Path Based Shift-Share Analysis: Using Additional Information in Decomposing Regional Economic Changes

Esteban Fernández Vázquez[1], Bart Los[2] and Carmen Ramos Carvajal[1]

[1]University of Oviedo, Department of Applied Economics, Faculty of Economics, Campus del Cristo, Oviedo, 33006, Spain. e-mail: evazquez@uniovi.es; cramos@uniovi.es

[2]University of Groningen, Groningen Growth and Development Center and SOM Research School, P.O. Box 800, NL-9700 AV Groningen, The Netherlands. e-mail: b.los@eco.rug.nl

ABSTRACT:

Shift-Share analysis is a well-known methodology frequently used to obtain insights into the determinants of regional growth processes. It can address many issues, such as output growth, employment growth and productivity growth. After the initial equation proposed by Dunn (1960), several extensions have been suggested in order to overcome some conceptual problems. One of the most important undesirable properties that have been mentioned is the so-called "non-uniqueness" of the results. That is, numerous decomposition forms are equivalent to the classical shift-share equation from a theoretical point of view, but the results often depend strongly on the choice of a specific one. In this paper, we propose a methodology based on maximum entropy econometrics to incorporate additional information to select the unique shift-share formula that fits this information best. We illustrate the method empirically by investigating the sources of change of employment growth in Spanish regions, 1986-2000.

Keywords: Shift-share analysis, maximum entropy econometrics, regional employment, Spain.

Acknowledgements: This paper is based on Chapter 4 of the first author's Ph.D. thesis (Fernández, 2004). The authors would like to thank Erik Dietzenbacher for his comments on parts of the thesis that are relevant for this paper.

1. Introduction

The classical Shift-Share technique has been traditionally applied in regional science to explain the influence of different factors on the observed change in a variable. Basically, the main idea in this kind of analysis is that the temporal variations in a variable z_{ij} (where *i* refers to the economic sector and *j* to the region) depend on three factors or effects: a National Effect measuring the influence of the national economic growth process, a Sectoral Effect reflecting the effect of differences between regions in the industry mix and, finally a Regional or Competitive Effect measuring the regional differences in the dynamics of sector *i*. This analysis can provide useful information to policy makers: for the design of policies for a region it could be interesting to know, for instance, what is the influence of its specific sectoral specialization on the economic growth[1].

The three effects commented are summarized in the following equation:

$$\Delta z_{ij} = NE_{ij} + SE_{ij} + RE_{ij} \qquad (1)$$

where $NE_{ij} = r z_{ij}^0$, $SE_{ij} = (r_i - r) z_{ij}^0$ and $RE_{ij} = (r_{ij} - r_i) z_{ij}^0$; this is the formulation of the so-called *Classical Shift –Share Equation.* In it, r is the growth rate of variable z in the whole nation, r_i is the growth rate of this variable for the national sector *i* and r_{ij} is the growth rate of z_{ij}, considering two time periods 0 (initial) and 1 (final). Therefore:

$$r = \frac{z^1}{z^0} - 1 \qquad (2)$$

$$r_i = \frac{z_i^1}{z_i^0} - 1 \qquad (3)$$

$$r_{ij} = \frac{z_{ij}^1}{z_{ij}^0} - 1 \qquad (4)$$

As it is well known, the *National Effect* (*NE*) measures the hypothetical growth that the variable z_{ij} would have had if its growth rate had been the same as the national average; the *Sectoral Effect* (*SE*) measures the differential growth rate of national sector *i* compared to

[1] As an example of this interest, the Foundation for the Research depending on Caixa Galicia (a Spanish bank, http://www.fundacioncaixagalicia.org/) and the European Fund for Regional Development co-supported the study *Sectoral Structure and Regional Convergence* (De la Fuente, 2000), which used Shift –Share techniques to identify the effects commented here on the income per capita of Spanish regions.

the national aggregate growth rate; and, finally, the *Regional Effect* (*RE*) measures the influence of the specific economic characteristics of region *j* comparing the growth rate of sector *i* in this region with the national growth of this sector.

The bases of Shift-Share were already established in the sixties by Dunn (see Dunn, 1960). This technique has been applied for decades to decompose temporal changes in the regional levels of sectoral output or employment. However, its use has been the cause of some debate: works that criticized the Shift-Share techniques are Richardson (1978) or Holden *et al.* (1989); a strong defence can be found in Fothergill & Gudgin (1979). During these years several modifications and extensions have been proposed. Among the most important advances, are the works of Esteban-Marquillas (1972), Arcelus (1984), Berzeg (1978, 1984) or Barff & Knight (1988).

Traditionally, one of the most criticised features of the Classical Shift-Share equation has been the existence of some asymmetries in the analysis that lead to a non-unique solution. On of them is related to the choice of reference periods to weight the effects (Klaassen & Paelinck, 1972 or Barff & Knight, 1988). The main objective of this paper is the suggestion of a different perspective. We argue that any available additional information for periods inbetween the initial and final time period considered can be used to divide the interaction terms in a way that fits the data better than implied by simply taking averages. The additional data are used in a Maximum Entropy (ME) estimation procedure to arrive at parameter estimates that together specify a unique division of the interaction terms.[2]

The paper is organized as follows. In Section 2, we present the "non-uniqueness" problem in Shift-Share analysis in formal terms. Section 3 introduces a decomposition "path-based" method which continuous temporal paths for the factors involved in decomposition problems and obtains a type of solutions that depend on unknown parameters. In Section 4, the principles of ME estimation are highlighted, and we show how ME estimation techniques can be used to estimate the parameter of interest in solving the "non-uniqueness" problem in Shift-Share. Section 5 is devoted to a discussion of additional information that can be used to implement the ME approach. In Section 6 we present an empirical illustration of the approach. We will study the sources of change of employment

[2] Maximum Entropy econometrics and strongly related Cross Entropy methods have been used in an intersectoral setting before. See, for example, Golan *et al.* (1994) and Robinson *et al.* (2001) for methods to estimate missing data in input-output tables and social accounting matrices.

growth in Spanish regions between 1986 and 2000. Our aim is to assess the importance of changes in the national economy as a whole on the one hand, the effects of specific sectoral specialization and the consequences of some regional advantages and disadvantages. Section 7 concludes the paper.

2. The Non-Uniqueness Problem

As we have commented in the introduction to this section, Shift-Share techniques measure the different effects of the change in a variable z_{ij} between two time periods, considering three effects: national, sectoral and regional, as expressions (2)-(4) show. In this section we will illustrate how this result is closely related to the decomposition of temporal change in a variable that can be expressed as a product of several factors.

So, considering variable z_{ij} as:

$$z_{ij} = xy_i w_{ij} \tag{5}$$

where x is the national value of the variable (the aggregate of all sectors and regions), *i. e.*, $x = z$. On the other hand, y_i is a ratio that measures the influence of the industry mix by the ratio $y_i = \frac{z_i}{z}$ that measures the weight of sector *i* over the national value. Finally, *w* shows the ratio of region *j* over the overall employment in sector *i* ($w_{ij} = \frac{z_{ij}}{z_i}$).

If we measure z_{ij} in the initial and final time periods, 0 and 1 respectively, the variation between them would be:

$$\Delta z_{ij} = z_{ij}^1 - z_{ij}^0 = x^1 y_i^1 w_{ij}^1 - x^0 y_i^0 w_{ij}^0 \tag{6}$$

Adding and subtracting $x^1 y_i^0 w_{ij}^0$ and $x^1 y_i^1 w_{ij}^0$ in (6) we obtain:

$$\Delta z_{ij} = \Delta x y_i^0 w_{ij}^0 + x^1 \Delta y_i w_{ij}^0 + x^1 y_i^1 \Delta w_{ij} \tag{7}$$

Each one of the terms of equation (7) shows the effects defined in (1):

$$NE_{ij} = \Delta x y_i^0 w_{ij}^0 = x^1 y_i^0 w_{ij}^0 - z_{ij}^0 = \left(\frac{x^1 y_i^0 w_{ij}^0}{z_{ij}^0} - 1 \right) z_{ij}^0 = \left(\frac{x^1 y_i^0 w_{ij}^0}{x^0 y_i^0 w_{ij}^0} - 1 \right) z_{ij}^0 = \left(\frac{x^1}{x^0} - 1 \right) z_{ij}^0 = r z_{ij}^0 \qquad (8)$$

$$SE_{ij} = x^1 \Delta y_i w_{ij}^0 = x^1 y_i^1 w_{ij}^0 - x^1 y_i^0 w_{ij}^0 = x^1 y_i^1 w_{ij}^0 - \frac{x^1}{x^0} x^0 y_i^0 w_{ij}^0 = \left(\frac{x^1 y_i^1 w_{ij}^0}{x^0 y_i^0 w_{ij}^0} - \frac{x^1}{x^0} \right) z_{ij}^0 =$$
$$= \left(\frac{x^1 y_i^1}{x^0 y_i^0} - \frac{x^1}{x^0} \right) z_{ij}^0 = \left(\frac{z_i^1}{z_i^0} - \frac{z^1}{z^0} \right) z_{ij}^0 = (r_i - r) z_{ij}^0 \qquad (9)$$

$$RE_{ij} = x^1 y_i^1 \Delta w_{ij} = x^1 y_i^1 w_{ij}^1 - x^1 y_i^1 w_{ij}^0 = z_{ij}^1 - \frac{x^1 y_i^1}{x^0 y_i^0} x^0 y^0 w_{ij}^0 = \left(\frac{z_{ij}^1}{z_{ij}^0} - \frac{x^1 y_i^1}{x^0 y_i^0} \right) z_{ij}^0 =$$
$$= \left(\frac{z_{ij}^1}{z_{ij}^0} - \frac{z_i^1}{z_i^0} \right) z_{ij}^0 = (r_{ij} - r_i) z_{ij}^0 \qquad (10)$$

