COMBINING REGIONAL AND NATIONAL INFORMATION FOR THE ESTIMATION OF REGIONAL INPUT-OUTPUT TABLES

Esteban Fernández-Vázquez[1*] and Carmen Ramos-Carvajal[2].

University of Oviedo, Department of Applied Economics, Faculty of Economics, Campus del Cristo, Oviedo, 33006, Spain. [1]evazquez@uniovi.es; [2]cramos@uniovi.es.

[*] Corresponding author.

ABSTRACT

Regional input-output tables are usually not constructed from survey data but are estimated. From known information about the row and column margins, the cells of the matrix are estimated using as a priori information either a regional matrix from a past period (updating) or a contemporaneous input-output table from the same time period (regionalization). This paper proposes the use of a composite Cross-Entropy approach that allows for introducing a mixture of both types of a priori information. An empirical application is included, where a regional input-output for Asturias (Spain) is estimated with this method and the result is analyzed and compared with other more traditional estimation techniques.

1. Introduction

The input-output modelling at a regional scale is a topic that has gained a lot of attention in the last decades, given the huge number potential applications of such models (including regional SAM's and CGE models) for economic researchers and policy makers. The well-known problem is that the compilation of the information required to build a survey-based input-output tables is extremely expensive and time-consuming for the statistical agencies, which causes an important lag between the compilation of the information in the surveys and the publication of the regional table. Therefore, the use of some non-survey method for estimating regional IO tables is becoming more and more popular for input-output users.

Basically, the non-survey techniques for obtaining a regional IO table consist in taking as point of departure an initial IO table, which is assumed to be similar to the table we want to estimate, together with known information on the row and column margins of the actual regional IO table. The basic idea of the estimation process is to choose as solution the table that, fulfilling the constraints imposed by the known information, is the closest to the prior matrix according to some divergence criterion. One of the most used adjusting procedures is the Cross-Entropy (CE) technique, which is based on the Kullback-Leibler divergence criterion. This adjusting technique has been proven to have an equivalent solution to the popular RAS scaling algorithm (for example, McDougall, 1999), although some authors claim that the CE procedure is preferable to some other alternatives because it allows for the inclusion of a wide range of initial information to be used efficiently in the estimation process (Robinson et al. 2001).

This paper explores from a new approach the role played by the initial information in the CE-based estimation process. Traditionally, the adjusting problem takes as point of departure either a regional matrix from a past period (updating) or a contemporaneous IO table from the same time period (regionalizing)[1]. The discussion on the characteristics of the different techniques of adjusting have attracted much attention in

---

[1] The same choice has to be made also with the RAS algorithm.

the literature (see, to name but a few of examples, Mello and Teixeira, 1993; Jackson, 1998; Gilchrist and StLouis, 1999; Jalili, 2000; Jackson and Murray, 2004; Oosterhaven et al., 2008, or Bonfiglio and Chelli, 2008). The novelty of our proposal is that it considers the possibility of including several initial matrices in the estimation process of a regional IO table, instead of choosing one of them.

The paper is organized in the following sections. Section 2 presents the basis of the CE solution to the estimation problem of a matrix with unknown cells but with information on its margins. In section 3, the details of the composite CE technique proposed in this paper is introduced. Section 4 shows a numerical MonteCarlo experiment where the performance of the proposed method is compared with other competing techniques. In Section 5 an empirical application with a real-world example is included, where the IO table for Asturias (a region of Spain) is estimated. Finally, section 6 concludes the paper.

2. The CE solution for the matrix balancing problem

We will base our explanations on the matrix-balancing problem depicted in Golan (2006, page 105), where the goal is to fill the (unknown) cells of a matrix of dimension using the information that is contained in the aggregate data of the row and column sums. Graphically, the point of departure of our problem is a matrix like Table 1.

Table 1: Known and unknown data in a matrix balancing problem.

z_{11}	...	z_{1j}	...	z_{1M}	$z_{1.}$
...		...		...	...
z_{i1}	...	z_{ij}	...	z_{i1}	$z_{i.}$
...		...		...	...
z_{N1}	...	z_{Nj}	...	z_{NM}	$z_{N.}$

$z_{.1}$	...	$z_{.j}$	...	$z_{.M}$

The z_{ij} cells of the matrix are the unknown quantities we would like to estimate (shaded in grey), where the aggregates $\sum_{j=1}^{M} z_{ij} = z_{i.}$, $\sum_{i=1}^{N} z_{ij} = z_{.j}$, and $\sum_{i=1}^{N} \sum_{j=1}^{M} z_{ij} = z$ are known. This is a familiar situation in the context of regional IO tables, where is very usual to have aggregate information (intermediate inputs and outputs per industry) earlier than the flows.

