tag{1.28}
\]

Thus, the filtered homothetic employment based on the non-spatial component of the variable would be obtained as follows:

\[
\bar{X}_{ij}^{**} = X_j \frac{\bar{X}_j}{X} \tag{1.29}
\]

leading to the following decomposition:

---

\(^8\) Although this new decomposition is initially referred to the competitive effect it could also be extended to another components (Arcelus, 1984).
\[ \Delta X_{ij} = X_{ij}r + X_{ij}(r_i - r) + \bar{X}_{ij}^* (r_i - r) + (X_{ij} - \bar{X}_{ij}^*)(r_i - r_i) \]  

(1.30)

from which two different effects can be identified: first, the filtered net competitive effect (FNCE) which describes the expected change in the variable assuming the national sectorial structure without spatial spillovers, and, second, the non-filtered locational effect (SLE) computing the difference between expected and real change of the variable due to the sectoral specialization of the region together with the spillover effects. In this case, it is verified that the sum of both effects leads to the same result as in the traditional shift-share.

The computation of dynamic effects would be very interesting in order to obtain large series of spatial and non-spatial competitive effects, thus allowing their modelling and forecasting.

5. Some findings for the Spanish case

The previously described models can be applied to the Spanish case, analyzing the sectoral evolution of the regional employment. More specifically, in this section we are focusing on the four main economic activities (agriculture, industry, construction and services) considering the European territorial units NUTS-III at a Spanish level leading to a total of 47 provinces.\(^9\)

The information has been provided by the Spanish Economically Active Population Survey (EPA), whose methodology was modified in 2005 due to three different reasons: the need to adapt to the new demographic and labour reality of Spain (due mainly to the increase in the number of foreign residents), the incorporation of new European regulations in accordance with the norms of the European Union Statistical Office (EUROSTAT) and the introduction of improvements in the information gathering method (changes in questionnaires and interviews carried out by the Computer Assisted Telephone Interviewing –CATI- method).

The shift-share analysis was carried out during the period 1999-2004 leading to some interesting findings related to sectoral and spatial patterns.

The Moran test was carried out in order to detect the spatial autocorrelation, leading to the conclusion that a slightly positive spatial autocorrelation exists among the Spanish provinces. More specifically, two different specifications of the spatial matrix have been considered in these tests: a binary exogenous matrix and a distance (km) based matrix. In the first case, the weights of the matrix are assumed to be 1 for neighbouring provinces and null in the remaining cases. Regarding the second option, the weights are obtained from the expression

\[ w_{ij} = d_{ij}^{-1}. \]

---

\(^9\) According to the methodology of our study, Ceuta and Melilla and the Balearic and Canary Islands are excluded since the definition of neighbouring region does not exactly fit to these cases.
The results obtained for the autocorrelation Moran test are summarized in table 1 and refer to the rate of growth in the considered period.10

**Table 1: Results of autocorrelation Moran test**

<table>
<thead>
<tr>
<th></th>
<th>Neighbour matrix</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>z</em>-value</td>
<td><em>p</em>-value</td>
</tr>
<tr>
<td><strong>Growth rates of gross employment</strong></td>
<td>Agriculture</td>
<td>3.757</td>
</tr>
<tr>
<td></td>
<td>Industry</td>
<td>-0.534</td>
</tr>
<tr>
<td></td>
<td>Construction</td>
<td>4.031</td>
</tr>
<tr>
<td></td>
<td>Services</td>
<td>0.726</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>4.151</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Neighbour matrix</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>z</em>-value</td>
<td><em>p</em>-value</td>
</tr>
<tr>
<td><strong>Growth rates of filtered employment</strong></td>
<td>Agriculture</td>
<td>-0.137</td>
</tr>
<tr>
<td></td>
<td>Industry</td>
<td>-0.613</td>
</tr>
<tr>
<td></td>
<td>Construction</td>
<td>1.192</td>
</tr>
<tr>
<td></td>
<td>Services</td>
<td>-0.159</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>-0.301</td>
</tr>
</tbody>
</table>