In other words, expression (7) actually is the classical Shift-Share equation. Note that the starting point of this section was a decomposition problem for a variable that can be defined by the product of three factors. The problem is that if in (6) we had made different mathematical transformations, the whole change in z_{ij} would had been expressed as the following sum:

$$\Delta z_{ij} = \Delta x y_i^1 w_{ij}^1 + x^0 \Delta y_i w_{ij}^1 + x^0 y_i^0 \Delta w_{ij} \qquad (11)$$

Consequently, the three effects can be measured by expressions different from the *classical* equation (7). Through this kind of transformation it would be possible to achieve six decomposition forms of the temporal change in z_{ij}. The expressions of all the decomposition forms are:

$\Delta x y_i^0 w_{ij}^0 + x^1 \Delta y_i w_{ij}^0 + x^1 y_i^1 \Delta w_{ij}$	Decomposition form 1 Equation (7)	(12a)
$\Delta x y_i^1 w_{ij}^1 + x^0 \Delta y_i w_{ij}^1 + x^0 y_i^0 \Delta w_{ij}$	Decomposition form 2 Equation (11)	(12b)
$\Delta x y_i^0 w_{ij}^0 + x^1 \Delta y_i w_{ij}^1 + x^1 y_i^0 \Delta w_{ij}$	Decomposition form 3	(12c)
$\Delta x y_i^0 w_{ij}^1 + x^1 \Delta y_i w_{ij}^1 + x^0 y_i^0 \Delta w_{ij}$	Decomposition form 4	(12d)
$\Delta x y_i^1 w_{ij}^1 + x^0 \Delta y_i w_{ij}^0 + x^0 y_i^1 \Delta w_{ij}$	Decomposition form 5	(12e)

$$\Delta x y_i^1 w_{ij}^0 + x^0 \Delta y_i w_{ij}^0 + x^1 y_i^1 \Delta w_{ij} \qquad \text{Decomposition form 6} \qquad (12f)$$

In general terms, in decomposition problems where variables are the product of n determinants, the number of possible decompositions is $n!$. All them are admissible as exhaustive (the sum of the effects equals the change in the variable) and taking one or another is a purely arbitrary choice that implies some variability in the results and the analysis conclusions could differ extensively.

3. The Path Based Approach

In this section, a framework for an alternative decomposition method will be sketched. It builds on earlier work by Hoekstra & Van den Bergh (2002) and in particular Harrison *et al.* (2000), who introduced the basics of what we will call the Path Based (PB) approach. The alternative setup starts from the premise that both the value of z_{ij} and the value of the determinants x, y_i, and w_{ij} have changed continuously over time, between time 0 and time 1. Hence, we can write:

$$z_{ij}(t) = x(t)\, y_i(t)\, w_{ij}(t) \qquad (13)$$

and, assuming differentiability of each factor an infinitesimal change in z can be expressed as

$$dz_{ij} = \frac{\partial z_{ij}}{\partial x}\frac{dx}{dt}dt + \frac{\partial z_{ij}}{\partial y_i}\frac{dy_i}{dt}dt + \frac{\partial z}{\partial w_{ij}}\frac{dw_{ij}}{dt}dt \qquad (14)$$

Finally, the total change in z_{ij} can be expressed as the sum of all the infinitesimal changes between time 0 and time 1:

$$\Delta z_{ij} = \int_{t=0}^{t=1} \frac{dx}{dt} y_i(t) w_{ij}(t) dt + \int_{t=0}^{t=1} x(t) \frac{dy}{dt} w_{ij}(t) dt + \int_{t=0}^{t=1} x y_i(t) \frac{dw_{ij}}{dt} dt \qquad (15)$$

Equation (14) shows that the derivatives of the determinants x, y_i, and w_{ij} to time t play an important role in the size of the effects attributed to changes in these determinants. Consequently, the choice of the functional forms of the functions, or in other words, the specification of the temporal paths that variables follow between initial and final periods, can have a big impact on the measurement of their effects that together add up to the variation in z_{ij}.

Harrison *et al.* (2000) proposed the solution arrived at by assuming straight-line paths of the factors:

$$x(t) = x^0 + \left(x^1 - x^0\right)t = x^0 + \Delta x t \tag{16a}$$

$$y_i(t) = y_i^0 + \left(y_i^1 - y_i^0\right)t = y_i^0 + \Delta y_i t \tag{16b}$$

$$w_{ij}(t) = w_{ij}^0 + \left(w_{ij}^1 - w_{ij}^0\right)t = w_{ij}^0 + \Delta w_{ij} t \tag{16c}$$

From this linear paths and equation (15), we will obtain the following effects:

$$\Delta x \;\; Effect = NE_{ij}^{\mu} = \Delta x y_i^0 w_{ij}^0 + \frac{1}{2}\Delta x \Delta y_i w_{ij}^0 + \frac{1}{2}\Delta x y_i^0 \Delta w_{ij} + \frac{1}{3}\Delta x \Delta y_i \Delta w_{ij} \tag{17}$$

$$\Delta y_i \;\; Effect = SE_{ij}^{\mu} = x^0 \Delta y_i w_{ij}^0 + \frac{1}{2}\Delta x \Delta y_i w_{ij}^0 + \frac{1}{2} x^0 \Delta y_i \Delta w_{ij} + \frac{1}{3}\Delta x \Delta y_i \Delta w_{ij} \tag{18}$$

$$\Delta w_{ij} \;\; Effect = RE_{ij}^{\mu} = x^0 y_i^0 \Delta w_{ij} + \frac{1}{2}\Delta x y_i^0 \Delta w_{ij} + \frac{1}{2} x^0 \Delta y_i \Delta w_{ij} + \frac{1}{3}\Delta x \Delta y_i \Delta w_{ij} \tag{19}$$

Actually, this approach yields the same solution as Sun's (1998) 'equal shares' method. Furthermore, this solution equals the average of the six decomposition forms of equations (12a)-(12f). In this paper, we propose a method to take such information explicitly into account in attributing parts of the interaction effects to the effects of the respective determinants.

The methodological innovation we propose is to relax the strict assumption of a straight line, by considering more flexible forms for the functions that describe the temporal behavior of factors x, y_i, and w_{ij}. In order to preserve possibilities to estimate the parameters that characterize the time-paths of the variables, we choose to consider a specific class of monotonic functions without inflexion points:

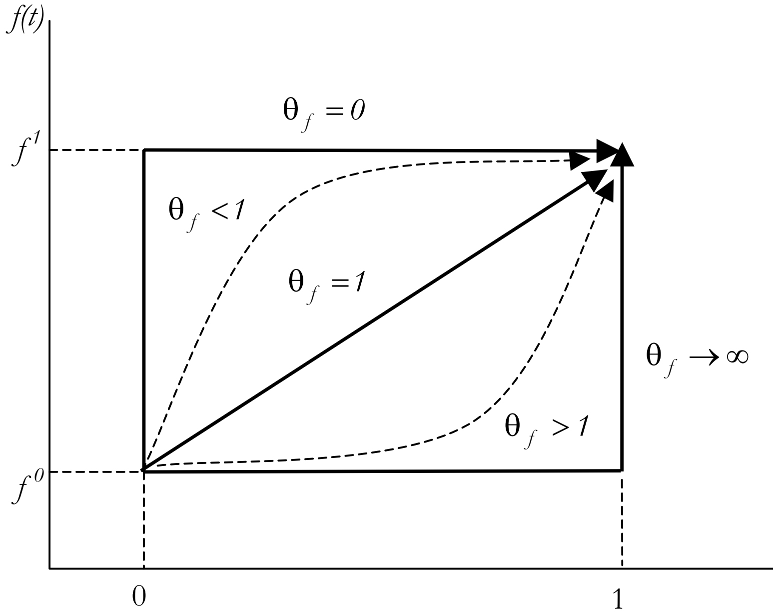
$$x(t) = x^0 + \Delta x t^{\theta_x}; \;\; \forall\, \theta_x > 0 \tag{20a}$$

$$y_i(t) = y_i^0 + \Delta y_i t^{\theta_{y_i}}; \;\; \forall\, \theta_{y_i} > 0 \tag{20b}$$

$$w_{ij}(t) = w_{ij}^0 + \Delta w_{ij} t^{\theta_{w_{ij}}}; \;\; \forall\, \theta_{w_{ij}} > 0 \tag{20c}$$

Obviously, the temporal path of a factor will be a straight line if the parameter θ equals 1, and the solution obtained by the method introduced here will be identical to Harrison's et al. (2000) solution. Figure1 indicates what the path for a generic factor *f* (which may refer to x, y_i, or w_{ij}) looks like if θ_f takes on a value different from 1.