Note that the z_{ij} elements can be expressed as sets of (column) probability distributions, simply dividing the quantities of the matrix by the corresponding column sums $z_{.j}$. In such a case, the previous matrix can be rewritten in terms of a new matrix P that is composed by a set of M probability distributions (Table 2).

Table 2: The matrix balancing problem in terms of probabilities.

p_{11}	...	p_{1j}	...	p_{1M}	y_1
...		...		...	...
p_{i1}	...	p_{ij}	...	p_{iM}	y_i
...		...		...	...

p_{N1}	...	p_{Nj}	...	p_{NM}	y_N
v_1	...	v_j	...	v_M	

Where the p_{ij}'s are defined as the proportions $\frac{z_{ij}}{z_{\cdot j}}$, and the new row and column margins as $v_j = \frac{z_{\cdot j}}{z}$ and $y_i = \frac{z_{i \cdot}}{z}$ respectively. Consequently, the followings equalities are fulfilled by the p_{ij} elements[2]:

$$\sum_{i=1}^{N} p_{ij} = 1 \; ; \; \forall j = 1, \dots, M \tag{1}$$

$$\sum_{j=1}^{M} p_{ij} v_j = y_i \; ; \; \forall i = 1, \dots, N \tag{2}$$

These two sets of equations reflect all we know about the elements of matrix P. Equation (2) shows the cross-relationship between the (unknown) p_{ij}'s in the matrix and the (known) sums of each row and column. Additionally, equation (1) indicates that the p_{ij}'s can be viewed as (column) probability distributions. Note that we have only $N + M$ pieces of information to estimate the $N \times M$ elements of matrix P, which makes the problem ill-posed.

The solution to this type of problems can be obtained by minimizing a divergence measure with a prior probability matrix Q subject to the set of constraints (1) and (2). This is called a Cross-Entropy (CE) problem, which can be written in the following terms:

$$\min_{P} D(P \| Q) = \sum_{i=1}^{N} \sum_{j=1}^{M} p_{ij} \ln \left(\frac{p_{ij}}{q_{ij}} \right) \tag{3}$$

---

[2] Note that in such a case, these p_{ij} elements can be seen as conditional probabilities to each column.

Subject to the same restrictions given by the set of equations () and (). The divergence measure $D(P\|Q)$ is the Kullback-Liebler entropy divergence between the posterior and prior distributions. The Lagrangian function for the CE problem is:

$$L = D(P\|Q) + \sum_{i=1}^{N} \lambda_i \left[y_i - \sum_{j=1}^{M} p_{ij} v_j \right] + \sum_{j=1}^{M} \mu_j \left[1 - \sum_{i=1}^{N} p_{ij} \right] \tag{4}$$

And the solutions are:

$$\tilde{p}_{ij} = \frac{q_{ij} \exp[\tilde{\lambda}_i v_j]}{\sum_{i=1}^{N} q_{ij} \exp[\tilde{\lambda}_i v_j]}; \ \forall i = 1, \dots N; j = 1, \dots M \tag{5}$$

The CE estimation procedure can be seen as an extension of the Maximum Entropy (ME) principle (or alternatively the ME can be considered as a particular case of the CE procedure), given that the solutions of both approaches are the same when the M *a priori* probability distribution contained in Q are all uniform.[3] It is well-known that depending on the choice made when specifying Q, the general matrix adjusting problem can be posed as an updating (if we take as prior a previous regional IO table Q^r) or a regionalizing problem (if the national IO table Q^n contemporaneous to the regional table we want to estimate is our prior).[4] In some cases there is no room for this choice, simply because only one of these tow priors is available. But it may well be that for one specific period and region we can both use as prior matrices Q^r or Q^n. This situation is becoming more and more frequent in countries with long tradition in building survey regional IO tables.

3. A composite CE method: the DWP estimation technique

---

[3] In other words, the ME solutions are obtained by minimizing the Kullback-Liebler divergence $D(P\|Q)$ between the unknown p_{ij} and the probabilities $q_{ij} = \frac{1}{N}; \ \forall i = 1, \dots N; j = 1, \dots M$.