As a first step the filtering process has been carried out to the variables (levels of sectoral employment in agriculture, industry, construction and services in the 47 Spanish NUTS III). In each spatial unit, the local spatial autocorrelation statistic $G_i(d)(1.17)$ is evaluated at a series of increasing distances (10km) together with its characteristics $E(G_i(d))$ and $\text{Var}(G_i(d))$, according to the previously considered expressions. The “optimal” distance11 was computed in order to obtain the filtered variable according to expression (1.21).

Once the filters have been identified, the previously explained models are applied in order to know the contributions of the spatial effects to the employment change.

With respect to the first model, the effects obtained according to (1.24) (the filtered national, sectoral and competitive effects) are compared to those related to the original (non filtered) variable in table 2.

The interpretation of these results should stress that the final effects of the spatial dependence are slightly negative, and therefore the elimination of the spatial effect would lead to an employment of 14451.86 versus 13671.75 with spatial effect in 1999. If we considered the filtered variables the change in the employment level in the period 99-04 would be 3291.95 instead of 2997.60.

10 Longhi and Nijkamp (2005) use the Moran test to detect autocorrelation in the employment levels and also in the absolute and relative changes of employment.

11 In this case we have considered the “optimal” distance, selecting the distance maximizing the spatial dependence according to expression (1.22). More specifically, the selected distances for the employment levels in year 1999 are 425 Km in Agriculture, 550 Km in Industry; 450 Km in construction and 550 Km in Services, while the distances in year 2004 are 425 Km in Agriculture, 550 Km in Industry; 450 Km in construction and 450 Km in Services. In this approach, as distance increases from one point, the local statistics also increase if spatial autocorrelation is detected.
As a consequence we can conclude that the aggregated national effect of the filtered value would increase in a 9% while the computation by provinces reflects the different spatial schemes.

