Article

Recovering Matrices of Economic Flows from Incomplete Data and a Composite Prior

Esteban Fernández-Vázquez

University of Oviedo, Department of Applied Economics, Faculty of Economics, Campus del Cristo, Oviedo, 33006, Spain; E-Mail: evazquez@uniovi.es

Abstract: In several socioeconomic applications, matrices containing information on flows-trade, income or migration flows, for example–are usually not constructed from direct observation but are rather estimated, since the compilation of the information required is often extremely expensive and time-consuming. The estimation process takes as point of departure another matrix which is adjusted until it optimizes some divergence criterion and simultaneously is consistent with some partial information-row and column margins–of the target matrix. Among all the possible criteria to be considered, one of the most popular is the Kullback-Leibler divergence [1], leading to the well-known Cross-Entropy technique. This paper proposes the use of a composite Cross-Entropy approach that allows for introducing a mixture of two types of *a priori* information–two possible matrices to be included as point of departure in the estimation process. By means of a Monte Carlo simulation experiment, we will show that under some circumstances this approach outperforms other competing estimators. Besides, a real-world case with a matrix of interregional trade is included to show the applicability of the suggested technique.

Keywords: cross-entropy estimation; data-weighted priors; matrices of flows; economic applications

1. Introduction

Research on socio-economic problems in general, and economic modeling in particular, often has to deal with information about the flows contained in a matrix of interaction between agents. Just to mention a couple of examples, international economists frequently analyze the flows of bilateral trade between a group of countries in order to measure the level of integration and demographers take data

from matrices of international or interregional migration flows and sociologists. The problem is that the compilation of the information required to build this type of matrices is often extremely expensive and time-consuming for the statistical agencies or research institutes. In this context, the use of estimation methods for recovering the cells of these matrices is becoming more and more popular; especially the estimation techniques based on information measures (see Golan *et al.*, [2] or Kapur and Kesavan [3]).

Basically, these techniques take as point of departure an initial matrix (the prior), which is assumed to be similar to the one we want to estimate, together with some limited information of the actual matrix, usually aggregated data -row and column margins-. The basic idea of the estimation process is to choose as solution the matrix that, fulfilling the constraints imposed by the known information, is the closest to the prior 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 paper explores from a new viewpoint the role played by the initial information in an estimation process based on information measures. Traditionally, the estimation problem takes as point of departure one single prior; for example, a previous matrix from a past period or a contemporaneous matrix for other geographical area. The novelty of our proposal is that it considers the possibility of including several initial matrices in the estimation process, instead of choosing only one of them. By means of a numerical simulation, we illustrate that the proposed technique that uses a composite of two priors obtain comparative better results than an adjustment from only one of them, provided that none of them is preferable to the other for all the probability distributions contained in the target matrix.

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 Monte Carlo 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 a matrix of interregional trade for the Spanish regions 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 ([4], page 105), where the goal is to fill the (unknown) cells of a matrix of dimension $N \times M$ using the information that is contained in the aggregate data of the row and column sums. This is a familiar situation in the context of economics, where the cells of some matrices containing information on flows between several agents are usually not observable directly. Instead, the researchers often limit to observe aggregate information (total sales or purchases per agent, for example), because these aggregates are much quicker and easier to obtain.

The z_{ij} cells of the matrix are the unknown quantities we would like to estimate, where the aggregates $\sum_{j=1}^{M} z_{ij} = z_{i\cdot}$, $\sum_{i=1}^{N} z_{ij} = z_{\cdot j}$, and $\sum_{i=1}^{N} \sum_{j=1}^{M} z_{ij} = z$ are known. 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 $\boldsymbol{P}$ that is composed by a set of M probability distributions (Table 1).

Table 1. The matrix balancing problem.

