MATRIX ADJUSTMENT WITH BOTH POSITIVE AND NEGATIVE ENTRIES: FROM GRAS TO GCE

Esteban Fernandez-Vazquez

RegioLab and Department of Applied Economics University of Oviedo, Spain

Abstract

The GRAS technique, proposed by Junius and Oosterhaven (2003), is the method commonly used to update or regionalize input-output matrices from a prior with positive and negative cells. This paper presents an adjustment technique based on Generalized Cross Entropy (GCE) as an alternative to GRAS. The basic idea of the proposed method is to assume each cell of the target matrix as a random variable for which we have partial information in the prior.

More specifically, 1) we assume each observation in the prior as a specific realization of the random process that generates the cells; and 2), we fix bounds for the maximum and minimum values that this random process can generate. From this information, together with the observed totals in the target matrix, the adjustment process is approached as a (constrained) minimization problem of the Kullback-Leibler divergence. This technique allows for potential changes in the sign of the cells, which can be something desirable in situations where sign-prevention in all the cases is too restrictive. We evaluate the performance of the proposed technique by means of numerical simulation and illustrate how can be applied with an empirical application.

1. Introduction

The general solution to the problem of adjusting a target IO table from a prior matrix is a new matrix that diverges least with respect to the prior and is consistent with the aggregate information observed for the target. The well-known biproportional RAS adjustment lies within this general problem and is the most frequently applied technique if all the cells in the matrix are positive. Adjusting a matrix with both positive and negative entries, however, implies some practical problems for RAS. If a RAS adjustment is applied to a matrix that contains negative cells, easily leads to a solution that may largely deviate from the structure of the prior matrix. Junius and Oosterhaven (2003) proposed the so-called generalized RAS (GRAS) as an alternative adjustment for such situations. GRAS is a signpreserving technique for adjusting a matrix, and it can be applied directly to positive and negative cells. This technique, as well as several other signpreserving adjusting techniques, defines an objective function written in terms of absolute values with respect to the matrix-entries.

In this paper we propose applying Generalized Cross-Entropy (GCE) estimation for this type of problems. The main advantage of the proposed technique is that it introduces more flexibility in the adjustment and allows for potential changes in the sign of the cells if we consider that sign preservation for all the cells could be too strict. In other words, the technique suggested here can make a change in the sign improbable but not absolutely impossible. The paper is divided into five additional sections. Section two presents the basic formulation of the proposed technique, whereas section three shows its solution. Section four compares the performance of the GCE estimation with other adjustment that are sign-preserving, GRAS included, by means of a numerical simulation. An empirical application is conducted in section five. Finally, section six presents the main conclusions and finishes the paper.

2. Formulation

Consider a prior matrix A with cells a_{ij} to be adjusted to a target matrix X with unknown cells x_{ij} , but with observable row and column totals u and v respectively. The traditional GRAS problem is to find the matrix X that deviates least from A and is consistent with the row and columns margins.

Junius and Oosterhaven (2003) formulated their proposed solution as a variant of the traditional RAS problem, but allowing for the presence of both positive and negative entries. Lenzen et al. (2007) suggested some modifications in the target function in order to account for the distance between the initial and the target matrix. The formulation proposed in Lenzen et al. (2007) is:

$$z_{ij} = \arg \min \sum_{i=1}^{T} \sum_{j=1}^{K} \left| a_{ij} \right| z_{ij} ln\left(\frac{z_{ij}}{e}\right) \qquad \text{being } z_{ij} = x_{ij}/a_{ij} \qquad (1)$$

And e is the base of the natural logarithm. This minimization is subject to the row and column constrains:

$$\sum_{i=1}^{T} a_{ij} z_{ij} = \sum_{i=1}^{T} x_{ij} = v_j;$$

$$\sum_{j=1}^{K} a_{ij} z_{ij} = \sum_{j=1}^{K} x_{ij} = u_i$$
(2)

In this paper an alternative updating method applicable for matrices with both positive and negative cells is proposed. The proposed technique can be seen as an extension of the paper by Golan et al. (1994). In that paper, a Generalized Cross Entropy (GCE) procedure was proposed to recover intersectoral information from incomplete data. The context for applying this idea was, however, somewhat restricted since it only considered matrices of coefficients (bounded between 0 and 1). In this article this method is extended to cases where the entries of the target matrix are flows instead of coefficients and can contain both positive (larger than 1) and negative cells.