Figure 1. Several temporal paths for factor f



As can be seen from Figure 1, the class of paths considered contains all possible monotonic paths for f^0 to f^1 that do not have inflexion points. This is a limitation for sure. An important category of paths not covered by our class of paths are those that contain values that are below the initial value or exceed the final value (assuming, without loss of generalization that f^1 is larger than f^0). Considering this type of temporal paths and taking into account expression (15), Δz_{ij} can be written as the sum of the following equations:

$$\Delta x \;\; Effect = NE^{*}_{ij} = \Delta x y_i^0 w_{ij}^0 + \frac{\theta_x}{\theta_x + \theta_{y_i}} \Delta x \Delta y_i w_{ij}^0 + \frac{\theta_x}{\theta_x + \theta_{w_{ij}}} \Delta x y_i^0 \Delta w_{ij} + \frac{\theta_x}{\theta_x + \theta_{y_i} + \theta_{w_{ij}}} \Delta x \Delta y_i \Delta w_{ij} \quad (21)$$

$$\Delta y_i \;\; Effect = SE^{*}_{ij} = x^0 \Delta y_i w_{ij}^0 + \frac{\theta_{y_i}}{\theta_x + \theta_{y_i}} \Delta x \Delta y_i w_{ij}^0 + \frac{\theta_{y_i}}{\theta_{y_i} + \theta_{w_{ij}}} x^0 \Delta y_i \Delta w_{ij} + \frac{\theta_{y_i}}{\theta_x + \theta_{y_i} + \theta_{w_{ij}}} \Delta x \Delta y_i \Delta w_{ij} \quad (22)$$

$$\Delta w_{ij} \;\; Effect = RE^{*}_{ij} = x^0 y_i^0 \Delta w_{ij} + \frac{\theta_{w_{ij}}}{\theta_x + \theta_{w_{ij}}} \Delta x y_i^0 \Delta w_{ij} + \frac{\theta_{w_{ij}}}{\theta_{y_i} + \theta_{w_{ij}}} x^0 \Delta y_i \Delta w_{ij} + \frac{\theta_{w_{ij}}}{\theta_x + \theta_{y_i} + \theta_{w_{ij}}} \Delta x \Delta y_i \Delta w_{ij} \quad (23)$$

Where NE^*_{ij}, SE^*_{ij} and RE^*_{ij} show respectively the National, Sectoral and Regional Effects obtained by the use of PBM technique for the sector *i* in the region *j*. The interpretation of these results is quite intuitive. Starting with the national effect, equation (4.17) shows that:

$$NE^*_{ij} = xy^0_i w^0_{ij} + \frac{\theta_x}{\theta_x + \theta_{y_i}} \Delta x \Delta y_i w^0_{ij} + \frac{\theta_x}{\theta_x + \theta_{w_{ij}}} \Delta x y^0_i \Delta w_{ij} + \frac{\theta_x}{\theta_x + \theta_{y_i} + \theta_{w_{ij}}} \Delta x \Delta y_i \Delta w_{ij} \qquad (21)$$

whereas, as appears in (8), the solution yielded by classical Shift-Share equation was:

$$NE_{ij} = \Delta x y^0_i w^0_{ij} \qquad (8)$$

We can see how the path based decomposition allows a share of the joint terms to be allocated to variable *x*. The respective portion of these interactions assigned to *x* is determined by the size of θ_x in relative terms to $\theta_{y_i}, \theta_{w_{ij}}$. Then, is easy to see that classical Shift-Share equation assigns a much lower to parameter θ_x value than the rest of the parameters. On the other hand, the Regional Effect expression obtained by the method proposed is:

$$RE^*_{ij} = x^0 y^0_i \Delta w_{ij} + \frac{\theta_{w_{ij}}}{\theta_x + \theta_{w_{ij}}} \Delta x y^0_i \Delta w_{ij} + \frac{\theta_{w_{ij}}}{\theta_{y_i} + \theta_{w_{ij}}} x^0 \Delta y_i \Delta w_{ij} + \frac{\theta_{w_{ij}}}{\theta_x + \theta_{y_i} + \theta_{w_{ij}}} \Delta x \Delta y_i \Delta w_{ij} \qquad (23)$$

whereas the classical Shift-Share solution that appears in expression (10) was:

$$RE_{ij} = x^1 y^1_i \Delta w_{ij} \qquad (10)$$

In other words, this technique gives an infinitely large value to $\theta_{w_{ij}}$ relative to the other parameters. Finally, regarding Sectoral Effect, with classical Shift-Share equation we obtain:

$$SE_{ij} = x^1 \Delta y_i w^0_{ij} \qquad (9)$$

that is the result of equation (22) when parameter θ_x takes a value close to zero and simultaneously $\theta_{w_{ij}}$ is infinitely bigger than the rest. Taking all this into account, we can conclude that in a context where the three factors of z_{ij} (the national value of the variable, the weight of sector *i* over the national aggregate and the weight of region *j* over sector *i*) are increasing[3], equation (7) gives to the national effect (NE_{ij}) its minimum value and, at the same time, assigns to the Regional Effect (RE_{ij}) its maximum value. Unless the true values of the unknown parameters coincide exactly with the values supposed by the classical Shift-Share equation (in other words, unless the arbitrary determination of temporal paths was the same as the true paths), the obtained effects by (7) will have underestimated the national effect and overestimated the Regional Effect.

An important fact that must be noted is that classical Shift-Share solutions match particular cases of functional forms (21)-(23). In other words, classical Shift-Share technique is not out of the solutions achieved by the path based technique, and the expressions of the effects are just results of a specific value assignation to parameters $\theta_x, \theta_{y_i}, \theta_{w_{ij}}$. The contributions of the factors to the change in z_{ij} will be determined by the size of these parameters. Therefore, the decomposition problem can be viewed as a problem of value assignation to unknown parameters; if we only have data about the factors in the initial and final period the solution would be the application of the indifference principle and the assumption that equality $\theta_x = \theta_{y_i} = \theta_{w_{ij}}$ holds. This solution equals the computation of the average for the six decomposition forms (12a-12f).

Other important point is that any positive finite values for parameters $\theta_x, \theta_{y_i}, \theta_{w_{ij}}$ provide an exhaustive decomposition where the sum of the effects equals the total change in z_{ij}. When Shift-Share techniques are applied usually the sum from all the regions of the Sectoral and Regional Effects are zero by definition. This happens because both effects are defined as differences to an average rate: in the case of the Sectoral Effect is a differential effect with respect to the whole national rate, the Regional Effect is based in the difference of a regional industry to the growth rate of the national sector. If one wants to obtain

[3] A context where $\Delta x, \Delta y_i, \Delta w_{ij} \geq 0$.

decomposition where the sum of the Regional Effects for every sector *i* equals zero, i. e. $\sum_j RE^*_{ij} = 0; \forall i$, from equation (23) can be derived that the condition is:

$$\theta_{w_{ij}} = c; \forall i, j \tag{24}$$

Moreover, to obtain decompositions where $\sum_j \sum_i SE^*_{ij} = \sum_j SE^*_j = 0$, from equation (22) and taking into account (24) can be easily obtained the following condition:

$$\sum_i \frac{\theta_{y_i}}{\theta_x + \theta_{y_i}} \Delta y_i = 0 \tag{25}$$

4. Maximum Entropy Econometrics

In the previous section, we found that taking the mean contributions of all decomposition forms is the most reasonable solution to the non-uniqueness problem if the researcher has no information at all about the time paths of the determinants. In many cases, however, more information than the values of the determinants at t=0 and t=1 is available, for example about values of one or more of the determinants at intermediate points in time. Estimation of the parameters θ_i is generally not possible by means of classical econometric estimation procedures like least squares estimation. The amount of data is quite limited, which precludes the use of least squares estimation procedures based on limit theorems. Such procedures require at least more observations than parameters to be estimated, which is problematic in the input-output context studied here.

In this section, we will give an introduction to maximum entropy (ME) econometrics, a collection of tools that can be very convenient to use scarce additional information in producing estimates for the temporal path parameters θ.[4] To start with, let us assume that an event can have K possible outcomes E_1, E_2, ..., E_K with the respective distribution of probabilities $\mathbf{p} = p_1, p_2, ..., p_K$ such that $\sum_{i=1}^{K} p_i = 1$. Following the formulation of Shannon (1948), the entropy of this distribution $\mathbf{p}$ will be

4 See Kapur & Kesavan (1992) or Golan *et al.* (1996) for a detailed analysis of properties of the estimators obtained by means of these techniques.

$$H(\mathbf{p}) = -\sum_{i=1}^{K} p_i \ln p_i \tag{26}$$

which reaches its maximum when **p** is a uniform distribution ($p_i = \frac{1}{K}, \forall \text{i} = 1,...,\text{K}$). The entropy measure H indicates the 'uncertainty' of the outcomes of the event. If some information (*i.e.*, observations) is available, it can be used to estimate an unknown distribution of probabilities for a random variable x which can get values $\{x_1,...,x_K\}$.