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 (shaded in grey) are defined as the proportions $\frac{z_{ij}}{z_{.j}}$, and the new row and column margins as $v_j = \frac{z_{.j}}{z}$ and $y_i = \frac{z_{i.}}{z}$ respectively. Consequently, the followings equalities are fulfilled by the p_{ij} elements:

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

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

These two sets of equations reflect all we know about the elements of matrix $\boldsymbol{P}$. 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 $\boldsymbol{P}$, 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 $\boldsymbol{Q}$ 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:

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

Subject to the same restrictions given by the set of equations (1) and (2). The divergence measure $D(\boldsymbol{P}\|\boldsymbol{Q})$ is the Kullback-Liebler entropy divergence between the posterior and prior distributions. The Lagrangian function for the CE problem is:

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

And the solutions are:

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

where $\tilde{\lambda}_i$ are the Lagrangian multipliers associated with the constraints (2). The CE estimation procedure can be seen as an extension of the Maximum Entropy (ME) principle, given that the solutions of both approaches are the same when the *M a priori* probability distribution contained in $\boldsymbol{Q}$ are all uniform. The accuracy of the solution obtained for this matrix adjusting problem will depend on the choice made when specifying $\boldsymbol{Q}$ (see, for example, Hewings [5] for a detailed discussion on the role played by the prior information in such estimation problems in a socio-economic context). In some cases there is no room for this choice, simply because only one possible prior is available. But it may well be that for some cases we have the possibility of using two alternative matrices $\boldsymbol{Q}^a$ or $\boldsymbol{Q}^b$ as prior. For example, if the objective was to estimate inter-industry trade flows for a country in a specific year, it would be possible to take as prior the observed flows for the same country in a past year or, alternatively, to take as prior the distribution of inter-industry flows for a simultaneous year but in a different country. The next section of the paper deals with situations where we include both priors at the same time to the estimation process.

3. A Composite Ce Method: The Dwp Estimation Technique

The above sketched CE procedure can be extended in order to develop a more flexible estimator that allows for including in the estimation process both prior matrices $\boldsymbol{Q}^a$ and $\boldsymbol{Q}^b$. Related to the Bayesian Method of Moments (see Zellner, [6,7]), the technique has been proposed in Golan [8] as a 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 regression linear models. Another recent empirical application of this method to the field of empirical economic analysis can also be found in Bernadini [9]. The point of departure of the estimation technique proposed in Golan [8] it is to consider two alternative priors for each coefficient in a linear model. One of the prior is a uniform distribution centered on zero and the other is a spike distribution with a unit mass prior on zero. When the spike prior takes over the uniform one, the coefficient is shrunk and the corresponding variable is classified as extraneous.

This idea will be adapted to a matrix balancing problem of a target matrix $\boldsymbol{P}$ from two possible priors $\boldsymbol{Q}^a$ and $\boldsymbol{Q}^b$ and our objective in this context will be twofold: a) to identify which of the two priors would be preferable for each column of the matrix and, simultaneously, b) to estimate the target matrix. If we denote with $\boldsymbol{q}^a$ and $\boldsymbol{q}^b$ the two options we have for the *a priori* (column) distributions respectively, the objective proposed can be achieved by modifying the previous CE program in the following way:

$$\text{Min}_{\boldsymbol{P},\boldsymbol{P}^\gamma} D\left(\boldsymbol{P},\boldsymbol{P}^\gamma \| \boldsymbol{Q}^a, \boldsymbol{Q}^b, \boldsymbol{Q}^\gamma\right) = \sum_{j=1}^{M}(1-\gamma_j)\sum_{i=1}^{N} p_{ij}\, ln\left(\frac{p_{ij}}{q_{ij}^a}\right)$$

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

$$+ \sum_{h=1}^{H}\sum_{j=1}^{M} p_{hj}^\gamma\, ln\left(\frac{p_{hj}^\gamma}{q_{hj}^\gamma}\right) \qquad (6)$$

subject to:

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

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

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

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 prior q^b for each column 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 $\left(q_h^{\gamma} = \frac{1}{H}; \ \forall j = 1, .., M\right)$. This means that the *a priori* value for each γ_j parameter is 0.5, but the sample information contained in constraint (7) allows for estimates that deviate from this initial point.

To understand the logic of this data-weighted prior (DWP) estimator some further explanations 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 matrix Q^a is chosen as prior, being this divergence weighted by $(1 - \gamma_j)$ for each column. On the contrary, the second element of (6) measures the divergence with the prior Q^b and it is weighted by γ_j. The third element in (6) quantifies the Kullback divergence for the weighting parameters γ_j.