The point of departure is considering each element of the prior and target matrices, A and X, as realizations of a random variable that can take a

range of M possible values which are contained in a vector $\mathbf{b'}_{ij} = [b_{ij1}, \dots, b^*_{ij}, \dots, b_{ijM}]$ with values that are set exogenously. Each support vector \mathbf{b}_{ij} can be different for every cell and contains an odd number of values that are centered on point b^*_{ij} symmetrically. The entries in the prior matrix determine the central points b^*_{ij} for each vector. More specifically, each cell a_{ij} in the initial matrix \mathbf{A} is assumed to be this particular point of its corresponding vector $(a_{ij} = b^*_{ij})$, although any of the other points contained in \mathbf{b}_{ij} could have been observed instead. These other values for each vector are specified arbitrarily by the researcher, depending on our beliefs about how much it is possible to deviate from b^*_{ij} .

For the sake of simplicity, let us illustrate this idea by considering the simplest case whit M = 3. In this situation, the support vector would be defined as $\mathbf{b'}_{ij} = [(1-r)a_{ij}, a_{ij}, (1+r)a_{ij}] = [b_{ij1}, b_{ij}^*, b_{ij2}]$. The scalar r represents a rate of variation imposed by the researcher with respect to a_{ij} , which determines the minimum and maximum value assumed as possible for this cell. Note that if we set any $|r| \leq 1$, we prevent the possibility that this element could change its sign from positive to negative or vice versa, but this sign-preserving character can be removed just by setting a scalar |r| > 1.¹

Once the possible realizations for each entry in the matrices have been specified, given that we assume that they are generated by a random process, some probability distribution should be assigned to them. Although the support vectors for the cells of A and X are common, the distribution probabilities are different. In the case of A, these probabilities are set a priori by the researcher, but they are unknown for our target matrix X.

¹ We assume here that this scalar is common to all the cells in the matrix, but this assumption can be relaxed easily.

Starting with the elements of matrix A, we need to specify a probability distribution as $q'_{ij} = [q_{ij1}, ..., q^*_{ij}, ..., q_{ijM}]$ for each element a_{ij} . Continuing with the simplest case with M = 3, one natural way of doing this is assuming that all the values are equally probable and setting $q_{ij1} = q^*_{ij} = q_{ij2} = 1/3$. This solution implies giving to the value actually observed (b^*_{ij}) the same probability as to the extreme cases b_{ij1} and b_{ij2} . An alternative could be to assign an arbitrarily high probability q^*_{ij} to it and assuming that the two extreme cases are equally probable to each other. Whatever the specific probabilities chosen, the general rule $q_{ij1} = q_{ij2} = (1 - q^*_{ij})/2$ guarantees that:

$$a_{ij} = \sum_{m=1}^{M} q_{ijm} b_{ijm} \tag{3}$$

We apply the same reasoning with the elements of the target matrix X, but now the probability distributions $p'_{ij} = [p_{ij1}, p_{ij2}, ..., p_{ijM}]$ are unknown. The value of each cell of this matrix is given by the expression:

$$x_{ij} = \sum_{m=1}^{M} p_{ijm} b_{ijm} \tag{4}$$

In this framework of analysis, the original problem of adjusting matrix X from matrix A, has been transformed in a new problem where a set of posterior probabilities P are estimated from the a priori probabilities Q.

3. The GCE solution

A GCE solution is applied in order to find the solution to our generalization of the estimation problem can be posed as a minimization program like:

$$\underset{\boldsymbol{P}}{Min} D(\boldsymbol{P} \| \boldsymbol{Q}) = \sum_{m=1}^{M} \sum_{i=1}^{T} \sum_{j=1}^{K} p_{ijm} \ln\left(\frac{p_{ijm}}{q_{ijm}}\right)$$
(5)

Subject to:

$$\sum_{j=1}^{K} x_{ij} = \sum_{j=1}^{K} \left(\sum_{m=1}^{M} p_{ijm} b_{ijm} \right) = v_i; \ \forall i$$
(6)

$$\sum_{i=1}^{T} x_{ij} = \sum_{i=1}^{T} \left(\sum_{m=1}^{M} p_{ijm} b_{ijm} \right) = u_j; \forall j$$
(7)

In the original paper by Junius and Oosterhaven (2003, pp. 90-91) and in the correction proposed by Lenzen et al. (2007, pp. 464-465) proofs of the biproportionality of the solution of the GRAS algorithm are presented. In a similar fashion, this section presents the solution of the GCE program contained in equations (5)-(7) and it shows that the solution achieved are biproportional to the information contained in the prior matrix.

