Input-Output and General Equilibrium: Data, Modeling and Policy Analysis

Brussels, Free University of Brussels
September 2 – 4, 2004

International Conference Jointly Organized by the Global Economic Modeling Network (EcoMod) and the International Input-Output Association (IIOA)

Step-Wise SAM Updating Approaches (*)

Casiano MANRIQUE DE LARA PEÑATE (ULPGC)
Dolores SANTOS PEÑATE (ULPGC)

(*): Partially supported by Ministerio de Ciencia y Tecnología and FEDER (grant BFM2002-04525-C02).

1. INTRODUCTION

A Social Accounting Matrix (SAM) is a square matrix, $X=\left(x_{ij}\right)_{1\le i,j\le n}$, whose columns and rows represent the expenditure and receipt accounts of economic sectors. The (i,j) element of this matrix, x_{ij}, is the flow from column account j to row account i. A SAM verifies the balance equations, that is, every row sum must be equal to the associated column sum. The classic updating problem consists of finding a SAM $X^T=\left(x_{ij}^T\right)_{1\le i,j\le n}$, corresponding to instant $t=T$, such that its *structure* is in some sense similar to the structure of the SAM observed in a previous instant $t=0$, $X^0=\left(x_{ij}^0\right)_{1\le i,j\le n}$, taking into account the known column totals at instant $t=T$. This problem can be formulated as follows:

$$\begin{aligned}
&\min D(X,X^0)\\
&\sum_{i=1}^{n} x_{ij}=y_j,\quad j\in\{1,2,...,n\}\\
&\sum_{j=1}^{n} x_{ij}=y_i,\quad i\in\{1,2,...,n-1\}\\
&x_{ij}\ge 0,\quad i,j\in\{1,2,...,n\}
\end{aligned}\qquad (1)$$

where the column sums, $y_j,\ j\in\{1,2,\ldots,n\}$, are given and $D\left(X,X^0\right)$ denotes the *disimilarity* between the structures of X and X^0.

In this work, "similar structure" means that $x_{ij}^T=0\Leftrightarrow x_{ij}^0=0$ and $\frac{x_{ij}^T}{y_j^T}\approx\frac{x_{ij}^0}{y_j^0}$. That is, X^T and X^0 have the same set of null elements and similar column coefficients. Function D can be defined according to different dissimilarity criteria, these dissimilarity functions can be classified into two groups, first group includes the functions defined using entropy measures and second group contains the distances defined as L_p metrics. Approaches as the RAS model (Stone and Brown, 1962), the biproportional estimation (Bacharach, 1965) and the minimization of the sum of cross entropies (MSCE) (Golan, Judge and Robinson, 1994; Robinson, Cattaneo and El-Said, 2001), belong to the first group. Relations between these adjustment-updating approaches are shown in Macgill (1977, 1978) and McDougall (1999). Measures such as those used by Matuszewski, Pitts and Sawyer (1964) and Harrigan and Buchanan (1984), are L_p metrics.

The assumption established in the classic updating problem according to which column totals are known, is not realistic. Most cases the knowledge of the column sums is partial, additionally some information about flows at instant $t=T$ is available. Sometimes the information about the economic transactions at instant $t=T$ is not compatible with the structure of the initial matrix, leading to unfeasible problems. This means that it is not possible to find a matrix adjusted to the known values with the same set of null flows than the initial SAM. This unfeasibility could be a consequence of errors in data, usually obtained from different sources, or it could represent a change in the structure of the SAM. The feasibility conditions for the biproportional estimation of a SAM, given by Bacharach (1965), can be adapted as follows.

Feasibility conditions

Let $X=\left(x_{ij}\right)_{1\leq i,j\leq n}$ be a SAM and $y_j,\ j\in I=J=\{1,2,\ldots,n\}$, the column totals.

Definiton 1. A null set is a set $I_1\times J_1\subseteq I\times J$ such that $x_{ij}=0,\ \forall\left(i,j\right)\in I_1\times J_1$.