Suppose that there are T observations $\{y_1, y_2,..., y_T\}$ available such that

$$\sum_{i=1}^{K} p_i f_t(x_i) = y_t,\ 1 \leq t \leq T \tag{27}$$

with $\{f_1(x), f_2(x),..., f_T(x)\}$ a set of known functions representing the relationships between the random variable x and the observed data $\{y_1, y_2,..., y_T\}$. In such a case, the ME principle can be applied to recover the unknown probabilities. This principle is based on the selection of the probability distribution that maximizes equation (26) among all the possible probability distributions that fulfill (27). The following constrained maximization problem is posed:

$$\underset{\mathbf{p}}{Max}\ H(\mathbf{p}) = -\sum_{i=1}^{K} p_i\, ln\, p_i \tag{28}$$

subject to:

$$\sum_{i=1}^{K} p_i f_t(x_i) = y_t,\ \forall t = 1,...,T$$

$$\sum_{i=1}^{K} p_i = 1$$

In this problem, the last restriction is just a normalization constraint that guarantees that the estimated probabilities sum to one, while the first T restrictions guarantee that the recovered distribution of probabilities is compatible with the data for all T observations. The Lagrangian function for problem (28) is

$$L = -\sum_{i=1}^{K} p_i \ln p_i + \sum_{t=1}^{T} \lambda_t \left[y_t - \sum_{i=1}^{K} p_i f_t(x_i) \right] + \mu \left[1 - \sum_{i=1}^{K} p_i \right] \tag{29}$$

and the corresponding estimates for the probabilities p_i are

$$\hat{p}_i = \frac{exp\left[-\sum_{t=1}^{T} \hat{\lambda}_t f_t(x_i) \right]}{\sum_{i=1}^{K} exp\left[-\sum_{t=1}^{T} \hat{\lambda}_t f_t(x_i) \right]}, \quad \forall i = 1,...,K \tag{30}$$

with $\hat{\lambda}_t$ the Lagrangian multipliers associated to the first T restrictions in the constrained maximization problem (28). It is important to note that even for T=1 (a situation with only one observation), the ME approach yields an estimate of the probabilities. Hence, in situations in which the number of observations is not large enough to apply econometrics based on limit theorems, this approach can be used to obtain robust estimates of unknown parameters.[5] A disadvantage of ME estimators is that comparisons of means and variances of estimators are not possible. Such comparisons are common practice in classical least squares and maximum likelihood econometrics.

For our current purposes, it is important that the above-sketched procedure can be generalized and extended to the estimation of unknown parameters for traditional linear models. Let us suppose that the problem at hand is the estimation of a linear model where a variable y depends on n explanatory variables x_i:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e} \tag{31}$$

where $\mathbf{y}$ is a $(T \times 1)$ vector of observations for y, $\mathbf{X}$ is a $(T \times n)$ matrix of observations for the x_i variables, $\boldsymbol{\theta}$ is the $(n \times 1)$ vector of unknown parameters $\boldsymbol{\theta}' = (\theta_1,...,\theta_n)$ to be estimated, and $\mathbf{e}$ is a $(T \times 1)$ vector reflecting the random term of the linear model. For each θ_i, it will be assumed that there is some information about its $M \geq 2$ possible realizations by means of a 'support' vector $\mathbf{b}' = (b_1,...,b^*,...,b_M)$, the elements of which are symmetrically distanced around a central value $\theta_i = b^*$ (the prior expected value of the

[5] Golan *et al.* (1996, p. 12) contains a simple, classic example of this technique, the so called "dice problem".

parameter), with corresponding probabilities $\mathbf{p}'_\mathbf{i} = (p_{i1},..., p_{iM})$. For the sake of convenient exposition, it will be assumed that the M values are the same for every parameter, although this assumption can easily be relaxed. Now, vector $\boldsymbol{\theta}$ can be written as

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \cdots \\ \theta_n \end{bmatrix} = \mathbf{Bp} = \begin{bmatrix} \mathbf{b}' & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \mathbf{b}' & . & \mathbf{0} \\ . & . & . & . \\ \mathbf{0} & \mathbf{0} & . & \mathbf{b}' \end{bmatrix} \begin{bmatrix} \mathbf{p_1} \\ \mathbf{p_2} \\ \cdots \\ \mathbf{p}_n \end{bmatrix} \tag{32}$$

with $\mathbf{B}$ and $\mathbf{p}$ of dimensions (nxnM) and (nMx1), respectively. The value for each parameter is then given by

$$\theta_i = \mathbf{b}'\mathbf{p_i} = \sum_{m=1}^{M} b_m p_{im} \;;\; \forall i = 1,...,n \tag{33}$$

For the random terms, a similar approach is chosen. To express the lack of information about the actual values contained in **e**, we assume a distribution for each e_t, with a set of $K \geq 2$ values $\mathbf{v}' = (v_1,...,v_K)$ with respective probabilities $\mathbf{q}'_t = (q_{t1}, q_{t2},..., q_{tK})$.[6] Hence, we can write

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \cdots \\ e_T \end{bmatrix} = \mathbf{Vq} = \begin{bmatrix} \mathbf{v}' & \mathbf{0} & . & \mathbf{0} \\ \mathbf{0} & \mathbf{v}' & . & \mathbf{0} \\ . & . & . & . \\ \mathbf{0} & \mathbf{0} & . & \mathbf{v}' \end{bmatrix} \begin{bmatrix} \mathbf{q_1} \\ \mathbf{q_2} \\ \cdots \\ \mathbf{q_T} \end{bmatrix} \tag{34}$$

and the value of the random term for an observation t equals

$$e_t = \mathbf{v}'\mathbf{q_t} = \sum_{k=1}^{K} v_j q_{tk} \;;\; \forall t = 1,..., T \tag{35}$$

And, consequently, equation (31) can be transformed into

$$\mathbf{y} = \mathbf{XBp} + \mathbf{Vq} \tag{36}$$

---

6 Usually, the distribution for the errors is assumed symmetric and centered about 0, therefore $v_1 = -v_K$.

Now, the estimation problem for the unknown vector of parameters $\boldsymbol{\theta}$ is reduced to the estimation of $n+T$ probability distributions, and the following maximization problem (similar to problem (28)) can be solved to obtain these estimates:

$$\underset{\mathbf{p},\mathbf{q}}{Max}\, H(\mathbf{p},\mathbf{q}) = -\sum_{i=1}^{n}\sum_{m=1}^{M} p_{im}\, ln(p_{im}) - \sum_{t=1}^{T}\sum_{k=1}^{k} q_{tk}\, ln(q_{tk}) \qquad (37)$$

subject to:

$$\sum_{i=1}^{n}\sum_{m=1}^{M} x_{it} b_m p_{im} + \sum_{k=1}^{K} v_j q_{tk} = y_t,\ \forall t = 1,...,T$$

$$\sum_{m=1}^{M} p_{im} = 1,\ \forall i = 1,...,n$$

$$\sum_{k=1}^{K} q_{tk} = 1,\ \forall t = 1,...,T$$

By solving the associated Lagrangian function, we find

$$\hat{p}_{im} = \frac{\exp\left[-\sum_{t=1}^{T}\hat{\lambda}_t x_{it} b_m\right]}{\sum_{m=1}^{M}\exp\left[-\sum_{t=1}^{T}\hat{\lambda}_t x_{it} b_m\right]},\ \forall i = 1,...,n;\ \forall m = 1,...,M \qquad (38)$$

$$\hat{q}_{tk} = \frac{exp\left[-\sum_{t=1}^{T}\hat{\lambda}_t v_k\right]}{\sum_{k=1}^{K} exp\left[-\sum_{t=1}^{T}\hat{\lambda}_t v_k\right]},\ \forall t = 1,...,T;\ \forall k = 1,...,K \qquad (39)$$

Finally, these estimated probabilities allow us to obtain estimations for the unknown parameters.[7] The estimated value of θ_i will be:[8,9]

---

7 Golan *et al.* (1996, Chapter 6) show that these estimators are consistent and asymptotically normal. In Golan *et al.* (1996, Chapter 7) the finite sample behavior of the ME estimators is numerically compared to traditional least squares and maximum likelihood estimators. In experimental samples with limited data, the ME estimators are found to be superior.

8 The construction of the vector **b** is based on the researcher's prior knowledge (or beliefs) about the parameter. Sometimes, the choice of minimum and maximum values b_1 and b_M is quite obvious, but in other cases a 'natural' choice does not exist. In such a situation, it will not be possible to obtain an accurate solution to the estimation problem if the actual parameter value is out of the fixed range, say $\theta_i > b_M$. Therefore, one should be careful in choosing the maximum and minimum values of **b**. Golan *et al.* (1996, chapter 8) devote more attention to consequences of choices concerning the elements of the

$$\hat{\theta}_i = \sum_{m=1}^{M} \hat{p}_{im} b_m \ , \ \forall i = 1,...,n \qquad (40)$$

This approach can be applied to the decomposition problem studied in the previous section, since limited additional information would enable us to obtain estimates of the parameters that determine the contribution of each determinant to the total change that has actually been observed. In other words, non-arbitrary solutions to the decomposition problem could be obtained. In the next section several situations with availability of various types of additional data will be considered, as well as the way to estimate the effects of the factors to the total change Δz_{ij} using this technique.

5. Incorporating Additional Information in Shift-Share Analysis

In the previous sections we have justified the crucial role of the parameters $\theta_x, \theta_{y_i}, \theta_{w_{ij}}$ for measuring the National, Sectoral and Regional Effects, so it would be desirable that their values were not assigned in an arbitrary way. One way to do this is computing the mean of the six decomposition forms (12a)-(12f). Nevertheless, taking an average solution is not the only way to obtain non-arbitrary values of the parameters. Assuming a context where some additional information between the initial and final periods is available (data for intermediate points), it would be possible to use this information to obtain their respective contributions in a not arbitrary way, since the computation of these effects can be viewed as an estimation problem for unknown parameters. Therefore, the initial decomposition problem can be approached as an estimation question.

In this section we will suppose a scenario in which we have some additional observations for intermediate periods. A "dynamic Shift-Share" in the more traditional sense is not possible, however, since we suppose that these intermediate observations are only available for some of the three factors x, y_i, and w_{ij} considered.[10] To assess the contribution of factor

---

vector **b**. An almost universal result is that wider bounds can be used without substantial consequences for the characteristics of the estimators.

9 Fernández (2004, pp. 69) proves that the solution of the constrained maximization problem (37) without additional information yields estimates equal to the expected value b^* of the prior distribution.