[4] See Hewings (1984) for a detailed discussion on the role played by the prior information in such estimation problems.

The above sketched procedure can be extended in order to develop a more flexible estimator that allows for including in the estimation process both prior matrices Q^r and Q^n. Related to the Bayesian Method of Moments (BMOM, see Zellner, 1996, 1997), the technique has been proposed in Golan (2001) as data-based method of estimation that uses both sample and non-sample information in determining a basis for coefficient reduction and extraneous variable identification. In other words, this technique allows for shrinking the coefficient of the explanatory variables that can be classified as irrelevant in a linear model. Another recent empirical application of this method can also be found in Bernadini (2008).

Our objective in the context of a matrix balancing problem will be to identify which of the two priors would be preferable for each industry contained in the IO table to be estimated and simultaneously to obtain estimated matrices with a good performance. For the sake of simplicity, let us assume that we want to estimate a symmetric industry-by-industry regional IO table ($N = M$). If we denote with q^r and q^n the regional and national *a priori* (column) distributions respectively, the objective proposed can be achieved by modifying the previous CE program in the following way:

$$\min_{P,P^\gamma} D(P, P^\gamma \| Q^r, Q^n, Q^\gamma) = \sum_{j=1}^{M} (1 - \gamma_j) \sum_{i=1}^{N} p_{ij} \ln \left(\frac{p_{ij}}{q_{ij}^r} \right)$$

$$+ \sum_{j=1}^{M} \gamma_j \sum_{i=1}^{N} p_{ij} \ln \left(\frac{p_{ij}}{q_{ij}^n} \right) \tag{6}$$

$$+ \sum_{h=1}^{H} \sum_{j=1}^{M} p_{hj}^\gamma \ln \left(\frac{p_{hj}^\gamma}{q_{hj}^\gamma} \right)$$

subject to:

$$\sum_{j=1}^{M} p_{ij} v_j = y_i \ \forall i = 1, \dots, N \tag{7}$$

$$\sum_{i=1}^{N} p_{ij} = 1 ; \ \forall j = 1, \dots, M \tag{8}$$

$$\sum_{h=1}^{H} p_{hj}^{\gamma} = 1 \; ; \; \forall j = 1, \dots, M \tag{9}$$

The γ_j parameters are estimated simultaneously with the unknown probabilities p_{ij} of the matrix. Each γ_j measures the weight given to the national prior q^n for each industry and it is defined as $\gamma_j = \sum_{h=1}^{H} b_{hj}^{\gamma} p_{hj}^{\gamma}$, where $b_{1j}^{\gamma} = 0$ and $b_{Hj}^{\gamma} = 1$ are respectively the lower and upper bound defined as the support of these parameters (note that this implies that $0 \leq \gamma_j \leq 1; \; \forall j = 1, .., M$). The a priori probability distributions fixed for them are uniform $\left(q_h^{\gamma} = \frac{1}{H}; \; \forall j = 1, .., M \right)$.

To understand the logic of this data-weighted prior (DWP) estimator an explanation on the objective function of the previous minimization program is required. Note that equation (6) is divided in three terms. The first term quantifies the divergence between the recovered probabilities and the *a priori* probabilities where a previous regional table is chosen as prior, being this divergence weighted by $(1 - \gamma_j)$ for each industry. On the contrary, the second element of (6) measures the divergence with a simultaneous national prior and it s weighted by γ_j. The third element in (6) relates to the Kullback divergence of the weighting parameters γ_j.

The solutions of this minimization program are:

$$\tilde{p}_{ij} = \frac{q_{ij}^{r \, (\tilde{\gamma}_j / A_j)} exp\left[(A_h{}^{-1})\tilde{\lambda}_i v_j\right]}{\sum_{i=1}^{N} q_{ij}^{r \, (\tilde{\gamma}_j / A_j)} exp\left[(A_h{}^{-1})\tilde{\lambda}_i v_j\right]} \; ; \; \forall i = 1, \dots, N; j = 1, \dots, M \tag{10}$$

Where:

$$\tilde{\gamma}_j = \sum_{h=1}^{H} b_{hj}^{\gamma} \tilde{p}_{hj}^{\gamma},$$

$$A_j = [1 - \tilde{\gamma}_j] / [(\tilde{\gamma}_j - 1) - \tilde{\gamma}_j \sum_{i=1}^{N} q_{ij}^r ln(q_{ij}^r) + \tilde{\gamma}_j]$$

and $\tilde{\lambda}_i$ are the Lagrangian multipliers associated with restrictions (7). The properties of this DWP estimator in the context of classical linear regression models haven tested in Golan (2001).[5]