The solutions of this minimization program are:

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

where:

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

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

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 have been tested in Golan [8] (under some mild assumptions, see Golan [8], 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, [8], page 179).

Simultaneously to the estimation of the p_{ij} cells of the matrix, the DWP estimator discriminates for each column 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. Without any sample information [*i.e.*, without the set of constraints (7) the estimates of these parameters would be $\tilde{\gamma}_j = 0.5$ for each and every column. The

more informative the constrains (7), the larger the deviation with this initial value of the parameter. Note that as $\tilde{\gamma}_j \to 0$ the prior $\boldsymbol{q^a}$ gains weight for column j and the estimates approach those of the CE updating process from $\boldsymbol{Q^a}$. On the contrary, large values of $\tilde{\gamma}_j$, the CE estimation from prior $\boldsymbol{q^b}$ takes over. Consequently, relatively large values of $\tilde{\gamma}_j$ ($\tilde{\gamma}_j \geq 0.5$) will be an indication of a column j characterized by a high weight of prior $\boldsymbol{q^b}$. In other words, in this specific column j it would be preferable to use an adjustment from the a *priori* matrix $\boldsymbol{Q^b}$ rather than updating $\boldsymbol{Q^a}$. On the contrary, comparatively small values of $\tilde{\gamma}_j$ (when $\tilde{\gamma}_j < 0.5$) are a signal of an column j where the updating process should be preferred.

4. Testing the Dwp Estimation Technique with a Numerical Experiment

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.

For the sake of simplicity, let us assume that we want to estimate a symmetric matrix ($N = M$). In the experiment we have fixed a target matrix $\boldsymbol{Z}$ with dimensions 15×15 where the only known information is the column and row margins. This matrix has been fixed as the actual matrix of annual interregional trade (in millions of €) for the 15 Spanish inland regions in 2006. The matrix was constructed by the Lawrence Klein Institute (from the Autonomous University of Madrid) inside the C-Intereg project (for more details, visit http://www.c-intereg.es). Such a matrix is normally quite difficult to construct, given the huge amount of information that it requires, being available matrices only for the short period from 2002 to 2006 constructed annually. This matrix $\boldsymbol{Z}$ has been transformed into a matrix of column coefficients $\boldsymbol{P}$ to be estimated from the information contained in vectors $\boldsymbol{v}$ and $\boldsymbol{y}$. The information contained in the margin vectors of the matrix (total imports and export per region) is much more accessible, given that it can be obtained from the Regional Accounts regularly published by the Spanish Statistical Institute (see http://www.ine.es/en/inebmenu/mnu_cuentas_en.htm for more details on the Spanish Regional Accounts).

We also defined several a priori matrices $\boldsymbol{Q}$ to be used in the estimation of $\boldsymbol{P}$. Firstly, we have generated a possible *a priori* matrix $\boldsymbol{Q^a}$, being the values of this matrix obtained as $q_{ij}^a = p_{ij} \cdot u_{ij}^a$ where u_{ij}^a is a perturbation term that distributes as $u_{ij}^a \sim N(1, \sigma)$ and $\sigma = 0.1$. Note that the value of scalar σ reflects the level of deviation between the prior and the posterior: the smaller its value, the more similar the *a priori* and the target matrix.

Additionally to this *a priori* matrix, we have also generated a matrix $\boldsymbol{Q^b}$ whose elements have been obtained as:

$$q_{ij}^b = \begin{cases} q_{ij1}^b = p_{ij1} \cdot u_{ij1}^b \text{ and } u_{ij1}^b \sim N(1, 2\sigma); \ j1 = 1, \dots, 10 \\ q_{ij2}^b = p_{ij2} \cdot u_{ij2}^b \text{ and } u_{ij2}^b \sim N(1, 0.5\sigma); \ j2 = 11, \dots, 15. \end{cases}$$

In other words, this new *a priori* matrix is characterized by having ten of their columns (from column number 1 to number 10) more dissimilar to the target matrix $\boldsymbol{P}$ than the competing prior $\boldsymbol{Q^a}$; which means that in these cases it is not a very informative prior and consequently would be preferable taking $\boldsymbol{Q^a}$ as initial matrix for the adjusting process of these specific columns. However, for the remaining five columns (number 11 to 15) it happens the opposite, given that the distribution is closer to the target matrix than the prior $\boldsymbol{Q^a}$.