10 The "dynamic Shift-Share analysis" was proposed by Barff & Knight III (1988). If observations for all the factors were available for a period s ($0<s<1$), dynamic Shift-Share would amount to decomposing z_s-z_0 and z_1-z_s in the classic way outlined in Section 2, and subsequently adding results for the corresponding

x_i, equations (20a)-(20b) will be used again, but in a slightly different form. They contains a stochastic component ε_{it} that allows the factors to diverge from the deterministic path that we would like to estimate[11]

$$x(t) = x^0 + \Delta x t^{\theta_x} e^{\varepsilon_t};\ \forall\, \theta_x > 0 \tag{41a}$$

$$y_i(t) = y_i^0 + \Delta y_i t^{\theta_{yi}} e^{\varepsilon_{it}};\ \forall\, \theta_{y_i} > 0 \tag{41b}$$

$$w_{ij}(t) = w_{ij}^0 + \Delta w_{ij} t^{\theta_{wij}} e^{\varepsilon_{ijt}};\ \forall\, \theta_{w_{ij}} > 0 \tag{41c}$$

Defining $g(t) = x(t) - x^0$, $g_i(t) = y_i(t) - y_i^0$ and $g_{ij}(t) = w_{ij}(t) - w_{ij}^0$ and taking logarithms, we have:

$$ln\left(\frac{g(t)}{\Delta x}\right) = \theta_x\, ln(t) + \varepsilon_t, \text{ or}$$

(42a)

$$x^*(t) = \theta_x t^* + \varepsilon_t$$

$$ln\left(\frac{g_i(t)}{\Delta y_i}\right) = \theta_{y_i}\, ln(t) + \varepsilon_{it}, \text{ or}$$

(42b)

$$y_i^*(t) = \theta_{y_i} t^* + \varepsilon_{it}$$

$$ln\left(\frac{g_{ij}(t)}{\Delta w_{ij}}\right) = \theta_{w_{ij}}\, ln(t) + \varepsilon_{ijt}, \text{ or}$$

(42c)

$$w_{ij}^*(t) = \theta_{w_{ij}} t^* + \varepsilon_{ijt}$$

Equations (42a)-(42c) are linear models with one parameter to be estimated. Hence, it is possible to apply the Maximum Entropy estimation technique for linear relationships analyzed in the previous section, and they can be written as

---

effects in the two decompositions to obtain the contributions for z_1-z_0. A discussion of transitivity problems in this approach is beyond the scope of this paper.

11 We assume that $\varepsilon_{it} = 0$ in the final period. This ensures that the factor has its final value in the final period 1.

$$x^*(t) = \sum_{m=1}^{M} b_m p_m t^* + \sum_{k=1}^{K} v_k q_{tk} \tag{43a}$$

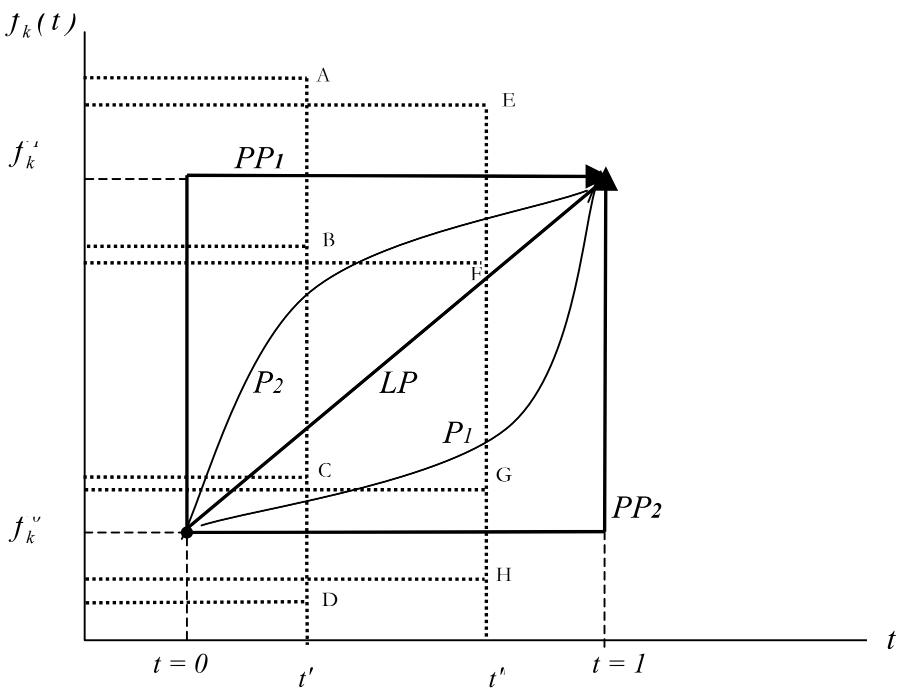
$$y_i^*(t) = \sum_{m=1}^{M} b_{im} p_{im} t^* + \sum_{k=1}^{K} v_{ik} q_{tik} \tag{43b}$$

$$w_{ij}^*(t) = \sum_{m=1}^{M} b_{ijm} p_{ijm} t^* + \sum_{k=1}^{K} v_{ijk} q_{tijk} \tag{43c}$$

which ends up as constraints in maximization problems like those depicted in (35). According to equation (38), solving these problems yield estimates for parameters $\theta_x, \theta_{y_i}, \theta_{w_{ij}}$. For those factors for which there is no additional information, the estimates should equal 1 to resemble the linear path. Hence, the central value b^* should be set to 1. Upon having obtained estimates for $\theta_x, \theta_{y_i}, \theta_{w_{ij}}$ substitution of these values and the observations for x, y_i, w_{ij} in equations (21)-(23) yields the estimated respective National, Sectoral and Regional Effects.

The use of additional information in the framework outlined above can cause nontrivial problems if the information is rather unlikely to be generated by a time path belonging to the class of paths defined by equations (41a)-(41c). This happens if observations for intermediate periods rule out a monotonic path. We deal with such observations by fitting the most appropriate monotonic paths. Figure 2 depicts all possibilities if two intermediate observations are available.

Figure 2. Estimated temporal paths with intermediate observations



For intermediate period *t'* observations for this determinant can be categorized as A, B, C or D, depending on whether they are above or below the linear path and inside or outside the rectangle. In the same vein, we have E, F, G or H for intermediate period *t''*. If the two observations are both like B, C, F and G, no problems are encountered. If A and E are observed, the closest monotonic path is PP_1, which corresponds to $\theta_f = 0$. If D and H are observed, PP_2 is most appropriate and $\theta_f = \infty$.[12] If A (above the rectangle) and H (below the rectangle) are observed, we opt for the linear path, since it is the average of PP_1 (implied by A) and PP_2 (implied by H). If points like B and E or B and H are observed, there will be an observation inside the rectangle (B) and another one in the outside (E or H). In such cases, to obtain valid estimates of the parameter it will be assumed that points E or H are not outside the rectangle but just on the border of the rectangle (given by PP_1 and PP_2 respectively). The same procedure is applied in situations with observations like A and F or D and F.

It should be noted that the important issue is that the flexibility of this estimation method allows including information even if there were not direct observations of the factors appearing in the decomposition problem. If there is some kind of knowledge about the behavior of other variables that are somehow related to these factors, this information can be used to obtain estimates of the parameters.

6. Illustration: Analysis of regional employment dynamics in Spain (1986-2000)

We apply the techniques developed in the previous sections to study the National, Sectoral and Regional Effects to changes in regional employment in Spain, over the period 1986-2000. It should be emphasized that the aim of this section is not so much to provide a "deep" analysis of the dynamics of Spanish regional employment, but rather to provide an illustration of the methods proposed in this paper. The required data were taken from the Spanish Regional Accounts published by the Spanish Statistical Institute (INE), considering a 15-sector classification and are detailed in tables A1, A2 and A3 in Appendix A.

Tables A2 and A3 show the values of the studied variable for every region and sector in the reference periods. Their rightmost columns show the overall regional levels of employment, and the bottom rows the sectoral levels. So, information about the three

---

12 If $\theta_f = \infty$, a "very big" value must be inserted in equations (21)-(23) to obtain numerical results. In the empirical application described in the next section, we used the value 10^{20} in such cases.

factors is available: the total national employment (factor x), the weight of sector i over this aggregate (factor y_i) and the weight of the variable in region j over the national value of sector i (factor w_{ij}). If the classical Shift-Share formulation is used to measure the effects of changes in these three factors over the variation observed in z_{ij}, the aggregate outcomes[13] for each of these effects will be the following, measured thousand of workers as well as a percentage over the total change in regional employment:

Table 1. Classical Shift-Share equation outcomes

		Absolute effect (thousand of workers)			% of total variation		
	TOTAL VARIATION	**National Effect**	**Sectoral Effect**	**Regional Effect**	**National Effect**	**Sectoral Effect**	**Regional Effect**
AND	825.30	649.28	-10.00	186.02	78.67	-1.21	22.54
ARAG	112.70	158.52	-13.95	-31.87	140.66	-12.38	-28.28
AST	12.20	148.01	-34.49	-101.32	1213.21	-282.73	-830.48
BAL	119.20	93.46	21.73	4.02	78.40	18.23	3.37
CAN	292.50	154.03	28.25	110.23	52.66	9.66	37.68
CANT	38.60	66.67	-7.97	-20.10	172.72	-20.64	-52.08
CAST-L	131.70	326.76	-65.35	-129.72	248.11	-49.62	-98.49
CAST-LM	137.70	194.99	-50.82	-6.46	141.60	-36.91	-4.69
CAT	943.20	756.42	79.99	106.79	80.20	8.48	11.32
CVAL	578.50	461.88	-15.07	131.69	79.84	-2.61	22.76
EXT	82.80	110.47	-20.25	-7.42	133.42	-24.46	-8.96
GAL	61.70	410.04	-226.66	-121.68	664.57	-367.36	-197.21
MAD	845.30	607.38	277.32	-39.39	71.85	32.81	-4.66
MUR	158.00	114.64	-3.83	47.19	72.55	-2.42	29.87
NAV	68.80	72.32	0.06	-3.57	105.11	0.08	-5.19
BC	213.60	277.69	48.80	-112.89	130.00	22.85	-52.85
LR	19.40	38.65	-7.73	-11.52	199.23	-39.85	-59.38

