

4 Balancing a SAM

As already noted, the data sources used in this process are numerous in disagreement, and in need of supplemental assumptions. Hence, it is no surprise that the resulting SAM is not balanced. The largest discrepancies are found inside accounts concerned with secondary distribution of income. This concerns, in particular, the disaggregated household account because of the assumptions that have been involved to spread the various incomes, transfers, and expenditures of households within its disaggregated components or between the disaggregated components and other accounts.

Once again, there are several ways of balancing inconsistent social accounting or any other matrix. Besides economic closure rules, one of the most commonly used techniques to balance matrices is the RAS approach. It is typically used for updating SAMs for which *new row* and *column sums* are known. As we have already noticed in the case of the input-output transactions matrix, the RAS technique produces a new transaction matrix that is consistent with the new row and column sums by interactively adjusting the row and column entries proportionately until the new totals are obtained. This approach has at least one severe drawback. The RAS technique assumes that the initial SAM is consistent and that there is no measurement error in the row and column sums. When dealing with social accounting matrices in general, the initial SAM will often not be consistent, there will typically be measurement errors, and there will certainly be some data entries that the analyst finds more reliable than others. Mainly for that reason, the cross-entropy approach seems to be better adapted for balancing the SAM. If a reader is interested in comparing the RAS approach with the cross-entropy technique, we refer him to Part III, section 3.3. The only question that remains concerns the distribution law of the model to be retained, capable of describing the above discrepancies and stochastic errors. This will be answered in the following section.

Since much has been said in Part II concerning relationships between the cross-entropy approach and the statistical information theory approach, only a concise presentation of this technique will be given here, and the reader is referred to the aforementioned references for further detail.

The entropy technique is a method of solving undetermined estimation problems. The problem is undetermined because, in the case of a SAM, for an $n \times n$ matrix dimension, we seek to identify $n(n - 1)$ independent, unknown, non-negative parameters, i.e., all the cells of the SAM but one column or row, in conformity with Walras's law. In other words, restrictions must be imposed on the estimation problem so that we have enough information to obtain a unique solution and to provide enough degrees of freedom. The underlying philosophy of entropy estimation is to use all the information at hand for the problem and only that information: the estimation procedure should not ignore any available information nor should it add any false information.³³

33 See Shannon (1948) and Theil (1967) for a discussion of the concept of 'information.'

In the case of a SAM estimation, ‘information’ may be the knowledge that there is measurement error concerning the variables, and that some parts of the SAM are known with more certainty than others. There may be a prior in the form of a SAM from a previous year, whereby the entropy problem is to estimate a new set of coefficients ‘close’ to the prior using new information to update it. Furthermore, ‘information’ could consist of moment constraints on, for instance, row and column sums, or the average of the column sums. In addition to the row and column sums, ‘information’ may also consist of certain economic aggregates such as total value-added, final demand components, and/or imports. In that way, it becomes possible³⁴ to maintain Walrasian conditions of equilibrium. Such information may be incorporated as linear adding-up restrictions on the relevant elements of the SAM. In addition to equality constraints such as these, information may also be incorporated in the form of inequality constraints to the macro-aggregates mentioned. In most cases, macroeconomic theory can be useful in suggesting signs or interval of variation of certain parameters or ratios. This information will then be incorporated among other constraining equations. Finally, one may want to restrict cells that are zero in the prior to remain so after the entropy balancing procedure. Similarly, some cell values belonging to the SAM to be updated may not need to be modified because they come from well documented sources. Such cell values could then be restricted to stay unmodified during all steps of information processing.

4.1 Shannon-Kullback-Leibler Cross-Entropy

Let us follow for the next estimation procedure found, for example, in Robinson et al. (2001), and let the SAM be defined as a matrix T with elements T_{ij} representing a payment from column account j to row account i . As mentioned above, each account is supposed to display Walrasian equilibrium. In other words, every row sum (tot_i) in the SAM must equal the corresponding column sum (tot_j):

$$tot_i = \sum_j T_{ij} = \sum_j T_{ji} \quad (4.1)$$

Dividing each cell entry in the matrix by its respective column total generates a matrix of column coefficients A :

$$A_{ij} = \frac{T_{ij}}{tot_j} \quad (4.2)$$

³⁴ This constitutes a necessary and not a sufficient condition for the achieving of Walras equilibrium.

It is assumed that the entropy problem starts with a prior A which plausibly is a SAM from a previous period or, as in this case, a raw and unbalanced SAM. A represents the starting point from which the cross-entropy balancing procedure departs in deriving the new matrix of coefficients A . The entropy problem is to find a new set of A coefficients which minimize the so-called Kullback-Leibler (1951) divergence measure of the ‘cross-entropy’ (CE) between the prior A^* and the *posteriori* coefficients matrix A .