The Lagrangean function related to (5)-(7) is:

$$\mathcal{L} = \sum_{m=1}^{M} \sum_{i=1}^{T} \sum_{j=1}^{K} p_{ijm} \ln\left(\frac{p_{ijm}}{q_{ijm}}\right) + \sum_{j=1}^{K} \lambda_j \left[u_j - \sum_{i=1}^{T} \left(\sum_{m=1}^{M} p_{ijm} b_{ijm}\right)\right] + \sum_{i=1}^{T} \pi_i \left[v_i - \sum_{j=1}^{K} \left(\sum_{m=1}^{M} p_{ijm} b_{ijm}\right)\right]$$
(8)

With corresponding derivatives:

$$\frac{\partial \mathcal{L}}{\partial p_{ijm}} = ln \left(\frac{p_{ijm}}{q_{ijm}} \right) + 1 - \lambda_j b_{ijm} - \pi_i b_{ijm} = 0$$
(9)

Imposing the optimality conditions in (9) yields:

$$ln(p_{ijm}) = \pi_i b_{ijm} + ln(q_{ijm}) + (\lambda_j b_{ijm} - 1);$$

or
$$p_{ijm} = exp \ (\pi_i b_{ijm}) q_{ijm} exp(\lambda_j b_{ijm} - 1) = \tilde{\rho}_{ijm} q_{ijm} \tilde{\sigma}_{ijm}$$
(10)

Being $\tilde{\rho}_{ijm} = exp(\pi_i b_{ijm})$ and $\tilde{\sigma}_{ijm} = exp(\lambda_j b_{ijm} - 1)$. Note that this biproportional relationship between the prior and posterior distributions Q and P does not hold for the prior and target matrices A and X. This means that the GCE solution is not necessarily sign-preserving, but depends on the absolute value of scalar r used to set the possible values included in the support vector.

4. Evaluation by numerical simulation

In this section the suggested GCE solution will be compared with the GRAS solution by a numerical simulation under several possible scenarios. As point of departure, we have taken as initial matrix A the same used by Junius and Oosterhaven (2003) to illustrate the GRAS procedure:

<<Insert Table 1 around here>>

The target matrix to be estimated is generated in each trial of the numerical simulations modifying each cell in the cells of Table 1 by introducing some noise:

$$x_{ij} = a_{ij} \times \varepsilon_{ij}, \text{ where } \varepsilon_{ij} \sim N(1, \sigma)$$
(11)

This generation process for the cells of the target keeps the zeros in the initial matrix. Additionally, the standard deviation σ conditions the distance between the initial and the final elements. Initially we set $\sigma = 0.1$, given that with such a standard deviation the possibility of changes in the sign of the cells is virtually prevented. Additionally we try with different values of σ (specifically $\sigma = 0.2$ and 0.5) in order to consider larger differences between A and X. Note that a standard deviation as 0.1 or 0.2 virtually prevents a change in the sign of the cell, but a standard deviation in ε_{ij} as big as 0.5 allows the possibility of such a change. The row and column margins marked in bold in Table 1 are assumed as observable in the target matrices generated and incorporated as constrains to the adjustment problem.

The application of the proposed GCE technique requires setting the *M* points contained in the support vectors $\mathbf{b'}_{ij}$ and that define the possible values taken in the target cells x_{ij} . We opted for the simple case with M = 3, where $\mathbf{b'}_{ij} = [(1-r)a_{ij}, a_{ij}, (1+r)a_{ij}]$, setting different values for the scalar *r*. Specifically, we set values r = 1, 2 and 10.

