

# **Estimating a Social Accounting Matrix Using Entropy Difference Methods**

Sherman Robinson

Moataz El-Said

International Food Policy Research Institute  
Washington, DC

September 1997

## Abstract

There is a continuing need to use recent and consistent multisectoral economic data to support policy analysis and the development of economywide models. Updating and estimating input-output tables and Social Accounting Matrices (SAMs) for a recent year is a difficult and a challenging problem. Typically, input-output data are collected at long intervals (usually five years or more), while national income and product data are available annually, but with a lag. Supporting data also come from a variety of sources; e.g., censuses of manufacturing, labor surveys, agricultural data, government accounts, international trade accounts, and household surveys. The problem in estimating a SAM for a recent year is to find an efficient (and cost-effective) way to incorporate and reconcile information from a variety of sources, including data from prior years. The traditional RAS approach requires that we start with a consistent SAM for a particular period and “update” it for a later period given new information on row and column sums. This paper extends the RAS method by proposing a flexible entropy difference approach to estimating a consistent SAM starting from inconsistent data estimated with error, a common experience in many countries. The method is flexible and powerful when dealing with scattered and inconsistent data. It allows incorporating errors in variables, inequality constraints, and prior knowledge about any part of the SAM (not just row and column sums). Since the input-output accounts are contained within the SAM framework, updating an input-output table can be viewed as a special case of the general SAM estimation problem. The paper presents the structure of a SAM and a mathematical description of the estimation problem. It then describes the classical RAS procedure and the entropy difference approach. An example of the entropy difference approach applied to the case of Mozambique is presented. In addition, an appendix includes a listing of the computer code in the GAMS language used in the procedure.

Paper presented at the MERRISA (Macro-Economic Reforms and Regional Integration in Southern Africa) project workshop. September 8 -12, 1997, Harare, Zimbabwe. Our thanks to George Judge, Amos Golan, Hans Löfgren, and workshop participants for helpful comments on earlier drafts

## Table of Contents

Introduction .....	1
Structure of a Social Accounting Matrix (SAM) .....	1
The RAS Approach to SAM estimation .....	3
An Entropy Difference Approach to SAM estimation .....	4
Equality and Inequality Constraints .....	5
Incorporating Measurement Error .....	6
An Example: Mozambique .....	8
Conclusion .....	9
References .....	12
Appendix A: Mathematical Representation .....	13
Appendix B: GAMS Code .....	15

## Introduction

There is a continuing need to use recent and consistent multisectoral economic data to support policy analysis and the development of economywide models. A Social Accounting Matrix (SAM) provides the underlying data framework for this type of model and analysis. A SAM includes both input-output and national income and product accounts in a consistent framework. Input-output data are usually prepared only every five years or so, while national income and product data are produced annually, but with a lag. To produce a more disaggregated SAM for detailed policy analysis, data are often supplemented by other information from a variety of sources; *e.g.*, censuses of manufacturing, labor surveys, agricultural data, government accounts, international trade accounts, and household surveys. The problem in estimating a disaggregated SAM for a recent year is to find an efficient (and cost-effective) way to incorporate and reconcile information from a variety of sources, including data from prior years.

Estimating a SAM for a recent year is a difficult and challenging problem. A standard approach is to start with a consistent SAM for a particular prior period and “update” it for a later period, given new information on row and column totals, but no information on the flows within the SAM. The traditional RAS approach, discussed below, addresses this case. However, one often starts from an inconsistent SAM, with incomplete knowledge about both row and column sums and flows within the SAM. Inconsistencies can arise from measurement errors, incompatible data sources, or lack of data. What is needed is an approach to estimating a consistent set of accounts that not only uses the existing information efficiently, but also is flexible enough to incorporate information about various parts of the SAM.

In this paper, we propose a flexible entropy difference approach to estimating a consistent SAM starting from inconsistent data estimated with error. The method is very flexible, incorporating errors in variables, inequality constraints, and prior knowledge about any part of the SAM (not just row and column sums). The next section presents the structure of a SAM and a mathematical description of the estimation problem. The following section describes the classical RAS procedure, followed by a discussion of the entropy difference approach. Next we present an example of the entropy difference approach applied to the case of Mozambique. An appendix presents the computer software used in the procedure.

### Structure of a Social Accounting Matrix (SAM)

A SAM is a square matrix whose corresponding columns and rows present the expenditure and receipt accounts of economic actors. Each cell represents a payment from a column account (actor) to a row account. Define  $T$  as the matrix of SAM transactions, where  $T_{ij}$  is a payment from column account  $j$  to row account  $i$ . Following the conventions of double-entry bookkeeping, the total receipts (income) and expenditure of each actor must balance. That is, for a SAM, every row sum must equal the corresponding column sum:

$$y_i = \sum_j T_{ij} = \sum_j T_{j,i} \quad (1)$$

where  $y_i$  is total income and expenditure of account  $i$ .

A SAM coefficient matrix,  $A$ , is constructed from  $T$  by dividing the cells in each column of  $T$  by the column sums:

$$A_{ij} = \frac{T_{ij}}{y_j} \quad (2)$$

By definition, all the column sums of  $A$  must equal one, so the matrix is singular. Since column sums must equal row sums, it also follows that (in matrix notation):

$$y = Ay \quad (3)$$

A typical national SAM includes accounts for production (activities), commodities, factors of production, and various actors (“institutions”) which receive income and demand goods. The structure of a simple SAM is given in Table 1. Activities pay for intermediate inputs and factors of production and receive payments for exports and sales to the domestic market. The commodity account buys goods from activities (producers) and the rest of the world (imports) and sells commodities to activities (intermediate inputs) and final demanders (households, government, and investment). In this simple SAM, tariffs, indirect taxes, and subsidies are left out. Gross domestic product (GDP) at factor cost (payments by activities to factors of production) equals gross domestic income and also equals GDP at market prices (consumption plus investment plus government demand plus exports minus imports).