$$\min I(A \| A^*) = \sum_i \sum_j A_{ij} \ln \frac{A_{ij}}{A^*_{ij}} = \sum_i \sum_j A_{ij} \ln A_{ij} - \sum_i \sum_j A_{ij} \ln A^*_{ij} \quad (4.3)$$

subject to

$$\sum_j A_{ij} tot_i = tot_i \quad (4.4)$$

$$\sum_j A_{ij} = 1 \text{ and } 0 \leq A_{ij} \leq 1 \quad (4.5)$$

Note that, according to Walras’s law in general equilibrium theory, one equation can be dropped in the second set of constraints: If all but one column and row sums are equal, the last one must also be equal. The solution of the above problem is obtained by setting up the Lagrangian. The k macro-aggregates can be added to the set of constraints on the problem above as follows:

$$\sum_i \sum_j H_{ij}^{(k)} T_{ij} = \gamma^{(k)} \quad (4.6)$$

where H is an $n \times n$ aggregator matrix with ones for cells that represent the macro-constraints and zeros otherwise, and γ is the value of the aggregate constraint. As mentioned above, in the real world one faces economic data measured with error. The cross-entropy problem can also be formulated as an ‘error-in-variables’ system where the independent variables are measured with noise e . If, for example, we assume the known column sums are measured with error, the row/column consistency constraint can be written as:

$$tot_j = x_j + e_j \quad (4.7)$$

where tot_j is the vector of row sums and x_j , the known vector of column sums, is measured with error e_j . The prior estimate of the column sums could be, for instance, the initial column sums, the average of the initial column and row sums, or the row sums.

Following Golan et al. (1996), the errors are written as weighted averages of known constants v defined over a finite discrete support space $m > 1, \dots, M$ with points:

$$e_i = \sum_{m=1, \dots, M} f_{im} v_{im} \quad (4.8)$$

where f_{im} is a set of weights that fulfil the following constraints:

$$\sum_{m>1, \dots, M} f_{im} = 1 \text{ and } 0 \leq f_{im} \leq 1 \quad (4.9)$$

In the estimation problem, the weights are thus treated as probabilities to be estimated, and the prior for the error distribution in this case is chosen to be a symmetric distribution around zero with predefined lower and upper bounds, and using either three or five weights. Naturally, not only the column and row sums can be measured with error, the macro-aggregates by which we constrain our estimation problem may also be measured with error, and so we can operate with two sets of errors with separate weights f_1 's on the column sum errors, and weights f_2 's on the macro-aggregate errors. The optimization problem in the 'errors-in-variables' formulation is now the problem of finding A 's, f_1 's, and f_2 's that minimize the cross-entropy measure, including terms for the error weights:

$$\begin{aligned} \min I(A \parallel A^* \cup f_1 \parallel f_2) = & \left(\sum_i \sum_j A_{ij} \ln A_{ij} - \sum_i \sum_j A_{ij} \ln A_{ij}^* \right) + \\ & + \left(\sum_i \sum_{m, \dots, M} f_{1im} \ln f_{1im} - \sum_i \sum_{m, \dots, M} f_{1im} \ln f_{1iJ} \right) + \left(\sum_i \sum_{m, \dots, M} f_{2im} \ln f_{2im} - \sum_i \sum_{m, \dots, M} f_{2im} \ln f_{2iJ} \right) \end{aligned} \quad (4.10)$$

Referring once again to the definition of information provided by Kullback and presented in the second part of this book, cross-entropy measurement reflect how much the information we have introduced has moved our solution estimates away from the inconsistent prior, while also accounting for the imprecision of the moments assumed to be measured with error. Hence, if the information constraints are binding, the distance from the prior will increase. If they are not binding, the cross-entropy distance will be zero. It becomes now clearer why we have proposed, while assessing forecasting performance of the entropy technique, the difference between the average error variance coefficients (AEVC) of the periods 2006 and 2007 as a benchmark, maximum divergence precision measurement.

4.2 Balancing a SAM Through Tsallis-Kullback-Leibler Cross-Entropy

In the following, we are going to generalize the Jaynes-Kullback-Leibler model (4.10) and thus reconsider all the implications of the above theorem on the power law property of economy.

Let us formally explain Tsallis relative entropy model (2.47-2.50) to be minimized together with the above suggested constraints. In this presentation, the Bregman form of relative entropy (2.47) will be used:

$$\begin{aligned} \min I_q [A \| A^* \cup w \| wo] = & \alpha \frac{1}{q-1} \sum_i A_{iJ} [(A_{iJ})^{q-1} - (A^*_{iJ})^{q-1}] - \sum_i (A_{iJ} - A^*_{iJ})(A^*_{iJ})^{q-1} + \\ & (1-\alpha) \frac{1}{q-1} \sum_h w_{ih} [(w_{ih})^{q-1} - (wo_{ih})^{q-1}] - \sum_h (w_{ih} - wo_{ih})(wo_{ih})^{q-1} \end{aligned} \quad (4.11)$$

subject to:

$$\sum_j A_{ij} tot_j = tot_i \quad (4.12)$$

$$\sum_{i>2...M}^N A_{ij} = 1 \quad (4.13)$$

$$\sum_{h>1...I} w_{ih} = 1 \quad (4.14)$$

Symbols are as in equ. (4.9), except w_{ih} , which takes the place of f_{2ij} , and both represent disturbance errors on parameters but, this time, of different distribution laws.