The probability distributions q'_{ij} associated to each element a_{ij} are the other important point in the GCE adjustment. They implicitly reflect our beliefs about how much deviation we assume between the observed realization in the cell a_{ij} and its unknown counterpart x_{ij} . If we believe that the "extreme" values $(1-r)a_{ij}$ or $(1+r)a_{ij}$ are not probable (i.e., the x_{ij} element has to be close to the initial cell a_{ij}), we can assign a prior distribution with a mass probability in the central point and $q_{ijm} \simeq 0$ for the rest of values. If, on the contrary, we consider that the entry x_{ij} is not necessarily very close to the initial a_{ij} but it can take values across all the parameter space defined in b'_{ij} with equal probability, we can assume an uniform distribution $q_{ij1} = q_{ij2}^* = q_{ij2} = 1/3$. In the experiment we apply this uniform distribution and also a "spike" one as $q'_{ij} = [0.025, 0.95, 0.025]$.

In order to extend the comparison, we consider adjustment techniques other than GRAS. The recent papers by Huang et al. (2008), Pavia et al. (2009) or Termushoev et al. (2010) evaluated alternative adjustment procedures to the GRAS objective function (1), suggesting the three following variants:

$$z_{ij} = \arg \min \sum_{i=1}^{T} \sum_{j=1}^{K} |a_{ij}| (z_{ij} - 1)^{2} \qquad \begin{array}{c} Improved Normalized \\ Squared Differences (INSD) \end{array}$$
(12)
$$z_{ij} = \arg \min \sum_{i=1}^{T} \sum_{j=1}^{K} (a_{ij})^{2} (z_{ij} - 1)^{2} \qquad \begin{array}{c} Improved Squared \\ Differences (ISD) \end{array}$$
(13)
$$z_{ij} = \arg \min \sum_{i=1}^{T} \sum_{j=1}^{K} |a_{ij}^{3}| (z_{ij} - 1)^{2} \qquad \begin{array}{c} Improved Weighted Squared \\ Differences (IWSD) \end{array}$$
(14)

To evaluate the performance of these five estimation approaches (GCE, GRAS, INSD, ISD and IWSD), 1,000 trials have been carried out. There are several different deviation measures that can be applied to evaluate the adjustment (see Lahr 2001, appendix 3, for a survey of the possible measures). In the experiment we opted for calculating the *Weighted Absolute Percentage Error* (WAPE), which has been largely used when evaluating the performance of adjusting techniques (see Jiang et al., 2010 and 2011 for recent examples). This measure averages the percentage error giving larger weights to errors in large cells than errors in small cells (Oosterhaven et al, 2008). It is defined as:

$$WAPE = \sum_{i=1}^{T} \sum_{j=1}^{K} 100 \frac{|x_{ij} - \hat{x}_{ij}|}{\sum_{i=1}^{T} \sum_{j=1}^{K} |x_{ij}|}$$
(15)

where the \hat{x}_{ij} elements denote the estimated entries. Additionally, the socalled *Standardized weighted absolute difference (SWAD)* is calculated as well:

$$SWAD = \sum_{i=1}^{T} \sum_{j=1}^{K} \frac{|x_{ij}| \times |x_{ij} - \hat{x}_{ij}|}{\sum_{i=1}^{T} \sum_{j=1}^{K} [x_{ij}]^2}$$
(16)

which is a deviation measure similar to WAPE, but now the absolute deviations are weighted by the size of the true transactions (Lahr, 2001). Table 2 shows the results.

<<Insert Table 2 around here>>

Deviation measures in Table 2 indicate a very similar performance between GRAS and INSD, which both clearly beat ISD and IWSD under any of the three scenarios simulated. These results are similar to those reported in Temurshoev et al. where several adjusting techniques were evaluated by means of an empirical application for The Netherlands and Spain (see Tables 2, 3 and 4 in Temurshoev et al., 2011). The proposed GCE technique, however, slightly outperforms GRAS and INSD and the gains in comparatively smaller deviations become larger when scalar σ grows. This result is not surprising, given that the GCE technique is not a strictly signpreserving: it departs from the cell present in the prior matrix but allows for a potential change of sign in the corresponding posterior cell. We can make this change more or less likely by setting the a priori probabilities q. Generally speaking, the higher the probability assigned to the central point in the support vectors (q_{ij}^*) , the smaller the probability of a change in the sign of the solution. The performance of the technique seems relatively insensitive to changes in the support vectors (by changing the scalar r) or to changes in the a priori distributions set in q.