Simultaneously to the estimation of the p_{ij} cells of the matrix, the DWP estimator discriminates for each industry j between the two priors considered. The proposed estimation strategy provides estimates of the weighting parameters γ_j, obtained as:

$$\tilde{\gamma}_j = \sum_{h=1}^{H} b_{hj}^{\gamma} \tilde{p}_{hj}^{\gamma}, \tag{11}$$

which can be used as a tool for this purpose. Note that as $\tilde{\gamma}_j \rightarrow 0$ the regional prior $\boldsymbol{q^r}$ gains weight for industry j and the estimates approach those of the CE updating process. On the contrary, large values of $\tilde{\gamma}_j$, the CE estimation (regionalizing) with a national prior $\boldsymbol{q^n}$ takes over. Consequently, relatively large values of $\tilde{\gamma}_j$ ($\tilde{\gamma}_j \geq 0.5$) will be an indication of an industry j characterized by a high weight of the national prior. In other words, in this industry would be preferable to use a regionalizing adjustment of the contemporaneous national IO table rather than updating a previous regional IO table. On the contrary, comparatively small values of $\tilde{\gamma}_j$ (when $\tilde{\gamma}_j < 0.5$) are a signal of an industry j where the updating process should be preferred.

4. Testing the DWP estimation technique with numerical experiments

In order to test the performance of the proposed estimation technique, we have carried out a numerical simulation exercise where the DWP estimation is compared with a more traditional adjusting process where only one prior matrix is considered.

---

[5] Under some mild assumptions (see Golan 2001, page 177) the consistency and asymptotic normality of the DWP estimates can be ensured. Additionally, these assumptions also guarantee that the approximate variances of the DWP estimator is lower than the approximate variance of the generalized CE estimator, which in turn is lower than the approximate variance of a ML-LS estimator (see Golan, 2001, page 179).

In the experiment we have fixed a target matrix $\mathbf{Z}$ of interindustry flows with dimension 30×30 where the only known information is the column and row margins. This matrix has been fixed as the actual industry-by-industry symmetric table for the region of Asturias (Spain) in 2005, published by SADEI (the regional statistical agency of Asturias). Moreover, the matrix $\mathbf{Z}$ has been transformed into a matrix of column coefficients $\mathbf{P}$ to be estimated from the information contained in vectors $\mathbf{v}$ and $\mathbf{y}$. We also defined several a priori matrices $\mathbf{Q}$ to be used in the estimation of $\mathbf{P}$. Firstly, we have generated a matrix $\mathbf{Q}^r$ that plays the role of a previous matrix for this region. The values of this matrix have been obtained as $q_{ij}^r = p_{ij} \cdot u_{ij}^r$; where u_{ij}^r is a perturbation term that behaves as $u_{ij}^r \sim N(1, \sigma)$ and $\sigma = 0.1$.

4.1. *First experiment*

Additionally to this *a priori* matrix, we have also generated a matrix $\mathbf{Q}^{nA}$ that is assumed to be a national matrix contemporaneous to our regional target matrix $\mathbf{P}$. The elements of $\mathbf{Q}^{nA}$ have been obtained as $q_{ij}^{nA} = p_{ij} \cdot u_{ij}^{nA}$, being $u_{ij}^{nA} \sim N(1, 2\sigma)$. Note that is clear that the prior contained in matrix $\mathbf{Q}^r$ is closer to matrix $\mathbf{P}$ than the prior contained in $\mathbf{Q}^{nA}$. In this scenario we have estimated matrix $\mathbf{P}$ by three different ways: updating the previous regional matrix $\mathbf{Q}^r$, regionalizing the national matrix $\mathbf{Q}^{nA}$ and using the proposed DWP estimation technique that takes both matrices as possible priors. These three estimation strategies correspond respectively with the minimization of the three following divergence measures:

$$\underset{P}{\text{Min}}\, D(\mathbf{P} \| \mathbf{Q}^r) = \sum_{i=1}^{30} \sum_{j=1}^{30} p_{ij}\, ln\left(\frac{p_{ij}}{q_{ij}^r}\right) \tag{12a}$$

$$\underset{P}{\text{Min}}\, D(\mathbf{P} \| \mathbf{Q}^{nA}) = \sum_{i=1}^{30} \sum_{j=1}^{30} p_{ij}\, ln\left(\frac{p_{ij}}{q_{ij}^{nA}}\right) \tag{12b}$$