Empirical, long practice with this class of economy-wide models provides some prior information on relevant ranges for parameter values and likely parameter estimates. Furthermore, while the support of any imposed prior distribution for a parameter is a maintained hypothesis (the estimate must fall within the support), the shape of the prior distribution over that support (e.g., the weights on each support point) is not. Unless the prior is perfect, the data will push the estimated posterior distribution away from the prior. The direction and magnitude of these shifts are, in themselves, informative. Also, note from Equations (4.11– 4.14) that, with increases in the number of data points, the second term of prediction in the objective function increasingly dominates the first term *precision*. In the limit, the first term in the objective becomes irrelevant. The prior distributions on parameters are only relevant when information is scarce.

4.3 A SAM as a Generalized Input-Output System

In the present paragraph, Kullback-Leibler (K-L) information divergence is extended to Tsallis non-ergodic systems and a q-Generalization of the K-L relative entropy criterion function (c.f.), with a priori consistency constraints, is derived for balancing a SAM as a generalized input-output transaction matrix.

On the basis of an unbalanced, Gabonese social accounting matrix (SAM) representing a generalized inverse problem input-output system, we propose to update and balance it following the procedure explained through the above section.

4.3.1 A Generalized Linear Non-Extensive Entropy Econometric Model

This section applies the results of, e.g., Jaynes (1957) and Golan et al. (1996) to present the model to be later implemented for updating and balancing input-output systems. While the argument in the criterion function is already known (see Equation 4.18), we need to reparametrize³⁵ the generalized linear model, to be introduced later into the model as restrictions in the spirit of Bayesian method of moments (e.g., Zellner, 1991). Note that such a linear restriction will be affected by a stochastic term expected to belong to the larger family of power law distribution. Let us succinctly present the general procedure for parameter reparametrization as it follows:

$$Y = X \cdot \beta + \varepsilon \quad (4.15)$$

Parameter β in general bears values not constrained between 0 and 1. When this is the case, reparametrization will no longer be necessary since parameter variation area fits well to probability definition area. The variable ε is an unobservable disturbance term with finite variance, owing to the economic data nature of exhibiting observation errors from empirical measurement or from random shocks. These stochastic errors are assumed to be driven by a large class of PL. As in classical econometrics, variable Y represents the system, the image of which must be recovered, and X accounts for covariates generating the system with unobservable disturbance ε to be estimated through observable error components e . Unlike classical econometric models, no constraining hypothesis is needed. In particular, the number of parameters to be estimated may be higher than the observed data points and the quality of collected information data low. Additionally, as already explained, to increase the accuracy of such estimated parameters from the poor quality of data points, the entropy objective function allows for incorporation of all constraining functions which act as Bayesian a priori information in the model.

Let us treat each $\beta_k (k = 1, \dots, K)$ as a discrete random variable with compact support and $2 < M < \infty$ possible outcomes. Thus, we can express β_k as:

$$B_k = \sum_{m=1}^M p_{km} v_{km} \quad \forall k \in K \quad (4.16)$$

where p_{km} is the probability of outcome v_{km} and the probabilities must be non-negative and sum up to one. Similarly, by treating each element e_i of e as a finite and discrete

³⁵ Reparametrization aims at treating parameters of the model as outputs of probability distribution to be estimated following the procedure presented by Golan et al. (1996) and later exploited for modelling many entropy econometric models (see, e.g., Bwanakare, et al. (2014, 2015, 2016). Since the same probabilities are related to entropy variable defining the criterion function, optimizing the whole model then leads to outputs taking into account stochastic a priori information owing to model restrictions.

random variable with compact support and $2 < M < \infty$ possible outcomes centred around zero, we can express e_i as:

$$e_i = \sum_{j>>1,\dots,J} r_{nj} z_{nj} \quad (4.17)$$

where r_n is the probability of outcome z_n on the support space j . Following Bwanakare (2014), we will use the commonly adopted index n , here and in the remaining mathematical formulations, to set the number of statistical observations. Note that the term e can be initially fixed as a percentage of the explained or endogenous variable, as an *a priori* Bayesian hypothesis. Posterior probabilities within the support space may display non-Gaussian distribution. The element v_{km} constitutes an *a priori* information provided by the researcher while p_{km} is an unknown probability whose value must be determined by solving a maximum entropy problem. In matrix notation, let us rewrite $\beta = V \cdot P$

$$\text{with } p_{km} \geq 0 \text{ and } \sum_{k=1}^K \sum_{m>>1,\dots,M} p_{km} = 1,$$

where again, is the number of parameters to be estimated and the number of data points over the support space. Also, let $e = r \cdot w$

$$\text{with } r_{nj} \geq 0 \sum_{n=1}^N \sum_{j>>1,\dots,J} r_{nj} = 1$$

and $r_{nj} = 1$ for N the number of observations and the K number of data points over the support space for the error term. Then, the Tsallis cross-entropy econometric estimator can be stated as:

$$\begin{aligned} MinH_q(p \| p^0, r \| r^0, w \| w^0) &\equiv \alpha \sum p_{km} \frac{[p_{km} / p_{o_{km}}]^{q-1} - 1}{q-1} + \beta \sum r_{nj} \frac{[r_{nj} / r_{o_{nj}}]^{q-1} - 1}{q-1} + \\ &+ \delta \sum w_{ts} \frac{[w_{ts} / w_{o_{ts}}]^{q-1} - 1}{q-1} \end{aligned} \quad (4.18)$$