The matrix of column coefficients,  $A$ , from such a SAM provide raw material for much economic analysis and modeling. For example, the intermediate-input coefficients (known as the “use” matrix) correspond to Leontief input-output coefficients. The coefficients for primary factors are “value added” coefficients and give the factor distribution of income. Column coefficients for the commodity accounts represent domestic and import shares, while those for the various final demanders provide expenditure shares. There is a long tradition of work which starts from the assumption that these various coefficients are fixed, and then develops various linear multiplier models. They also provide the starting point for estimating parameters of nonlinear, neoclassical production functions, factor-demand functions, and household expenditure functions.

In principle, it is possible to have negative transactions, and hence coefficients, in a SAM. Such negative entries, however, can cause problems in some of the estimation techniques described below and also may cause problems of interpretation in the coefficients. A simple

Table 1. A national SAM

Receipts	Expenditure				
	Activity	Commodity	Factors	Institutions	World
Activity	Domestic sales				exports
Commodity	Intermediate Inputs	Final demand			
Factors	Value added (wages/rentals)				
Institutions	Factor Income			Capital inflow	
World	Imports				
Totals	Total costs	Total absorption	Total factor income	Gross domestic income	Foreign exchange inflow

approach to dealing with this issue is simply to treat a negative expenditure as a positive receipt and a negative receipt as a positive expenditure. For example, if a tax is negative, treat it as a subsidy. That is, if  $T_{i,j}$  is negative, we simply set the entry to zero and add the value to  $T_{j,i}$ . This “flipping” procedure will change row and column sums, but they will still be equal.

### The RAS Approach to SAM estimation

The classic problem in SAM estimation is the problem of “updating” an input-output matrix when we have new information on the row and column sums, but do not have new information on the input-output flows. The generalization to a full SAM, rather than just the input-output table, is the following problem. Find a new SAM coefficient matrix,  $A^*$ , that is in some sense “close” to an existing coefficient matrix,  $\bar{A}$  but yields a SAM transactions matrix,  $T^*$ , with the new row and column sums. That is:

$$T_{i,j}^* = A_{i,j}^* y_j^* \tag{4}$$

$$\sum_j T_{i,j}^* = \sum_j T_{j,i}^* = y_i^* \tag{5}$$

where  $y^*$  are known new row and column sums.

A classic approach to solving this problem is to generate a new matrix  $A^*$  from the old matrix  $A$  by means of “biproportional” row and column operations:

$$A_{i,j}^* = R_i \bar{A}_{i,j} S_j \quad (6)$$

or, in matrix terms:

$$A^* = \hat{R} \bar{A} \hat{S} \quad (7)$$

where the hat indicates a diagonal matrix of elements of  $R$  and  $S$ . Bacharach (1970) shows that this “RAS” method works in that a unique set of positive multipliers (normalized) exists that satisfy the biproportionality condition and that the elements of  $R$  and  $S$  can be found by a simple iterative procedure.<sup>1</sup>

Note that it is easy to generalize the RAS procedure to work with rectangular matrices that are not SAMs. That is, the method can be adapted to work with any matrix with known row and column sums: for example, an input-output matrix that includes final demand columns (and is hence rectangular). In this case, the column coefficients for the final demand accounts represent expenditure shares and the new data are final demand aggregates.

## An Entropy Difference Approach to SAM estimation

The fundamental estimation problem is that, for an  $n$ -by- $n$  SAM, we seek to identify  $n^2$  unknown non-negative parameters (the cells of  $T$  or  $A$ ), but have only  $2n-1$  independent row and column adding-up restrictions. The RAS procedure imposes the biproportionality condition, so the problem reduces to finding  $2n-1$   $R$  and  $S$  coefficients (one being set by normalization), yielding a unique solution. In a recent book, Golan, Judge, and Miller (1996) suggest a variety of Bayesian estimation techniques using what they describe as “maximum entropy econometrics”. Golan, Judge, and Robinson (1994) apply this approach to estimating a new input-output table given knowledge about row and column sums of the transactions matrix – the classic RAS problem discussed above. In this section, we extend this methodology to situations where there are different kinds of prior information than knowledge of row and column sums.

Golan, Judge, and Robinson (1994) use an entropy difference formulation, and set up the problem as finding a new set of  $A$  coefficients which minimizes the entropy distance between the prior  $\bar{A}$  and the new estimated coefficient matrix

---

<sup>1</sup>“Generally” exists because there are situations where the method fails. For example, a column or row of zeros cannot be proportionately adjusted to sum to a non-zero number. For the method to work, the matrix must be “connected”, which is a generalization of the notion of “indecomposable” [Bacharach (1970, p. 47)].