The outcomes in Table 1 show the direction and the intensity of the three effects and allow some conclusions to be made about the regional employment dynamics. So, it is possible to see a growth in the employment levels over all the regions but, on the other hand, different features in other aspects. For example, let us suppose that the objective of the study was to identify the regions with some kind of regional characteristics that make their growth in employment be especially dynamic. These regions would be those where the Regional Effects was positive. The outcomes that Table 1 show would lead to the conclusions that these regions are Andalusia, Canary and Balearic Islands, Catalonia, the Valencian Community and Murcia. In the remaining regions some regional disadvantages make that

[13] All the results that appear in the tables of this chapter are expressed in aggregate terms by region, i. e., $NE_j = \sum_{i=1}^{15} NE_{ij}$, $SE_j = \sum_{i=1}^{15} SE_{ij}$ and $RE_j = \sum_{i=1}^{15} RE_{ij}$.

their employment growth has been smaller than expected, taking into account their industry mix and the growth in the national employment.

As argued before, these solutions are determined by a specific assignation of values to parameters $\theta_x, \theta_{y_i}, \theta_{w_{ij}}$ where we suppose that $\theta_x = 0$, $\theta_{w_{ij}} \to \infty$ and $\theta_x < \theta_{y_i} < \theta_{w_{ij}}$. If we assume that parameters θ are the same for all the factors (i.e., $\theta_x = \theta_{y_i} = \theta_{w_{ij}}$) for all region *i* and sector *j* would yield an average solution. This would be a natural thing to do if no information would be available for the years in-between 1986 and 2000. To illustrate the techniques outlined in the previous sections, we will estimate some of the θ parameters by employing additional information. Specifically, we incorporate information about national employment levels by sector for the intermediate periods from 1992 to 1994. The data are the following:

Table 2. Additional information about sectoral employment levels (1992-1994)

	Sector															
Year	**s1**	**s2**	**s3**	**s4**	**s5**	**s6**	**s7**	**s8**	**s9**	**s10**	**s11**	**s12**	**s13**	**s14**	**s15**	**TOTAL**
1992	1211.1	131.7	894.6	138.9	264.1	424.4	392.8	182.8	390.2	1204.1	2938.6	722.7	320.7	1208.5	2530.9	**12956.1**
1993	1149.8	127.2	841.5	137.5	230.9	428.2	363.2	168.5	375.1	1096.3	2881.8	715.7	314.8	1206.1	2536.7	**12573.3**
1994	1103.8	122.1	830.9	138.4	224.3	418.3	354.2	158.1	367.8	1066.5	2929.6	726.8	309.6	1229.9	2527.7	**12508**

Note that these data provide supplementary information about factors *x* and *y* for these years, but there is no information about the share of sectoral employment over the regions in the intermediate periods considered, *i. e.*, there is no additional information for factor w_{ij}. Consequently, the corresponding estimates for the parameter of this factor will be $\hat{\theta}_{wij} = 1; \forall i = 1,...,15; \ j = 1,...,17$. The use of these non-informative estimates $\hat{\theta}_{wij} = 1; \forall i, j$ guarantee that $\sum_{j=1}^{17} RE^*_{ij} = 0; \forall i$ (see equation (24)).

The supplementary information of Table 2 will be used to estimate the parameter θ_x, which is common for all the sectors, and the 15 parameters $\theta_{yi}; i = 1,..,15$. For the estimation of θ_x we need to decide on the values to be assigned to the a priori distributions contained in the support vector $\mathbf{b}^x$ (see equation 31) and the possible realizations for the random term in vectors $\mathbf{v}^x$ (see equation 34). The following vectors were used throughout the empirical analyses below:[14]

[14] Fernández (2004, p. 142-143) tested the assertion by Golan *et al.* (1996, p. 138) that the estimation results are generally not very sensitive to the choice of a particular set in the specific context of a path-based shift-share analysis. His results strongly confirmed this assertion by Golan *et al.*

$\mathbf{b}^x$=[-5.0, -3.0, -1.0, 1.0, 3.0, 5.0, 7.0]' and $\mathbf{v}^x$=[-0.4, -0.2, 0, 0.2, 0.4]'

Consequently, the ME program for estimate θ_x will be:

$$\underset{\mathbf{p},\mathbf{w}}{Max}\, H(\mathbf{p},\mathbf{w}) = -\sum_{m=1}^{M} p_m\, ln(p_m) - \sum_{t=1}^{T}\sum_{k=1}^{K} q_{tk}\, ln(q_{tk}) \quad (44)$$

Subject to:

$$x^*(t) = \sum_{m=1}^{M} b_m^x p_m t^* + \sum_{k=1}^{K} v_k^x w_k;\ \forall t = 1,...,T$$

$$\sum_{m=1}^{M} p_m = 1$$

$$\sum_{k=1}^{K} q_{tk} = 1;\ \forall t = 1,...,T$$

From the intermediate values of x in Table 3, the estimate for the parameter will be:

$$\hat{\theta}_x = \sum_{m=1}^{M} \hat{p}_m b_m^x = 1.64 \quad (45)$$

This estimate indicates that the best monotonic approximation (given the additional data considered) to the temporal path of the national employment would be like P_1 (see Figure 2). The estimation procedure for the 15 parameters θ_{yi} will be very similar. The supporting vector $\mathbf{b}^y$ will have the same $M = 7$ values as in the previous case that are common for the 15 sectors. However, as the temporal evolution of the sectoral employment weights in those years has been quite different among the industries, it has been necessary to increase the bounds of vectors $\mathbf{v}^y$ to get estimates that fit to the observed temporal behaviour. Specifically, the supporting vectors employed have been:

$\mathbf{b}^y$=[-5.0, -3.0, -1.0, 1.0, 3.0, 5.0, 7.0]' and $\mathbf{v}^y$=[-2, -1, 0, 1, 2]'

With these vectors we need to solve the following ME program:

$$\underset{\mathbf{p},\mathbf{w}}{Max}\, H(\mathbf{p},\mathbf{q}) = -\sum_{i=1}^{15}\sum_{m=1}^{M} p_m \, ln(p_m) - \sum_{i=1}^{15}\sum_{t=1}^{T}\sum_{k=1}^{K} q_{itk} \, ln(q_{itk}) \tag{46}$$

Subject to:

$$y_i^*(t) = \sum_{m=1}^{M} b_m^y p_{im} t^* + \sum_{k=1}^{K} v_k^y q_{itk};\ \forall t = 1,...,T;\ \forall i = 1,...,15$$

$$\sum_{m=1}^{M} p_{im} = 1;\ \ \forall i = 1,...,15$$

$$\sum_{k=1}^{K} q_{itk} = 1;\ \forall t = 1,...,T;\ \forall i = 1,...,15$$

plus the additional constraint to guarantee that $\sum_{j=1}^{17}\sum_{i=1}^{15} SE_{ij}^* = \sum_{j=1}^{17} SE_j^* = 0$ (see equation (25)):

$$\sum_{i=1}^{15} \frac{\sum_{m=1}^{M} b_m^y p_{im}}{\theta_x + \sum_{m=1}^{M} b_m^y p_{im}} \Delta y_i = 0 \tag{47}$$

The estimates will be expressed as:

$$\hat{\theta}_{yi} = \sum_{m=1}^{M} \hat{p}_{im} b_m^y;\ \forall i = 1,..,15 \tag{48}$$

The following table summarizes the estimated values for the parameters obtained:

Table 3. Estimates for parameters θ_{yi}

	Sector														
Sector	**s1**	**s2**	**s3**	**s4**	**s5**	**s6**	**s7**	**s8**	**s9**	**s10**	**s11**	**s12**	**s13**	**s14**	**s15**
$\hat{\theta}_{yi}$	0.81	0.75	1.67	1.14	0.9	2.35	0.78	0.49	1.23	0.93	0	0	2.4	1.52	0.37

Once we have obtained estimates for parameters $\theta_x, \theta_{y_i}, \theta_{w_{ij}}$, we include them in the decomposition forms (21)-(23), in order to obtain the following effects:

Table 4. Effects obtained by the PB approach with additional information

		Absolute effect (thousand of workers)			% of total variation		
	TOTAL VARIATION	National Effect	Sectoral Effect	Regional Effect	National Effect	Sectoral Effect	Regional Effect
AND	825.3	685.72	-38.45	178.03	83.09	-4.66	21.57
ARAG	112.7	150.18	-10.76	-26.72	133.26	-9.55	-23.71
AST	12.2	122.29	-20.28	-89.81	1002.39	-166.23	-736.16
BAL	119.2	97.96	19.41	1.82	82.18	16.28	1.53
CAN	292.5	179.48	20.35	92.67	61.36	6.96	31.68
CANT	38.6	61.21	-2.76	-19.85	158.56	-7.14	-51.43
CAST-L	131.7	290.09	-42.74	-115.65	220.27	-32.45	-87.81
CAST-LM	137.7	184.59	-42.35	-4.54	134.05	-30.75	-3.30
CAT	943.2	789.38	64.47	89.34	83.69	6.84	9.47
CVAL	578.5	482.96	-13.31	108.86	83.48	-2.30	18.82
EXT	82.8	105.91	-19.30	-3.81	127.91	-23.30	-4.61
GAL	61.7	340.72	-152.05	-126.97	552.22	-246.44	-205.79
MAD	845.3	652.63	220.82	-28.15	77.21	26.12	-3.33
MUR	158	124.08	-14.74	48.66	78.53	-9.33	30.80
NAV	68.8	72.11	-2.87	-0.45	104.82	-4.17	-0.65
BC	213.6	266.57	42.00	-94.97	124.80	19.66	-44.46
LR	19.4	35.37	-7.50	-8.46	182.30	-38.67	-43.63