S. to

$$Y = X \cdot \beta + e = X \cdot \sum_{m=1}^M v_m \left(\frac{p_{km}^q}{\sum_{m=1}^M p_{km}^q} \right) + \sum_{j=1}^J z_j \left(\frac{r_{nj}^q}{\sum_{j=1}^J r_{nj}^q} \right) \quad (4.19)$$

$$\sum_{k=1}^K \sum_{m>2,\dots,M} p_{km} = 1$$

$$\sum_{n=1}^N \sum_{j>2,\dots,J} r_{nj} = 1$$

$$\sum_{t=1}^T \sum_{s=2,\dots,S}^S w_{ts} = 1 \quad (4.20)$$

Additionally, k macro-aggregates can be added to the set of constraints as follows:

$$\sum_i \sum_j H^{(d)}_{ij} T_{ij} = \gamma^{(d)} + \sum_{s=1}^S g_s \left(\frac{w_{ts}^q}{\sum_{t=1}^T w_{ts}^q} \right), \quad (4.21)$$

where H is an $d \times d$ aggregator matrix with ones for cells that represent the macro-constraints and zeros otherwise, and γ is the expected value of the aggregate constraint. Once again, g_s stands for a discrete point support space from $s = 2, \dots, S$. Probabilities w_{ts} stand for point weights over g_s . The real q , as previously stated, stands for the Tsallis parameter.

Above, $H_q(p||p^0, r||r^0, w||w^0)$ is nonlinear and measures the entropy in the model. Relative entropies of three independent systems (three posteriors p , r , and w and corresponding priors p^0 , r^0 , and w^0) are then summed up using weights $\alpha\beta\delta$. These are positive reals summing up to unity under the given restrictions. We need to find the minimum divergence between the priors and the posteriors while the imposed stochastic restrictions and normalization conditions must be fulfilled. As will be the case in the application below, the first component of the criterion function may concern the parameter structure of the table; the second component errors on column (or row) totals and the last component may concern errors around any additional consistency variable, such as the GDP in the case below. As it has been shown by Tsallis (2009), this form of entropy displays the same basic properties as K-L information divergence index or relative entropy. The estimates of the parameters and residual are sensitive to the length and position of support intervals of β parameters (Equations 4.16 and 4.17) in the context of the Bayesian prior. When parameters of the proposed model are expressed under the form of elasticity or ratio—as will be the case in the example below—then the support space should be defined inside the interval between zero and one and will fit that of the usual probability variation interval. In such a case, no reparametrization of the model is needed. In general, support space will be defined between minus and plus infinity, according to the prior belief about the parameter area variation by the modeller. Additionally, within the same support space, the model estimates and their variances should be affected by the support space scaling effect, i.e., the number of affected point values (Foley, 1994). The higher the number of these points, the better the prior information about the system. The weights $\alpha\beta\delta$ are introduced into the above dual objective function. The first term of “precision” accounts for deviations of the estimated parameters from the prior (generally defined under a support space). The second and the third terms of “prediction

ex-post” account for the empirical error term as a difference between predicted and observed data values of the model. As expected, the presented entropy model is an efficient information processing rule that transforms, according to Bayes’s rule, prior and sample information into posterior information (Ashok, 1979).

4.4 Input-Output Power Law (PI) Structure

It is time now to come back to the fundamental problem concerning the true statistical nature of input-output data used in the above studies or those below. In recent years, as already explained in Part I, many studies (Champernowne, 1953; Gabaix, 2008) have shown that a large array of economic laws take the form of a PL, in particular macroeconomic scaling laws, distribution of income and wealth, size of cities, firms³⁶, and the distribution of financial variables, such as returns and trading volume. Stanley and Mantegna (2007) have studied the dynamics of a general system composed of interacting units each with a complex internal structure comprising many subunits where the latter grow in a multiplicative way over a period of twenty years. They found the system follows a PL distribution. Such outputs should present similarities with the internal mechanism of national accounts tables, such as an input output table or a SAM. A PL displays, besides its well-known scaling law, a set of interesting characterizations related to aggregative properties of a PL according to which a power law is conserved under addition, multiplication, polynomial transformation, and minimum and maximum. As far as the PL hypothesis for a SAM is concerned, taking into consideration the above literature and using PL properties, it should not be difficult to prove the PL character of a SAM, including the Gaussian trivial case. About SAM construction and components, see for example, Pyatt (Pyatt, 1985). General equilibrium (Wing Ian Sue , Sept 2004) implies that respective row and column totals are expected to balance. Conceptually, this model is based on the laws of product and value conservation (Serban and Blake (Serban Scriciu & Blake., 2005.)) which guarantee conditions of zero profits, market clearance, and income balance. However, different stages of statistical data processing remain concomitant with human errors and the SAM will not balance. It is generally assumed that the main sources of these imbalances remain different sources of documentation and different time of data collecting. This means that an unknown number of economic transaction values within the matrix are inconsistent with the data generating macroeconomic system. For clarity, let us use Table 13 to explain these imbalances, noting, for instance, a difference between the institution row and column totals as follows:

$$(iT + e_q) - (iT + \varepsilon_q) = (e_q - \varepsilon_q) \quad (4.22)$$

36 See Bottazzi et al. (2007) for a different standpoint on the subject.

Table 11: General structure of a stochastic non-balanced SAM

	Activities	Commodities	Factors	Institutions	Capital	World	Total
Activities	0	Ac	0	Ai	0	aw	aT+ ε_1
Commodities	Ca	0	0	Ci	cc	0	cT+ ε_2
Factors	Fa	0	0	0	0	0	fT+ ε_3
Institutions	Ia	Ic	If	ii	0	iw	iT+ ε_4
Capital	0	0	0	ci	0	cw	cT+ ε_5
World	0	Wc	0	wi	0	0	wT+ ε_6
Total	aT+ e_1	cT+ e_2	fT+ e_3	iT+ e_4	cT+ e_5	wT+ e_6	

Source: own elaboration.

The term on the left hand side of the above expression represents the difference between two erroneous and unequal totals of *institution* account. The origin of that difference results from difference between plausibly different stochastic errors e_4 and ε_4 , respectively, on column and row totals. In Table 11, the first alphabetical letter of symbols inside each cell represents the first letter of the row (supply) account, and the second letter represents the first letter of the corresponding (demand) column. In the SAM prototype below, e.g., the symbol “Ca”, explains purchases by the activity sector of goods and services from the commodity sector.

The targeted purpose is to find, out of all probability distributions, a set of *a posteriori* probabilities closest to *a priori* initial probabilities and insure the balance of the SAM table while satisfying other imposed consistency moments and normalization conditions. Following Shannon terminology, one may consider post-entropy structural coefficients and disturbance errors, respectively, as signal and noise. The first step consists in computing *a priori* coefficients by column from real data from Table 11 by dividing each cell account by the respective column total. Next, we treat these column coefficients as analogous to probabilities and column totals as expected column sums, weighted by these probabilities (see Equation 4.19). These coefficient values will serve as the starting, best prior estimates of the model. The other two types of priors to initialize the solution concern errors on column totals (Equation 4.17) and on gross domestic product (GDP) at factor and market prices (Equation 4.21). GDP variables are added to the model with the purpose of binding the latter to meet consistency macroeconomic relationships for different accounts inside the SAM. Other macroeconomic relations like those affecting interior or global consumptions could be added. The proposed approach combines non-ergodic Tsallis entropy with Bayes’ rule to solve a generalized random inverse problem. We may optionally consider only

some cell values as certain³⁷ while the rest of the accounts are unknown. This is one of the strongest points of the entropy approach over other mechanical techniques of balancing the national accounts table through a stochastic framework. All row and column totals are known with uncertainty. It is straightforward to notice that the potential freedom degree number of parameters to estimate $(n - 1) (n - 1)$ remains significantly higher than n observed data points (column totals). In a particular case of a SAM, and due to empty cells, that number of unknown parameters may be much lower. In any event, that will not generally prevent us from dealing with an ill-behaved inverse stochastic problem. The next important step is initializing the above defined errors through a reparametrizing process. A *five point support space symmetric around zero* is defined. To scale the error support space to real data, we apply Chebychev's inequality and three sigma rules (Serban Scriciu & Blake, 2005). Corresponding optimal probability weights are then computed so as to define the prior noise component (Robinson et al., 2001).

4.5 Balancing a SAM of a Developing Country: the Case of the Republic of Gabon

In our analysis of the last cases, we have rather underscored technical aspects of entropy for balancing input-output tables. However, when statistical data from different sources are available and sufficiently consistent, applying a complex procedure as the one relying on entropy formalism can be more time consuming than relatively easier techniques like the RAS (e.g., Pukelsheim, 1994, Bacharach (1970)). This is the case for many developed countries where statistical data gathering is generally efficient³⁸. In reverse, as we are going to see in the coming pages, this is not the case for the majority of developing countries in which statistical data are not only scarce but also of bad quality.

Thus, to complete an array of empirical advantages of the proposed entropy approach, we are going to analyse the case of developing countries where complete

³⁷ Only transaction accounts with the rest of the world (import, export, external current balance), plus government commodity consumption accounts are concerned.

³⁸ The statistical data gathering system in Poland can be seen as relatively efficient in comparison with those of most of developing countries. Availability of data on a large scale and their quasi-consistency though from various unrelated sources is the criterion retained here for giving such an appraisal. As a result, it should be relatively easier to balance national account tables without using complicated procedures such the entropy-related one. Thus, inconsistencies displayed in Table 1 may not reflect outputs from other publications on the same subject. The purpose of the present example is just to show the performance of the cross-entropy procedure in balancing a system under constraining, a priori information, like different macroeconomic identities characterizing national account tables.

statistical information is generally unavailable. Not only does such information not fully exist, what does should be approached with a high level of uncertainty. Applying traditional balancing techniques, like the RAS approach, becomes in practice difficult.