<table>
<thead>
<tr>
<th>NUTS</th>
<th>Total</th>
<th>Filtered</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Albacete</td>
<td>28.3</td>
<td>26.4</td>
</tr>
<tr>
<td>2</td>
<td>Alicante</td>
<td>117.2</td>
<td>90.1</td>
</tr>
<tr>
<td>3</td>
<td>Almeria</td>
<td>40.6</td>
<td>37.7</td>
</tr>
<tr>
<td>4</td>
<td>Ávila</td>
<td>11.9</td>
<td>15.4</td>
</tr>
<tr>
<td>5</td>
<td>Badajoz</td>
<td>433.0</td>
<td>46.9</td>
</tr>
<tr>
<td>6</td>
<td>Barcelona</td>
<td>90.9</td>
<td>480.8</td>
</tr>
<tr>
<td>7</td>
<td>Bilbao</td>
<td>28.4</td>
<td>119.8</td>
</tr>
<tr>
<td>8</td>
<td>Burgos</td>
<td>27.2</td>
<td>38.7</td>
</tr>
<tr>
<td>9</td>
<td>Cáceres</td>
<td>27.2</td>
<td>32.1</td>
</tr>
<tr>
<td>10</td>
<td>Cádiz</td>
<td>67.5</td>
<td>86.6</td>
</tr>
<tr>
<td>11</td>
<td>Castellón</td>
<td>41.7</td>
<td>36.0</td>
</tr>
<tr>
<td>12</td>
<td>Ciudad Real</td>
<td>32.9</td>
<td>37.1</td>
</tr>
<tr>
<td>13</td>
<td>Córdoba</td>
<td>47.0</td>
<td>42.8</td>
</tr>
<tr>
<td>14</td>
<td>Coruña (A)</td>
<td>85.9</td>
<td>151.1</td>
</tr>
<tr>
<td>15</td>
<td>Cuenca</td>
<td>13.6</td>
<td>14.8</td>
</tr>
<tr>
<td>16</td>
<td>Girona</td>
<td>54.9</td>
<td>38.7</td>
</tr>
<tr>
<td>17</td>
<td>Granada</td>
<td>50.1</td>
<td>42.3</td>
</tr>
<tr>
<td>18</td>
<td>Guadalajara</td>
<td>12.9</td>
<td>15.3</td>
</tr>
<tr>
<td>19</td>
<td>Huelva</td>
<td>29.5</td>
<td>37.8</td>
</tr>
<tr>
<td>20</td>
<td>Huesca</td>
<td>17.0</td>
<td>16.9</td>
</tr>
<tr>
<td>21</td>
<td>Jaén</td>
<td>42.6</td>
<td>41.6</td>
</tr>
<tr>
<td>22</td>
<td>León</td>
<td>35.8</td>
<td>46.0</td>
</tr>
<tr>
<td>23</td>
<td>Lleida</td>
<td>32.4</td>
<td>31.1</td>
</tr>
<tr>
<td>24</td>
<td>Logroño</td>
<td>22.6</td>
<td>24.4</td>
</tr>
<tr>
<td>25</td>
<td>Lugo</td>
<td>29.9</td>
<td>33.6</td>
</tr>
<tr>
<td>26</td>
<td>Madrid</td>
<td>455.9</td>
<td>577.5</td>
</tr>
<tr>
<td>27</td>
<td>Málaga</td>
<td>86.0</td>
<td>70.7</td>
</tr>
<tr>
<td>28</td>
<td>Murcia</td>
<td>90.3</td>
<td>76.0</td>
</tr>
<tr>
<td>29</td>
<td>Orense</td>
<td>25.4</td>
<td>28.8</td>
</tr>
<tr>
<td>30</td>
<td>Oviedo</td>
<td>74.6</td>
<td>90.1</td>
</tr>
<tr>
<td>31</td>
<td>Palencia</td>
<td>49.6</td>
<td>17.3</td>
</tr>
<tr>
<td>32</td>
<td>Pamplona</td>
<td>71.8</td>
<td>51.3</td>
</tr>
<tr>
<td>33</td>
<td>Pontevedra</td>
<td>24.7</td>
<td>71.1</td>
</tr>
<tr>
<td>34</td>
<td>Salamanca</td>
<td>60.5</td>
<td>31.7</td>
</tr>
<tr>
<td>35</td>
<td>San Sebastián</td>
<td>60.5</td>
<td>73.2</td>
</tr>
<tr>
<td>36</td>
<td>Santander</td>
<td>39.1</td>
<td>48.4</td>
</tr>
<tr>
<td>37</td>
<td>Segovia</td>
<td>13.0</td>
<td>16.4</td>
</tr>
<tr>
<td>38</td>
<td>Sevilla</td>
<td>111.4</td>
<td>111.2</td>
</tr>
<tr>
<td>39</td>
<td>Soria</td>
<td>8.2</td>
<td>9.3</td>
</tr>
<tr>
<td>40</td>
<td>Tarragona</td>
<td>56.2</td>
<td>45.1</td>
</tr>
<tr>
<td>41</td>
<td>Teruel</td>
<td>10.9</td>
<td>10.0</td>
</tr>
<tr>
<td>42</td>
<td>Toledo</td>
<td>40.8</td>
<td>49.1</td>
</tr>
<tr>
<td>43</td>
<td>Valencia</td>
<td>175.4</td>
<td>154.1</td>
</tr>
<tr>
<td>44</td>
<td>Valladolid</td>
<td>42.2</td>
<td>56.8</td>
</tr>
<tr>
<td>45</td>
<td>Vitoria</td>
<td>27.1</td>
<td>33.2</td>
</tr>
<tr>
<td>46</td>
<td>Zamora</td>
<td>11.9</td>
<td>14.3</td>
</tr>
<tr>
<td>47</td>
<td>Zaragoza</td>
<td>74.2</td>
<td>72.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sectoral Effect</th>
<th>Competitive Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>2997.6</td>
</tr>
<tr>
<td>Filtered</td>
<td>3291.9</td>
</tr>
</tbody>
</table>