These results (denoted with superscripts *) will be compared to the effects obtained by other solutions that do not take into account additional information, namely, the classical Shift-Share equation (superscripts *s*) and the mean of all the decomposition equations (12a)-(12f) (with superscripts μ):

Table 5. Comparison to other solutions

	Classical Shift-share			Mean solution		
	$100\left[\frac{NE^*}{NE^s}\right]$	$100\left[\frac{SE^*}{SE^s}\right]$	$100\left[\frac{RE^*}{RE^s}\right]$	$100\left[\frac{NE^*}{NE^\mu}\right]$	$100\left[\frac{SE^*}{SE^\mu}\right]$	$100\left[\frac{RE^*}{RE^\mu}\right]$
AND	105.61	384.49	95.71	101.03	88.21	93.63
ARAG	94.74	77.11	83.85	98.95	95.35	96.18
AST	82.62	58.80	88.64	96.09	97.29	95.32
BAL	104.82	89.34	45.42	100.98	95.17	102.00
CAN	116.52	72.04	84.08	102.85	93.95	96.19
CANT	91.80	34.58	98.74	98.47	93.95	96.24
CAST-L	88.78	65.41	89.16	97.67	97.64	95.16
CAST-LM	94.67	83.33	70.23	98.88	96.42	89.68
CAT	104.36	80.60	83.66	100.72	95.35	97.29
CVAL	104.56	88.32	82.66	100.85	104.52	96.88
EXT	95.87	95.27	51.40	99.15	92.38	121.67
GAL	83.09	67.08	104.35	95.89	96.51	93.31
MAD	107.45	79.63	71.47	101.54	95.76	100.34
MUR	108.24	385.04	103.11	101.45	93.57	94.58
NAV	99.72	-5136.79	12.47	99.97	95.47	135.87
BC	96.00	86.08	84.13	99.43	95.89	96.60
LR	91.50	97.04	73.47	98.33	96.51	96.24
Average	**104.01**	**67.66**	**81.69**	**100.72**	**95.23**	**97.69**

The three rightmost columns show how the differences between the classical Shift-Share equation and the PB approach are substantial in some cases. In Catalonia, for instance, whereas the measurements of National Effect is quite similar, the Sectoral and Regional Effect obtained by the classical Shift-Share equation is approximately 20% smaller than in the PB technique. Something similar happens to Aragon, the Valencian Community and the Basque Country. For the other regions, the differences are even bigger, changing sometimes the sign of the effect (such Navarra[15]). In general terms, although the differences are not remarkable in the measurement of the National Effect, for the other two effects one can observe major divergences. Using the PB approach would lead to obtaining Sectoral Effects approximately 30% less important than in the classical Shift-Share equation; furthermore the average Regional Effect would be around 20% smaller. Actually, this is a theoretical result that has been advanced in previous sections: the classical Shift-Share equation "overestimates" the Regional Effect if all the factors of the decomposition problem grow between the initial and the final period.

---

[15] The extremely big values that appear in this table for the Sectoral Effects in Murcia and Navarra are a consequence of the results obtained by classical Shift-Share equation being very small (-3.8 and 0.1, respectively). So, not very big differences (these same effects are –22.30 and –4.75, respectively, by equation 4.11) produce these extreme values of the ratios.

The three leftmost columns compare the solutions obtained under the PB approach to the average of decomposition forms (12a)-(12f), which equals the solution of the method proposed without additional information. This comparison is useful to measure to what extent this supplementary information changes the outcomes obtained. In general terms the divergences to the mean solution are not as large as before, but still considerable in some specific cases. Note that in some regions the PB method yields outcomes approximately 20% larger (see the Regional Effect for Extremadura) or around 10% smaller (Sectoral Effect for Andalusia). The conclusion is that the use of this additional information leads to substantially different outcomes from those obtained by the application of a mean solution.

As an additional test, a yearly dynamic average decomposition will be computed to compare its results with the effects obtained by the PB method suggested, the classical Shift-Share analysis and the mean solution. When a "dynamic" decomposition is computed, the variability in the results of decomposition forms is reduced[16]. This type of decomposition in several stages can be accomplished in the empirical example studied: although we have supposed a scenario where the only additional information were the levels of sectoral employment form 1992 to 1994, yearly data of regional employment levels for sectors are available from 1986 to 2000. If the results obtained by the PB approach are close to those obtained by a yearly decomposition, this means that using only the piece of information considered, it is possible to obtain similar results to this dynamic decomposition. Consequently, a yearly mean decomposition has been computed, *i. e.,* a decomposition applying the equations (12a)-(12f) year by year through the *T* years from 1986 to 2000. Its results will be compared to those yielded by the classical Shift-Share equation, the mean solution of all decompositions and the PB method with additional information from 1992 to 994. The following table shows the squared differences for each effect:

Table 6. Differences to the yearly mean decomposition

	National Effect	Sectoral Effect	Regional Effect	TOTAL
Classical Shift-Share	100.45	86.54	63.76	**250.75**
Mean decomposition	23.30	31.92	38.94	**94.16**
PB approach	27.72	30.68	30.45	**88.85**

Although the variability in the results obtained by the yearly decomposition is small, since this "dynamic" decomposition reduces the interaction term which is split up among the

[16] See Fernández (2004, pp. 134-136) for a more detailed explanation.

factors, it requires information for many periods. The objective of this comparison is to see which of the three other decomposition forms yields more similar outcomes to the yearly decomposition. Note that, in general terms, when all the effects are taken into account the results by the PB technique solution are the closest to the ones yielded by the yearly decomposition. On the other hand, the main gain is given by the consideration of the mean decomposition (the differences go from 250.75 with the classical Shift-Share equation to 94.16, 62%). The application of the PBM with these additional data obtains a more modest reduction from this mean decomposition (from 94.16 to 88.85, a 7%). Only for the case of the National Effect the mean of all decomposition forms would obtain results closer to the yearly decomposition. The conclusion would be that, using only a limited amount of additional information (sectoral employment levels from 1992 to 1994), it is possible to obtain similar results to a very flexible decomposition form.

7. Conclusions

Classical Shift-Share suffers from the "non-uniqueness" problem. Since many decomposition formulae are equally valid from a theoretical point of view, the substantial differences in outcomes noted by Klaasen & Paelinck (1972) pose a serious problem. This paper does not challenge the theoretical equivalence of decomposition formulae but proposes a methodology using Maximum Entropy econometrics to select the decomposition formula that provides an optimal 'fit' to additional empirical information.

The point of departure is a class of monotonic time paths for variables, which led us to label our method the "path based" (PB) method. It was shown that taking the average over all decomposition formulae is equivalent to one specific member of this class, i.e. the linear path. Next, we showed how the parameters that characterize the paths can be estimated, even if the available data is very limited. If information about the values of the determinants contained in the analysis is completely absent, the estimation procedure yields the linear path. If some information is available for some periods between the initial period and the final period of the analysis, the selected path is a different one. Together, the estimated parameters define a decomposition formula. From an empirical point of view, this formula is to be preferred over other decomposition formulae that can be constructed by means of the monotonic times paths considered.

We applied the methodology to quantify the National, Sectoral and Regional Effects to changes in sectoral employment levels in Spanish regions between 1986 and 2000. As additional information we considered the actual levels of regional employment for three intermediate years. The results indicate that the use of additional information in the PB approach can well yield results that differ substantially from the mean over all traditional decomposition formulae, or equivalently, the linear path. For some sectors and effects, the differences amount to more than 20%. Differences of this size lead us to believe that the PB method provides an interesting alternative to computing averages over decomposition formulae.

A couple of challenges remain to be solved, however. In a considerable number of cases, the additional information did not fit the class of monotonic paths we defined. We opted for a rather pragmatic solution if the value of a determinant in an intermediate period exceeded the values in both the initial and the final period (or if it was lower than both), which implies non-monotonicity. We would of course prefer an approach in which non-monotone paths could be estimated. More research should be done in this respect, because a more general class of time paths would complicate the construction of the constrained maximization problems characteristic of maximum entropy estimation procedures. It could also be interesting to see whether estimation results would change if we would estimate the parameters in a way that takes the continuous time character of the temporal paths explicitly into account. In this paper, we do implicitly assume that final demand levels are constant over a year, which is not really in line with the continuous nature of the temporal paths considered. This seems to be a very ambitious task, however.

References

Barff, R..A. and P.L. Knight III (1988), "Dynamic Shift-Share Analysis", *Growth and Change*, vol. 19, pp. 2-10.

Dietzenbacher, E. and B. Los (1998), "Structural Decomposition Techniques: Sense and Sensitivity", *Economic Systems Research*, vol. 10, pp. 307-323.

Dietzenbacher, E. and B. Los (2000), "Structural Decomposition Analyses with Dependent Determinants", *Economic Systems Research*, vol. 12, pp. 497-514.