10

$$\underset{P,P^{\gamma}}{\text{Min}} \, D(P,P^{\gamma} \| Q^r, Q^n, Q^{\gamma}) = \sum_{j=1}^{30} (1-\gamma_j) \sum_{i=1}^{30} p_{ij} \, ln \left(\frac{p_{ij}}{q_{ij}^r} \right)$$

$$+ \sum_{j=1}^{30} \gamma_j \sum_{i=1}^{30} p_{ij} \, ln \left(\frac{p_{ij}}{q_{ij}^{nA}} \right) \tag{12c}$$

$$+ \sum_{h=1}^{H} \sum_{j=1}^{30} p_{hj}^{\gamma} \, ln \left(\frac{p_{hj}^{\gamma}}{q_{hj}^{\gamma}} \right)$$

subject to the same type of constraints explained before. To evaluate the performance of these estimation approaches, 1,000 trials have been carried out and we have computed the average of a measure of overall deviation between the target matrix and the estimates. Specifically, we obtained the mean absolute error percentage (MAPE), given that is frequently used in the studies that evaluate the performance of updating IO tables. It is defined as:

$$MAPE = \sum_{i=1}^{30} \sum_{j=1}^{30} 100 \frac{|z_{ij} - \hat{z}_{ij}|}{\sum_{i=1}^{30} \sum_{j=1}^{30} z_{ij}} \tag{13}$$

where the $\hat{z}_{ij}$ elements denote the estimated flows. To extend this evaluation, we also obtained the deviation MAPE measure between the actual and estimated technical coefficients a_{ij}, where $a_{ij} = {z_{ij}}/{x_j}$ and x_j is the total output of industry j in the target IO matrix. These output values are assumed as known in the experiment and they have been fixed as the actual output values of the regional IO table of Asturias in 2005. Finally, we paid also attention to the accuracy in the estimation of the output multipliers, obtaining the MAPE for the l_{ij} elements of the matrix L, being $L = [I - A]^{-1}$ the inverse of Leontief.[6] The following table summarizes the results obtained:

---

[6] We also follow the usual procedure of not taking into account the element on the main diagonal of L in this comparison, given that these cells cannot be smaller than one by definition.

Table 3. First experiment: average MAPE for flows, technical coefficients and output multipliers

Technique (prior used)	MAPE		
	z_{ij}	a_{ij}	l_{ij}
Regionalizing (Q^{nA})	10.56	3.85	9.56
Updating (Q^{r})	5.42	1.91	4.87
DWP (Q^{nA}, Q^{r})	6.00	2.15	5.48

Not surprisingly, the best results in this experiment are those obtained by the updating approach: given that the matrix taken as prior is closer to the target matrix than the matrix chosen in the regionalizing approach, the accuracy of the former technique is larger. It is important to note that, in such a situation, there are not gains by using the proposed DWP estimation. This is given by the fact that there is one a priori matrix that in all the cases (industries) is closer to the target matrix than the other alternative. Irrespectively on the industry, it would be always better to use as prior the previous regional table Q^{r} than the competing national prior Q^{nA} and it would be also preferable to any possible combination of Q^{r} and Q^{nA}. The weighting parameters γ_j, estimated by the DWP technique are not very helpful either in this case
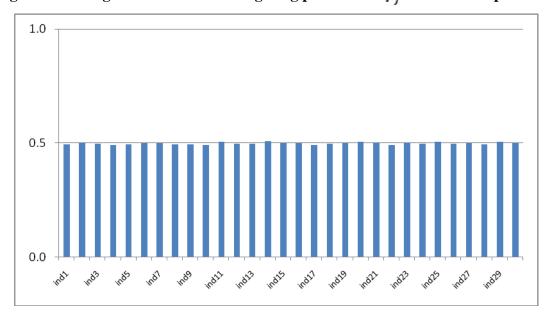
Figure 1. Average estimates of the weighting parameters γ_j in the first experiment

The figure shows that the average estimates are around the threshold value 0.5 under the conditions specified in this first experiment. This is a signal that points out that the

DWP technique does not manage to discriminate clearly between the two options for priors. All in all, the results obtained from this first experiment seems to suggest, finally, that the DWP option does not perform comparatively better than an adjusting technique with only one prior if one of the initial matrix considered is always preferable than the other.