Based on the Shannon entropy approach, a large number of studies—particularly from developing countries—designed to balance SAM tables have been prepared in the last two decades. The already cited paper of Robinson et al. (2001), consecutive to the publications of Golan et al. (1996), has become a reference work for having shown an algorithm—in GAMS code (General Algebraic Modelling System)—for balancing a SAM in the case of uncertainty. One can list other studies with identical purpose, such as those of of Salem (2004) for Tunisia and Kerwat et al. (2009) for Libya. Murat (2005), using Shannon cross-entropy formalism, has balanced a Turkish Financial Social Accounting Matrix and, more recently, Miller et al. (2011) has built and balanced a disaggregated SAM for Ireland. Note that these last two countries belong, respectively, to the category of intermediary developed and developed countries. Many other entropy-based studies have been presented for various countries like Malawi, South Africa, Zimbabwe, Ghana, Gabon, and Vietnam. The results shown below generalizes, once again, Shannon formalism by applying a non-extensive entropy divergence formalism.

4.5.1 Balancing the SAM of GABON by Tsallis Cross-Entropy Formalism

A complete description of data sources or others details concerning the methodology of building the aggregated and disaggregated SAM of Gabon can be found in Bwanakare (2013).³⁹ That methodology has been proposed by Robinson et al. (2001) for balancing the SAM of Mozambique. Briefly, it consists of two steps in building the final SAM. In the first step, an aggregate and unbalanced SAM is built on the basis of official macroeconomic data. The later will serve as a control in building a much more disaggregated SAM in which accounts will be obtained by splitting out aggregated accounts of the balanced⁴⁰ SAM of the first step. Table 12 below represents the initial aggregated and unbalanced SAM of Gabon. Statistical data come from three sources: the Ministry of Planning, the Ministry of Economy and Finance, and the Bank of Central Africa States.

³⁹ This document was prepared with the help of the Directorate of National Accounting at the Ministry of Planning and Development of the Republic of Gabon. A copy of the outputs of the balancing of this SAM has been transmitted to the Ministry. See the document at <http://www.numilog.com/236150/Methodologie-pour-la-balance-d-une-matrice-de-comptabilite-sociale-par-l-approche-econometrique-de-l-entropie--le-cas-du-Gabon> (ebook: Paris: Editions JePublie)

⁴⁰ We have used the cross-entropy technique for balancing such an aggregated SAM.

Table 12: Aggregated non-balanced basic SAM of Gabon

Activity	Product	Factors	Agents	enterprises	government	ROW	Accumulation Total
Activities	-	4 784 615	-	103 019	-	-	4 887 634
Product	1 984 702	0	-	969 422	-	666 164	1 879 676
Factors	2 618 186	-	-	-	-	-	698 293
Agents	-	-	1 291 908	-	47 159	133 781	11 006
Enterprises	-	-	1 326 100	6 018	38 994	49 224	-
Government	284 746	215 181	1 029	104 043	484 284	742 353	9 621
ROW	-	1 198 460	-850	201 885	241 948	145 179	-
Accumulation	-	-	-	99 467	559 791	152 715	-113 680
Total	4 887 634	6 198 256	2 618 186	1 483 854	1 372 177	1 889 417	1 786 622
							698 293
							20 934 438

Sources: own calculations on the basis of official statistics of Gabon. (millions of CFA)

Thus, one may observe that all columns are not balanced with respective rows. This is the case for *the enterprise* and *government* accounts. This means that the general equilibrium is not attained. The total revenue of enterprises is higher than their total expenditures and the situation is reverse in the case of government.

In the case of Gabon, we target disaggregate accounts from 8 macro-accounts of the aggregated SAM to 82 subaccounts of the disaggregated new SAM. Among these subaccounts, we have 31 activity accounts, 33 product accounts (including 3 margin accounts which replace wholesale and retail trade), 3 accounts of labour factors and 1 of social capital, 4 classes of households, 3 business institutions, 4 government accounts, including public investment. The rest of the accounts are private investment, change in stock and rest of the world. Disaggregating a SAM requires much and often non-consistent information from different sources and periods. In the case of Gabon, this is particularly true for labour or the household accounts since reliable information allowing to tease out such accounts is scarce.

In statistical theory of information terms, the problem to be solved is one of finding a new disaggregated, *a posteriori* balanced SAM as close as possible to the initial unbalanced and disaggregated *a priori* SAM, while fulfilling imposed statistical and/or macroeconomic restrictions. To implement the model, we use the mathematical expression of non-extensive relative entropy under the next additional macroeconomic restrictions related to the targeted period (i.e., 1997 in the case of Gabon):

- Nominal GDP = consumption + investment + government expenditures + export – import
- GDP at factor cost = Nominal GDP – indirect taxes + subsidies
- Nominal GNP = Nominal GDP + net foreign income
- Fixing the input-output coefficients inside the SAM to the level of the previous period, implying that the structure of the Gabonese economy has not changed during the preceding years. Such a hypothesis remains realistic in the case of most developing countries, over a relatively long period.
- All accounts concerning business with the rest of the world have been fixed to the known level from international sources. This is so because, generally, data on international business remain reliable even in the case of developing countries.