Esteban-Marquillas, J. M. (1972): "Shift-Share analysis revisited", *Regional and Urban Economics*, 3, pp.249-256.

Fernández, E. (2004), *The Use of Entropy Econometrics in Decomposing Structural Change*, unpublished PhD thesis, University of Oviedo (Spain).

Golan, A., G. Judge and D. Miller (1996), *Maximum Entropy Econometrics: Robust Estimation with Limited Data* (Chichester UK, John Wiley).

Golan, A., G. Judge and S. Robinson (1994), "Recovering Information from Incomplete or Partial Multisectoral Economic Data", *Review of Economics and Statistics*, vol. 76, pp. 541-549.

Harrison, W.J., J.M. Horridge and K.R. Pearson (2000), "Decomposing Simulation Results with Respect to Exogenous Shocks", *Computational Economics*, vol. 15, pp. 227-249.

Hoekstra, R. and J.C.J.M. Van den Bergh (2002), "Structural Decomposition Analysis of Physical Flows in the Economy", *Environmental and Resource Economics*, vol. 23, pp. 357-378.

INE (1987), *Contabilidad Nacional de España 1986,* (Madrid, Instituto Nacional de Estadística).

INE (1993), *Contabilidad Nacional de España 1992,* (Madrid, Instituto Nacional de Estadística).

INE (1994), *Contabilidad Nacional de España 1993,* (Madrid, Instituto Nacional de Estadística).

INE (1995), *Contabilidad Nacional de España 1994,* (Madrid, Instituto Nacional de Estadística).

INE (2001), *Contabilidad Nacional de España 2000* (Madrid, Instituto Nacional de Estadística).

Kapur, J.N. and H.K. Kesavan (1993), *Entropy Optimization Principles with Applications* (New York: Academic Press).

Klaassen L. H. and J. H. P. Paelinck (1972): "Asymmetry in Shift-Share analysis", *Regional and Urban Economics*, 3, pp. 256-261.

Shannon, J. (1948), "A Mathematical Theory of Communication", *Bell System Technical Bulletin Journal*, vol. 27, pp. 379-423.

Sun, J.W. (1998), "Changes in Energy Consumption and Energy Intensity: A Complete Decomposition Model", *Energy Economics*, vol. 20, pp. 85-100.

Appendix A: Data for Empirical Illustration

Table A1. Sectoral classification applied

Sector	Name
s1	Agriculture
s2	Energy
s3	Primary metals, metal products, electrical machinery and instruments
s4	Chemical products
s5	Transport equipment
s6	Food, drinks and tobacco
s7	Textiles, clothing and leather
s8	Paper and derived products
s9	Industries not classified elsewhere
s10	Building materials
s11	Commerce, restaurants and repair services
s12	Transport and communications
s13	Finance and insurance
s14	Other commercial services
s15	Non commercial services

Table A2. Sectoral employment in Spanish regions (1986, thousand of workers)

Region /Sector	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15	*TOTAL*
ANDALUSIA	279.6	14.4	55.7	9.9	30.6	67.2	32.9	10	23.6	125.2	378	95.2	30.7	94.3	328.4	**1575.7**
ARAGON	63.2	7.1	32.4	4.8	12.7	13.2	16.6	3.8	11.2	26.5	70.6	21	8.9	25.1	67.6	**384.7**
ASTURIAS	68.7	34	38.6	1.6	4.9	8.7	2.7	1.8	3.7	25.5	67.7	20.1	5.7	18	57.5	**359.2**
BALEARIC ISLANDS	13.8	2.7	4.1	0.2	0.2	5.4	12.6	1.4	8	27	78.9	16.9	6	16.5	33.1	**226.8**
CANARY ISLANDS	50.2	5.1	5.8	0.5	0.9	10.9	0.7	2.8	5	42.6	124.4	26.4	6.8	28	63.7	**373.8**
CANTABRIA	31.1	1.3	18.7	3.9	3.6	7.5	1.3	1.3	4.9	10.4	28.7	8.7	3.4	10.8	26.2	**161.8**
CASTILLA Y LEON	180.7	21	30.3	6.4	24.9	36.9	12.9	6.4	23	60	154.2	45.1	15.3	41.6	134.3	**793**
CASTILLA-LA MANCHA	119.1	4.9	23.4	4.4	1.6	18.8	26.2	1.6	12.8	44.3	85.9	23.3	8.6	20.2	78.1	**473.2**
CATALONIA	101.9	22	173	55.9	48.2	68.4	152.2	34.8	70.4	119.1	384.8	127.7	59.7	180.6	237	**1835.7**
VALENCIAN COM.	132.7	6.2	74.9	7.7	13.4	39.3	100.1	12.6	66.9	78	270.7	58.1	24.1	76.8	159.4	**1120.9**
EXTREMADURA	72.8	2	5.7	0.3	0.2	9.9	6	0.7	3	25.1	55.3	11.1	5.1	11.1	59.8	**268.1**
GALICIA	404.1	10.3	33.3	3.7	30.1	27.4	11.5	4.1	19.2	62.4	165.1	43.7	14.7	42.2	123.3	**995.1**
MADRID	20.3	13.6	101.8	26.9	38.8	30.5	27.7	31.2	34.8	109.7	310.5	123.8	70.2	175.5	358.7	**1474**
MURCIA	46.7	3	10	3	6.4	18.5	7.4	2	9.8	20.7	56.6	15.4	4.7	18.3	55.7	**278.2**
NAVARRA	19.3	0.8	24.7	1.4	7.5	11.9	4.2	5.6	7.7	10.5	32.5	10.7	3.5	14.1	21.1	**175.5**
BASQUE COUNTRY	28.8	6.9	159.2	10.1	20	14.9	6.6	14.5	35.5	37.1	127.3	34.8	16.3	61.8	100.1	**673.9**
LA RIOJA	13.3	0.4	6.1	0.5	0.8	6.7	12.3	1.4	5.5	6.2	16.5	2.7	2.5	4.9	14	**93.8**
TOTAL	**1646.3**	**155.7**	**797.7**	**141.2**	**244.8**	**396.1**	**433.9**	**136**	**345**	**830.3**	**2407.7**	**684.7**	**286.2**	**839.8**	**1918**	**11263.4**

Table A3. Sectoral employment in Spanish regions (2000, thousand of workers)

Region /Sector	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15	*TOTAL*
ANDALUSIA	262.8	17.3	70.8	11.8	23.5	60.1	25.9	14.1	40.8	269.1	626	115.1	41.3	258.1	564.3	**2401**
ARAGON	36.8	4.3	39.9	3.6	21.5	12.8	11.8	6.2	16.3	40.3	94.5	27.8	11.7	64.2	105.7	**497.4**
ASTURIAS	33.9	11	31.2	1.8	3.5	8.9	3.4	3.5	6.5	38.8	77.7	22.4	6.1	42.6	80.1	**371.4**
BALEARIC ISLANDS	6.8	2.9	7.6	0.3	0.9	7	6.1	3	6.5	45.8	120.3	27.2	7.6	46.9	57.1	**346**
CANARY ISLANDS	43.9	4.7	13.8	1	1.6	15.2	0.7	4.3	6	84.1	212.2	44.6	9.2	83.7	141.3	**666.3**
CANTABRIA	14.2	1.9	14.9	2.4	3.3	7.1	1.7	1.2	5.2	24.2	41.1	11.6	3.5	27.9	40.2	**200.4**
CASTILLA Y LEON	95.7	12.3	41	4.8	23.5	36.2	10	7.7	28.8	105.6	159.3	50.7	19.4	107.5	222.2	**924.7**
CASTILLA-LA MANCHA	71.6	4.5	31.5	5.8	2.7	21.6	29.3	3.6	20.5	75.8	109.8	31.1	11.4	52.8	138.9	**610.9**
CATALONIA	76.2	17	223.1	61.6	67.4	86.4	121.8	66	106.5	244.3	574.2	163.6	69.3	480.6	420.9	**2778.9**
VALENCIAN COM.	83.1	8.9	129.3	9.5	21.1	42.5	95.1	20.7	78	175.7	420.4	87.4	31.8	200.4	295.5	**1699.4**
EXTREMADURA	53.2	3.1	7.7	0.4	0.3	9.3	4.5	1	4	50.4	73.8	14	7	30.9	91.3	**350.9**
GALICIA	183.9	9.8	46.2	3.3	31	28.5	22.5	6.2	28.5	114.9	203.5	50.1	17.5	106.9	204	**1056.8**
MADRID	16.7	17.5	118.1	23.4	36.2	30.1	26.7	57.4	44.8	199.7	423.8	178.8	91.5	452.6	602	**2319.3**
MURCIA	50.6	3	17.3	3.8	3.9	20.4	9.2	3.1	15.1	47.4	95.7	23.8	7	40.6	95.3	**436.2**
NAVARRA	17	1.4	29.9	2	14.2	10.3	3.7	5.5	8.3	22.5	35	13.4	4.7	33	43.4	**244.3**
BASQUE COUNTRY	15.2	6.1	151.9	8.3	23.2	14	4.7	13.5	39.6	77.2	162.7	47.3	15.5	146.5	161.8	**887.5**
LA RIOJA	10.6	0.5	9.8	0.8	2.2	7.6	6.5	1.8	6.4	10	16.7	4.9	2.8	11.2	21.4	**113.2**
TOTAL	**1072.4**	**126.4**	**984.2**	**144.6**	**280**	**418.3**	**383.7**	**218.9**	**461.8**	**1629.1**	**3456.3**	**916**	**357.7**	**2190.9**	**3314.4**	**15954.7**