4.2. Second experiment

In the second numerical experiment we try to replicate other type of situations, where the use of a specific prior is not preferable to other alternative in all the cases. This happens in situations where the researcher wants to estimate a regional IO table characterized by having, in the one hand, part of its industries with in the (column) coefficients that present a similar distribution to a previous regional table. However, in the other hand, some other industries are more similar to a contemporaneous national IO table.

Keeping the rest of conditions identical to the first one, in this second experiment we introduce a new national *a priori* Q^{nB}; where the elements have been defined as:

$$q_{ij}^{nB} = \begin{cases} q_{ij1}^{nB} = U[0,0.02]; \ j1 = 1, \ldots, 15 \\ q_{ij2}^{nB} = p_{ij2} \cdot u_{ij1}^{nB} \text{ and } u_{ij2}^{nB} \sim N(1,0.5\sigma); \ j2 = 16, \ldots, 30. \end{cases}$$

In other words, this new a priori matrix is characterized by having one half of their columns (from industry number 1 to number 15) completely random; which means that in these cases it would be preferable taking Q^r as initial matrix. However, for the other one half of the industries (number 16 to 30) it happens the opposite, given that the distribution is closer to the target matrix than the prior Q^r. Under these new conditions we repeated the simulation and computed the MAPE measures for the flows, technical coefficients and output multipliers. Table 4 reports the average results we found in the 1,000 trials:

Table 4. Second experiment: average MAPE for flows, technical coefficients and output multipliers

	MAPE

Technique (prior used)	z_{ij}	a_{ij}	l_{ij}
Regionalizing ($\boldsymbol{Q}^{nB}$)	60.38	22.41	54.87
Updating ($\boldsymbol{Q}^{r}$)	5.42	1.91	4.87
DWP ($\boldsymbol{Q}^{nB}$, $\boldsymbol{Q}^{r}$)	4.84	1.82	4.72

Now we observe some interesting results. Firstly, we can see how using only the new national IO table as prior for the estimation of our target matrix would increase the deviation between estimates and the actual values. The erratic behaviour we specified for 15 out of the 30 industries contained in $\boldsymbol{Q}^{nB}$ causes this result. But this does not mean that all the information contained in this matrix should be neglected, because in the other one half of the industries the column coefficients distribute closer to the target matrix than the regional prior $\boldsymbol{Q}^{r}$. If we incorporate both matrices of *a priori* information in the adjusting process by using the DWP estimation, we let the data speak for themselves and choose the most appropriate prior for each industry, which in the end obtains smaller deviation measures.

The average results obtained for the weighting parameters γ_j also show a clear picture of how the DWP estimation works:
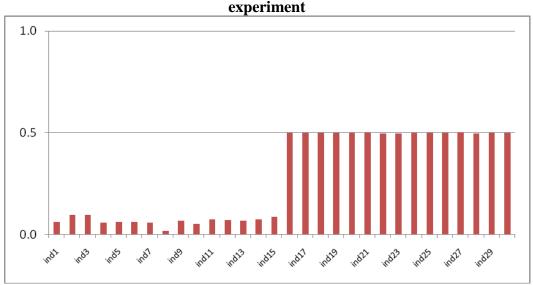
Figure 2. Average estimates of the weighting parameters γ_j in the second experiment

Note that now we have a complete different picture form the results obtained in the first experiment. Under the conditions described in this second experiment, the DWP estimation technique allows for identifying the industries where the distribution of the coefficients should be taken from one of the initial matrices (industries from 1 to 15). Even when the *a priori* expected value is $\tilde{\gamma}_j$=0.5, the information included into the estimation process leads the DWP technique to give much bigger weights to one of the priors in this first group of industries. Note also that for the second group (industries from 16 to 30) this information does not seem to be enough to obtain estimates of γ_j significantly larger than 0.5, although the overall results in terms of our MAPE deviation measure is satisfactory.

5. An empirical application: estimating the regional IO table for Asturias, 2005

As a complement to the numerical simulation made in the previous section, this section presents an empirical application of the proposed DWP technique and compares the results obtained with other competing techniques. For this purpose, we took again the industry-by-industry symmetric IO table for the region of Asturias (Spain) in 2005. Assuming that the only known information of this matrix are the row and column margins (intermediate outputs and inputs respectively) and the vector **x** of total output, we try to estimate the inter-industry flows matrix **Z**, the matrix **A** of technical coefficients and the Leontief inverse **L**. For this purpose, we will apply and adjusting process to obtain the column-coefficients matrix **P** from different initial matrices, all they having the same industry classification at 30 branches. Details of the industry classification are given in the Appendix.