When analysing the discrepancy between the prior and the posterior SAM, important modifications are observed. In particular, important discrepancies take place in the case of the institution and factor accounts. For instance, we note that wage assessment in the petroleum sector is probably underestimated for the senior executive category by around 300% of the real value deriving from post entropy modelling. In the period 1977, It had been pointed out many times by international institutions and media that financial transparency in Gabon needed to be improved. We note large modifications in factor inputs for bank and insurance activities.

Finally, it is important to note that outputs from Shannon cross-entropy (reported in Bwanakare, 2013) are identical to those from Tsallis cross-entropy formalism. As

pointed out many times, this suggests that we are dealing with a Gaussian distribution model. Thus, since similar outputs have been published in the above reference, outputs from the non-extensive entropy technique are not presented in this monograph. Nevertheless, interested readers can obtain more details on these outputs.

To conclude, the question of assessing the performance of the approach could be posed here. In fact, since no previous, benchmark SAM exists in the case of Gabon⁴¹, it is difficult to know to what extent we have deviated from values representing the true level of economy. Fortunately enough, entropy approach allows additional information embodied by the macroeconomic restrictions to be easily incorporated into the model. Next, when optimum solution is reached, we then get the best results, generally conforming to our expectations. This should be the case in the present Gabon model.

4.6 About the Extended SAM

A SAM can be extended in different ways and for different purposes. Generally a SAM is extended to incorporate monetary aspects of the economy or to take into account the natural environment. In this document, we will deal with this last case only.

A SAM can be extended and incorporate auxiliary accounts concerning the environment and natural resource sectors, so that it becomes possible to analyse interactions between them and the economy. In fact, an environmentally extended SAM (ESAM) usually captures the relationships among economic activities, pollution abatement activities, and pollution emissions. The multiplier and structural path analyses are applied to the ESAM for assessing environmental impacts of pollution-related economic policies.

Recent literature shows that an ESAM can be a useful tool for environmental policy analysis. Interested readers can find rich and detailed information in the monograph of Plich (2002).

Table 13 presents a Polish unbalanced ESAM. The particularity of that matrix is that we have added four new sectors related to ecological activity. The first sector is the abatement ecological activity sector. In that sector, firms carry out depollution activities. The second sector is the abatement ecological commodity sector, which offers the produced services and products to the market. The third sector is that of pollution fees. Firms (or households) must pay tradable pollution permits or other forms of tax to government as a cost of using the polluting engines. In this context, this sector is considered as a (negative) factor of production Plich (2002). The last sector concerns the environmental capital accumulation for depolluting activity. It is

⁴¹ As in many developing countries, even if it existed, it would not necessarily represent a good reference.

worth emphasising that Table 13 may display more or less weakness since any benchmark table for Poland has not been found in the existing literature. The sole source of data used is the Polish Central Office of Statistics (GUS). Under these circumstances, let us suppose that more reliable information does not exist, as it often happens for this kind of research. Finding results using traditional approaches may take time and reliability of outputs is generally limited since environmental data assessment is a difficult task. The purpose of this section is to apply entropy formalism to update the unbalanced Table 13. As we already know, the more significant moment restrictions are, the less significant precision errors will result. There are many reasons to consider the non-extensive entropy model to be—in this case of an ecological model—an ideal *balancing rule* since it has been proven to display multidisciplinary properties in many application areas. After having applied cross-non-extensive entropy formalism (see Equations 4.18–4.21), we present below Table 14 an environmentally extended (aggregated) balanced Polish SAM (2005). In this experiment, accounts related to government and to foreign operations are supposed to be known with a random error. Such restrictions seem acceptable for a country like Poland, where statistical data on government incomes and expenses or operations with the rest of the world remain sufficiently reliable. The next restriction has concerned matrix cells with zero values in the initial Table 13. These zero value cells have been supposed to be known with certainty so that no change has modified them after computation. Information divergence between the two tables is reported in Table 15. Values inside the cells are in percent. The reader may notice the precision of the model, through the retained constraining variables in the minimization entropy model.

Table 13: An environmentally extended (aggregated) *unbalanced* Polish SAM (2005).

	aACT	aPOLLA- BAT	pCOM	PPOLLA- BAT	Labor	Capital	POLLFEES Hou	Ent	GRE	CapAc	CapAcEnv Row	Total
aACT	0	0	19448,87	0,00	0,00	0,00	0,00	69,36	0,00	0,00	0,00	19518,23
aPOLLA- BAT	0	0	0	234,53	0,00	0,00	0,00	0,00	0,00	0,00	0,00	234,53
pCOM	10796,32	0,00	0,00	0,00	0,00	0,00	0,00	6989,91	0,00	1834,48	59,87	24111,74
PPOLLA- BAT	65,41	0,49	0,00	0,00	0,00	0,00	0,00	167,70	0,00	0,00	0,00	233,60
Labor	3520,11	1,58	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	3521,69
Capital	5048,40	5,21	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5053,61
POLLFEES	0,00	227,25	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	227,25
Hou	0,00	0,00	0,00	0,00	3503,16	2573,64	0,00	0,00	464,61	2711,41	0,00	9628,10
Ent	0,00	0,00	0,00	0,00	0,00	2126,84	0,00	0,00	42,46	0,00	0,00	2247,19
GRE	87,99	0,00	942,48	0,00	0,00	353,13	227,25	2241,24	0,00	353,13	0,00	4263,85
CapAc	0,00	0,00	0,00	0,00	0,00	0,00	0,00	652,48	1125,76	24,14	0,00	1889,86
CapAcEnv	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	29,93	29,93	0,00	59,87
Row	0,00	0,00	3709,90	9,56	18,53	0,00	0,00	196,31	271,56	39,99	0,00	4245,85
Total	19518,23	234,53	24101,25	244,09	3521,69	5053,61	227,25	10317,00	1891,86	3985,55	59,87	4245,85

Source: own calculations on the basis of data from Polish Statistical Office (GUS) and from EUROSTAT.