$$\begin{aligned} & \min \left[ \sum_i \sum_j A_{ij} \ln \frac{A_{ij}}{\bar{A}_{ij}} \right] \\ & = \min \left[ \sum_i \sum_j A_{ij} \ln A_{ij} - \sum_i \sum_j A_{ij} \ln \bar{A}_{ij} \right] \end{aligned} \quad (8)$$

subject to:

$$\sum_j A_{ij} y_j^* = y_i^* \quad (9)$$

Which implies that

$$\sum_j A_{j,i} = 1 \quad (10)$$

Bacharach (1970, pp. 83-85) shows that, for a particular scaling of the coefficients, the solution  $A^*$  for this problem is a biproportional estimate – that is, the solution of the RAS method satisfies the first order conditions for minimizing the entropy distance.<sup>2</sup>

In using the entropy difference approach, we generalize the RAS method in four areas: (1) add additional linear constraints on the new T matrix in addition to knowledge of row and column sums; (2) specify constraints on T as inequalities rather than equalities; (3) assume that row and column sums are not fixed parameters but involve errors in measurement; and (4) assume that the initial estimate,  $\bar{A}$ , is not based on a balanced SAM.

### Equality and Inequality Constraints

In addition to row and column sums, one often has additional knowledge about the new SAM. For example, aggregate national accounts data may be available for various macro aggregates such as value added, consumption, investment, government, exports, and imports. There also may be information about some of the SAM accounts such as government receipts and expenditures. This information can be summarized as additional linear adding-up constraints on various elements of the SAM. Define an n-by-n aggregator matrix, G, which has ones for cells in the aggregate and zeros otherwise. Assume that there are k such aggregation constraints, which are given by:

---

<sup>2</sup>Thissen and Löfgren (1997) also note this result



$$\sum_i \sum_j G_{ij}^{(k)} T_{ij} = \gamma^{(k)} \quad (11)$$

where  $\gamma$  is the value of the aggregate. These conditions are simply added to the constraint set in the entropy difference formulation.

While one may not have exact knowledge about values for various aggregates, including row and column sums, it may be possible to put bounds on some of these aggregates. Such bounds are easily incorporated by specifying inequality constraints in equation (11). It is also feasible to specify an error structure, either on equations or on variables, which can be interpreted as arising from either measurement error or some “noise” in the equation system.

### Incorporating Measurement Error

There are two general ways to incorporate noise into the system. One is to assume an “errors in equations” formulation where the process is assumed to include random noise and, second, an “errors in variables” formulation where the variables are assumed to be measured with noise. We will follow the errors-in-variables approach here, since the SAM is not viewed as a model but as a statistical framework where the issue is not in specifying an error generating process but as a problem of measurement.<sup>3</sup>

Rewrite the SAM equation as:

$$\begin{aligned} y &= A [\bar{x} + e] = A\bar{x} + Ae \\ y &= \bar{x} + e \end{aligned} \quad (12)$$

where  $y$  is the vector of row sums and  $x$ , measured with error  $e$ , is the initial known vector of column sums. Following Golan, Judge, and Miller (1994, chapter 6), we write the errors as a weighted average of known constants as follows:

$$e_i = \sum_{jw} W_{ijw} \bar{v}_{ijw} \quad (13)$$

subject to the weights summing to one:

---

<sup>3</sup>The problem is analogous to the distinction between errors in equations and errors in variables in standard regression analysis. See, for example, Judge *et al.* (1985). Golan and Vogel (1997) describe an errors in equations approach to the SAM estimation problem.

$$\sum_{jwt} W_{i,jwt} = 1 \quad (14)$$

where  $jwt$  is the set of weights,  $w$ . In the estimation, the weights are treated as variables bounded between zero and one subject to (14). The constants,  $v$ , are usually chosen to yield a symmetric distribution with moments depending on the number of elements in the set  $jwt$ . For example, if the error distribution is assumed to be rectangular and symmetric around zero, with known upper and lower bounds, the error equation becomes:

$$e_i = w_i \bar{v}_i - (1 - w_i) \bar{v}_i \quad (15)$$

In general, one can add more  $v$ 's to incorporate more information about the error distribution (e.g., more moments, including variance and kurtosis).

The procedure is simple. Given knowledge about the error bounds, equation (13) and (14) are added to the constraint set and equation (12) replaces the SAM equation. The problem is messier in that the SAM equation is now nonlinear, involving the product of  $A$  and  $e$ . The minimization problem is to find a set of  $A$ 's and  $w$ 's that minimize the entropy difference including a term in the errors:

$$\min \left[ \sum_i \sum_j A_{i,j} \ln A_{i,j} - \sum_i \sum_j A_{i,j} \ln \bar{A}_{i,j} + \sum_i \sum_{jwt} W_{i,jwt} \ln W_{i,jwt} \right] \quad (16)$$

subject to the constraint equations on column and row sums being equal, that the  $w$ 's fall between zero and one, and any other linear aggregation inequalities or equalities. Note that if the distribution is symmetric, then when all the  $w$ 's are equal, which is the default prior, all the errors are zero.<sup>4</sup>

Another source of measurement error may arise such that the initial SAM,  $\bar{A}$ , may not itself be a balanced SAM. That is, its corresponding rows and columns may not be equal. This

---

<sup>4</sup>When the error distribution is assumed to be rectangular between the upper and lower bounds, and is symmetric around zero (that is two  $w$  s), equation (16) is written as follows:

$$\min \left[ \sum_i \sum_j A_{i,j} \ln A_{i,j} - \sum_i \sum_j A_{i,j} \ln \bar{A}_{i,j} + \sum_i [ W_i \ln W_i + (1 - W_i) \ln (1 - W_i) ] \right]$$

situation does not change the entropy difference estimation procedure, but implies that it is not possible to achieve an entropy difference measure of zero because the prior is not feasible. The idea is to find a new feasible SAM that is entropy-close to the infeasible prior.