Definiton 2. A null set $I_1\times J_1$ is maximal if $I_1\times J_1\subset I_2\times J_2\Rightarrow\exists\left(i,j\right)\in I_2\times J_2$ such that $x_{ij}\neq 0$.

Assumption 1. Each maximal null set $I_1\times J_1$ verifies $\sum_{i\in I_1}y_i\leq\sum_{j\in J\setminus J_1}y_j$.

Proposition 1. Consider the updating problem of a SAM X^0 and their null sets, and let $y_j=y_j^T,\ j\in\{1,2,\ldots,n\}$, the given column totals at instant $t=T$. The RAS, biproportional and MSCE approaches, verify the following results:

1. A solution of problem (1) exists if and only if assumption 1 holds.
2. There exists a matrix X^T such hat $x_{ij}^0=0\Rightarrow x_{ij}^T=0$ if and only if assumption 1 holds.
3. $x_{ij}^0>0$ and $x_{ij}^T=0$ if and only if assumption 1 holds as equations for some null set.

Example 1.

Consider the initial matrix $X^0 = \begin{bmatrix} 0 & 0 & 25.14 & 30.50 & 0.15 \\ 0 & 0 & 12.46 & 72.14 & 77.68 \\ 1.58 & 13.42 & 0 & 20.12 & 2.48 \\ 7.24 & 98.86 & 0 & 86.72 & 16.66 \\ 47.01 & 50 & 0 & 0 & 0 \end{bmatrix}$. The

column totals are $y_1^0 = 55.83,\ y_2^0 = 162.28,\ y_3^0 = 37.60,\ y_4^0 = 209.48,\ y_5^0 = 97.01.$ The maximal null sets are $I_1 \times J_1 = \{1,2\} \times \{1,2\}$, $I_2 \times J_2 = \{3,4,5\} \times \{3\}$ and $I_3 \times J_3 = \{5\} \times \{3,4,5\}$. The feasibility conditions are:

$$\begin{aligned} y_1 + y_2 &\le y_3 + y_4 + y_5 \\ y_3 &\le y_1 + y_2 \\ y_5 &\le y_1 + y_2 \end{aligned}$$

The following two sets of column totals, $y_1 = 60,\ y_2 = 200,\ y_3 = 38,\ y_4 = 210,\ y_5 = 100,$ and $y_1 = 60,\ y_2 = 288,\ y_3 = 38,\ y_4 = 210,\ y_5 = 100,$ are feasible. The case $y_1 = 60,\ y_2 = 408,\ y_3 = 38,\ y_4 = 210,\ y_5 = 100,$ is unfeasible. For the feasible values, the matrix X^T is

$$X^T = \begin{bmatrix} 0 & 0 & 25.07 & 34.74 & 0.19 \\ 0 & 0 & 12.92 & 102.05 & 85.02 \\ 1.86 & 19.31 & 0 & 14.83 & 1.99 \\ 8.33 & 130.49 & 0 & 58.38 & 12.80 \\ 49.81 & 50.19 & 0 & 0 & 0 \end{bmatrix}$$

and

$$X^T = \begin{bmatrix} 0 & 0 & 24.27 & 35.55 & 0.19 \\ 0 & 0 & 13.73 & 174.44 & 85.02 \\ 2.02 & 35.97 & 0 & 0 & 0 \\ 8.76 & 201.24 & 0 & 0 & 0 \\ 49.21 & 50.78 & 0 & 0 & 0 \end{bmatrix}$$

respectively. In the second case, the null value for initial nonzero flows is a consequence of the equalities in the feasibility conditions for the final column sums. In contrast to the entropy measures, the L_2 metric produces new null flows for certain column totals which satisfy the feasibility conditions as inequalities. For example, for $y_1 = 57, y_2 = 312, y_3 = 36, y_4 = 245, y_5 = 104,$ the final matrices X^T obtained using the entropy and L_2 measures are:

$$X^T = \begin{bmatrix} 0 & 0 & 22.97 & 33.87 & 0.16 \\ 0 & 0 & 13.03 & 200.90 & 98.06 \\ 1.80 & 31.47 & 0 & 1.97 & 0.76 \\ 8.20 & 223.52 & 0 & 8.25 & 5.02 \\ 46.99 & 57.00 & 0 & 0 & 0 \end{bmatrix}$$

and

$$X^T = \begin{bmatrix} 0 & 0 & 22.09 & 34.70 & 0.20 \\ 0 & 0 & 13.91 & 202.19 & 95.90 \\ 1.62 & 23.92 & 0 & 8.10 & 2.36 \\ 7.68 & 231.78 & 0 & 0 & 5.54 \\ 47.70 & 56.30 & 0 & 0 & 0 \end{bmatrix}.$$

respectively.

We now consider different situations:

1. The updating problem is feasible.
2. The updating problem is unfeasible.
    (a) According to the decision maker's knowledge the structure has not changed, the unfeasibility is produced by errors in the information used or in the equations incorporated in the model.
    (b) The structure of the SAM has changed. The new null sets must be identified.

If the problem is unfeasible but the structure of the new SAM is assumed to be equal to the initial matrix, the information available must be revised in order to find errors and correct them. If we admit that a new structure exists, we must determine the null sets for the final SAM. The following linear program allows us to identify the critical flows (Möhr, Crown and Polenske (1987)):

$$\min \sum_{i,j=1}^{n} c_{ij} x_{ij}$$

$$\sum_{i=1}^{n} x_{ij} = y_j, \quad j \in \{1, 2, ..., n\}$$

$$\sum_{j=1}^{n} x_{ij} = y_i, \quad i \in \{1, 2, ..., n-1\}$$

$$x_{ij} \geq c_{ij}\varepsilon, \quad i, j \in \{1, 2, ..., n\}$$

where $c_{ij} = 1$ if $x_{ij} > 0$, $c_{ij} = 0$ if $x_{ij} = 0$, and ε is a small positive number. The set of new nonzero flows identifies a sufficient completion of X^0.

In this work we consider feasible problems and present a procedure to obtain updated matrices taking into account the existence of very dissimilar changes between the different the new given values to be used in the updating process, with respect to the initial SAM. We analyse the effect of this *dispersion* on the solution to the updating problem.

2. UPDATING A SAM IN THE PRESENCE OF DISPERSION

Consider an updating problem where $X^0 = \left(x_{ij}^0\right)_{1\le i,j\le n}$ is the initial matrix and $X_T = \left\{\hat{x}_{ij}^T : (i,j) \in S_T\right\}$ is the set of strictly positive given values for the flows in $S_T \subset I \times J$, being $I = J = \{1,2,\dots,n\}$. Since matrices $kX^0, k \in \mathbb{R}_+$ have the same structure than X^0, we can interpret the quotients $\frac{\hat{x}_{ij}^T}{x_{ij}^0}, (i,j) \in S_T$, as an indicator of the differences between the initial and final structures. It seems reasonable to assume that similar quotients $\frac{\hat{x}_{ij}^T}{x_{ij}^0} \approx k$, are associated to kX^0 and dispersed values increase the separation from the initial structure.

Taking into account that a few flows with isolated quotients can affect the updating process, we analyse the effect of using subsets of S_T in the objective function of the optimization problem. These subsets can be selected by means of cluster analysis. Additionally, we could have applied some allowed deviation from the given values.

Updating approach (feasible problems)

1. Calculate the quotients $\frac{\hat{x}_{ij}^T}{x_{ij}^0}, (i,j) \in S_T$.

   Obtain a sequence $S_{T1} \subset \dots \subset S_{Tp} \subset S_T$ such that

   $$\rho\left(S_{T,1}\right) < \dots < \rho\left(S_{T,p-1}\right) < \rho\left(S_{T,p}\right) = \rho\left(S_T\right)$$

   where $\rho\left(S_{T,q}\right)$ represents a dispersion measure of the quotients $\frac{\hat{x}_{ij}^T}{x_{ij}^0}, (i,j) \in S_{T,q}$, and $S_{T,p} = S_T$.

   Set $q = 1$.

2. For $S = S_{T,q}$ solve the updating problem

$$
\begin{aligned}
&\min D_{/S}(X, X^0) \\
&\sum_{i=1}^{n} x_{ij} = y_j, \quad j \in \{1,2,...,n\} \\
&\sum_{j=1}^{n} x_{ij} = y_i, \quad i \in \{1,2,...,n-1\} \\
&x_{ij} = \hat{x}_{ij}^{T}, \quad (i,j) \in S_T \\
&x_{ij} = 0, \quad (i,j) \in S_0 \\
&x_{ij} \geq 0, \quad i,j \in \{1,2,...,n\}
\end{aligned}
$$

where $D_{/S}$ is the objective function restricted to set S, and

$S_0 = \{(i,j) : x_{ij}^0 = 0\}$. Let $X_q^T = \left(x_{ijq}^T\right)_{1 \leq i,j \leq n}$ be the solution obtained.

3. Set $q = q+1$. If $q = p+1$ finish the loop, 4 else go to step 2.

During this process, we can use different strategies to define the starting points of the sequence of updating problems solved in step 2. We could also use the column totals obtained as solutions to previous updating steps in order to obtain new solutions to a certain adjustment problem, assuming that column sums are known.

3.- AN APPLICATION

In order to illustrate our argumentation we present an example using a 32×32 regional accounting matrix for the Canary Islands constructed starting from its 1992 Input Output Table.

Both Table 1 and Figure 1 show the ratios $\frac{\hat{x}_{ij}^{T}}{x_{ij}^{0}}, (i,j) \in S_T$ in increasing order. Figure 2 shows the different groups obtained using cluster analysis (leaving out the two highest values of these ratios for graph scaling reasons).

Table 1

CELL ROW	COLUMN	RATIO X^T / X^0	GROUP
13	7	0.253	1
10	13	0.640	1
10	11	0.938	1
19	15	1.340	2
15	21	1.342	2
6	13	1.480	2
11	10	1.507	2
1	32	1.531	2
11	4	1.566	2
13	5	1.821	2
23	15	2.013	2
13	10	2.136	3
32	1	2.151	3
11	3	2.223	3
17	22	2.230	3
1	31	2.380	3
31	32	2.380	3
17	18	2.467	3
28	25	2.621	3
19	16	2.764	3
30	25	3.442	4
22	17	3.548	4
19	17	4.548	5
17	19	4.634	5
22	15	5.562	6
15	22	5.998	6
18	17	11.781	7
32	17	21.171	8

Figure 1. Quotients XT/X0 for preset values

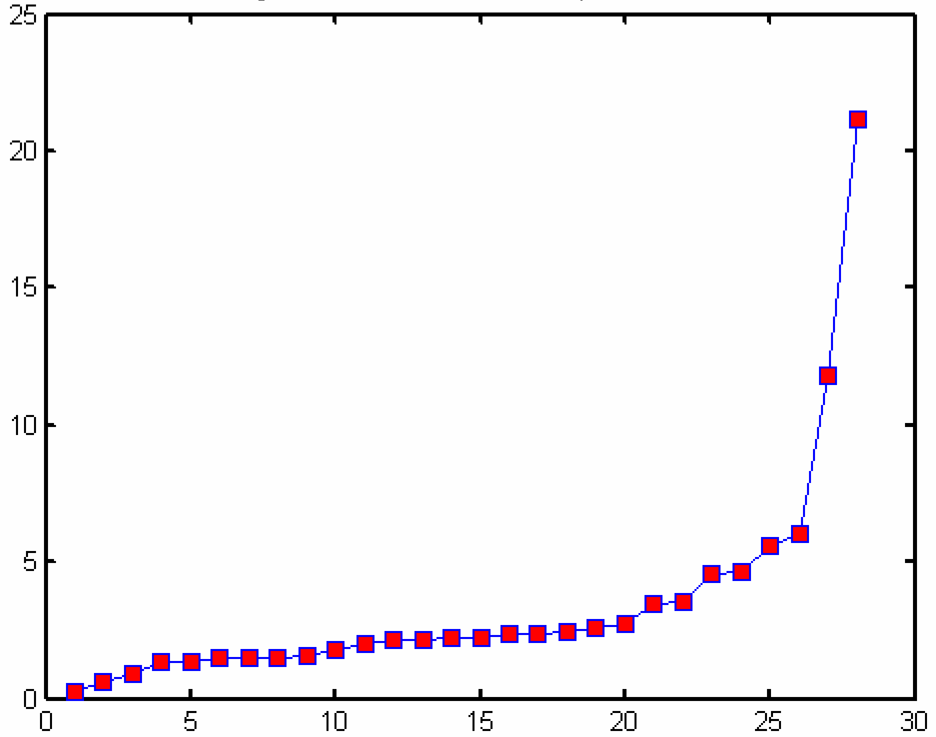

Figure 2. Ratios XT/X0

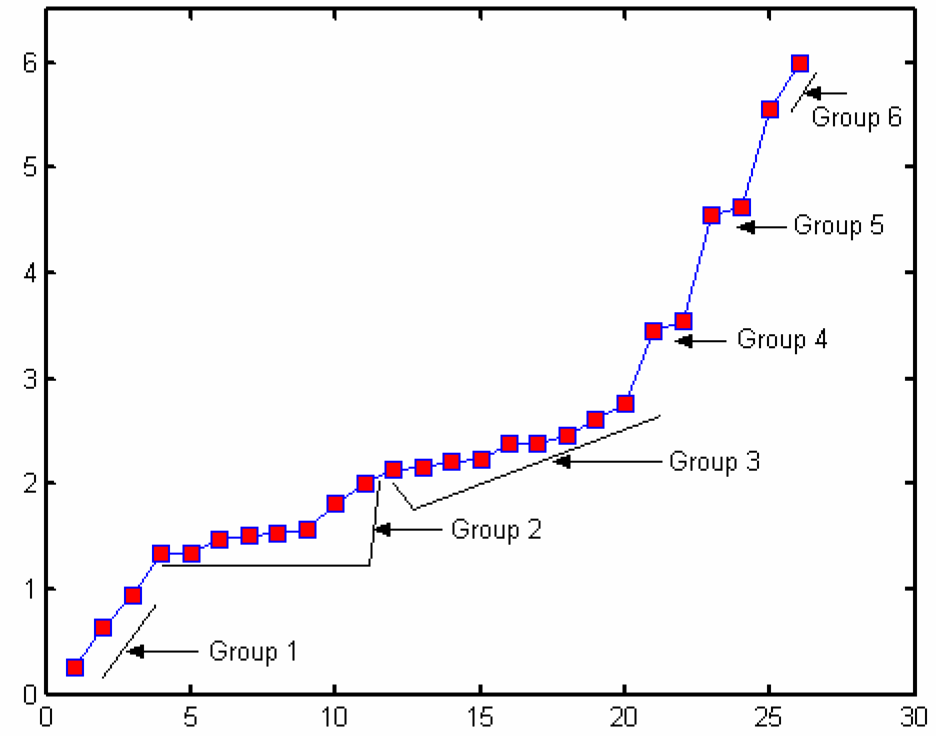

Table 2 shows the dispersion measure (diameter) values for the different groups previously identified.

Table 2

GROUP	NUMBER OF ELEMENTS n	DIAMETER D	$D/n^{3/2}$
G_1	3	0.685	0.131
G_2	8	0.673	0.029
G_3	9	0.628	0.023
G_4	2	0.106	0.037
G_5	2	0.086	0.030
G_6	2	0.436	0.154
G_7	1	0	0
G_8	1	0	0

As the minimum value of $\frac{D}{n^{3/2}}$ is reached by group G_3, we set $S_1 = G_3$. We calculate the distances $D(S_1, G_i) = \max\{d(x,y) : x \in S_1,\, y \in G_i\}$ for $i \neq 3$, and determine G_i such that $D(S_1, G_i) = \min_{k \neq 3} D(S_1, G_k)$. Then $S_2 = S_1 \cup G_i$. The rest of the sets in the sequence are obtained similarly. The sets selected in each of the steps are shaded in grey in table 3.

Table 3

GROUP	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8
S_1	2.511	1.424		1.412	2.498	3.862	9.645	19.038
S_2	3.295	2.208			2.498	3.862	9.645	19.038
S_3	3.295				3.294	4.650	10.441	19.834
S_4	4.381					4.650	10.441	19.834
S_5						5.745	11.528	20.921
S_6							11.528	20.921
S_7								20.921
S_8								

During this process, we used three different strategies to define the starting points of the sequence of updating problems solved in the different steps described in the previous section. These strategies consisted, first, in using the previous solution as the starting point of each of the steps. The second and third strategies used the average and the median of all the previous solutions, respectively. The first of the steps in each of the strategies used always the initial values of the SAM as a starting point. Table 4 shows the value of the objective function obtained in each of the steps, for the different strategies used.

Table 4

STEP	Previous Solution	Average	Median
1	4,01	4,01	4,01
2	9,05	4,09	4,27
3	9,14	4,16	4,34
4	7,61	4,84	4,55
5	10,02	4,52	4,44
6	10,07	4,29	4,73
7	8,71	6,11	4,67
8	11,69	5,38	4,68

It can be easily observed that the last two strategies are sensibly better than the first one, being the median strategy the one that shows the best optimal values. We can also observe a sequence of values for each of the strategies that is not very consistent. Since each of the steps includes extra restrictions, the optimal values should increase, what is not always the case. This result is due to the fact that the solver gets “stuck” in local optimums.

4.- CONCLUSIONS

When we know the new column totals, the adjustment problem can be formulated as a convex program with linear restrictions. In this case the local optimums are global ones. When, on the contrary, the new column totals are not known, the restrictions are still linear but the objective function gets more complicated and the solution becomes very sensible to the values used as starting points in the optimization process. In fact, the solver gets “stuck” in local optimums. It is therefore necessary to somehow “direct” the search in order to obtain god solutions.

A way to proceed could consist in considering different column totals selected following a certain criteria. One of those criteria could use the different column totals obtained from the original problem by taking different starting points.

We hope to be able to undertake in the near future the study of the properties of this optimization problem with the aim of obtaining some theoretical results in connection with the solutions and the sensitivity analysis.

5.- REFERENCES

Bacharach, M. (1965), 'Estimating Nonnegative Matrices from Marginal Data', *International Economic Review.* Vol.6. No.3, 294 – 310.

Golan, A., Judge, G. and Robinson, S. (1994), 'Recovering Information from Incomplete or Partial Multisectoral Economic Data', *The Review of Economics and Statistics* 76, 541 – 9.

Harrigan, F.J. and Buchanan, I. (1984), 'A Quadratic Programming Approach to IO Estimation and Simulation', *Journal of Regional Science*, Vol. 24, 3, 339 – 358.

Macgill, S.M., (1977), 'Theoretical properties of biproportional matrix adjustment', *Environment and Planning A*, Vol. 9, 687 – 701.

Matuszewski T.I., Pitts P.R. and Sawyer J.A. (1964), 'Linear Programming Estimates of Changes in Input Coefficients', *Canadian Journal of Economics and Political Science*, Vol. XXX,2 , 203 – 211.

McDougall, R.A. (1999), 'Entropy Theory and RAS are Friends' (http://www.sjfi.dk/gtap/papers/McDougall.pdf)

Polenske, Karen R.; Crown, William H. y Möhr, Malte (1986) *A Critical Review of RAS Literature.* Report Nº36 (MIT. Department of Urban Studies and Planning. Multiregional Planning Staff).

Robinson, S. Cattaneo, A. and El-Said, M (2001), 'Updating and Estimating a Social Accounting Matrix Using Cross Entrophy Mehods', *Economic Systems Research.* Vol. 13. No. 1. 47 - 64.

Stone, R. and Brown, A. (1962), *A computable Model of Economic Growth. (A Programme for Growth 1),* London, Chapman and Hall.