Under these conditions we have estimated matrix $\boldsymbol{P}$ by three different ways: updating the *a priori* matrix $\boldsymbol{Q}^a$, updating the *a priori* matrix $\boldsymbol{Q}^b$ and using the proposed DWP estimation technique that construct a composite of both matrices as possible priors. These three estimation strategies correspond respectively with the minimization of the three following divergence measures:

$$\text{Min}_{\boldsymbol{P}}\, D(\boldsymbol{P}\|\boldsymbol{Q}^a) = \sum_{i=1}^{15}\sum_{j=1}^{15} p_{ij}\, ln\left(\frac{p_{ij}}{q_{ij}^a}\right) \tag{12a}$$

$$\text{Min}_{\boldsymbol{P}}\, D(\boldsymbol{P}\|\boldsymbol{Q}^b) = \sum_{i=1}^{15}\sum_{j=1}^{15} p_{ij}\, ln\left(\frac{p_{ij}}{q_{ij}^b}\right) \tag{12b}$$

$$\text{Min}_{\boldsymbol{P},\boldsymbol{P}^\gamma}\, D\left(\boldsymbol{P},\boldsymbol{P}^\gamma\|\boldsymbol{Q}^a,\boldsymbol{Q}^b,\boldsymbol{Q}^\gamma\right) = \sum_{j=1}^{15}(1-\gamma_j)\sum_{i=1}^{15} p_{ij}\, ln\left(\frac{p_{ij}}{q_{ij}^a}\right)$$
$$+ \sum_{j=1}^{15}\gamma_j \sum_{i=1}^{15} p_{ij}\, ln\left(\frac{p_{ij}}{q_{ij}^b}\right) \tag{12c}$$
$$+ \sum_{h=1}^{H}\sum_{j=1}^{15} p_{hj}^\gamma\, ln\left(\frac{p_{hj}^\gamma}{q_{hj}^\gamma}\right)$$

subject to the same type of constraints explained before.

To evaluate the performance of these alternative estimation approaches, 1,000 trials have been carried out and we have computed the average of three measures of overall deviation between the target matrix and the estimates. Specifically, we obtained the total absolute error (TAE), the total squared error (TSE) and the total Kullback divergence (TKL), being respectively defined as:

$$TAE = \sum_{i=1}^{15}\sum_{j=1}^{15}|p_{ij}-\hat{p}_{ij}| \tag{13a}$$

$$TSE = \sum_{i=1}^{15}\sum_{j=1}^{15}\left(p_{ij}-\hat{p}_{ij}\right)^2 \tag{13b}$$

$$TKL = \sum_{i=1}^{15}\sum_{j=1}^{15}\hat{p}_{ij}\, ln\left(\frac{\hat{p}_{ij}}{p_{ij}}\right) \tag{13c}$$

where the $\hat{p}_{ij}$ elements denote the estimated probabilities under the three different approaches. Table 2 summarizes the average results obtained:

Table 2. Deviation measures between the target and estimated matrices in the numerical experiments.

Technique (prior used)	Deviation measures		
	TSE	TAE	TKL
Adjusting from $\boldsymbol{Q}^a$	0.014	0.936	0.052
Adjusting from $\boldsymbol{Q}^b$	0.037	1.425	0.142
DWP (mixture of $\boldsymbol{Q}^a$, $\boldsymbol{Q}^b$)	0.013	0.882	0.049