### **An Example: Mozambique**

To illustrate the use of the proposed entropy difference procedure, we apply it to a 1994 macro SAM for Mozambique.<sup>5</sup> We start from an inconsistent SAM in which some row and column sums are not equal. We assume that the row sums are measured with error, and apply the entropy difference method discussed above. The estimation and balancing procedures were implemented using the GAMS programming language. Appendix A shows the model equations in standard algebra notation and the GAMS code is given in Appendix B. Inequality constraints include upper and lower bounds on the coefficient matrix  $A$  ( $0 < A_{ij} < 1$ ) and bounds on the error variables (within the range of plus or minus the difference between column and row totals). The model also includes equality constraints on various cells of the SAM, including GDP at factor cost, GDP at market prices, and the government revenue row (tax receipts).

The prior unbalanced SAM for Mozambique in millions of 1994 meticaïs is presented in Table 2. Tables 3 and 4 present the estimated balanced SAM and the cell-by-cell percent deviation of the estimated from the initial (inconsistent) SAM. The largest changes occur in the savings-investment and international trade accounts. The initial row-column sum deviations were largest in the activity, commodity, household, and capital (savings-investment) accounts. In the model, the solution was achieved with essentially zero error variables. In this case, using an errors in variables specification does no damage since the constraints are simply not binding. That is, in a Bayesian approach, imposing non-binding constraints requires no additional updating of the prior.

We also did some sensitivity analysis, perturbing the initial SAM until the solution was infeasible. Using the errors in variables approach resulted in a feasible (and robust) solution in which the errors in variables were not zero. In the perturbed problem, some of the percent deviations in individual cells changed a lot and the solution tended to generate wide variations in cell entries. Of course, in the perturbed model, the entropy difference was taken with respect to the initial unbalanced SAM, and the solution deviated a lot from the earlier unperturbed SAM. In this case, the inconsistent prior had to be adjusted a lot. It is clearly useful to impose appropriate constraints – a general result in econometrics that certainly also applies to this essentially Bayesian method.

---

<sup>5</sup>Arndt, Cruz, Jensen, Robinson, and Tarp (1997) describe the Mozambique SAM in detail and the use of the entropy difference approach to estimate it.

## **Conclusion**

The entropy difference approach provides a flexible and powerful method for estimating a social accounting matrix (SAM) when dealing with scattered and inconsistent data. The method represents a considerable extension of the standard RAS method, which assumes that one starts from a consistent prior SAM and has knowledge only about row and column totals. The entropy difference framework allows a wide range of prior knowledge to be used efficiently in estimation. The prior knowledge can be in a variety of forms, including linear and nonlinear inequalities, equalities, errors in equations, and errors in variables. One also need not start from a balanced or consistent SAM.

Table 2. Prior unbalanced 1994 Macro SAM for Mozambique

(millions of 1994 meticaís)

Receipts	Expenditure									Totals
	activities	Commodities	factors	enterprise	households	Rec. govt.*	govt. invest	priv. invest	world	
activities	0.0	14827.4	0.0	0.0	2101.0	0.0	0.0	0.0	1488.2	18416.6
commodities	7917.5	0.0	0.0	0.0	6753.3	1764.5	2118.5	2197.8	0.0	20751.6
factors	9805.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9805.4
enterprises	0.0	0.0	3699.7	0.0	0.0	33.0	0.0	0.0	0.0	3732.7
households	0.0	0.0	6031.3	3417.5	0.0	29.6	0.0	0.0	209.5	9687.9
Rec. govt.	733.9	357.4	74.4	165.2	139.5	0.0	0.0	356.7	0.0	1827.1
govt. invest	0.0	0.0	0.0	0.0	0.0	0.0	0.0	406.2	1712.3	2118.5
priv. invest	0.0	0.0	0.0	150.0	649.2	0.0	0.0	0.0	2163.9	2963.0
world	0.0	5573.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5573.8
Totals	18456.5	20758.6	9805.4	3732.7	9643.0	1827.1	2118.5	2960.7	5573.8	

Source: Arndt, C. et al., 1997.

\* Recurrent government expenditures

Table 3. Estimated balanced 1994 Macro SAM for Mozambique

(millions of 1994 meticaïs)

Receipts	Expenditure									Totals
	activities	Commodities	factors	enterprise	households	Rec. govt.	govt. invest	priv. invest	World	
activities	0.0	14823.9	0.0	0.0	2110.4	0.0	0.0	0.0	1502.4	18436.7
commodities	7897.4	0.0	0.0	0.0	6774.2	1766.0	2118.5	2199.0	0.0	20755.1
factors	9805.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9805.4
enterprises	0.0	0.0	3700.5	0.0	0.0	32.2	0.0	0.0	0.0	3732.7
households	0.0	0.0	6030.5	3410.7	0.0	28.9	0.0	0.0	195.4	9665.5
Rec. govt.	733.9	357.4	74.4	165.2	139.5	0.0	0.0	356.7	0.0	1827.1
govt. invest	0.0	0.0	0.0	0.0	0.0	0.0	0.0	406.2	1712.3	2118.5
priv. invest	0.0	0.0	0.0	156.8	641.3	0.0	0.0	0.0	2163.7	2961.8
world	0.0	5573.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5573.8
Totals	18436.7	20755.1	9805.4	3732.7	9665.5	1827.1	2118.5	2961.8	5573.8	

Table 4. Differences between the estimated and the original unbalanced 1994 Macro SAM for Mozambique

(% )

Receipts	Expenditure									Totals
	activities	Commodities	factors	enterprise	households	Rec. govt.	govt. invest	priv. invest	world	
activities		-0.02			0.45				0.96	0.11
commodities	-0.25				0.31	0.09	0.00	0.05		0.02
factors	0.00									0.00
enterprises			0.02			-2.44				0.00
households			-0.01	-0.20		-2.44			-6.71	-0.23
Rec. govt.	0.00	0.00	0.00	0.00	0.00					0.00
govt. invest									0.00	0.00
priv. invest				4.55	-1.21				-0.01	-0.04
world		0.00								0.00
Totals	-0.11	-0.02	0.00	0.00	0.23	0.00	0.00	0.04	0.00	

## References

- Arndt, C., A. Cruz, H. Jensen, S. Robinson, and F. Tarp (1997). "A Social Accounting Matrix for Mozambique: Base Year 1994." Institute of Economics, University of Copenhagen.
- Bacharach, Michael (1970). Biproportional Matrices and Input-Output Change. Cambridge University Press. University of Cambridge, Department of Applied Economics.
- Brooke, A., D. Kendrick, and A. Meeraus (1988). GAMS a User's Guide, the Scientific Press, San Francisco.
- Golan, Amos, George Judge, and Douglas Miller (1996). Maximum Entropy Econometrics, Robust Estimation with Limited Data. John Wiley & Sons.
- Golan, Amos, George Judge, and Sherman Robinson (1994). "Recovering Information from Incomplete or Partial Multisectoral Economic Data." *Review of Economics and Statistics* 76, 541-9.
- Golan, Amos, and Stephen J. Vogel (1997). "Estimation of Stationary and Non-Stationary Accounting Matrix Coefficients With Structural and Supply-Side Information." ERS/USDA. Unpublished.
- Pindyck, R. S., and Daniel L. Rubinfeld (1991). Econometric Models & Economic Forecasts. McGraw Hill.
- Judge, George G. et al. (1985). The Theory and Practice of Econometrics. John Wiley & Sons.
- Thissen, Mark, and Hans Löfgren (1997). "A New Approach to SAM updating with an application to Egypt". Forthcoming in Environment and Planning A.

## Appendix A: Mathematical Representation



**Table A.1: Entropy Difference Equations**

#	Equation	Description
1	$T_{ij} = A_{ij} (X_i + e_i)$	SAM equation
2	$Y_i = X_i + e_i$	Row/column sum consistency
3	$e_i = \sum_{jwt} W_{i,jwt} \bar{v}_{i,jwt}$	Error definition
4	$\sum_{jwt} W_{i,jwt} = 1$	Sum of weights
5	$H = \left[ \sum_i \sum_j A_{ij} \ln A_{ij} - \sum_i \sum_j A_{ij} \ln \bar{A}_{i,j} + \sum_i \sum_{jwt} W_{i,jwt} \ln W_{i,jwt} \right]$	Entropy Difference
6	$\sum_j T_{i,j} = X_i + e_i$	Row sum
7	$\sum_i T_{i,j} = Y_j$	Column sum
8	$\sum_i \sum_j G_{ij}^{(k)} T_{ij} = \gamma^{(k)}$	Additional Constraints

**Inequalities**

$$0 < A_{ij} < 1$$

$$0 < W_i < 1$$

**Notation**

**Set**

i and j SAM accounts  
 jwt weights on errors in variables

**Variables**

$A_{ij}$  SAM coefficient matrix  
 $e_i$  Error variable  
 $H$  Entropy Difference (objective)  
 $T_{ij}$  Transactions SAM  
 $X_i$  Column sum  
 $W_{i,jwt}$  Error weights  
 $Y_i$  Row sum

**Parameters**

$\bar{A}_{i,j}$  Prior SAM coefficient matrix  
 $G_{ij}^{(k)}$  k'th aggregator matrix  
 $\gamma^{(k)}$  k'th control total  
 $\bar{v}_{i,jwt}$  Error bounds

## Appendix B: GAMS Code

## Appendix B: GAMS code

What follows is a listing of the GAMS program used in illustrating the entropy difference method discussed above. A quick list of some of GAMS features are listed below. For additional information about GAMS syntax see Brooke, Kendrick, and Meeraus (1988).

In the GAMS language:

- Parameters are treated as constants in the model and are defined in separate "PARAMETER" statements.
- "SUM" is the summation operator, sigma.
- "\$" introduces a conditional "if" statement.
- The suffix ".FX" indicates a fixed variable.
- The suffix ".L" indicates the level or solution value of a variable.
- The suffix ".LO" and ".UP" indicate the lower and upper bounds, respectively of a variable.
- An asterisk "\*" in the first column indicates a comment. Alternative treatments in the model Code are shown commented out.
- An "ALIAS" statement is used to give another name to a previously declared set.
- A semicolon (;) terminates a GAMS statement.
- Items between slashes (/) are data or set elements.

```

$TITLE Entropy Difference. Mozambique Macro SAM
$OFFSYMLIST OFFSYMXREF OFFUPPER
#####
#####
*
*
* MOZAM2 balances a Mozambique macro SAM assuming that the
initial data
* are measured with error and that the initial SAM is not
balanced.
* Some macro control totals are assumed known, including some
cells in
* the SAM. Row and column totals are assumed known only with
error.
* The error is specified by a rectangular distribution with
known
* bounds.
*
* Programmed by Sherman Robinson and Moataz El-Said, August
1997.
* Trade and Macroeconomics Division
* International Food Policy Research Institute (IFPRI)
* 1200 Seventeenth Street, N.W.
* Washington, DC 20036-3006 USA
* Email: S.Robinson@CGNET.COM
* M.El-Said@CGNET.COM
*
* Method described in S. Robinson and M. El Said, "Estimating
a Social
* Accounting Matrix Using Entropy Difference Methods."
September 1997.
* See also A. Golan, G. Judge, and D. Miller, Maximum Entropy
* Econometrics, John Wiley & Sons, 1996.
*
* Based on program used in C. Arndt, A. S. Cruz, H. T.
Jensen,
* S. Robinson, and F. Tarp, "A Social Accounting Matrix for
Mozambique:
* Base Year 1994." Institute of Economics, University of
Copenhagen,
* March 1997.
*
* Original version programmed by Sherman Robinson and Andrea
Cattaneo.
*
#####
#####
SETS

```

```

i      sam accounts / ACT   Activities
                        COM   Commodities
                        FAC   Factors
                        ENT   Enterprises
                        HOU   Households
                        GRE   Govt recurrent expenditures
                        GIN   Govt investment
                        CAP   Capital account
                        ROW   Rest of world
                        TOTAL /

ii(i)  all accounts in i except TOTAL

* For a uniform distribution, set jwt to only two entries. Error
* range set below with the vbar parameter.

jwt     weights on errors in variables
        / 1*3 / ;

ii(i)   = YES;
ii("Total") = NO;

ALIAS (i,j), (ii,jj);

#####          SAM DATABASE
#####
TABLE SAM(i,j)  social accounting matrix
                ACT          COM          FAC
ENT
ACT              0.0        14827.4240      0.0
0.0
COM              7917.5040      0.0        0.0
0.0
FAC              9805.4140      0.0        0.0
0.0
ENT              0.0          0.0        3699.7060
0.0
HOU              0.0          0.0        6031.3080
3417.5060
GRE              733.6000      357.4000      74.4000
165.2000
GIN              0.0          0.0        0.0
0.0
CAP              0.0          0.0        0.0
150.0000
ROW              0.0          5573.8150      0.0
0.0
Total            18456.5180     20758.639     9805.414
3732.706

```

	HOU	GRE	GIN
ACT	2101.0490	-0.3270	0.0
COM	6753.3320	1764.5000	2118.5000
2197.7980			
*COM	6953.3320	1564.5000	2518.5000
2597.7980			
FAC	0.0	0.0	0.0
ENT	0.0	33.0000	0.0
HOU	0.0	29.6000	0.0
GRE	139.5000	0.0	0.0
GIN	0.0	0.0	0.0
CAP	649.1560	-356.6730	-406.2000
ROW	0.0	0.0	0.0
Total	9643.037	1470.1	1712.3
2197.798			

	ROW	Total
ACT	1488.1570	18416.303
COM	0.0	20751.634
FAC	0.0	9805.414
ENT	0.0	3732.706
HOU	209.5010	9687.915
GRE	0.0	1470.1
GIN	1712.3000	1712.3
CAP	2163.8570	2200.14
ROW	0.0	5573.815
Total	5573.815	

```

##### Parameters and Scalars
#####
PARAMETER

SAM0(i,j)      Base SAM transactions matrix
T0(i,j)        Matrix of SAM transactions (flow matrix)
T1(i,j)        Adjusted matrix of SAM transactions for
negative coefficients
Abar0(i,j)     Prior SAM coefficient matrix

```

```

Abar1(i,j)     Adjusted prior SAM coefficient matrix for
negative coefficients
Target0(i)     Targets for macro SAM column totals
Vbar(i,jwt)    Error bounds
DELTA          Tolerance to allow zero entries in new SAM
;

```

```

SCALARS

sumtarg0       sum of targets
gdp0           base GDP
gdp00          GDP from final SAM
gdpfc0         GDP at factor cost
;

```

```

##### Initializing Parameters
SAM("TOTAL",jj) = sum(ii, SAM(ii,jj));
SAM(ii,"TOTAL") = sum(jj, SAM(ii,jj));

sam0(i,j)       = sam(i,j);

##### Divide SAM entries by 1000 for better scaling
sam(i,j)        = sam(i,j)/1000;

Abar0(ii,jj)$SAM(ii,jj) = SAM(ii,jj)/SAM("TOTAL",jj) ;

```

```

T0(ii,jj)       = SAM(ii,jj);
T0("TOTAL",jj) = sum(ii, SAM(ii,jj));
T0(ii,"TOTAL") = sum(jj, SAM(ii,jj));

DELTA           = .000001;

Display T0, Abar0 ;

```

```

##### CROSS ENTROPY
#####

```

```

##### RED ALERT!!!
#####

```

```

* The ENTROPY DIFFERENCE procedure uses LOGARITHMS: negative flows
in
* the SAM are NOT GOOD!!!
*
* The option used here is to detect any negative flows and net them
out
* of their respective symmetric cells, e.g.
*
*           negative flow ACT ---> GRE is set to zero

```

```

*          and ADDED to GRE ---> ACT as a positive number.
* The entropy difference method can then be implemented.
* After balancing, the negative SAM values are returned to
their
* original cells for printing.

SET
red(i,j)          Set of negative SAM flows
;

Parameter
redsam(i,j)      Negative SAM values only
rtot(i)          Row total
ctot(i)          Column total
;

rtot(ii)          = sum(jj, T0(ii,jj));
ctot(jj)          = sum(ii, T0(ii,jj));

red(ii,jj)$ (T0(ii,jj) LT 0) = yes ;
redsam(ii,jj)     = 0;
redsam(ii,jj)$red(ii,jj)    = T0(ii,jj);
redsam(jj,ii)$red(ii,jj)    = T0(ii,jj);

*Note that redsam includes each entry twice, in corresponding
row
*and column. So, redsam need only be subtracted from T0.
Tl(ii,jj)         = T0(ii,jj) -
redsam(ii,jj);
Tl("Total",jj)   = sum(ii, Tl(ii,jj));
Tl(ii,"Total")   = sum(jj, Tl(ii,jj));

redsam("total",jj) = sum(ii,
redsam(ii,jj));
redsam(ii,"total") = sum(jj,
redsam(ii,jj));

sam(ii,"total")   = sum(jj, Tl(ii,jj));
sam("total",jj)   = sum(ii, Tl(ii,jj));

rtot(ii)         = sum(jj, Tl(ii,jj));
ctot(jj)         = sum(ii, Tl(ii,jj));

Abar1(ii,jj)     =
Tl(ii,jj)/sam("total",jj);

display "NON-NEGATIVE SAM" ;

```

```

display redsam, Tl, Abar0, Abar1, rtot, ctot ;

##### Initializing Parameters after accounting for negative values
#####
*SR Note that target column sums are being set to average of
initial
* row and column sums. Initial column sums could have been used
instead,
* depending on data quality and prior knowledge.

target0(ii)      = (sam(ii,"total") + sam("total",ii))/2 ;
sumtarg0         = sum(ii, sam(ii,"total")) ;
gdpfc0           = Tl("fac","act");
gdp0             = Tl("fac","act") + Tl("gre","act")
                 - Tl("act","gre") + Tl("gre","com") ;

Display gdpfc0, gdp0;

##### VARIABLES
#####
VARIABLES
A(i,j)           Post SAM coefficient matrix
TSAM(i,j)        Post matrix of SAM transactions
Y(i)             row sum of SAM
X(i)            column sum of SAM
ERR(i)          Error value
W(i,jwt)        Error weight
DENTROPY        Entropy difference (objective)
GDPFC           GDP at factor cost
GDP             GDP at market prices
;

##### INITIALIZE VARIABLES
#####

A.L(ii,jj)       = Abar1(ii,jj) ;
TSAM.L(ii,jj)    = Tl(ii,jj);
Y.L(ii)          = target0(ii) ;
X.L(ii)          = target0(ii) ;
ERR.L(ii)        = 0.0 ;
W.L(ii,jwt)      = 1/card(jwt) ;
DENTROPY.L       = 0;
GDPFC.L          = gdpfc0;
GDP.L            = gdp0;

##### CORE EQUATIONS
EQUATIONS

SAMEQ(i)         row and column sum constraint

```

```

SAMMAKE(i,j)   make SAM flows
ERROREQ(i)    definition of error term
SUMW(i)       Sum of weights
ENTROPY       Entropy difference definition
ROWSUM(i)     row target
COLSUM(j)     column target

##### EQUATIONS IMPOSING KNOWN INFORMATION

GDPFCDEF      define GDP at factor cost
GDPDEF        define GDP
;

*CORE
EQUATIONS=====
==

SAMEQ(ii)..   Y(ii)           =E= X(ii) + ERR(ii) ;

SAMMAKE(ii,jj)$Abar1(ii,jj)..
ERR(jj)) ;    TSAM(ii,jj) =E= A(ii,jj) * (X(jj) +

ERROREQ(ii).. ERR(ii)       =E= SUM(jwt,
W(ii,jwt)*vbar(ii,jwt)) ;

SUMW(ii)..   SUM(jwt, W(ii,jwt)) =E= 1 ;

ENTROPY..    DENTROPY       =E= SUM((ii,jj)$Abar1(ii,jj)),
delta)       A(ii,jj)*(LOG(A(ii,jj) +
delta)))     - LOG(Abar1(ii,jj) +
W(ii,jwt)*LOG(W(ii,jwt) +
delta)) ;    + SUM((ii,jwt),

ROWSUM(ii).. SUM(jj, TSAM(ii,jj)) =E= Y(ii);

COLSUM(jj).. SUM(ii, TSAM(ii,jj)) =E= (X(jj) + ERR(jj))
;

*ADDITIONAL MACRO CONTROL-TOTAL
EQUATIONS=====

GDPFCDEF..   GDPFC =E= TSAM("fac","act") ;

GDPDEF..     GDP      =E= TSAM("fac","act") + TSAM("gre","act")
              - TSAM("act","gre") +
TSAM("gre","com") ;

##### Defining bounds for cell values
#####

A.LO(ii,jj)$ABAR1(ii,jj)      = 0 ;
A.UP(ii,jj)$ABAR1(ii,jj)      = 1 ;
A.FX(i,j)$NOT Abar1(i,j)      = 0;

TSAM.lo(ii,jj)                = 0.0 ;
TSAM.up(ii,jj)                = +inf ;
TSAM.FX(ii,jj)$NOT Abar1(ii,jj) = 0 ;

##### Fix other known cell values
#####

SET fixrow(i) /act, com, fac, ent, hou /
fixrow2(i) /act, com, fac, ent / ;

TSAM.fx("gre",fixrow) = Tl("gre",fixrow) ;
TSAM.fx("gre","cap")  = Tl("gre","cap") ;
TSAM.fx("gin","cap")  = Tl("gin","cap") ;

GDP.FX                  = GDP0 ;
GDPFC.FX                = GDPFC0 ;

##### Define variables bounds on errors
#####
* VBAR parameter defines upper and lower bounds on rectangular
error
* distribution on variable X. Here they are set at the difference
between
* the min and max column and row sums.

*vbar(ii,"1")           = .01*target0(ii) ;
*vbar(ii,"3")           = -.01*target0(ii) ;
vbar(ii,"1")            = 0.5*abs(rtot(ii)-ctot(ii)) ;
vbar(ii,"3")            = -0.5*abs(rtot(ii)-ctot(ii)) ;

*SR to use only two weights, delete set element "2" in jwt and
*comment out next statement.
vbar(ii,"2")            = 0.0 ;
W.LO(ii,jwt)            = 0 ;
W.UP(ii,jwt)            = 1 ;

*SR fix errors to zero by fixing weights at 1/3.
*W.FX(ii,jwt)           = 1/card(jwt) ;