With regard to the sectoral and competitive effects, the analysis is more complex, since changes can be found both in the amount of the effects and in their signs. In fact, the observed variations are caused by two different factors: the considered variable (original or filtered employment) and the new filtered rates of growth. Sectoral effects are compared in table 3:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>-274.4</td>
<td>-370.6</td>
</tr>
<tr>
<td>Filtered</td>
<td>-284.3</td>
<td>-468.1</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.96</td>
<td>0.79</td>
</tr>
</tbody>
</table>
These results show positive interactions in most of the sectors (agriculture, industry and construction), services being the only activity with no significant positive spatial contribution and thus leading to a reduction of 26% in the sectoral employment.

As we have previously explained, the second model, with the consideration of a spatial competitive effect separated from the filtered competitive effect, shows the advantage of being strictly comparable to the traditional competitive effect. A graphical representation is shown in figure1.

Regarding the third model we must point out that the filtered net competitive effect (FNCE) reflects the variation in employment due to the advantages (disadvantages) of each sector in each different region when assuming a sectoral structure similar to the national one (homothetic) once the spillover effects have been discounted. On the other side, spatial locational effect (SLE) measures the deviation with respect to the previous hypothesis due to spatial effects and the mobility of labour market in response to comparative advantages. We compare the FNCE with the net competitive effect (NCE, Esteban-Marquillas, 1972) where the spatial are not considered and the spatial net competitive effect (SNCE**, Mayor and López, 2005) is based on the homothetic spatial employment. Table 4 summarizes the results of these effects by sectors showing a positive spatial influence in the sectoral employment except in the case of the services This conclusion is coincident with the SNCE** with a binary specification with the exception of the industrial employment.

<table>
<thead>
<tr>
<th>Table 4: Comparison of the net competitive effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCE (E-M, 1972)</td>
</tr>
<tr>
<td>SNCE**/NCE(E-M, 1972)</td>
</tr>
<tr>
<td>SNCE**_Boarnet</td>
</tr>
<tr>
<td>SNCE**/NCE(E-M, 1972)</td>
</tr>
<tr>
<td>FNCE/ ECEN (E-M, 1972)</td>
</tr>
</tbody>
</table>

It must be pointed out that the new effects related to model 3 and those associated to the approach by Mayor and López (2005) have different interpretations, since the first one refers to a local spatial dependence while the second one responds to a more general perspective. The spatial locational effect (ELE) shows changes in its value for each combination sector-region but with a certain stability as it can be observed in figure 2.
Figure 1: Decomposition of the competitive effect into filtered competitive effect (FCE) and spatial competitive effect (SCE)
Figure 2: Decomposition of the competitive effect between filtered net competitive effect (FNCE) and the spatial locational effect (SLE)
5. Concluding remarks

In this paper we have analyzed the influence of the spatial effects in the evolution of the regional employment, with the aim of improving the explanation of the existing differences. The proposed method considers each sector separately, thus allowing changes in the sectoral structure and also between the initial and final periods. From the conceptual point of view, this approach assumes that the considered value is the result of the spatial and non-spatial relations.

The advantage of the proposed models is the possibility of measuring the spatial spillovers for each region in terms of employment. Time series of these new effects could be obtained and modelled by means of the dynamic shift-share analysis in order to obtain the corresponding future values.

One of the main problems of the filtering processes consists on the election of the optimal distance in order to obtain the filter, since the obtained results are sensitive to the different considered screens. The underlying idea that the intensity relation is reduced with the distance is not always true and, therefore, the definition of new spatial weights based not only on the distance could be a suitable solution.

References:


