Team Research Report, No. MM-2018-1
(Műhelytanulmány, MM-2018-1)
Centre for Public Administration Studies
Corvinus University of Budapest
Hungary

ADDITIVE RAS AND OTHER MATRIX ADJUSTMENT TECHNIQUES FOR MULTISECTORAL MACROMODELS

Krisztián Koppány[1] - Tamás Révész[2]

[1] Széchenyi István University, Győr, Hungary (Széchenyi István Egyetem)

[2] Corvinus University of Budapest (Budapesti Corvinus Egyetem)

Content

1. Introduction 4
2. The matrix adjustment problem and its most commonly used solution methods ...6
2.1. The matrix adjustment problem 6
2.2. The RAS method 8
2.3. Other matrix adjustment methods 12
3. Adjustment methods for matrices with negative and zero cells and margins 14
3.1. Non-sign-preserving methods 21
3.2. The additive RAS method 24
3.3. The model estimating and transforming the 2010 EU I-O tables to GTAP format 32
4. The most important matrices to be adjusted in multisectoral macroeconomic analyses 37
4.1. Input-output tables "type A" 38
4.2. Matrices of taxes less subsidies on products 40
4.3. Transformation matrices of consumption and investment 42
4.4. Further matrices to be adjusted in the national economy statistics......................44
5. Combined and sequential applications of biproportional matrix adjustment methods 48
6. Summary 49
References 50

Additive RAS and other matrix adjustment techniques for multisectoral macromodels

Tamás Révész and Krisztián Koppány

Abstract

This paper gives a brief overview of the biproportional matrix adjustment problem. We focus on the most frequent case with certain and consistent row and column sums, and no special conditions to the cells of the matrix. After the definition and mathematical formulation of the problem, we describe the so-called distance and entropy functions assigning a non-negative real number to the difference of the estimated and reference matrices. These functions are to be minimized subject to given row and columns sums, and in special cases, some non-negativity and sign-preserving conditions. For these models, we present some iterative solution methods, among them the so-called additive RAS algorithm developed and used first by Révész (2001). On one hand, in the case of non-negative reference matrix and positive marginal conditions, one version of additive RAS gives the same solution as the standard RAS, and on the other hand, in the case of negative cells, but sign-preserving margins, another version gives the same solution as the improved normalized squared differences (INSD) model without penalties for sign-switching. We demonstrate that additive RAS is more efficient and more aesthetic than the GRAS and other iterative solution methods used by previous authors. In the case of small differences, additive RAS, especially the flexible version, tends to be sign-preserving, unless it is forced by sign-switches of the margins. Using the example of Lemelin (2009), we demonstrate that additive RAS performs very well even in such an extreme sign-switching case, moreover, it gives the best solution compared to other algorithms. The paper overviews some standard matrix balancing problems in practice, where such methods can be used. The most important conditions for the successful application are the knowledge about the economic phenomena under investigation and the deep understanding of the related reference matrix. For an example of this, a current research project is presented, where both additive RAS and other more complex adjusting models were used and showed a good performance.

Keywords: biproportional matrix adjustment; mathematical programming; distance function; reference matrix; RAS-method; sign flip; zero margins; least squares; Lagrangian; first order conditions; mean average deviation; input-output table;

Subject classification codes: include these here if the journal requires them

1. Introduction

Estimating the elements of a matrix, when only the margins (row and column sums) are known, is a standard problem in many disciplines. Certain cells can be known (in this case subtracting them from the relating row and column sums, the problem can be converted to the case of unknown cells) or we have only indirect information about them. Generally, this indirect information is a reference matrix, for which we assume that it has the row and column structure 'similar' to the target matrix. The reference matrix can be known counterpart of the target matrix for some previous period or different unit of observation. Similarity, or the opposite of this, can be measured by a 'distance' function with the objective to minimize its value.

A typical example is the so-called trip matrix estimation problem. Here, a general x_{ij} element of the **X** matrix denotes the quantity of goods, the number of people transported, or the number of trips between *i*th and *j*th places. Performing a full-survey to find out the **X** for each period can be very expensive and time-consuming. But if we have **A**, the counterpart of **X** for some previous period, and the row and column sums of the current **X** matrix, that is the number of trips originating and terminating in each zone are known for current period, we can try to bend the reference matrix **A** to these new margins, preserving its structure the best as possible.

Many researchers from several disciplines developed solution strategies to this kind of matrix adjustment problems, in some cases independently, being unaware of other's achievements. That's why mathematically equivalent methods are called diversely in distinct disciplines. In transport science these are known as Fratar or Furness (Ortúzar and Willumsen, 2011). In economics the same procedure mainly used for balancing input-output tables is called RAS. This is in the focus of this paper.