The deviation measures shown on Table 2 provide some interesting results. Firstly, we can see the estimation of matrix $\boldsymbol{P}$ taking $\boldsymbol{Q}^b$ as point of departure presents a comparatively worse performance than an adjustment from $\boldsymbol{Q}^a$. Not surprisingly, the comparatively more dissimilar distributions from matrix $\boldsymbol{P}$ we specified for ten out of the fifteen columns contained in $\boldsymbol{Q}^b$ cause this result. But this does not necessarily mean that all the information contained in this matrix should be neglected. Note that in the remaining columns the elements of $\boldsymbol{Q}^b$ distribute closer to the target matrix than their counterparts in the competing prior $\boldsymbol{Q}^a$. Therefore, matrix $\boldsymbol{Q}^b$ contains also valuable *a priori* information that could

be useful in the estimation problem. 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 column, which in the end obtains smaller deviation measures.
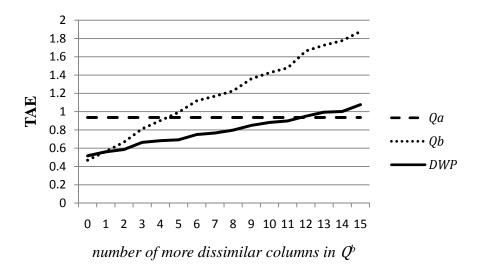
The average results obtained for the weighting parameters γ_j also show how the DWP estimation works. Under the conditions described in the experiment, the DWP estimation technique identifies the columns where the distribution of the coefficients should be taken from one specific prior of the two initial matrices considered. Without any sample information the *a priori* expected value is $\tilde{\gamma}_j = 0.5$, but the information included into the estimation process leads the DWP technique to give on average smaller weights to the first group of columns (column from 1 to 10) in matrix $\boldsymbol{Q}^b$. The weights estimated for this first group are in all the cases not larger than 0.5; which means that in this case the priors contained in $\boldsymbol{Q}^a$ take over. On the contrary, for the second group of columns (from 11 to 15) the estimates of γ_j are in all the cases equal or larger than 0.5, pointing out that for these columns prior $\boldsymbol{Q}^b$ should be preferred.

The relative performance of the DWP technique depends to a great extent on the degree of comparative similarity of the auxiliary prior $\boldsymbol{Q}^b$ with the target matrix $\boldsymbol{P}$. If the prior $\boldsymbol{Q}^b$ is closer for every column than the prior $\boldsymbol{Q}^a$, there would not be gains from using the composite prior between both because it would be always better to use as prior $\boldsymbol{Q}^b$ than the competing $\boldsymbol{Q}^a$ and it would be also preferable to any possible combination of $\boldsymbol{Q}^b$ and $\boldsymbol{Q}^a$ (unless that the estimate of γ_j for every column equals exactly one). A similar conclusion would be obtained in a symmetric case when for each column $\boldsymbol{Q}^a$ is more similar to $\boldsymbol{P}$ than $\boldsymbol{Q}^b$. It is on intermediate situations when the DWP estimator outperforms the adjustment form one single prior, given that takes the specific columns for each one that should be selected. In other words, when the number of column that behave like j1 (more dissimilar to $\boldsymbol{P}$ than their counterparts in $\boldsymbol{Q}^a$) in the numerical experiment takes intermediate values between 0 and 15. Figure 1 illustrates this idea extending the definition of the column of matrix $\boldsymbol{Q}^b$ to all the possible cases in the terms of the previous numerical simulation.

The horizontal axis of the figure contains different numbers of columns that behave like $q_{ij1}^b = p_{ij1} \cdot u_{ij1}^b$ and $u_{ij1}^b \sim N(1, 2\sigma)$ in the prior $\boldsymbol{Q}^b$. The vertical axis shows the mean of the absolute errors of the three competing adjusting technique that have been obtained along 1,000 simulations. The discontinuous line shows the average absolute deviation between the target matrix and the estimates when they are obtained by and adjustment from $\boldsymbol{Q}^a$. Obviously, this deviation does not depend on the characteristics of $\boldsymbol{Q}^b$, so it is a constant value. The dotted line represents the absolute deviations between the cells of $\boldsymbol{P}$ and the estimates obtained from prior $\boldsymbol{Q}^b$. Not surprisingly, it takes very low values when all the columns are more similar to $\boldsymbol{P}$ than the other prior $\boldsymbol{Q}^a$ and it grows as long as the number of more dissimilar columns also increases.