```

```

*Set target column sums, X.
X.FX(ii) = TARGET0(ii) ;

Display vbar ;
##### DEFINE MODEL
#####

MODEL SAMENTROP / ALL /

##### SOLVE MODEL
#####

OPTION ITERLIM = 5000;
OPTION LIMROW = 0, LIMCOL = 0;
OPTION SOLPRINT = ON;

* SAMENTROP.holdfixed = 1 ;
* SAMENTROP.optfile = 1 ;
option NLP = MINOS5 ;
* OPTION NLP = CONOPT;
* SAMENTROP.WORKSPACE = 25.0;

##### Solve statement
#####

SOLVE SAMENTROP using nlp minimizing dentropy ;

#####
*----- Parameters for reporting results
Parameters
Macsam1(i,j) Assigned new balanced SAM flows from
entropy diff
Macsam2(i,j) Balanced SAM flows from entropy diff
x 1000
SEM Squared Error Measure
percent1(i,j) percent change of new SAM from
original SAM
PosUnbal(i,j) Positive unbalanced SAM
PosBalan(i,j) Positive balanced SAM
Diffnrnce(i,j) Differnce between original SAM and
Final SAM in values
NormEntrop Normalized Entropy a measure of total
uncertainty
;

```

```

macsam1(ii,jj) = TSAM.l(ii,jj);
macsam1("total",jj) = SUM(ii, macsam1(ii,jj)) ;
macsam1(ii,"total") = SUM(jj, macsam1(ii,jj)) ;
macsam2(i,j) = macsam1(i,j) * 1000 ;
SEM = Sum((ii,jj), SQR(A.L(ii,jj) -
Abar1(ii,jj))) /
SQR(9) ;
percent1(i,j)$ (T1(i,j))= 100*(macsam1(i,j)-T1(i,j))/T1(i,j);
PosUnbal(i,j) = T1(i,j) * 1000;
PosBalan(i,j) = macsam2(i,j);
Diffnrnce(i,j) = PosBalan(i,j) - PosUnbal(i,j);
NormEntrop = SUM((ii,jj)$ (Abar1(ii,jj)), A.L(ii,jj))*
LOG
(A.L(ii,jj)) / SUM((ii,jj)$ (Abar1(ii,jj)),
Abar1(ii,jj)* LOG (Abar1(ii,jj)))
;
display macsam1, macsam2, percent1, sem, dentropy.l, PosUnbal,
PosBalan;
display NormEntrop, Diffnrnce ;

##### Return negative flows to initial cell position
#####

macsam1(ii,jj) = macsam1(ii,jj) + redsam(ii,jj) ;
macsam1("total",jj) = SUM(ii, macsam1(ii,jj)) ;
macsam1(ii,"total") = SUM(jj, macsam1(ii,jj)) ;
macsam2(i,j) = macsam1(i,j) * 1000 ;

gdp00 = macsam1("fac","act") +
macsam1("gre","act")
- macsam1("act","gre") +
macsam1("gre","com") ;

display macsam1, macsam2 ;
display gdp0, gdp00, gdp.l, gdpfc0, gdpfc.l ;

#####
$ontext
##### Export Imbalanced & balanced SAM to a lotus spreadsheet
#####
* The libinclude statement requires additional gams programs
written
* by Tom Rutherford. See webpage:
http://www.gams.com/contrib/sslink/ssdoc.htm

$LIBINCLUDE SSDUMP Posunbal sam0.WK1 B1..B1
$LIBINCLUDE SSDUMP Posbalan macsam2.WK1 B1..B1
$LIBINCLUDE SSDUMP percent1 percent1.WK1 B1..B1

```



```
$LIBINCLUDE SSDUMP Diffnrnce      Diffnrnce.WK1 Bl..Bl
```

```
$offtext
```

```
*#####
```

```
*SR print some stuff
```

```
A.L("total",jj) = SUM(ii, A.L(ii,jj)) ;
```

```
A.L(ii,"total") = SUM(jj, A.L(ii,jj)) ;
```

```
ABAR1("total",jj) = SUM(ii, ABAR1(ii,jj)) ;
```

```
ABAR1(ii,"total") = SUM(jj, ABAR1(ii,jj)) ;
```

```
Display A.L, ABAR1 ;
```

```
*##### THE END
```

```
*#####
```