The most common methods of the above matrix adjustment problems can be classified into two large groups: the so-called 'entropy' models (which contain the logarithm function from the information theory, the classical RAS is also like this) and the models with quadratic objective functions based on the principle of least squares. In the literature a vast discussion emerged about which the method and under what circumstances is more efficient and more reliable. According to experience so far, the best choice depends on the mathematical properties of the reference matrix, target margins, and expectations about the target matrix (non-negativity, zero values, sign switching, sparse matrix etc.), and the economic content of the matrix. Negative and zero values, for example, impede the use of standard methods. Often happens that one of the margins, or some cells of the margins of the adjusted matrix must be zero, while the matrix should continue to contain positive and negative values as well.

The discussion between the authors covered not only the practical application (which method gives better estimates or works more reliably, ensures that the elements of the reference matrix are preserved, for instance), but also the mathematical properties of the techniques (are they biased, do they work in special cases, do they give a unique solution, the same are the solutions of models defined in different ways[3] etc.).

This paper briefly overviews these methods and attributes highlighting the mathematical background first, then focusing on the statistical problems and estimation

[3] The problem (or the objective function) should be defined in the context of transactions or coefficients, for example, see Mesnard (2011) chapter 4.2.

features of matrices in multisectoral modelling practice (various transaction and transformation matrices needed for computable general equilibrium (CGE) models). In this context, special emphasis is given to the 'additive RAS' method of Révész (2001) developed for estimating matrices with a negative reference elements and/or non-positive target margins, and the illustration of its efficiency with numerical examples. Based on practical experience and the lessons learned from the literature, we propose ways of further development and future application of estimation methods.

2. The matrix adjustment problem and its most commonly used solution methods

This Section first formalizes the standard matrix adjustment problem, then overviews the two major groups of the proposed solution methods for non-negative matrices, the 'entropy models' and the quadratic objective function models, and their relationship.

2.1. The matrix adjustment problem

The matrix adjustment problem most commonly discussed in the literature can be formulated as follows (see for example Lahr and Mesnard (2004), which is the basis for the following review).

Let $\mathbf{X}^*$ be an m x n unknown matrix, for which row sums are equal to the known $\mathbf{u}$ column vector, and column sums are equal to the also known $\mathbf{v}$ column vector (that is, $\mathbf{X}^*\mathbf{1} = \mathbf{u}$, $\mathbf{1}^T\mathbf{X}^* = \mathbf{v}$, where $\mathbf{1}$ is the summation vector and ^{T} denotes transpose).

If we also have an m x n reference matrix $\mathbf{A}$ (also known as prior), which is similar to $\mathbf{X}^*$ in its structure, then it can be estimated with the matrix $\mathbf{X}$, also of the size m x n, which has row sums equal to the column vector $\mathbf{u}$ and column sums equal to $\mathbf{v}$ (i.e., $\mathbf{X1} = \mathbf{u}$, $\mathbf{1}^T\mathbf{X} = \mathbf{v}$), such that $\mathbf{X}$ is the most similar to the reference matrix $\mathbf{A}$ in some sense.

In the paragraph above, the use of the term ‘structure’ is justified by the following considerations. Of course, depending on the definition of the ‘similarity’ (or the ‘deviation’ or ‘distance’) of the two matrices (**A** and **X**), the solution to the problem (**X**) may be different. Even if we lock the formula for comparing the two matrices, the solution will naturally continue to depend on the reference matrix **A**. If, however, it depends not only on the structure of **A** (the internal proportions of its elements), but also on its ‘level’ (that is, multiplying **A** by a γ scalar yields different solution), then it is advisable to first modify **A** (proportionally adjust by multiplying by the γ scalar) that grand total ($\mathbf{1}^T\mathbf{A}\mathbf{1}$) are equal to the grand total of ($\mathbf{1}^T\mathbf{X}^*\mathbf{1}$), i.e. $\mathbf{1}^T\mathbf{A}\mathbf{1} = \mathbf{1}^T\mathbf{v} = \mathbf{u}\mathbf{1}$. This makes it only possible that if with this equal scaling, conditions $\mathbf{A}\mathbf{1} = \mathbf{u}$ and $\mathbf{1}^T\mathbf{A} = \mathbf{v}$ are also fulfilled, the modified reference matrix **A** is the solution **X** itself. ‘Levelling’ of the reference matrix **A** reduces or (if it meets the conditions) eliminates the need for further correction.

Of course, with a given formula of the ‘similarity’ of two matrices, it is possible that the problem has several solutions for which this formula gives the same value. However, if **A** is irreducible, the set of possible solutions is compact, and the objective function can be continuously differentiated over that set, there is only one solution, which is true for the methods involved here (Mesnard, 2011). This problem is not discussed in general in this paper, however, we will return to the point when discussing the concrete formulae of ‘similarity’.