Specifically, we considered the symmetric Spanish IO table for 2005 (Q^n, obtained from the Spanish Statistical Institute INE), and two previous regional matrices for Asturias in 1995 and 2000 (Q^{r95} and Q^{r00}, respectively) published by SADEI. The specification of these two previous regional tables requires an explanation: if we only had considered as prior the regional table for 2000, we would have a situation as the one described in the first numerical experiment (because the coefficients in 2005 are probably close to those of the 2000 table). We opted for including the 1995 table as

initial matrix to illustrate how the DWP technique works in a case of a large time lag between the published regional tables. Consequently, it is probable that the regional economy experiences a structural change in part of their industries between 1995 and 2005. In such a situation, taking the 1995 table as prior would not be the best option for all the industries, because the contemporaneous national IO table for 2005 could be preferable as initial distribution in the estimation of the coefficients for some industries (the situation studied in the second experiment).

The following table summarizes the results obtained in this study case:

Table 5. MAPE for flows, technical coefficients and output multipliers for several estimation strategies

	MAPE		
Technique (prior used)	z_{ij}	a_{ij}	l_{ij}
Regionalizing (Q^n)	36.35	15.08	38.41
Updating1 (Q^{r00})	24.47	9.61	24.72
DWP1 (Q^n, Q^{r00})	25.35	10.50	27.17
Updating2 (Q^{r95})	38.53	16.07	39.88
DWP2 (Q^n, Q^{r95})	25.45	10.58	27.39

In the estimation problem studied here, we have available a previous regional IO table from 2000 sufficiently close to the target matrix, so this option is always preferable to any other of the possible initial matrices. In line with the results obtained in the first numerical experiment on previous section, the proposed DWP technique does not offer clear gains from combining *a priori* matrices when we have such a scenario where one of them is clearly preferable to the other (comparing the DWP1 with the Updating1 results).

However, it could happen that we had not a recent previous regional IO table to update. This is relatively frequent for the case of the Spanish regions, where the lag between the published IO regional tables is sometimes 10 years or even more (in Catalonia, for example, the two most recent regional tables have been published in 1987 and 2001). What would happen for the case of Asturias if we did not have the table of 2000, but we had the 1995 IO table instead? We can observe in the table how a "pure" updating yields larger deviations than the regionalizing strategy. In such a case, where it is

possible that some industries had experienced a structural change (but some others had not) is where the DWP technique can be helpful. Note that in this case we can detect an improvement in the overall error measures (compare the DWP2 with the Updating2 MAPE values).

The estimates of the weighting parameters γ_j from the DWP estimations in the two cases are also in line with the results obtained in the Monte Carlo simulations:
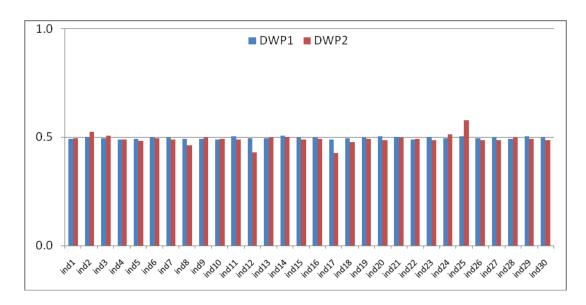
Figure 3. Estimates of the weighting parameters γ_j in the empirical application

We can see how the DWP estimator discriminates more clearly between the national and the regional priors in the case where the national table for 2005 is combined with the regional table for 1995 (DWP2): the technique manages to identify some industries (industry 2, Extractive and mining activities; or in particular industry 25, Computer services) where the contemporaneous national prior is specially favored in the estimation process. Oppositely, there are other industries where the previous regional *a priori* matrix of 1995 is clearly preferred (industry 12, Steel; and industry 17, Energy). This discrimination is almost inexistent when the DWP combine the national 2005 table with the regional 2000 matrix.