The solid line represents the deviation measures for the estimates obtained by the DWP technique. When the prior $\boldsymbol{Q}^b$ has very few columns (less than two) more dissimilar to $\boldsymbol{P}$ than $\boldsymbol{Q}^a$, the estimation with the DWP yields worse result than an adjustment from $\boldsymbol{Q}^b$. Conversely, if $\boldsymbol{Q}^b$ has many columns (twelve or more) more dissimilar to $\boldsymbol{P}$ than $\boldsymbol{Q}^a$, although the DWP technique outperforms a CE estimation from $\boldsymbol{Q}^b$, it yields comparative higher deviations than and adjustment from $\boldsymbol{Q}^a$. It is on intermediate situations where the DWP approach obtains better results than the CE estimation from only one of the priors, given that in such situations taking a composite of both priors allows for choosing the most valuable information contained in each one.

Figure 1. Absolute deviations between the target and estimates matrices under different levels of similarity between $\boldsymbol{P}$ and $\boldsymbol{Q}^b$.

number of more dissimilar columns in Q^b

5. An Empirical Application: Estimating the Interregional Trade Matrix in Spain, 2006

As a complement to the numerical simulation made in the previous section, this section presents an empirical application of the DWP technique and compares the results obtained with other competing techniques. For this purpose, we took again the matrix of interregional trade for fifteen Spanish regions in 2006. Let us assume that the only known information of this matrix are the row and column margins (total sales and purchases per region respectively), and from this partial information we want to estimate the inter-industry flows matrix $\boldsymbol{Z}$. For this purpose, we will apply an adjusting process to obtain the column-coefficients matrix $\boldsymbol{P}$ from two different initial matrices. Moreover, we also assume that we have some information on the expected structure of this matrix, obtained from the observed matrices of interregional trade column coefficients for two consecutive years in the past, specifically 2004 and 2005 ($\boldsymbol{Q}^{04}$ and $\boldsymbol{Q}^{05}$ respectively).

Although intuitively one could think that taking the closest matrix (*i.e.*, the 2005 matrix) as prior would be always preferable, it might also happen that for some uncontrolled reasons the structure of interregional trade in 2006 was more similar to a more distant *a priori* matrix in time. For example, it could happen that some unnoticed phenomenon happened in 2005 that disrupted the usual configuration of the matrix for some of the columns – for example, a massive strike in one region, a natural disaster that spoils the crops in a specific place, *etc.* - but the normality returns in 2006. In such a case, taking all the columns of this matrix as point of departure could be not the best option, and alternative previous matrix could be preferable as initial distribution in the estimation of the coefficients for some regions.

Consequently, we include as alternative prior in the estimation problem the matrix for 2004 and again we compare the performance of three estimation techniques of the target matrix $\boldsymbol{P}$: an adjustment considering from $\boldsymbol{Q}^{04}$, from the prior $\boldsymbol{Q}^{05}$ and the DWP estimator that takes both possible priors. Table 3 summarizes the results obtained in this study case, applying the same criteria as before for comparing the estimated and the target matrix:

Table 3. Deviation measures between the target and estimated matrices in the empirical application.

Technique (prior used)	Deviation measures		
	TSE	TAE	TKL
Adjusting from Q^{04}	0.063	2.424	1.096
Adjusting from Q^{05}	0.083	2.428	0.783
DWP (mixture of Q^{04}, Q^{05})	0.047	1.999	0.672

In the estimation problem studied here, the results seem to be in line with the results obtained in the numerical experiment. Firstly, we can see that there is not an *a priori* matrix always superior to the other, but it depends on the deviation criterion applied: an adjustment from Q^{05} would be preferred to an estimation from Q^{04} under the Kulback divergence, but the opposite decision would be concluded if we pay attention to the squared or absolute divergence criterion. However, the results obtained applying the DWP estimation yields the smallest deviations under the three. The comparatively better performance of the DWP estimation is a consequence of a situation like the explained above: matrix Q^{04} contains *a priori* information that should be preferred to the prior contained in Q^{05} for some regions, but the contrary situation happens in the remaining cases. In such a case, if we incorporate both matrices by using the DWP estimation, we let the data speak for themselves and choose the most appropriate prior for each column, which in the end obtains smaller deviation measures (the use of the DWP estimator does not imply a significant difference in terms of time cost. Using an average personal computer and the CONOPT solver of the GAMS 2.0 software, the estimations made in this section took less than five seconds).