Table 14.: An environmentally extended (aggregated) *balanced* Polish SAM (2005). (10 billion zł).

	aACT	aPOLLA- BAT	pCOM	pPOLLA- BAT	LABOR	CAPITAL	POLLFEES	HOU	ENT	GRE	CAPAC	CAPACENVROW	Total
aACT	0,00	0,00	19432,32	0,00	0,00	0,00	0,00	67,27	0,00	0,00	0,00	0,00	19499,60
aPOLLA- BAT	0,00	0,00	0,00	232,75	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	232,75
pCOM	10820,75	0,00	0,00	0,00	0,00	0,00	0,00	6785,67	0,00	818,21	1896,08	59,93	24034,99
pPOLLA- BAT	71,47	0,48	0,00	0,00	0,00	0,00	0,00	170,09	0,00	0,00	0,00	0,00	242,04
LABOR	3543,13	1,55	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	3544,68
CAPITAL	4977,47	5,12	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	4982,59
POLLFEES	0,00	225,60	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	225,60
HOU	0,00	0,00	0,00	0,00	3526,05	2637,61	0,00	0,00	525,12	2827,23	0,00	0,00	9891,98
ENT	0,00	0,00	0,00	0,00	0,00	1987,15	0,00	0,00	0,00	41,00	0,00	0,00	2100,26
GRE	86,79	0,00	925,35	0,00	0,00	357,83	225,60	2126,06	0,00	364,76	0,00	0,00	4144,57
CAPAC	0,00	0,00	0,00	0,00	0,00	0,00	0,00	553,12	1236,46	23,79	0,00	0,00	1896,08
CAPACENVROW	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	31,99	27,95	0,00	0,00	59,93
ROW	0,00	0,00	3677,31	9,29	18,63	0,00	0,00	189,77	306,70	41,64	0,00	0,00	4243,34
Total	19499,60	232,75	24034,99	242,04	3544,68	4982,59	225,60	9891,98	2100,26	4144,57	1896,08	59,93	4243,34

Source: own calculations on the basis of data from Polish Statistical Office (GUS) and from EUROSTAT.

Table 15: Percentage of Information divergence between tables 13 and 14.

aACT	aPOLLA- BAT	pCOM	pPOLLA- BAT	LABOR	CAPITAL	POLLFEES	HOU	ENT	GRE	CAPAC	CAPACENVROW	Total
aACT	0	0,00	0,09	0,00	0,00	0,00	3,01	0,00	0,00	0,00	0,00	0,10
aPOLLA- BAT	0	0,00	0,00	0,76	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,76
COM	-0,23	0,00	0,00	0,00	0,00	0,00	2,92	0,00	-4,30	-3,35	-0,11	0,32
pPOLLA- BAT	-9,27	1,63	0,00	0,00	0,00	0,00	-1,42	0,00	0,00	0,00	0,00	-3,61
LABOR	-0,65	1,77	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-0,65
CAPITAL	1,40	1,79	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,41
POLLFEES	0,00	0,73	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,73
HOU	0,00	0,00	0,00	-0,65	-2,49	0,00	0,00	-13,02	-4,27	0,00	0,00	-2,74
ENT	0,00	0,00	0,00	0,00	6,57	0,00	0,00	0,00	3,45	0,00	0,00	6,54
GRE	1,36	0,00	1,82	0,00	-1,33	0,73	5,14	0,00	-3,29	0,00	0,00	2,80
CAPAC	0,00	0,00	0,00	0,00	0,00	0,00	15,23	-9,83	1,46	0,00	0,00	-0,33
CAPACENVROW	0,00	0,00	0,00	0,00	0,00	0,00	0,00	-6,87	6,64	0,00	0,00	-0,11
ROW	0,00	0,00	0,88	2,82	-0,53	0,00	3,33	-12,94	-4,12	0,00	0,00	0,06
Total	0,10	0,76	0,27	0,84	-0,65	1,41	0,73	4,12	-11,02	-3,99	-0,11	0,06

Source: own calculations.

Glossary of table abbreviations:

aAct: activity sector

aPOLLABAT: abatement ecological activity sector

pCom: commodity sector

pPOLLABAT: abatement ecological commodity sector

Labor: labor sector (factor of production)

Capital: capital sector (factor of production)

Pollfees: pollution fees sector (factor of production)

Hou: households institution

Ent: enterprise institution

GRE: government institution

CapAc: capital accumulation sector (private investment).

CAPACENV: abatement -oriented capital accumulation sector (private investment)

RoW: rest of the world institution.