5. An empirical application: updating the intermediate and final demand tables for Spain

The numerical simulation made in the previous section is complemented in this section, which presents an empirical application of the proposed GCE technique and compares the results obtained with the GRAS adjustment. For this purpose, we take as target matrix the total (domestic and imported) product-by-industry Use tables for Spain in 2007, elaborated by the Spanish Statistical Institute (INE) and given at basic prices. The intermediate demand tables are classified into 118 commodities. After a cleaning process to remove commodities with zeros in the margins, we ended up with a classification into 102 products plus "Net taxes" (with both positive and negative entries) and 75 industries. We distinguish 3 categories for the final demand: "Total final consumption" (private and public), "Gross capital formation" (with both positive and negative entries, because it includes changes in inventories), and "Total exports".²

For the adjustment of the intermediate demand table we assume that the totals of the 2007 intermediate and final demand tables are observable and we adjust the cells of this target matrix (X) on the basis of two different prior matrices (A): the use tables for 2002 and 2006. In order to account the huge variability that could be present in the cells of the matrix between two time periods, the GCE technique is based on support vectors again with M = 3 values, as in the numerical experiment in the previous section, but now allowing for much larger extreme values by considering a scalar r = 100. In the same way as in the numerical simulation, two a priori q probability distributions (one uniform and one spike) have been considered as well. Table 3 presents the results of these adjustments:

² Data are available at <u>http://www.ine.es/en/daco/daco42/cne00/cneio2000_en.htm</u>.

<<Insert Table 3 around here>>

The results corresponding to the estimation of the intermediate demand table show that a traditional GRAS adjustment is preferable to a GCE estimation under any of the two a priori probability distributions considered. This result is partially conditioned by the fact that most of the cells of this intermediate matrix are always positive, so the sign-preserving nature of GRAS results in a comparative advantage with respect to the GCE estimation.

The projection of the final demand matrix, however, depicts a different picture. In This case GCE outperform clearly GRAS both in the adjustment from 2002 as in the one from 2006. The explanation is that in the final demand categories, the changes in sign can be more frequent (e.g., between 2006 and 2007 Gross capital formation changed from positive to negative and vice versa for seven commodities) which implies a problem for GRAS but not for GCE.

6. Concluding remarks

An adjustment technique for matrices with positive and negative cells has been proposed in this paper. The suggested GCE method has as main advantage a higher flexibility when compared with other traditional signpreserving techniques, as GRAS. Given that it requires the specification of a supporting vector containing all the possible realizations of each cell, it allows for preventing changes in the sign simply by not considering values that imply such a change. Alternatively, supporting vectors with values that change the sign of a cell can be included with an arbitrarily low a priori probability. This situation can reflect researcher's belief about the behavior of a specific entry in a matrix, where a change in its sign can be improbable but not totally impossible. The numerical experiment conducted in the paper, as well as the empirical application, suggest that the proposed GCE technique can be considered as an alternative to other adjustment techniques in situations when we cannot be completely sure about a potential change in the sign of a cell.

References

Golan, A., Judge, G. and Robinson, S. (1994): Recovering information from incomplete or partial multisectoral economic data, Review of economic and Statistics, 76, pp. 541-549.

Huang, W., S. Kobayashi and H. Tanji (2008) Updating an Input–Output Matrix with Sign-Preservation: Some Improved Objective Functions and their Solutions. Economic Systems Research, 20, pp. 111–123.

Jiang, X., Dietzenbacher, E. and Los, B. (2010) Targeting the collection of superior data for the estimation of the intermediate deliveries in regional input -output tables. Environment and Planning A 42, pp. 2508-2526.

Jiang, X., Dietzenbacher, E. and Los, B. (2011) Improved estimation of regional Input–Output tables using cross-regional methods. Regional Studies (forthcoming).

Junius, T. and Osterhaven, J. (2003) The solution of updating or regionalizing a matrix with both positive and negative entries, Economic Systems Research, 15, pp. 87–96.

Lahr, M.L. (2001) A Strategy for Producing Hybrid Regional Input–Output Tables. In: M.L. Lahr and E. Dietzenbacher (eds.) Input–Output Analysis: Frontiers and Extensions. New York, Palgrave, pp. 211–242.

Lenzen, M., Wood, R. and Gallego, B. (2007) Some comments on the GRAS method, Economic Systems Research, 19, pp. 461–465.