The estimates of the parameters γ_j in this empirical application measures the respective weights given to the columns of prior Q^{05}, and their graphical representation can be useful to understand how the DWP discriminates between the two priors in this problem:

Figure 2 shows how the DWP estimator discriminates between the two priors: for all the regions (columns) in the matrix the initial guess of parameter γ_j is 0.5. From this *a priori* value, the technique manages to identify some regions where the prior contained in the respective column of the matrix Q^{05} (region 4, Cantabria; region 9, Extremadura; and region 11, Madrid) where this prior is more strongly weighted in the estimation process. Oppositely, there are other regions where the *a priori* matrix Q^{04} is clearly preferred and the estimated value of the corresponding weighting parameter γ_j is smaller than 0.5 (region 3, Asturias; and region 15, La Rioja). This discrimination between priors allows for optimizing the use of the two priors and yields smaller deviations between the estimated matrix and the actual values of interregional trade.
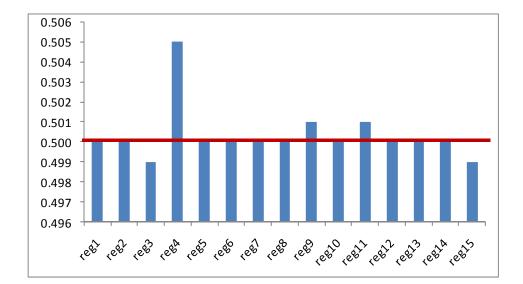
Figure 2. Estimates of the weighting parameters γ_j .

6. Conclusions

Economic analysis and modeling often requires using some non-survey method for estimating matrices of economic flows. Traditionally, these techniques take an initial matrix that is considered somehow similar to the one to be estimated. This *a priori* matrix is adjusted until it fulfills the constraints imposed by the known information and at the same time minimizes some divergence criterion with respect to the initial values. The Cross Entropy technique is a well-known example of such a procedure, when the divergence measure used is the Kullback-Leibler divergence.

This paper suggests a new approach of dealing with this initial information. Based on previous work by Golan, the so-called DWP estimation strategy considers the possibility of including several *a priori* matrices in the estimation process of the cells of an unknown matrix. By means of a Monte Carlo simulation, the performance of the proposed DWP method is compared with the CE technique when only one prior is considered. The findings of this experiment highlights this proposed technique as a useful tool in situations where we have several possible *a priori* matrices and none of them is preferable to the other for all the cases (columns). The empirical application with a real-world example, where a matrix of interregional trade for the Spanish region in 2006 is estimated, seems to confirm this conclusion.

References and Notes

1.   Kullback, J. *Information Theory and Statistics*; Wiley: New York, NY, USA, 1959.
2.   Golan, A.; Judge, G.; Miller, D. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*; John Wiley & Sons: New York, NY, USA, 1996.
3.   Kapur, J.N.; Kesavan, H.K. *Entropy Optimization Principles with Applications*; Academic Press: New York, NY, USA, 1992.
4.   Golan, A. Information and entropy econometrics—A review and synthesis. *FnT. in Econometrics* **2006**, *2*, 1–145.

5. Hewings, G.J.D. The role of prior information in updating input-output models. *Socio-Econ. Plann. Sci.* **1984**, *18*, 319–339

6. Zellner A. Models, prior information and Bayesian analysis. *J. Econometrics* **1996**, *75*, 51–68.

7. Zellner A. The Bayesian method of moments (BMOM): Theory and applications. *Adv. Econom.* **1997**, *12*, 85–105.

8. Golan A. A simultaneous estimation and variable selection rule. *J. Econometrics* **2001**, *101*, 165–193.

9. Bernardini R.A Composite generalized cross entropy formulation in small samples estimation. *Econometric Rev.* **2008**, *27*, 596–609.