

### 3 The Input-Output (IO) Table and its Main Application

Leontief (1941, 1966), the 1973 Economics Noble laureate and the father of the I-O table approach to economic analysis, began his first book about I-O analysis with the following words:

*This modest volume describes an attempt to apply the economic theory of general equilibrium—or better, general interdependence—to an empirical study of interrelations among the different parts of a national economy as revealed through covariations of prices, outputs, investments, and incomes.*

Leontief tried to apply neo-classical (Walras) general equilibrium to practical economic life. This suggests that subsequent analyses based on I-O tables or their extensions could have economic interpretations within the Walrasian framework apart from a few particular cases—e.g., those that consider the environment—violating Pareto optimum conditions.

The objective of this chapter is to present a consistent methodology of updating, forecasting, and economic modelling on the basis of I-O tables—for which underlying matrices are ill-behaved or data are not reliable. The proposed maximum entropy methodology can dynamically assess I-O multipliers and update and forecast I-O table information by combining the generalized maximum entropy principle and macroeconomic theory. The procedure remains in line with multiplier-accelerator analysis, assuming that induced investment is a function of expected growth. The only required condition to apply the proposed techniques is the availability of statistical information on final demand or value-added accounts which allow for updating under some constraining information (macroeconomic or not) obtained earlier, according to the traditional approach I-O table.

In the following sections, classical structure of an I-O table will be reviewed. The next step will describe I-O multipliers and their usage before trying to solve the more complex aspects of their estimation. Next, the proposed methodology for updating and forecasting an ill-behaved IO table will be described and the model presented.

#### 3.1 The I-O Table and Underlying Coefficients

The I-O method (Tomaszewicz 1992, 2005) represents a quantitative research approach that helps to understand how economic global product is created and shared with particular reference to connections within different sectors of production at this intermediate stage of product processing (Almon, & Clopper 2000; Avonds & Luc 2007; Robinson, Cattaneo & El-Said, 2001).

Use of this method of analysis is built upon I-O tables, the construction of which assumes the existence of constant returns to scale in the production process and the existence of general Walrasian equilibrium in the overall economy. In fact, as suggested earlier, the intention of Wassily Leontief was to apply general equilibrium theory previously proposed by Warlas (Leontief, 1970, 1986). In the context of the methodology of this work, let us first present the familiar I-O table structure, known for many decades, with only a slight modification that allows it to retain its square form. The fundamental concept of the I-O model is the concept of a direct technical financial coefficient:

$$a_{ij} = \frac{x_{ij}}{X_j} \quad \text{for } i, j = 1, \dots, n. \quad (3.1)$$

Matrix  $A = (a_{ij})$  is called a matrix of coefficients of inputs. It expresses the proportion of the value of product from sector  $i$  to be involved (sold) in the sector  $j$  for production of one unit value (e.g., 1 euro). The elements of this table are also called technical coefficients when expressed in quantitative flows between industries.

**Table 4:** Simplified inputs-outputs table structure.

Sector	Flows					The demand of the final	Total
1	$x_{11}$	$x_{12}$	$x_{13}$	...	$x_{1n}$	$Y_{1f}$	$X_1$
2	$x_{21}$	$x_{22}$	$x_{23}$	...	$x_{2n}$	$Y_{2f}$	$X_2$
...	...	...	...	...	...	...	....
N	$x_{n1}$	$x_{n2}$	$x_{n3}$	...	$x_{nn}$	$Y_{nf}$	$X_n$
	$D_1$	$D_2$	$D_3$	...	$D_n$		$D$
Total	$X_1$	$X_2$	$X_3$	...	$X_n$	$Y_f$	

Source: own elaboration.

#### Indications:

$X_j$  is the value of the global product of  $j$ - branch,  $j = 1, \dots, n$ .

$x_{ij}$  is the flow from the branches  $i$  to  $j$ , i.e., the value of the product manufactured in branch  $i$ -th and consumed by a branch of the  $j$ -th,  $i, j = 1, \dots, n$ .

$Y_{if}$  is the value of the final demand,  $i = 1, \dots, n$  and  $f = 1, \dots, F$ . Index  $f$  represents final demand composition, such as households, investment, exports, etc. The number  $F$  depends on the degree of table aggregation.

$D_j$  is the value added from branches of the  $j$ -th,  $j = 1, \dots, n$ .

Note that the above table can be split into four main parts:

The first part (upper-left part of the table) is composed of sub-matrix  $A$  showing the structure of intermediary demand of products “ $i$ ” by sector “ $j$ ”.

In the second part (top right), we have the final demand ( $Y_{if}$ ). Its traditional elements are: household consumption, government consumption, investment and stocks and the export sector.

The third part (lower left) shows the revenue created in the modes of production, i.e., the value added ( $D_j$ ), i.e., remuneration of labour and gross profit of capital in production sectors. If in the second part the “export” sector is explicitly shown, then in the third part one must, consequently, show the import sector.

Part four of the table refers to the secondary division of generated revenues. In the case of the I-O matrix it remains empty. Only construction of the social accounts matrix allows for completing this information.

Here, one should recall macroeconomic balance between final demand and value added, i.e.:

$$\sum_j D_j = \sum_i Y_i.$$

A further important equation is the definition of the Leontief model:

$$(I - A)X = Y \quad (3.2a)$$

or

$$(I - A)^{-1}Y = X \quad (3.2b)$$

where:

$X = [X_j]$  n-dimensional relationships to the column vector of global product

$Y = [Y_i]$  n-dimensional relationships to the row vector of the final product and  $I$  is the identity matrix. This relationship between the final product  $Y$ , the global product  $X$ , and cost structure matrix  $A$  is useful for the calculation of one of these elements when information about the other items is available. In the literature, this is known as a forecast of the first type (3.2a) or the second type (3.2b).

## 3.2 Input-Output Multipliers

### 3.2.1 The principal models

Multiplier coefficients play a central role in economic analysis (Leontieff, 1970). They make it possible to measure the impact of the exogenous variable or a shock on the whole system in which elements are interconnected. On these grounds, systemic models like these based on I-O tables—or their extension, or on computable general equilibrium models—have proved decisively superior to the classical *ceteris paribus* approach, using the classical econometric models.

The above defined relations (3.2a) and (3.2b) are very useful in empirical research. (3.2b) explains, in (input) multiplier terms, responses of producing sectors to a one

unit value increase in sectorial demand. For example, if the government spends one additional euro for buying a product from a given sector, what will the (backward) production impact be on the whole economy? The response is given by the level of multiplier coefficients along the sector column under consideration and total impact is given by its multiplier sum. The term  $(I - A)^{-1}$  in relation (3.2b) represents the multiplier. This formula can be decomposed as follows:

$$(I - A)^{-1} = \sum_{i=0}^{i=\infty} A^i \text{ with } A^0 = I \text{ and } A^i = A^{i-1} A \quad (3.2c)$$

Such a multiplier displays three impacts:

- initial impact equal to one,
- direct impact equal to A, and
- indirect impact summing up to  $A^2 + A^3 + \dots + A^k + \dots + A^n$ . Note that this geometric progression is quickly convergent; after a few steps, the last term vanishes to zero.

If we try to make a thorough analysis regarding the probabilistic nature of the multiplier, then we can rewrite (3.2c) as follows:

$$e^A i! \cong \sum_{i=0}^{\infty} A^i = (I - A)^{-1}, \quad (3.2d)$$

after having used the Taylor development. Thus, due to (3.2d), economic multiplier structure displays the exponential family of laws, which may constitute—as indicated in the first part of this book—a transitional law between power law and Gaussian law when power law-related higher frequency data are progressively aggregated towards the low frequency series. The purpose of this short discussion is to draw attention to the use of formula (3.2c). Not only should the frequency level of data have an impact on results, but also the functional relation of matrix A could modify the convergence transition path between the three probabilistic laws above.

Equation (3.2b) is generalized by the following, one of the most important relations of I-O modelling theory, referred to as *the central model of input-output*:

$$Z = B (I - A)^{-1} Y \quad (3.3)$$

$B$  = matrix of input coefficients for specific variables in economic analysis (intermediates, labour, capital, energy, emissions, etc.)

$I$  = Identity matrix

$A$  = matrix of technical financial coefficients  $[a_{ij}]$

$Y$  = diagonal matrix of final demand

$Z$  = matrix with results for direct and indirect requirements (intermediates, labour, capital, energy, emissions, etc.)

Three of the most frequently used types of multipliers in I-O analysis are those that estimate the effects of the exogenous changes of final demand (consumption, investment, exports) on:

- a) outputs of the sectors in the economy,
- b) value added and income earned by the households, and
- c) employment that is expected to be generated by the new activity levels.

However, due to interesting potential applications of this theory, it is worthwhile to be more complete about I-O multipliers. In empirical research, the I-O models used are based on input coefficients. Nevertheless, there is also a family of I-O models that are based on output coefficients. These models are sometimes called Gosh-models (Gosh, 1964). These models can be used to study price and cost effects or forward linkages of industries. Input coefficients reflect production functions or cost structures of activities. In contrast, output coefficients are distribution parameters for products reflecting market shares.

Presented below are only the four basic I-O models with input and output multipliers. The four I-O models have a dual character with an underlying symmetry. Each I-O model with input coefficients has a complement with output coefficients.

$$d_j = D_j / X_j, \text{ i.e., input coefficient for value added.} \quad (3.4)$$

Input coefficients for intermediates ( $a_{ij}$ ) (3.1) reflect the requirements for the use of product  $i$  in industry  $j$  for one unit of output of industry  $j$ . The capital and labour requirements are defined in the same way.

$$b_{ij} = X_{ij} / X_i, \text{ output coefficients for products} \quad (3.5)$$

$$y_i = Y_i / X_i, \text{ output coefficients for final demand} \quad (3.6)$$

Output coefficients for intermediates ( $b_{ij}$ ) identify the share of deliveries of sector  $i$  for sector  $j$ , ( $x_{ij}$ ) in the total output of sector  $i$ .

#### **Model 1: quantity model with input coefficients**

This model has been already defined in (3.2a) and (3.2b).

#### **Model 2: price model with input coefficients**

$$A'p + d = p \quad (3.7)$$

$$(I - A')p = d \quad (3.8)$$

$$p = (I - A')^{-1}d \quad (3.9)$$

$A'$  = transposed matrix of input coefficients for intermediates with  $A = [a_{ij}]$  for  $i, j = 1, \dots, m$ .

$I$  = identity matrix

$x$  = column vector of unit product price indexes for sectors 1 to  $m$ .  $w$  = column vector of exogenous input coefficients for value added  $w_1, \dots, w_m$ .

**Model 3: price model with output coefficients**

$$Bp + y = p \quad (3.10)$$

$$(I - B)p = y \quad (3.11)$$

$$p = (I - B)^{-1}y \quad (3.12)$$

$B$  = matrix of output coefficients for intermediates with  $B = b_{ij}$  for  $i, j = 1, \dots, m$

$I$  = identity matrix

$p$  = column vector of unit product price indexes for sectors 1 to  $m$

$y$  = column vector of exogenous output coefficients for final demand by product, with  $y_1, \dots, y_m$ .

**Model 4: Quantity model with output coefficients**

$$B'x + Z = X \quad (3.13)$$

$$(I - B')X = Z \quad (3.14)$$

$$x = (I - B')^{-1}Z \quad (3.15)$$

$B'$  = Transposed matrix of output coefficients for intermediates with  $B = b_{ij}$  for  $i, j = 1, \dots, m$

$I$  = identity matrix

$X$  = Column vector of product output for sectors 1 to  $m$

$Z$  = Column vector of exogenous value added by sector, with  $Z_1, \dots, Z_m$ .

However, due to lack of a proper microeconomic foundation, I-O models with output coefficients are rarely used in empirical research. The question often put to empirical researchers concerns the extent to which I-O multipliers are stable in time and behave according to expectations. Naturally, we can ask the same—but on a different scale—concerning the stability of I-O coefficients. We do not pretend to have the answers to these questions, which remain beyond the scope of this book.

### 3.2.2 A Model of Recovering the Sectorial Greenhouse Gas Emission Structure

Starting from insufficient information, let us combine below the central model of I-O (3.3) with the statistical information theory approach to predict emission multipliers of an extended I-O table (Miller & Blair, 1985). As above, let vector  $X_t$  be the global product of a given economy,  $A_t$  a matrix of I-O coefficients,  $Y_t$  a diagonal matrix of final demand, and  $B_t$  a matrix of outputs emission. Let us now suppose that these pieces of information have been obtained for a given period zero  $t_0$  and that this has made it possible to derive multipliers  $Z_0$  from the equation explaining emission content from final demand.

**Table 5:** Example of matrix  $B_1$  of ecological emissions.

Themes	Sector 1	Sector 2	...	Sector $n$	Themes total (period 1)
$CO_2$					$E_1$
$CH_4$					$E_2$
$N_2O$					$E_3$
$HFC_s$					$E_4$
$PFC_s$					$E_5$
$SF_6S$					$E_6$
Total global product	$X_1$	$X_2$	...	$X_n$	T

Next, we need to predict these multipliers for the next period one  $t_1$  to eventually compare changes within economic sectors and/or environmental themes. As very often happens in the real world, the only information concerning each emission theme is the total  $E_i$  (where index  $i$  refers to each type of theme) estimated at the end of the period  $t_1$ , for example, a certain number of tons of  $CO_2$  emitted during that period is equal to  $E_1$  without, however, any knowledge concerning the industrial branch being the source of that emission. We must then estimate matrix  $B_1$  (see Table 5) on the basis of the *priori* information of the initial period 0 and from data on measured sector totals of each ecological theme in the current period 1.

Evidently, we are confronted with an inverse problem since there may exist infinite combinations of emission levels by theme, related to global product, the total of which could lead to  $E_i$ . Without additional imaginative assumptions, any classical approach could solve this class of problem. As we will see later, some techniques exist like the bi-proportional approach, known as the RAS method (Bacharach, 1971), which could offer a solution with sufficiently limited effects as it does not take into account additional information on the investigated stochastic model. The Bayesian approach could be used in this class of problem. It can be shown that the approach presented below may be seen as an enhanced Bayesian technique to incorporate the second law of thermodynamics, which is a natural principle of organization. The neuronal class of model could also be suggested. However, it is not based on any theory, its application is time-consuming, and its outputs are not always guaranteed.

Using minimum information and without additional assumption, we suggest solving the problem by applying minimum entropy formalism, according to (2.43 or 2.44)<sup>23</sup>. Shannon-Gibbs entropy has been applied with success for updat-

<sup>23</sup> At the end of Part IV of the book, a theorem is presented proposing the power law properties of

ing and balancing matrices. However, on theoretical grounds, this assumes that entropy is a positive function of the number of possible states and is extensive (see 2.21), and it then neglects the possibility of inter-correlations among the states and their impact.

Let us then minimize the criterion function of the next non-extensive entropy model:

$$\text{Min}(p, p_0) \rightarrow I_q[p \| p_0] = \frac{1}{q-1} \sum_j p_j [(p_j)^{q-1} - (p_{0j})^{q-1}] - \sum_j (p_j - p_{0j})(p_{0j})^{q-1} \quad (3.16)$$

subject to

$$\sum_{j=1}^n \sum_{h=1}^M p_{jh} = 1$$

$$E_i = \sum_{j=1}^n \sum_{h=1}^M \frac{p_{jh}^q}{\sum_{j=1..n} p_{jh}^q} Z_h X_j \quad (3.17)$$

Probabilities  $r_{ij}$  distribution of theme emission  $i$  on column  $j$  is defined on a discrete support space  $Z_h$  where  $H$  is the number of points minimally equal to two,  $z = [z_1, z_2, \dots, z_H]$ . Support space is added since  $r_j$  are not probabilities and do not sum—by column—to unity. On a support space  $Z$ , parameters  $r_j$  are distributed with probabilities

$$P_{jk} = [p_{jk1}, p_{jk2}, \dots, p_{jkH}]'$$

Thus, in matrix formulation, we have:

$$E = ZP,$$

with

$$Z = \begin{pmatrix} z' & \dots & & \\ & z' & \dots & \\ \dots & & & z' \end{pmatrix}$$

$$P = \begin{pmatrix} p_{11} & p_{21} \dots & \dots & p_{H1} \\ p_{12} \dots & p_{22} \dots & & \dots p_{H2} \dots \\ \dots & \dots & \dots & \dots \\ p_{1H} & \dots p_{2H} & \dots & p_{HH} \end{pmatrix}$$

an input-output or a SAM. For the time being, it is suggested that the existence of these properties be assumed and non-extensive entropy formalism be applied. As we already know, non-extensive entropy formalism generalizes Shannon-Kullback-Leibler, which means at least we cannot lose the advantages of that generalization.



Moments expression in (3.17), as already seen in Part II of this work, contains a term of probability coefficients referred to as *escort distribution*<sup>24</sup>.

In that way we get quantity emission  $e_{ij}$ , being elements of matrix B, and on this basis, we can immediately derive the matrix of direct and indirect requirement emissions consecutive to one unit final demand  $Y_i$  according to the relation (3.3).

In this example, we have supposed the global product  $x_j$  and transactions matrix  $x_{ij}$  or I-O coefficients are known. At the end of this book, in Annex C (Tables 17 and 18), an instructional example is provided in which we are asked to recover total pollutants emission by industrial sector and by region on the basis of moment information. In the next section, we will remove these assumptions and suppose that the only information we have about the current I-O table is the final demand vector  $Y$  and a priori components, a previous I-O matrix.

### 3.3 Updating and Forecasting I-O Tables

#### 3.3.1 Generalities

Methods for updating I-O tables (Snower, 1990; Toh, 1998) can be categorized into univariate, bivariate, econometric, and stochastic procedures (e.g., Miller & Blair, 1985:266–316). All methods attempt to solve the following problem: row and column sums of an I-O table should correspond to the exogenous projection, and negative inputs should be avoided.

The basic idea of univariate methods to update I-O tables is to correct the matrix of input coefficients row-wise with a diagonal matrix of correction factors. An example of a version of such a method is the Statistical Correction Method.

The Bivariate approach, in contrast to univariate methods, which work with corrections of rows only, corrects rows and columns of an I-O table at the same time. The well-known RAS approach (Stone, 1984; Toh, 1998) represents an example of such an approach. However, a simple RAS procedure will normally fail to produce an acceptable projection of the structural change of an I-O table when change in relative prices and change in technology are substantial. Nevertheless, the incorporation of other exogenous data in the modified RAS procedure will tend to improve the quality of the projection. Several variations of the RAS technique can be found, for example, in Miller and Blair (1985:276–313).

Stochastic procedures assume that many independent variables may influence changes of input coefficients. The changing coefficients do not follow homogenous

---

<sup>24</sup> As discussed in (II.2.5.2), note that there is now an ongoing discussion of whether or not this form of constraint is appropriate in this case of Bergman relative entropy.

row and column multipliers, but rather the complex features of stochastic elements. The Lagrange method applied by the Central Statistical Office of the Netherlands (CBS) is an example of that method. The Least Squares Method can constitute another example of such an approach. It is worthwhile to add to this category of stochastic procedure: the EURO method used by EUROSTAT. The basic idea of this approach is to derive I-O tables that are consistent with official macroeconomic forecasts for GDP but avoid arbitrary adjustments of input coefficients to ensure the consistency of supply and demand. More specifically, this method should only use official macroeconomic forecasts as exogenous input for the iterative procedure. Column and row vectors for intermediate consumption and final demand should be derived as endogenous variables rather than accepted as exogenous variables from unspecified sources. The EURO method, like any other method presented above, presents advantages related to simplification of numerically and conceptually complex problems with substantial cost advantages. However, these advantages come at the price of certain disadvantages. In fact, primary forecasts for output levels normally not being available, it should be noted that the structural composition of final demand estimates in the Euro procedure is not based on econometric functions. Moreover, the impact of relative prices and other important economic variables such as innovation, technical progress, and productivity gains become difficult to fully anticipate.

### 3.3.2 RAS Formalism and its Limits

While the RAS method was implemented in the 1930s, Stone adapted the technique in 1961 for use in updating I-O tables from the work of Deming and Stephan (1940). The method is used when new information on the matrix row and column sums becomes available and we need to update a fully existing matrix.

Thus, following Lemelin, Fofana & Cockburn (2005) and Robinson, Cattaneo & El-Said (2001), the basic problem is generating a new  $n \times n$  matrix of  $A^1$  from an existing matrix  $A^0$  of the same dimension under restriction of the new given row and column totals  $X$  and  $Y$ . We then need to apply row and column multipliers  $r$  and  $s$ , respectively. The  $(2_n - 1)$  unknown multipliers are determined by the  $(2_n - 1)$  independent row and column restrictions using an iterative adjustment procedure.

Suppose we need to update a social accounting matrix (SAM). If we define  $T$  as a SAM transactions matrix, where  $t_{ij}$  is a cell value that satisfies the next restrictions:

$$T_{\bullet j} = \sum_i t_{ij}$$

to construct the coefficient matrix  $[a_{ij}] = A$  of a SAM, we divide each cell  $t_{ij}$  by the row total  $t_{\bullet j}$  of a corresponding column, i.e.:

$$a_{ij} = t_{ij} / t_{\bullet j}$$

In this case, if we indicate the unknown coefficients of  $A^1$  by  $[a_{ij}^1]$  and known coefficient elements of  $A^0$  by  $[a_{ij}^0]$ , the RAS procedure is as follows :

$$a_{ij}^1 = r_i a_{ij}^0 s_j,$$

In matrix notation, we have:

$$A^1 = \tilde{R} A^0 \tilde{S},$$

Where  $\tilde{\phantom{x}}$  indicates the diagonal matrix of elements  $R$  and  $S$ . Then, this last equation shows that the RAS method successfully constitutes an iterative algorithm of bi-proportional adjustments.

Let us indicate the serial number of different steps of the RAS algorithm by 1,2,...t and then define each algorithm as the following:

Step 1:

$$a_{i\bullet}^1 = \hat{x}_{i\bullet} / \sum_j x_{ij}^0 \Rightarrow x_{ij}^1 = a_{i\bullet}^1 x_{ij}^0 \Rightarrow b_{\bullet j}^1 = \hat{x}_{\bullet j} / \sum_i x_{ij}^1 \Rightarrow x_{ij}^2 = b_{\bullet j}^1 x_{ij}^1,$$

Step 2:

$$a_{i\bullet}^2 = \hat{x}_{i\bullet} / \sum_j x_{ij}^2 \Rightarrow x_{ij}^3 = a_{i\bullet}^2 x_{ij}^2 \Rightarrow b_{\bullet j}^2 = \hat{x}_{\bullet j} / \sum_i x_{ij}^3 \Rightarrow x_{ij}^4 = b_{\bullet j}^2 x_{ij}^3,$$

Step t:

$$a_{i\bullet}^t = \hat{x}_{i\bullet} / \sum_j x_{ij}^{2t-2} \Rightarrow x_{ij}^{2t-1} = a_{i\bullet}^t x_{ij}^{2t-2} \Rightarrow b_{\bullet j}^t = \hat{x}_{\bullet j} / \sum_i x_{ij}^{2t-1} \Rightarrow x_{ij}^{2t} = b_{\bullet j}^t x_{ij}^{2t-1}.$$

Thus, the step  $t$  corresponds to the last iteration ensuring a final convergence solution.

Compared to the cross-entropy approach presented above, the RAS procedure presents an obvious advantage of being relatively simple to use. However, it presents the following severe drawbacks:

*Lack of underlying economic theory and limited possibility of constraining information*

For instance, there is a need to fix some cell values inside an I-O matrix during its updating procedure when these values are known with sufficient certainty. The RAS procedure is not well suited for tackling such problems.

*No possibility to treat the problem on the stochastic side*

For instance, if the new known row and column totals,  $X$  and  $Y$ , are known with uncertainty—a realistic hypothesis—one will need to add a random component to the model. The RAS procedure is not suited for tackling such problems.

*The risk of starting with a basic I-O matrix  $A^0$  characterized by “spurious consistency”*

This is the case when the matrix has been updated or balanced on the basis of a theoretical hypothesis, e.g., macroeconomic closure rules. In such a case, the matrix appears well balanced despite possibly containing systematic errors.

Because of the above empirical problems, many researchers have tried the extension of the RAS procedure in the hope of rendering it more flexible. Lemelin et al. (2005) show that the RAS procedure can be apprehended as a Bayesian information processing rule with the new known row and column totals  $X$  and  $Y$  taken for new data in the Bayesian sense. In the process, after having compared the Lagrange multipliers of both procedures, the authors show conditions of equivalence between the RAS procedure and the Kullback-Leibler cross-entropy approach. Similar work has been presented by McDougall (1999). He concluded that the RAS approach corresponds to a generalized Shannon cross-entropy technique, suggesting that the latter cannot supplant the former. Nevertheless, according to McDougal, the cross-entropy approach can extend and adapt the RAS technique to problems that do not fit well with the traditional matrix balancing framework. Interestingly enough, Robinson et al. (2001), have conducted a comprehensive experiment on a 1994 SAM of Mozambique. Starting from the balanced SAM and randomly imposing new row and column totals, the authors have operated a Monte Carlo experiment in which they have simultaneously updated the matrix using RAS and Shannon cross-entropy procedures. They found RAS and CE to be equivalent measures—meaning that RAS is an entropy theoretic—if the CE method uses a single cross-entropy measure as an objective instead of attempting to use the sum of column cross-entropies. They concluded by confirming the findings of many previous researchers according to whom the RAS procedure remains less flexible in the case of new information in comparison with the cross-entropy technique, which is better at processing new information for optimal consistency of the updated SAM.

However, due to its popularity, researchers have proposed new extensions of the RAS approach [e.g., <http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-RA-07-013-EN.pdf>].

One of the interesting extensions has been the Model of Double Proportional Patterns (MODOP) developed by Stäglin (1972:69–81). This model consists of estimating all the existing cells of transaction. The resulting inconsistent matrix is then estimated by the RAS approach. The basic idea is to calculate the geometric mean of row and column multipliers and then to apply this factor to each element of the matrix. Following the above author, outputs from the MODOP are often similar to those from

the RAS approach. Some trials that allow the RAS approach to constrain targeted cells inside a matrix or to render the cells stochastic have been undertaken in recent years.

In a recent, thorough study on the comparative performance of the cross-entropy and RAS techniques, Chisari et al. (2012) concluded that cross-entropy had a more general character for the following reasons:

- a) It does not need all the new totals of rows or columns (although prediction will be less accurate).
- b) It does not need a balanced initial matrix (the sum of rows could be more/less than the sum of columns).
- c) New rims could contain an error term.
- d) New rims can be non-fixed parameters.
- e) Many values on the final matrix could be fixed (not necessarily a parameter).
- f) It allows non-linear constraints.

Referring to their simulation outputs, the authors propose *a rule of thumb consisting in preferring the RAS method if and only if any constraint or one constraint is enforced*. This seems to explain why the RAS approach continues to be successfully applied in different prediction studies. Furthermore, comparing the starting point to the RAS method, the above authors observe that the purchasing method is preferred to the supplying matrix because the aggregate bias is in this lower case. Furthermore, note that this suggestion does not seem to be consistent with the above investigations done by Robinson et al. (2001) on the Mozambique economy according to which the RAS and Shannon entropy approaches produce the same performance when no additional restriction is imposed. We will come back to this point when we present an example of I-O updating at the end of this section.

Trying to assess the cross-entropy approach in a dynamic, stochastic multi-objective optimisation problem, Bekker & Aldrich (2011) have concluded that acceptable results can be obtained while doing relatively few evaluations. Such an empirical fact tends to confirm a large area of cross-entropy application.

Finally, the central point to focus on through this section has been the limit—at least according to existing literature—of the RAS method compared to Shannon-Kullback-Leibler cross-entropy. Thus, since the latter is itself a converging case of Tsallis non-extensive entropy, the RAS procedure can be seen as relatively less attractive with respect to both entropy approaches.

### 3.3.3 Application: Updating an Aggregated EU I-O Matrix

#### 3.3.3.1 The RAS Approach

Tables 8 and 9 represent 27 aggregated EU symmetric I-O tables for domestic output at basic prices for years 2006 and 2007. In this simplified example, let us suppose that we have no information—though we do—on the elements of the intermediate

consumption sub-matrices for the period 2007 in Table 2. Using both the RAS and cross-entropy approaches, we are asked to predict those elements on the basis of the 2006 I-O matrix and sectorial accounts totals of the period 2007, which are supposed to be known.

In the example below, we directly use the transaction matrix in current price value. Thus, instead of coefficient matrix  $A$  presented in the above RAS formula, we use the matrix  $X$  of transactions. Using the formalism explained in the above section for the RAS approach, we present below the algorithm to solve the problem:

### Iteration 1

#### Calculation of row multipliers

1.098105 1.068032 1.066862 1.044765 1.052929 1.037053

							Actual	Row multipliers
50495.53	184777.3	3012.483	24612.5	3822.238	6453.238	273173.2	280691	1.027522
84702.14	2585662	382824.5	542838.5	230943.7	266151.9	4093122	4128898	1.008741
2941.775	45081.56	361192.4	47169.66	121551.4	47356.15	625293	598576	0.957274
38756.27	770976.1	124401.9	865048.6	264081.2	185633.2	2248897	2228709	0.991023
27106.73	694435.6	18950.7	753990.3	1314546	370177.4	3349764	3236274	0.96612
5269.119	83635.91	11198.4	72001.38	108426.3	267284.2	547815.4	538038	0.982152

where, e.g., starting multiplier 1.098105 to be later multiplied by the first column elements of the initial transaction matrix of 2006 is obtained from the quotient  $419340/381876$ , i.e., the first column output of 2007 divided by the first column output of 2006, both respectively from Tables 9 and 8.

With elements of row multipliers on the diagonal matrix premultiplied by  $X^{0ij}$  obtained in the above transformation, we get:

$\tilde{R}^0 X_{ij}$  equal to:

1.027522	0	0	0	0	0
0	1.008741	0	0	0	0
0	0	0.957274	0	0	0
0	0	0	0.991022853	0	0
0	0	0	0	0.96612	0
0	0	0	0	0	0.982152

X

50495.53	184777.3	3012.483	24612.5	3822.238	6453.238
84702.14	2585662	382824.5	542838.5	230943.7	266151.9
2941.775	45081.56	361192.4	47169.66	121551.4	47356.15
38756.27	770976.1	124401.9	865048.6	264081.2	185633.2
27106.73	694435.6	189507.7	753990.3	1314546	370177.4
5269.119	83635.91	11198.4	72001.38	108426.3	267284.2

equal to:

	51885.25	189862.6	3095.392	25289.88	3927.432	6630.843
	85442.48	2608262	386170.6	547583.2	232962.2	268478.2
	2816.084	43155.38	345759.9	45154.27	116358	45332.79
	38408.35	764055	123285.2	857282.9	261710.5	183966.7
	26188.35	670908	183087.1	728445	1270009	357635.7
	5175.075	82143.16	10998.53	70716.28	106491.1	262513.7
	209915.6	4358386	1052397	2274472	1991459	1124558
Actual	210592	4373650	1056144	2280581	1974301	1115919
Column multiplier	1.00322	1.003502	1.003561	1.002686	0.991384	0.992318

$X^1 = \tilde{R} X^0 \tilde{S}$  equal to:

51885.25	189862.6	3095.392	25289.88	3927.432	6630.843
85442.48	2608262	386170.6	547583.2	232962.2	268478.2
2816.084	43155.38	345759.9	45154.27	116358	45332.79
38408.35	764055	123285.2	857282.9	261710.5	183966.7
26188.35	670908	183087.1	728445	1270009	357635.7
5175.075	82143.16	10998.53	70716.28	106491.1	262513.7

X

1.00322	0	0	0	0	0
0	1.003502	0	0	0	0
0	0	1.003561	0	0	0
0	0	0	1.002685984	0	0
0	0	0	0	0.991384	0
0	0	0	0	0	0.992318

equal to:

52052.34	190527.6	3106.414	25357.81	3893.595	6579.903
85717.63	2617396	387545.8	549054	230955.1	266415.7
2825.153	43306.52	346991.2	45275.55	115355.5	44984.53
38532.04	766730.8	123724.2	859585.6	259455.7	182553.5
26272.69	673257.6	183739.1	730401.6	1259068	354888.3
5191.74	82430.84	11037.69	70906.23	105573.6	260497

*Iteration 2. ...j....:*

*Repeat Iteration 1 algorithm*

In this example, we reach the optimal value at the seventh iteration and then get the following values of the new supposed unknown transaction matrix of the year 2007:

0.999998	0	0	0	0	0
0	0.999999	0	0	0	0
0	0	1.000001	0	0	0
0	0	0	0.999999944	0	0
0	0	0	0	1.000001	0
0	0	0	0	0	1.000006

X

51929.5	190003.4	3095.379	25249.69	3870.157	6543.312
85635.77	2613871	386712.9	547482.9	229888.1	265307.2
2829.519	43356.44	347112.1	45258.99	115109.9	44909.49
38562.37	767033.6	123673.6	858620.7	258707.4	182111
26405.53	676396.6	184447.4	732693.9	1260792	355538.3
5228.733	82985.65	11103.05	71275.19	105935.8	261511.4

X

1.000001	0	0	0	0	0
0	1.000001	0	0	0	0
0	0	1	0	0	0
0	0	0	0.999999724	0	0
0	0	0	0	0.999999	0
0	0	0	0	0	0.999998



equal to:

51929.54	190003.5	3095.379	25249.68	3870.153	6543.302
85635.84	2613873	386712.9	547482.7	229887.9	265306.8
2829.521	43356.47	347112	45258.98	115109.8	44909.42
38562.4	767034	123673.6	858620.4	258707.2	182110.7
26405.55	676397	184447.4	732693.7	1260790	355537.7
5228.738	82985.69	11103.05	71275.17	105935.7	261511

**Table 6:** Error(%) prediction from the RAS procedure

3	0	1	-4	-7	4
1	0	0	-2	0	0
-7	5	-1	6	-4	3
-6	-1	-2	1	0	3
1	0	2	0	0	-1
-4	1	4	2	2	-2

The above errors are calculated as the error discrepancy percentage between the true matrix of transaction representing the period of year 2007 and the matrix updated by the RAS procedure on the basis of the 2006 I-O matrix.

### 3.3.3.2 The Entropy Approach

For comparative purposes, let us apply entropy formalism for updating the same IO transactions table as in the above example related to the RAS approach. Thus, using the Tsallis entropy formalism presented in (3.17) and in (2.48–2.50) under the hypothesis that transaction totals of the targeted period are known (and without any additional restriction), we get the outputs presented in Table 7.

Comparison of Tables 6 and 7 show slightly higher performance of the RAS approach. This seems to confirm the rule of thumb proposed by Chisari et al. (2012): *if and only if any constraint or one constraint is enforced* as in the present case. However, such a conclusion, as already mentioned above, is not in line with the one proposed by investigations conducted by Robinson et al. (2001) which lead to equivalent performance in the same conditions between the cross-entropy and RAS approaches. More investigations are needed to contradict or confirm the study results of the above authors. Of course, following the results of several investigations presented above, cross-entropy naturally presents higher performance than the RAS approach, particularly when statistical data are known with uncertainty. In the next section, we are going to propose the forecasting of a higher dimension I-O table through cross-entropy formalism. Later, when treating the case of a SAM with higher dimension, we

**Table 7:** Error (%) prediction of the I-O table by the Tsallis (or Shannon) cross-entropy procedure

Products (CPA)	P1	P2	P3	P4	P5	P6
P1	7.985	3.141	1.934	-0.049	-7.602	4.579
P2	8.954	2.214	-0.722	-0.099	-2.772	-0.158
P3	-8.035	2.042	-5.767	3.010	-11.670	-2.940
P4	-4.044	0.691	-4.533	7.906	-4.444	0.523
P5	0.370	-2.219	-2.198	-1.980	-6.118	-0.147
P6	-2.688	0.587	1.827	1.848	-2.704	-4.190

**Table 8:** Symmetric I-O Tables for domestic output at basic prices (year 2006; EU27, Mio. EUR current prices)

Inputs of products								
Products (CPA)	P1	P2	P3	P4	P5	P6	Others	Total
P1	45984	173007	2824	23558	3630	6223	126650	381876
P2	77135	2420959	358832	519579	219335	256642	2999462	6851944
P3	2679	42210	338556	45149	115441	45664	1130576	1720275
P4	35294	721866	116605	827984	250806	179001	2465270	4596826
P5	24685	650201	177631	721684	1248467	356951	1932724	5112343
P6	4798	78308	10497	68916	102976	257734	3054393	3577624
Others	191301	2765392	715330	2389956	3171688	2475408	1499777	13208852
Total	381876	6851944	1720275	4596826	5112343	3577624	13208852	

Source: own calculations.

Source: based on <http://ec.europa.eu/eurostat/web/esa-supply-use-input-tables/overview>

will present the methodology of updating it in the presence of uncertainty and with a free number of restrictions.

### 3.3.3.3 I-O table Forecasting

By “updating,” we compute operations on table rows and columns with the purpose of balancing its row and column totals, but a forecasting operation implies the use of a certain theoretical model—whether deterministic or not. To forecast an I-O table, the statistical data concerning final demand  $Y_i$  and the value added  $D_j$  are generally available or could be obtained on the basis of existing information. For example, in the case of Eurostat, this information exists for the time horizon year 2013 while

**Table 9:** Symmetric I-O Tables for domestic output at basic prices(year 2007)

Products (CPA)	Inputs of products							Total
	P1	P2	P3	P4	P5	P6	Others	
P1	53529	189315	3126	24311	3626	6784	138648	419340
P2	86528	2624065	386717	535924	229362	266302	3189197	7318095
P3	2656	45671	344941	48062	111100	46147	1236719	1835296
P4	36335	756877	121086	869135	258074	187202	2573895	4802604
P5	26539	673893	188669	730656	1264384	352132	2146659	5382932
P6	5005	83827	11606	72494	107755	257351	3172149	3710187
Others	208748	2944446	779152	2522023	3408631	2594266	1599374,605	14056642
Total	419340	7318096	1835296	4802604	5382933	3710185	14056642	

Source: based on <http://ec.europa.eu/eurostat/web/esa-supply-use-input-tables/overview>

where :

P1	Products of agriculture, hunting and fishing
P2	Industrial products (incl Energy)
P3	Construction work
P4	Trade, transport and communication services
P5	Financial services and business services
P6	Other services
Others	Use of imported products, cif Taxes less subsidies on products Value added at basic prices
Total	Row or column totals

neither the value of global product  $X_j$  nor the matrix of cost structure  $[a_{ij}]$  are known for that forecasting period. Here, the last updating of IO tables concerns the period 2007. Building an I-O table is time consuming and a lot of information, particularly concerning the transaction matrix (including import elements), is not easy to assess. Economic information from different enterprises or industries must be updated on the basis of new flows of additional information. As a result, getting a final version of that table comes long after with many lag years. Under such conditions, finding a workable methodology for setting up a robust prediction technique of an I-O table should bear precious advantages.

Therefore, the key question is: Based on the I-O table of the previous period, on vectors  $Y_i^p$  and  $D_j^p$  (both from the forecasting period), is it possible, using the connection (3.2), to estimate the unknown  $IOtable^p$  of the forecasting period? Of course, the

answer is not affirmative if we use the classical statistical-mathematical approach to solve that inverse problem. On the one hand, we do not have information about matrix  $A^p$  of the forecasting period to derive the value  $X^p$  using relation (3.2) and, in this way, to determine  $IOtable^p$ . On the other hand, as an effect of the possibility of estimating  $X^p$ , for example using a dynamic investment model, the new question that arises is: Disposing of  $X^p$  and of  $Y^p$  of the forecasting period, could one possibly determine the matrix coefficients  $A^p$  on the basis of the same relation (3.2)? Since the IO table has size  $N \times N$ , and we have by assumption only information about final demand  $Y^p$  and global product  $X^p$  of each sector, this means that we have  $(N - 2) \times (N - 2)$  degrees of freedom, where  $N$  is, once again, the number of branches in the IO table. Such a problematic belongs then to the category of inverse problems, which suggests that there may be an infinity of matrices  $A^p$  that could reproduce the identical values of final demand  $Y^p$  and the global product  $X^p$ . Among them, we will choose the one that best maximizes consistency information between the prior, data, and the posterior. We can retrieve a solution proposal to that problem in the second part of this monograph about the maximum entropy principle or relative (cross) entropy. Let us again present it below in the context of updating and forecasting an I-O table on any other form of its extension.

#### 3.3.3.4 The Non-Extensive Cross-Entropy Approach and I-O Table Updating

As suggested above, in the recent literature there are several methods for updating and balancing elements of national accounts balance sheets, for instance, an I-O table when equality of corresponding sums of columns and rows is required. Some of their limits have been emphasised here. Preference is then given to methods based on statistical theory of information for their capacity to adjust information when structural changes affect the economy or when additional consistent information has to be added to what already exists. The most frequently used theoretic-information methods are the Maximum Likelihood, Bayesian method of moments, and methods based on the maximum entropy principle. Through their article, Giffin & Caticha (2007) have proven that the principle of maximum entropy represents a generalization of the Bayesian approach as a method of inference on the basis of an *a priori* information. Probably for these reasons, application of the cross-entropy approach to balance the social accounting matrix has been widely adopted in empirical application during recent years (e.g., Robinson et al., 2001).

As we know from Part II of the monograph, on the basis of the results of Shannon (1948) and Jaynes (1957), Kulback-Leibler (KL) (1951, 1957) and Good (1963) have proposed the principle of minimum (relative) entropy. This principle aims at assessing *a posteriori* parameters (probabilities  $P$ ) of the most plausible, shortest divergence in relation to *a priori* parameters (probabilities  $Q$ ), under restrictions related to data moments, normalization condition, or any other *a priori* information presenting con-

sistency with probabilities in the criterion function. The formulation in the case of discrete events is like in (3.16) and, thus, we have:

$$\text{Min}(r_p, r_o) \rightarrow I_q[r \| r_o] = \frac{1}{q-1} \sum_j r_j [(r_j)^{q-1} - (r_{0j})^{q-1}] - \sum_j (r_j - r_{0j})(r_{0j})^{q-1} \quad (3.16')$$

or, as traditionally done in the case of Shannon-Gibbs entropy,

$$\text{Min}_{KL}(P \| Q) = \sum_{i=1}^n p_i \ln \left( \frac{p_i}{q_i} \right) = P' \ln P - P' \ln Q, \text{ for } i = 1, \dots, n$$

under restrictions:

$$X_j = \sum_{j=1..n} \frac{(r_j)^q}{\sum_{j=1..n} (r_j)^q} X_j \quad (3.18)$$

$$\sum_{j=1..n} D_j \frac{(r_j^{'d})^q}{\sum_{j=1..n} (r_j^{'d})^q} = \sum_{i=1..n} \frac{(r_i^{y'})^q}{\sum_{i=1..n} (r_i^{y'})^q} Y \quad (3.19)$$

$$\Omega(h) = F \quad (3.20)$$

$$P'I = 1 \text{ and } P \geq 0. \quad (3.21)$$

We adopt Shannon-Gibbs symbols in the criterion function above. Correspondence with the Tsallis criterion function in (3.16) is easy to do. The symbol P corresponds to r and Q to r<sub>o</sub>. Matrix P stands for posterior probabilities guaranteeing the balance of a previously unbalanced table, the elements of which sum up to unity by column (sector). Q is a matrix of coefficients from a known table. In the case of the I-O table (see Table 2) elements of Q are derived by dividing each column element by its total. They then represent input coefficients except the case of the final demand column, where these coefficients explain the structure of product consumption for a given final demand institution. Thus, its elements must satisfy the additivity condition. Equation (3.18) demonstrates that the column total must match with corresponding row elements multiplied by corresponding probabilities (coefficients) P<sub>j</sub>. Equation (3.19) states equality between value added and final demand totals, with  $p_j^{'d}$  and  $P_i^{y'}$  being transposed respective vector of sectorial value added components and vector of structural coefficients of final demand. Probabilities are presented in escort distribution formulation presented in footnote 17. Functional h in (3.20) gives a piece of information which has a significant relationship (consistency) with probabilities in the criterion function. This may be, for example, a macroeconomic balance equation or any distribution of a treatment of errors. Equation (3.21) is one of the additivity conditions of probabilities and reminds us that any probability can take a negative value.

### 3.3.3.5 Forecasting an I-O Table

Let us first formulate the model. As previously stated, a fundamental problem results from the lack of information about sectorial global product  $X_j^p$  or equivalent input transactions  $x_{ij}$  since we assume final demand to be known. Thus, before implementing entropy formalism (3.16), we must first estimate the values  $X_j^p$ . For this purpose, we propose using a dynamic model<sup>25</sup> of I-O in which investment is the endogenous variable in the context of accelerator analysis of macroeconomic theory. According to macroeconomic theory, it is expected that investment is induced if final demand is expected to grow. Based on the preceding period, traditional assumptions of such a model are as follows:

- the investment is a function of the expected growth,
- information about the coefficients of material and production factors are available,
- information about capital ratios are available,
- all economic sectors are in full effect,
- capital has an infinite lifetime.

We then have the following dynamic equation of global product:

$$X^{t+1} = B^{-1}[(I - A + B)X^t - Y^t] \quad (3.22)$$

and

$$Inv^t = B(X^{t+1} - X^t) \quad (3.23)$$

where:

- $X^{t+1}$ : global product,
- $Inv^t$ : induced investment,
- $B$ : coefficients of capital production,
- $A$ : coefficients of material,
- $Y^t$ : final demand,
- $t$ : the index of the time.

It should be added that value of global product derived in this way does not constitute the ultimate result of the whole process of forecasting. This value only provides the information a priori in terms of a Bayesian viewpoint and it will be changed in the process of entropy minimization, as previously explained. At the end of the procedure, applying relation of relative entropy in (3.16), we upgrade I-O coefficients for the period  $T + 1$  to obtain a new post entropy I-O table. Thus let us consider a known I-O matrix  $BT$  of period  $T$  in the form of Table 2.1 displaying coefficient structure, obtained

---

<sup>25</sup> See EUROSTAT.

after dividing each column element by the corresponding total column. The problem is retrieving new matrix  $B_{T+1}$  of period  $T + 1$  for which we already know global product from (3.22) and sectorial aggregates of final demand. As is often the case in real life, let us suppose sectorial value added is known, too. If our I-O table has a dimension  $n \times n$  (with  $n \gg 2$ ), then we have  $(n - 1)(n - 2)$  degree of freedom. Building on the formulation (2.40–2.41) involving an inverse problem, we obtain:

$$\text{Min}(b_1, b_0) \rightarrow I_q[B_1 \| B_0] = \frac{1}{q-1} \sum_i \sum_j b_{ij} [(b_{ij})^{q-1} - (b_{0ij})^{q-1}] - \sum_i \sum_j (b_{1ij} - b_{0ij})(b_{0ij})^{q-1} \quad (3.24)$$

under restrictions:

$$X_{1i} = \sum_{j=1..n} \frac{(b_{1ij})^q}{\sum_i (b_{1ij})^q} X_{1j} \quad (3.25)$$

with  $i=j$  in the case  $X_1$ ,

$$\sum_{j=1..n} D_j \frac{(b'_{1ij})^q}{\sum_{j=1..n} (b'_{1ij})^q} = \sum_{i=1..n} \frac{(b_{1ij})^q}{\sum_{i=1..n} (b_{1ij})^q} Y \quad (3.26)$$

$$\sum_{i=1..n} b_{1ij} = 1 \quad (3.27)$$

$$\Omega(h) = F \quad (3.28)$$

$$b_{1ij} \geq 0. \quad (3.29)$$

In the present model  $B_t[b_{ijt}]$  with  $(i = 0 \text{ and } 1)$  is a matrix of I-O coefficients for the period  $T_t$ . Thus,  $B_0$  is *a priori* known, and we seek to estimate  $B_1$ , that is, for  $t = 1$ . Since both coefficients display probability properties of continuously belonging to interval zero-one, of summing up to unity (by column), and of additivity, we do not need to reparameterize  $B$ , and it will be taken for probability. Tsallis complexity index  $q$  informing about departure from Gaussian to power law distribution appears in the criterion function (3.24) and in constraints (3.25) and (3.26) under formulation of escort distribution, as earlier explained in footnote 9. Index one on variable  $X_{1j}$  (explaining “global product”) in (3.25) refers to the period of forecast. It is equal to the vector column of coefficients multiplied by the total column, taking then into account weights related to escort distribution. A total obtained this way is equal to the total corresponding line  $X_{1i}$ . This applies equally well to (3.26), which means that totals of value-added  $D_j$  and of final demand  $Y_i$  are equal. The rest of the restrictions, i.e., (3.27–3.29) are as in (3.21).

### 3.3.4 Application: Forecasting the Aggregated 27 EU IO Coefficients

In this example, we will present two cases where we apply the non-extensive entropy principle<sup>26</sup> to forecast the aggregated 27 EU I-O coefficient table for the period 2007. The basis of that forecasting is the I-O table of the previous year, 2006, and some information related to the targeted I-O coefficients of 2007. Both tables can be found at the Eurostat statistics site: <http://ec.europa.eu/eurostat/web/esa-supply-use-input-tables/overview>. Since we already know the true values<sup>27</sup> of the I-O table of 2007, it will be much easier to assess the performance of the applied technique.

In the first case, besides information from the 2006 table, we are supposed to additionally have data on sectorial value added and final demand of the targeted period 2007. In this example, sectorial global product is neither known nor assessed. We then consider here a case which is more unfavourable than it appears in empirical investigations where, in general, many values of the targeted I-O table are assessable on the basis of existing theory or through intuition.

In the second case, besides the information used in the first case, we are supposed to additionally have the possibility of estimating some reliable inter-sector transactions of the targeted period 2007 and to incorporate them into the model as new data, according to the Bayesian model. Then, the following list of accounts which have been arbitrarily selected from the true 2007 table and information related to them have to be incorporated into the entropy model:

tobac	Tobacco products
pulpap	Pulp, paper, and paper products
Insur	Insurance and pension funding
compserv	Computer and related services
uranores	Uranium plus ores

Among the 60 accounts in the table—we have a NACE A60 breakdown—these five accounts represent 5% of all accounts and around 2% of their transaction total. Thus, in spite of using that additional piece of information, we are still dealing with an inverse problem consisting of recovering  $(n - 6) \times (n - 6)$  coefficients of the I-O table on the basis of the coefficient table from the previous period, 2006, and some additional, random information from the forecasted period. Not included<sup>28</sup> in this book,

<sup>26</sup> In this study, outputs from Shannon and Tsallis entropies are similar.

<sup>27</sup> As can be verified at the electronic address [http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=naio\\_15\\_agg\\_60\\_r2&lang=en](http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=naio_15_agg_60_r2&lang=en), this version of the 2007 table does not yet seem to be definitive since some accounts are not balanced. The implication is that input-output coefficients related to that table cannot be considered known with certainty. At the same electronic address, we can find input-output tables of other periods, such as for 2006. However, since we have made small changes to these two tables, they will be displayed in the annex.

<sup>28</sup> Due to lack of space, these tables cannot be presented in this book. Nevertheless, they are here



two output tables allow for assessing the performance of the proposed technique. The first table F1 displays the precision error for the ex post forecasted inter-sector transactions for the period 2007 and related to a minimum of hypotheses, that is, the knowledge of sectorial values added and final demand. Thus, that table displays the forecasted I-O transaction table for the period 2007 with the minimum a priori information. The second table F2 displays the precision error for the forecasted inter-sector transactions of the period 2007 in the second hypothesis, that is, with some knowledge on inter-sector transactions of the five above listed accounts.

What could we learn from these outputs? Here, we will limit our comments to general aspects of cross-entropy formalism. More details will be provided in the coming chapter concerning the procedure of updating a social accounting matrix, as a generalized case of an I-O table. The table “F2”, as already mentioned, shows the degree of precision of the entropy procedure in forecasting the I-O coefficients of the period 2007. The only a priori information remains the I-O transaction matrix of 2006, sectorial values added and final demands of the period 2007. All computations have been carried out with the GAMS code. To measure the precision of the I-O coefficient, we have used the next average error variance coefficient (AEVC):

$$AEVC = \frac{\sum_i \sum_j (a_{ij} - \hat{a}_{ij})^2}{(n-1)(n-1)}$$

The coefficients  $a_{ij}$  represent the true coefficients displayed by the 2007 table of the I-O matrix prepared by Eurostat. Though the table is not fully balanced, we still consider it sufficient to correspond to a true data generating system. The forecasted coefficients  $\hat{a}_{ij}$  are supposed to be affected by a certain margin of error. The denominator explains the degree of freedom of a Warlasian equilibrium table, i.e., the number of accounts<sup>29</sup> but one.

Let us first consider a simple benchmark measure of the information divergence between the 2006 and 2007 tables. In these circumstances, values of coefficients  $\hat{a}_{ij}$  represent the period 2006 and values of the coefficients  $a_{ij}$ , that of 2007. Thus, the derived variance value AEVC is 0.00003846. In the context of information theory, this value corresponds to the minimum entropy result when any restriction (except normality conditions) to the criterion function has been enforced. It corresponds, too, to the maximum entropy outputs in the same conditions.

Now, if we consider the first case when we know the inter-sector values added and final demands, we get a new AEVC equal to 0.00002932 and representing a discrepancy between the true and the estimated values. In comparison with the benchmark coefficient, the new piece of information has brought about an overall coeffi-

---

numbered F1 and F2 so that the interested reader can, on request, get a copy of them from the author.

<sup>29</sup> For computational reasons, two accounts concerning the mining sectors have been aggregated into one account.

cient variance reduction of around 24%. Though this coefficient is very small, a look at the non reported in this book outputs (table “F1”) could reveal a high variation in these coefficients. As it has been noticed in Shannon entropy econometrics literature, entropy formalism tends to shrink small probabilities in favour of those higher. Through this example, we discover the same stylised fact found in the case of Tsallis entropy formalism. Inside that table, we notice the presence of a 100% shift of many values. This is the result of small probability shrinkage, as explained above. To reduce these variations, we need to impose additional restrictions on the model. On theoretical grounds, the formal causes of small probability shrinkage is well presented in the early presented work (see previous chapters) of Golan et al., (1996) as an extension of the work of Greenberg et al., (1989) on the family of Stein-rule estimators<sup>30</sup> proposed by Stein & James (1961).

In the last experiment carried out, we additionally consider that some transaction values of the above listed accounts are known. We then get a new AEVC equal to 0.00001042502. In comparison with the benchmark coefficient, the new piece of information has brought about an overall coefficient variance reduction of around 74%. Noting that the small variance AEVC naturally represents the average for all values inside the table, this seems to contrast with the relatively higher precision error affecting many cell values of the tables F1 and , F2. In general, as underscored above, higher imprecisions tend to affect smaller values (then with smaller weights) inside the table cells.

### 3.3.5 Emission Coefficients Forecasting: A Theoretical Model

One may now combine the two previous procedures of retrieving information in the case of the inverse problem. The first procedure has assessed emission coefficients when sectorial global product and the transactions matrix were available (see section III.3.2.2). The second has assessed the I-O table on the basis of knowledge of sectorial final demand of current period and a priori information about the table of preceding period. Now, the interesting problematic could be assessing sectorial emission coefficients on a knowledge basis of:

- total emission by theme of the forecast period,
- sectorial final demand of the forecast period,
- an input-output table of a recent period.

---

**30** The main idea is that if three or more unrelated parameters are measured using the James-Stein estimator, their total square error will be lower than in the case of the least square (LS) estimator well known to provide the lowest variance among all other linear estimators. However, when each parameter is estimated separately, the LS estimator leads to higher performance. This is so because of the tendency of the Stein family of estimators to shrink small probabilities of the estimated system.

**Table 10:** A Simplified Environmentally extended I-O table

i	Sector	Flows					Final demand		Total
1	$X_1$	$x_{11}$	$x_{12}$	$x_{13}$	...	$x_{1m}$	$Y_1$	0	
2	$X_2$	$x_{21}$	$x_{22}$	$x_{23}$	...	$x_{2m}$	$Y_2$	0	$X_2$
...	...	...	...	...	...	...	...	0	...
m	$X_m$	$x_{m1}$	$x_{m2}$	$x_{m2}$		$x_{mm}$	$Y_m$	0	$X_m$
m+1	$D_j$ Value-added	$D_1$	$D_2$	$D_3$	...	$D_n$	0	0	$D$
...	$E_j$	$e_{11}$	$e_{12}$			$e_{1m}$	0	$e_{y1}$	$E1$
...	...	...	...	...	...	...	0	...	...
n	(Energy themes)	$e_{f1}$	...	...	...	$e_{fm}$	0	$e_{yf}$	$E_f$
	Total	$X_1$	$X_1$	$X_1$	...	$X_m$	$Y_v$	$Y_e$	Total

Source: own elaboration.

Solution of such an inverse problem can be set up as follows:

$$\text{Min} \rightarrow I_q[p||p_0] = \frac{1}{q-1} \sum_j p_j [(p_j)^{q-1} - (p_{0j})^{q-1}] - \sum_j (p_j - p_{0j})(p_{0j})^{q-1} \quad (3.17')$$

subject to

$$X_{1i} = \sum_{i=1..n} \frac{(p_{1ij})^q}{\sum_i (p_{1ij})^q} Z_h X_{1j} \quad (3.32)$$

with  $i=j$  in the case  $X_1$ ,

$$\sum_{j=1..n} D_j \frac{(p'_{1ij})^q Z_h}{\sum_{j=1..n} (p'_{1ij})^q} = \sum_{i=1..n} \frac{(p'_{1ij})^q Z_h}{\sum_{i=1..n} (p'_{1ij})^q} Y \quad (3.33)$$

$$\sum_{j=1}^n \sum_{h=1}^M p_{1jh} = 1$$

$$E_i = \sum_{j=1}^n \sum_{h=1}^M \frac{p_{1jh}^q}{\sum_{j=1..n} p_{1jh}^q} Z_h X_{1j} \quad (3.34)$$

$$\sum_{m=1}^h e_{fm} = 1$$

with  $h=f+2$ , i.e., additional columns due to the demand of emissions by institutions (one of them having zero value since demand of emissions must have a separate column).

All symbols are as before. Note that in this case the environmentally extended I-O matrix has the form presented in Table 10 below. Input coefficients are derived from global column totals, including, thus, quantities (instead of values) representing emissions.

### 3.4 Conclusions

As suggested above, generally, new information should render more homogeneous divergence between the true coefficients and those forecasted. According to the above results, we observe that new restrictions added to the model lead to a significant reduction of errors. This is in accordance with the rule of thumb proposed by Chisari et al. (2012) in section III.3.3.1 about particular conditions explaining the superiority of entropy approaches over the RAS technique.

The obtained AEVC coefficients naturally present a random character and different experiments would produce different values. However, as expected through Bayesian formalism, new data evidence will always tend to reduce the level of uncertainty or entropy.

The true difficulty in assessing a new methodology to assess a complex information system—like the one represented by an I-O table—is that, due to instruments of measure or adopted economic hypotheses, a part of the observed data may not be accurate. This can be even worse in the case of general equilibrium systems in which the balance of the whole system—or accounts—may be more or less forced. This observation is particularly true in developing countries where statistical information management can be more challenging. What we intend to explain here is that, faced with such circumstances, the output performance of the non-extensive entropy approach should, consequently, be taken with a certain margin of error. The quality of priors and model data will always play a central role.

## Bibliography – Part III

- Almon, & Clopper. (2000). Product-to-product tables via product-technology with no negative flows. *Economic Systems Research*, Vol. 12, No.1,, 12(1), pp. 27–43.
- Avonds, & Luc. (2007). *The input-output framework and modelling assumptions: considered from the point of view of the economic circuit*. The 16th International Input-Output Conference. Istanbul: International Input-Output Association.

- Bacharach, M. (1971). *Biproportional matrices and input-output change*. Cambridge: Cambridge University Press.
- Bekker, J., & Aldrich, C. (2011). The cross-entropy method in multi-objective optimisation: An assessment. *European Journal of Operational Research*, 211(1):112–121.
- Beutel, J., & De March, M. (1998). Input-output framework of the European System of Accounts (ESA 1995). *Paper presented at the Twelfth International Conference on Input-Output Techniques*. New York: International Input-Output Association.
- Caticha, A., & Giffin, A. (2007). Updating probabilities with data and moments. *Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*. NY, USA: Department of Physics, University at Albany–SUNY.
- Chisari, O.O., Mastronardi, L.J., & Romero, C.A. (2012). Building an input-output Model for Buenos Aires City. *MPRA* (40028).
- Deming, W.E., & Stephan, F.F. (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. *The Annals of Mathematical Statistics*, 11(4):427–444.
- Eurostat. (2002). *The ESA 1995 Input-Output Manual - compilation and analysis (draft)*. Luxembourg.
- Adom Giffin, A. & Caticha, A. (2007). *Updating Probabilities with Data and Moments*. Presented at the 27th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Saratoga Springs, NY, July 8–13, 2007.
- Gosh, A. (1964). *Experiments with Input-Output Models, An Application to the Economy of the United Kingdom, 1948–1955*. Cambridge: Cambridge University Press.
- Greenberg, E., Denzau, A.T., & Gibbons, P.C. (1989). Bayesian Estimation of Proportions with a Cross-Entropy Prior. *Communications in Statistics - Theory and Methods*, 18(5):1843–1861.
- Lemelin, A., Fofana, I., & Cockburn, J. (2005). *Balancing a social accounting matrix: Theory and application*. Centre Interuniversitaire sur le Risque des Politiques Economiques et l'Emploi (CIPREE), Université de Laval.
- Leontief, W. (1941). *The Structure of American Economy, 1919–1929*. New York: Cambridge, (mors): Harvard University Press, (Second Ed. 1951, , Oxford University Press).
- Leontief, W. (1966). *Input-Output Economics*. New York: Oxford University Press.
- Leontief, W. (1970). The Dynamic Inverse. *Contributions to Input-Output Analysis*, pp. 17–46.
- Leontief, W. (1986). *Input-Output Economics, Second Edition*. New York: Oxford University Press.
- McDougall, R.A. (1999). Entropy Theory and RAS are Friends. *Working Paper No. 6, GTAP*.
- Miller, R., & Blair, P. (1985). *Input-output analysis: Foundations and extensions*. Englewood Cliffs, N.J.: Prentice-Hall.
- Robinson, S., Cattaneo, A., & El-Said, M. (2001). Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods,. *Economic Systems Research*, 13(1):47–64.
- Snower, D.J. (1990). New methods of updating input-output matrices. *Economic System Research*, 2:27–38.
- Stäglin, Reiner (1972), MODOP - Ein Verfahren zur Erstellung empirischer Transaktionsmatrizen, in: Münzner, H. and Wetzel, W.: Anwendung statistischer und mathematischer Methoden auf sozialwissenschaftliche Probleme, Würzburg, pp. 69–81.
- Stein, C., & James, W. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp.*:361–379).
- Stone, R. (1955). Input-Output and the Social Accounts. *Proceedings of an International Conference on Input-Outputs Analysis*. New York: University of Pisa.
- Stone, R. (1981). Input-Output-Analysis and Economic Planning: A Survey. *Mathematical Programming and its Applications*. Milan: Angeli.
- Stone, R. (1984). Balancing the National Accounts: The Adjustment of Initial Estimates. *Demand, Equilibrium and Trade*.
- Toh, M.H. (1998). The RAS approach in updating input-output matrices: An instrumental variable interpretation and analysis of structural change. *Economic Systems Research*, 10(1):63–78.

- Tomaszewicz, Ł. (1992). *Przepływy międzygaleziowe. Elementy teorii*, Wydawnictwo UŁ (rozdz. 1–3, 7,).  
Łódź: University of Łódź.
- Tomaszewicz, Ł. (2005). Metody analizy input-output. *Przegląd Statystyczny*, 52(4), strony 15–22.
- United Nations. (1993). *Integrated Environmental and Economic Accounting, Handbook of National Accounting*. New York, USA: United Nations.
- Eurostat: <http://ec.europa.eu/eurostat/web/esa-supply-use-input-tables/overview>.