Temurshoev U, Webb C. and Yamano, N. (2011) Projection of supply and use tables: Methods and their empirical assessment. Economic Systems Research, 23, pp. 91–123.

	Goods	Services	Consumption	Net exports	Total output
Goods	7	3	5	-3	12
Services	2	9	8	1	20
Net Taxes	-2	0	2	1	1
Total Use	7	12	15	-1	33
Value added	5	8	0	0	13
Total input	12	15	20	-1	

Table 1. Initial matrix for the numerical simulation

Source: Junius and Oosterhaven (2003, page 94)

		$\varepsilon_{ij} \sim N(1,0.1)$	
	Technique	WAPE (%)	SWAD
	GRAS	4.03	0.004
	INSD	4.04	0.004
	ISD	5.96	0.006
	IWSD	9.42	0.009
	CGE (r = 1)	3.61	0.004
q' _{ij} = [0.333, 0.333, 0.333]	CGE (r = 2)	3.61	0.004
	$CGE \ (r=10)$	3.61	0.004
	CGE (r = 1)	3.60	0.004
q ′ _{<i>ij</i>} = [0.025, 0.95, 0.025]	CGE (r = 2)	3.60	0.004
	CGE (r = 10)	3.61	0.004
		$\varepsilon_{ij} \sim N(1,0.2)$	
	Technique	WAPE (%)	SWAD
	GRAS	8.09	0.009
	INSD	8.10	0.009
	ISD	11.92	0.012
	IWSD	18.56	0.018
	$CGE \ (r=1)$	7.22	0.008
$q'_{ij} = [0.333, 0.333, 0.333]$	CGE (r = 2)	7.22	0.008
	CGE (r = 10)	7.22	0.008
	CGE (r = 1)	7.21	0.008
$q'_{ij} = [0.025, 0.95, 0.025]$	CGE (r = 2)	7.20	0.008
	CGE (r = 10)	7.21	0.008
		$\varepsilon_{ij} \sim N(1, 0.5)$	
	Technique	WAPE (%)	SWAD
	GRAS	21.04	0.022
	INSD	20.39	0.022
	ISD	28.33	0.029
	IWSD	40.32	0.041
	$CGE \ (r=1)$	18.65	0.020
$q'_{ij} = [0.333, 0.333, 0.333]$	CGE (r = 2)	18.20	0.019
	$CGE \ (r=10)$	18.18	0.019
	$CGE \ (r=1)$	18.32	0.019
q' _{ij} = [0.025, 0.95, 0.025]	CGE(r=2)	18.16	0.019
	$CGE \ (r=10)$	18.14	0.019
	· · ·		

Table 2. Deviations between target and estimates in the numericalsimulation (1,000 trials)

	Intermediate use table (from 2002 to 2007)			
	Technique	WAPE (%)	SWAD	
	GRAS	14.95	0.0033	
q ' _{ij} = [0.333, 0.333, 0.333]	CGE (r = 100)	20.76	0.0050	
$q'_{ij} = [0.025, 0.95, 0.025]$	$CGE \ (r=100)$	20.83	0.0051	
	Final demand table (from 2002 to 2007)			
	Technique	WAPE (%)	SWAD	
	GRAS	26.00	0.0660	
q ' _{ij} = [0.333, 0.333, 0.333]	CGE (r = 100)	9.84	0.0105	
$q'_{ij} = [0.025, 0.95, 0.025]$	CGE (r = 100)	9.85	0.0106	
	Intermediate use table (from 2006 to 2007)			
	Technique	WAPE (%)	SWAD	
	GRAS	4.11	0.0013	
$q'_{ij} = [0.333, 0.333, 0.333]$	$\mathrm{CGE}\;(r=100)$	6.17	0.0015	
$q'_{ij} = [0.025, 0.95, 0.025]$	$\mathrm{CGE}\;(r=100)$	6.16	0.0014	
	Final demand table (from 2006 to 2007)			
	Technique	WAPE (%)	SWAD	
	GRAS	7.34	0.0130	
$q'_{ij} = [0.333, 0.333, 0.333]$	$\mathrm{CGE}\;(r=100)$	3.30	0.0025	
$q'_{ij} = [0.025, 0.95, 0.025]$	$\mathrm{CGE}\;(r=100)$	3.30	0.0024	

Table 3. Results of updating Spanish Use tables at basic prices.