MATRIX ADJUSTMENT WITH NON RELIABLE MARGINS: A GENERALIZED CROSS ENTROPY APPROACH

Esteban Fernandez Vazquez[1a*], Johannes Bröcker[b] and Geoffrey J.D. Hewings[c].

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

[b]Institute for Regional Research and Department of Economics. Christian-Albrechts-Universität zu Kiel, Olshausenstr 40, 24098 Kiel Germany. broecker@economics.uni-kiel.de

[c]University of Illinois at Urbana-Champaign, Regional Economics Application Laboratory, 607 S. Matthews, Urbana, IL 61801, U.S. hewings@illinois.edu

[*] Corresponding author.

ABSTRACT

When survey input-output tables are not available, some non-survey technique is applied to estimate the target matrix. From known information about the row and column margins, the cells of the matrix are estimated using as a priori information other table supposedly similar to the target matrix. The adjustment process, however, usually lies on the assumptions that we have perfect knowledge on the row and column margins of the target matrix, which could be considered as a non-realistic supposition. This paper explores the possibilities of changing this assumption and proposes alternatives matrix adjustments that can base on non-reliable margins.

1. Introduction

The process of adjusting an IO table bases on an a priori matrix and some piece of information of the target matrix. The solution of the estimation process is a matrix that diverges least with respect to the prior and is consistent with the aggregate information observed for the target. These aggregate data on row and column margins of the target are usually assumed as perfectly known. Note that this implies assuming that, even when the cells of the target are unknown, their sums are perfectly observable. Given that this assumption can be unrealistic, some adjustment techniques account for introducing non reliable margins in the estimation process. For example, in Golan and Vogel (2000) or Robinson et al. (2001) this possibility is considered when estimating a SAM by including an error term in the constraints of a Generalized Cross Entropy (GCE) adjustment.

Basing on this idea, this paper suggests an alternative approach for introducing different levels of reliability of the aggregate information we include in the adjustment. Usually, the way for allowing different levels of error in GCE estimation is by means of changes in the tolerable bounds. However, this paper proposes paying attention to the a priori probability distribution assumed for the values of the error. For any given bounds for the random noise present in the observed margins of the target, the specification of a specific prior distribution implies fixing a certain degree of reliability on these aggregate data. Next section describes the basic problem of matrix adjustment basing on the Cross Entropy method and section 3 generalizes it and allows for introducing noisy margins in the estimation. In section 4, a numerical experiment is carried out in order to compare the performance of different alternatives under several levels of noise in the observed aggregates.

2. The general matrix adjustment 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 using the information that is contained in the aggregate data of the row and column sums.

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

x_{11}	…	x_{1j}	…	x_{1N}	$R_{1.}$
…		…		…	…
x_{i1}	…	x_{ij}	…	x_{iN}	$R_{i.}$
…		…		…	…
x_{N1}	…	x_{Nj}	…	x_{NN}	$R_{N.}$
$C_{.1}$	…	$C_{.j}$	…	$C_{.N}$	

The x_{ij} cells of the matrix are the unknown quantities we would like to estimate (shaded in grey), where the row and column aggregates ($R_{i.}$ and $C_{.j}$ respectively) are known.

Consequently, the followings equalities are fulfilled by the x_{ij} elements:

$$\sum_{j=1}^{N} x_{ij} = R_{i.} \tag{1}$$

$$\sum_{i=1}^{N} x_{ij} = C_{.j} \tag{2}$$

These two sets of equations reflect all we know about the elements of matrix $\boldsymbol{X}$. Therefore, we have only $2N$ pieces of information to estimate the N^2 elements of $\boldsymbol{X}$. The solution to this type of problems can be obtained by minimizing a divergence measure with a prior probability matrix $\boldsymbol{B}$ subject to the set of constraints (1) and (2). If the specific divergence criterion

applied is the Kullback-Leibler measure a Cross-Entropy (CE) problem is posed, which can be written in the following terms:

$$\underset{\boldsymbol{X}}{\text{Min}}\, D(\boldsymbol{X}\|\boldsymbol{B}) = \sum_{i=1}^{N}\sum_{j=1}^{N} x_{ij}\, ln\left(\frac{x_{ij}}{b_{ij}}\right) \tag{3}$$

Subject to the restrictions given by the set of equations (1) and (2), which is equivalent to a RAS adjustment of the prior $\boldsymbol{B}$.

This general program will be modified in order to consider the possibility that the aggregate information contained in the margins do not correspond exactly with the margins of the target matrix $\boldsymbol{X}$.

3. A flexible CE estimation with non-reliable margins

The above-sketched procedure is widely used to estimate IO tables because aggregate information on the margins of the target matrix is available generally earlier than its individual cells. Usually it is supposed that we have perfectly reliable information about $\boldsymbol{X}$ in the margins $\boldsymbol{R}$ and $\boldsymbol{C}$, which can be considered as an unrealistic assumption.

Suppose that we observe row and column margins as $\widetilde{\boldsymbol{R}}$ and $\widetilde{\boldsymbol{C}}$, where:

$$\tilde{R}_{i.} = R_{i.} + \varepsilon_i;\ \forall i \tag{4}$$

$$\tilde{C}_{.j} = C_{.j} + \epsilon_j;\ \forall j \tag{5}$$

Where ε_i and ϵ_j are random errors that make the observed margins diverge from the real margins of the target matrix. In this situation is still possible to adjust our prior $\boldsymbol{B}$ with row and column margins not perfectly reliable by means of a Generalized Cross Entropy approach (GCE), following a similar approach to the ideas suggested in Golan and Vogel (2000) or Robinson et al. (2001).

The basic idea is to re-parameterize the errors ε_i and ϵ_j in terms of unknown probability distributions. The uncertainty about the realizations of these errors is introduced in the problem by considering each element ε_i and ϵ_j as discrete random variables with $K \geq 2$ possible outcomes (for the sake of simplicity K is assumed common for both). These values will be contained in two convex sets $\boldsymbol{v'} = \{v_1, \dots, 0, \dots, v_K\}$ and $\boldsymbol{u'} = \{u_1, \dots, 0, \dots, u_K\}$ respectively. We also assume that these possible realizations are symmetric ($-v_1 = v_K$; $-u_1 = u_K$) and centered on zero. The unknown probability distributions for the support vectors will be denoted as $\boldsymbol{w_\varepsilon}$ and $\boldsymbol{w_\epsilon}$ and, consequently, the random errors are defined as:

$$\varepsilon_i = \boldsymbol{v'}\boldsymbol{w_{\varepsilon i}} = \sum_{k=1}^{K} w_{\varepsilon ik} v_k \; ; \; \forall i$$

$$\epsilon_j = \boldsymbol{u'}\boldsymbol{w_{\epsilon j}} = \sum_{k=1}^{K} w_{\epsilon jk} u_k \; ; \; \forall j$$

Consequently, the GCE problem can be written in the following terms:

$$\underset{\boldsymbol{X},\boldsymbol{w_\varepsilon},\boldsymbol{w_\epsilon}}{\text{Min}} \; D\left(\boldsymbol{X}, \boldsymbol{w_\varepsilon}, \boldsymbol{w_\epsilon} \middle\| \boldsymbol{B}, \boldsymbol{w_\varepsilon^0}, \boldsymbol{w_\epsilon^0}\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} x_{ij} \, ln\left(\frac{x_{ij}}{b_{ij}}\right) + \sum_{i=1}^{N}\sum_{k=1}^{K} w_{\varepsilon ik} \, ln\left(\frac{w_{\varepsilon ik}}{w_{\varepsilon ik}^0}\right) + \sum_{j=1}^{N}\sum_{k=1}^{K} w_{\epsilon jk} \, ln\left(\frac{w_{\epsilon jk}}{w_{\epsilon jk}^0}\right) \tag{6a}$$

subject to:

$$\sum_{j=1}^{N} x_{ij} = R_{i\cdot} + \sum_{k=1}^{K} w_{\varepsilon ik} v_k \; ; \; \forall i \tag{6b}$$

$$\sum_{i=1}^{N} x_{ij} = C_{\cdot j} + \sum_{k=1}^{K} w_{\epsilon jk} u_k \; ; \; \forall j \tag{6c}$$

$$\sum_{k=1}^{K} w_{\varepsilon ik} = 1; \; \forall i \tag{6d}$$

$$\sum_{k=1}^{K} w_{\epsilon jk} = 1;\ \forall j \qquad (6e)$$

Note that both the bounds specified in the support vectors as well as the a priori probability distributions ($\boldsymbol{w}_{\boldsymbol{\varepsilon}}^{\boldsymbol{0}}$ and $\boldsymbol{w}_{\boldsymbol{\epsilon}}^{\boldsymbol{0}}$) reflect our assumptions on the way the errors are affecting the observed margins. Larger bounds in $\boldsymbol{v}$ and $\boldsymbol{u}$ would allow, obviously, for larger errors. In the context of GCE problems, the values of the supporting vectors for the errors are usually fixed following the three-sigma rule (Pukelsheim, 1994), which in this case implies to take as upper and lower bound $\pm$ three times the standard deviation of $\widetilde{\boldsymbol{R}}$ and $\widetilde{\boldsymbol{C}}$ respectively.

The a priori probabilities assumed for these bounds also play a role. Once the supports are fixed, the larger the a priori probability we assign to the central point, the greater our confidence on the observed margins. Note that if, for example, we fix $w_{\varepsilon ik}^{0} \simeq 1$ for the central value in the support $v_k = 0$, we are strongly assuming that $\tilde{R}_{i\cdot} = R_{i\cdot}$ and we impose a high penalty for diverging from this prior. In the other hand, a uniform distribution still imply assuming $\tilde{R}_{i\cdot} = R_{i\cdot}$ as the point of departure, but the penalty for obtaining solutions far from this prior is smaller than in the previous case.[1]

4. A numerical experiment

We will compare the performance of a CE estimation that does not consider the possibility of measurement errors in the margins with a GCE estimation under two alternative a priori distributions for the errors: a uniform one and a spike distribution with the probability mass concentrated in zero. For the sake of simplicity, only three points ($K = 3$) are included in the support

[1] See Appendix for details.

vectors of the errors, which have been fixed using the three sigma rule and being always the central point equal to zero. [2]

The objective of the three estimators will be to obtain the target matrix $\boldsymbol{X}$ from the prior matrix $\boldsymbol{B}$ and the noisy margins $\widetilde{\boldsymbol{R}}$ and $\widetilde{\boldsymbol{C}}$ under different levels of noise. For this purpose, we have taken as target matrix the inter-industry matrix of flows for the US in 2002 being the prior the IO table for the US in 1997, both classified in seven industries. The tables used in the experiment are available online as supplementary resources of Miller and Blair (2009) handbook.[3]

In order to introduce errors in the observed margins, the elements ε_i and ϵ_j of equations (3) and (4) are generated as:

$$\varepsilon_i \sim N[0, \alpha];\ \forall i$$
$$\epsilon_j \sim N[0, \alpha];\ \forall j$$

Being α a scalar specified as certain proportion of the average value of the cells in matrix $\boldsymbol{X}$. Obviously, the bigger the value of α, the more noise is introduced into the observed margins. In the generation of the noisy margins, additional constraints are imposed in order to assure that not negative values are generated and that the equality $\sum_{i=1}^{N} \tilde{R}_{i\cdot} = \sum_{j=1}^{N} \tilde{C}_{\cdot j}$ is hold. In this scenario, matrix $\boldsymbol{X}$ is estimated by the three alternative methods mentioned under several values of noise α. For each value of α used in the experiment, 100 simulations are carried out.

[2] More specifically, a probability of 99% has been assigned to zero, and the remaining 1% is split in two equal parts between the lower and upper bounds.

[3] See http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=9780521517133

In order to evaluate the relative performance of the three methods, the mean absolute percentage errors (MAPE) with respect to the real values are obtained for the technical coefficients and the elements of the Leontief inverse. Additionally, in each simulation one vector of final demand (y_i^s) is generated from the actual vector of final demand for the US IO table of 2002 (y_i^*) as $y_i^s = (y_i^*)(\mu_i)$, where μ_i is a stochastic term that distributes normally with mean one and a standard deviation set to 0.05. By multiplying y_i^s by the respective estimate of the Leontief inverse we have obtained the vector of estimated output by industry under the three estimation techniques studied here. The estimates have been compared in each trial of the simulation with the output obtained by multiplying y_i^s by the actual Leontief inverse and we have obtained the average absolute error as a proportion of the total actual output in the US IO table for 2002. Table 1 reports the results obtained for the three methods under consideration:

Table 1: Mean absolute errors for technical coefficients, Leontief inverse and simulated output

Matrix A (technical coefficients)									
α	**0**	**0.2**	**0.4**	**0.6**	**0.8**	**1**	**1.2**	**1.4**	**1.6**
No error	0.110	0.167	0.243	0.314	0.403	0.495	0.562	0.686	0.723
Uniform	0.241	0.242	0.245	0.254	0.265	0.279	0.297	0.324	0.342
Spike	0.112	0.161	0.230	0.296	0.366	0.436	0.504	0.577	0.649
Matrix L (Leontief inverse)									
α	**0**	**0.2**	**0.4**	**0.6**	**0.8**	**1**	**1.2**	**1.4**	**1.6**
No error	0.073	0.154	0.261	0.370	0.641	0.708	1.098	1.872	2.715
Uniform	0.294	0.297	0.304	0.320	0.343	0.373	0.414	0.465	0.533
Spike	0.077	0.148	0.247	0.351	0.503	0.668	0.975	1.403	1.551
Simulated output									
α	**0**	**0.2**	**0.4**	**0.6**	**0.8**	**1**	**1.2**	**1.4**	**1.6**
No error	0.001	0.013	0.028	0.040	0.054	0.068	0.090	0.169	0.280
Uniform	0.025	0.026	0.026	0.032	0.037	0.042	0.048	0.056	0.065
Spike	0.001	0.013	0.026	0.039	0.052	0.065	0.084	0.131	0.163

The results indicate, not surprisingly, that only if the observed margins equal the actual margins of the target matrix ($\alpha = 0$), it is preferable considering and adjustment that does not introduce any error. When some error is present in the aggregates we observe, a GCE estimation that allows for discrepancies between observed and actual margins yield lower deviation measures: the larger the size of the error, the better its comparative performance.

Interestingly enough, even if we opt for a GCE adjustment that includes an error term and we fix specific boundaries for it, still there is room for introducing a priori expectations about the reliability of the margins. If we do not have much confidence in the reliability of the margins we observe, we can specify a uniform distribution for the error values (second rows in each part of the table). This would yield lower error measures with respect to the real values when the observed margins diverge largely with the actual ones, but behaves comparatively worse than an adjustment without error if the size of the noise is reduced ($\alpha \leq 0.4$).

However, it is still possible to work under the assumption that the margins are reliable by imposing a spike prior distribution for the errors. Keeping the same error bounds[4], but assigning much probability to the central point, we allow for the presence of some noise in the margins although we a priori assume that the values on the bounds are improbable. Doing that, we obtain comparatively worse results than a GCE estimator with a uniform a priori distribution under large levels of noise ($\alpha \geq 0.6$). However, this adjustment seems to yield lower deviations than a pure CE estimation without error except in the case that the margins are perfectly reliable.

[4] Note that the bounds are fixed by the three sigma rule in any case.

In other words, the GCE estimation with a spike a priori distribution with the probability mass on zero can be seen as intermediate solution: it is not necessary to assume perfect knowledge on the margins although the possibility of large (positive or negative) errors is supposed very low. Obviously, it also allows for introducing different levels of reliability depending on the industry, for example, simply by changing the a priori probability we assign to the central point.

References

Bera, A.K. and S. Y. Park (2008): Optimal portfolio diversification using the Maximum Entropy principle, Econometric Reviews, 27: 484 — 512

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.

Golan, A. and S. J. Vogel (2000): Estimation of non-stationary social accounting matrix coefficients with supply-side information, Economic Systems Research, 12: 447-471.

Robison, S., Cattaneo, A. and M. El-Said (2001): Updating and estimating a social accounting matrix using cross entropy methods. Economic Systems Research 13: 47-64.

Appendix: CE penalty function with three support points for the error

The penalty function defined for deviating from the a priori expected value of the error term can be obtained in the following way. Without loss of generalization, we will focus on one of the two types of errors (for the row margins), that in a case with a three-point support $\boldsymbol{v}' = (v_1, v_2, v_3)$ centered on zero ($v_2 = 0$) and symmetric ($v_1 = -v_3$) can be written as:

$$\varepsilon_i = \boldsymbol{v}'\boldsymbol{w}_{\boldsymbol{\varepsilon i}} = v_1 w_1 + v_3 w_3 = -vw_1 + vw_3 \tag{A.1}$$

Note that the GCE formulation specifies for this type of errors the following divergence measure:

$$\begin{aligned} Pen &= w_1 ln\left(\frac{w_1}{w_1^0}\right) + w_2 ln\left(\frac{w_2}{w_2^0}\right) + w_3 ln\left(\frac{w_3}{w_3^0}\right) \\ &= w_1 ln\left(\frac{w_1}{w_1^0}\right) + [1 - w_1 - w_3] ln\left(\frac{[1 - w_1 - w_3]}{[1 - w_1^0 - w_3^0]}\right) + w_3 ln\left(\frac{w_3}{w_3^0}\right) \end{aligned} \tag{A.2}$$

Let us denote the negative component of each error ($-vw_1$) as a and the positive part (vw_3) as b. So, (A.2) can be rewritten as:

$$Pen = \frac{-a}{v} ln\left(\frac{1}{w_1^0}\frac{-a}{v}\right) + \left[\frac{v + a - b}{v}\right] ln\left(\frac{v + a - b}{v[1 - w_1^0 - w_3^0]}\right) + \frac{b}{v} ln\left(\frac{1}{w_3^0}\frac{b}{v}\right) \tag{A.3}$$

This penalty function depends on the divergence from the center of the support vector (a and b) and on the initial probabilities assigned to each value of the support vector (w_1^0, w_2^0, w_3^0). If an a priori uniform distribution ($w_1^0 = w_2^0 = w_3^0 = \frac{1}{3}$) is assumed, (A.3) turns into:

$$Pen = \frac{-a}{v} ln\left(\frac{-3a}{v}\right) + \left[\frac{v + a - b}{v}\right] ln\left(\frac{3[v + a - b]}{v}\right) + \frac{b}{v} ln\left(\frac{3b}{v}\right) \tag{A.3}$$

More generally, for any a priori symmetric distribution we can define a parameter like $\mu = \frac{1}{w_1^0} = \frac{1}{w_3^0}$. Therefore, $\frac{1}{w_2^0} = \frac{\mu}{\mu-2}$, given that $w_2^0 = 1 - w_1^0 - w_3^0$. Equation (A.3) can be then rewritten in terms of μ as:

$$Pen = \frac{-a}{v} ln\left(\frac{-\mu a}{v}\right) + \left[\frac{v+a-b}{v}\right] ln\left(\frac{\mu[v+a-b]}{[\mu-2]v}\right) + \frac{b}{v} ln\left(\frac{\mu b}{v}\right) \quad (A.4)$$

Deriving this penalty function with respect to μ we obtain:

$$\frac{\partial Pen}{\partial \mu} = \frac{1}{\mu} - \frac{v+a-b}{v}\frac{1}{\mu-2} \quad (A.5)$$

(A.5) can be written in terms of the posterior weight for the central point as:

$$\frac{\partial Pen}{\partial \mu} = \frac{1}{\mu} - \frac{w_2}{\mu-2} \quad (A.6)$$

Which is positive if the condition $w_2^0 > w_2$ holds. As the CE formulation is a form of shrinkage estimator (see Golan et al., 1996, p.31; Bera and Park, 2008, p. 491), a smaller weight will be assigned to the values associated with the largest prior probabilities. This means that this condition is fulfilled and that the penalty function increases in μ. In other words, the penalty for any given ε_i grows if the a priori probability mass assigned to zero increases (and, consequently, the prior weights w_1^0 and w_3^0 assigned to the bounds become smaller).