6. Concluding remarks

Regional input-output modelling often requires using some non-survey method for estimating IO tables at this scale. Traditionally, these techniques take an initial IO table, similar to the table to be estimated, which is somehow adjusted to fulfil the constraints imposed by the known information, and at the same time is the closest to the prior matrix according to some divergence criterion (the CE technique is a well-known example of such a procedure). The adjusting problem takes as point of departure either a regional matrix from a past period (updating) or a contemporaneous IO table from the same time period (regionalizing).

This paper, however, suggest a new approach of dealing with this initial information. Based in a previous work by Golan (2001), the so-called DWP estimation strategy considers the possibility of including several *a priori* matrices in the estimation process of a regional IO table, instead of choosing only one of them. By means of a Monte Carlo simulation, the performance of the proposed DWP method is compared with other adjusting techniques. The findings of this experiment suggest that the proposal can be useful in situation where none of the available prior matrices is preferable to the other for all the cases (industries). The empirical application with a real-world example, where the IO table for Asturias (a region of Spain) is estimated, seems to confirm this conclusion.

References

Bernardini R, (2008) A Composite Generalized Cross Entropy formulation in small samples estimation, Econometric Reviews, 27, 596-609.

Bonfiglio, A. and Chelli, F. (2008) Assessing the behaviour of non-survey methods for constructing regional input-output tables through a Monte Carlo simulation, Economic Systems Research, 20: 243-258.

Golan A, (2001) A simultaneous estimation and variable selection rule, Journal of Econometrics, 101, 165-193.

Golan, A. (2006) Information and Entropy Econometrics. A review and synthesis, Foundations and Trends in Econometrics, 2, pp. 1-145.

Golan, A., Judge, G. and D. Miller, (1996): Maximum Entropy Econometrics: Robust Estimation with Limited Data, Nueva York, John Wiley & Sons.

Goldberger, A.S., (1991) A Course in Econometrics, Cambridge-USA, Harvard University Press.

Hewings, G.J.D. (1984) The role of prior information in updating input-output models, Socio Economic Planning Science, 18: 319-339.

Jackson, R. W. (1998) Regionalizing national commodity-by-industry accounts. Economic Systems Research 10: 223-238.

Jackson, R.W. and Murray, A. (2004) Alternative Input-Output Matrix Updating Formulations. Economic Systems Research, 16: 135-148.

Jalili, A.R. (2000). Evaluating Relative Performances of Four Non-Survey Techniques of Updating Input-Output Coefficients. Economics of Planning 33: 221-237.

Kapur, J. N. and Kesavan, , H. K. (1992) Entropy Optimization Principles with Applications. Academic Press. New York.

Kullback, J. (1959) Information Theory and Statistics, New York: Wiley.

McDougall, R.A. (1999) Entropy Theory and RAS are Friends, GTAP Working Paper No. 6. Center for Global Trade Analysis, Purdue University.

Mello, L.R.J. and Teixeira, J.R. (1993) The updating of input-output matrices: a systematic survey, Economics Letters, 41: 241–245.

Oosterhaven, J., Stelder, D. and Inomata, S. (2008) Estimating international interindustry linkages: non-survey simulations of the Asian-Pacific economy, Economic Systems Research, 20: 395-414.

Robison, S., Cattaneo, A. and El-Said, M. (2001) Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods. Economic Systems Research 13: 47-64.

Zellner A, (1996) Models, prior information, and Bayesian analysis, Journal of Econometrics, 75, 51-68.

Zellner A, (1997) The Bayesian method of moments (BMOM): theory and applications. In: Fomby T, Hill R, (eds), Advances in Econometrics, 12, 85-105.

Appendix: industry classification

Industry number	Industry description
ind1	Agriculture
ind2	Extractive and mining activities
ind3	Mineral products
ind4	Food, beverages and tobacco
ind5	Textile manufactures
ind6	Wood and cork
ind7	Paper, publishing and editing industries
ind8	Coke and fuels
ind9	Chemistry industry
ind10	Rubber and plastics
ind11	Other non-metallic products
ind12	Steel
ind13	Machinery and mechanical equipment
ind14	Office, optical and electronic products
ind15	Transport equipment
ind16	Manufacturing n.e.c.
ind17	Energy
ind18	Building materials and construction
ind19	Commerce
ind20	Hotels and restaurants
ind21	Transport services
ind22	Post and telecommunications
ind23	Banking and insurance services
ind24	Renting and real estate services
ind25	Computer services
ind26	Services to companies
ind27	Public administration
ind28	Education services
ind29	Health
Ind30	Social services n.e.c.