In any case, the matrix adjustment task can be defined as a mathematical programming problem, where the goal is to find the optimal value of the target function (the maximum of the similarity formula or the minimum of a monotonous increasing function of the deviation), subject to the constraints $\mathbf{X}\mathbf{1} = \mathbf{u}$ and $\mathbf{1}^T\mathbf{X} = \mathbf{v}$ (and possibly some nonnegativity or sign-preservation conditions). If the constraints are linear, the optimum can be determined by the method of the Lagrangian multipliers. In the

Lagrangian function, these multipliers penalize for deviations from constraints and express how much a unit increase in constraints changes the optimum value. Some authors initially start from the Lagrangian function, but modify it, for example, by introducing deviations from constraints not only by multiplying the corresponding Lagrange multiplier, but by logarithm of the resulting multiplications, see, for example, Günlük-Şenesen and Bates (1988). This has benefits in the entropy models with logarithmic objective functions, if so the relations between the element of the reference matrix and the λ_i and τ_j Lagrangian multipliers for the deviations from the prescribed row and column sums can be expressed in a simple multiplication form of $x_{i,j} = a_{i,j} \cdot \lambda_i \cdot \tau_j$ in the solution (although not optimal for the original program) generated by the first-order conditions.

One can raise that instead of looking at how much the initial structure is preserved; the method should rather be looked at as to how much it changes elements of the prior in line with the expected changes in the margins. For example, if both the row and the column sums increased, then it is a legitimate expectation that the element at the interSection of them also increases (of course, only to the justified extent). This issue will be concerned in the next chapters, but we cannot undertake general discussion. In any case, it is worth thinking about compiling a criteria system that examines such and similar aspects.

2.2. The RAS method

The most obvious solution algorithm for the matrix adjusting problem is the RAS method, which was first documented in the 1930s, used in input-output modelling in the 1940s and was disseminated in economic literature by Sir Richard Stone (Stone 1961, Stone and Brown 1962).

The first step of the RAS iteration process is to multiply the rows of matrix **A** by the corresponding ratio of the prescribed and the actual row sums u_i/b_i (where **b** denotes the column vector of actual row sums, and b_i is the ith element of **b** for row i), hence adjusting the matrix horizontally to **u**, the vector of target row sums. Then the columns of the resulting matrix should be multiplied by the corresponding ratio of the prescribed and the actual column sums v_j/s_j (where **s** denotes the column vector of actual column sums, and s_j is the jth element of **s** for column j), hence adjusting the matrix vertically to **v**, the vector target column sums.[4] The second iteration starts with $\mathbf{A}^1$, the resulting matrix of the first iteration, and so on and on. Thus, in the ith step one can obtain matrix $\mathbf{A}^i$ performing the row and column direction adjustment referred to above on the matrix $\mathbf{A}^{i-1}$, the result of the (i-1)th iteration. This process is usually convergent.[5]

The limit of the matrix sequence $\mathbf{A}^i$ yields **X**, which is the solution of the following mathematical programming problem (Bacharach 1970)[6]:

$$\mathbf{X1} = \mathbf{u},\ \mathbf{1}^{\mathbf{T}}\mathbf{X} = \mathbf{v},\ \sum_{i=1}^{m}\sum_{j=1}^{n} x_{i,j}\ln(x_{i,j}/\mathrm{a}_{i,j}) \rightarrow \min. \quad (1)$$

The solution of (1) using the method of Lagrangian multipliers one can obtain

$$\mathbf{X} = \hat{\mathbf{r}}\,\mathbf{A}\,\hat{\mathbf{s}}, \quad (2)$$

---

[4] The order of row and column adjustments can be interchanged, the final solution is not affected.

[5] The necessary and sufficient conditions for convergence were demonstrated by MacGill (1977), see also Lemelin et al. (2013).

[6] According to Schneider and Zenios (1990) this corresponce was proved by Bregman (1967) before Bacharach (1970).

where ˆ denotes the diagonal matrix of a vector, and **r** and **s** are the vectors generated from the shadow prices of the constraints $\mathbf{X1} = \mathbf{u}$ and $\mathbf{1}^T\mathbf{X} = \mathbf{v}$, respectively (Bacharach 1970).

Since the target function is convex and can be continuously differentiated on a compact set, if matrix **A** is indecomposable (Zalai 2012), then the solution $\mathbf{X} = \hat{\mathbf{r}}\mathbf{A}\hat{\mathbf{s}}$ is unique, except for any arbitrary δ and $1/\delta$ scalar multipliers for **r** and **s** vectors (Bacharach 1970, Mesnard 2011). This means that if a **r** and **s** vector pair is a solution, then the $\mathbf{r}\cdot\delta$ and $\mathbf{s}/\delta$ vector pairs are also.

Thus, the RAS method is a biproportional technique, which adjusts the original matrix on one hand, by rows, and on the other hand, by columns with uniform multipliers (within the given row or column).

The objective function (1) of the RAS method is also known as an information loss formula of the information theory.[7] Lemelin et al. (2013) clearly shows that RAS is equivalent to the following task

$$\sum_{i=1}^{m} p_{i,j} = p_{\cdot,j}, \ \sum_{j=1}^{n} p_{i,j} = p_{i,\cdot}, \ \sum_{i=1}^{m}\sum_{j=1}^{n} p_{i,j}\ln(p_{i,j}/p^a_{i,j}) \rightarrow \min, \qquad (3)$$

where $p^a_{i,j} = a_{i,j}/\mathbf{1}^T\mathbf{A1}$, $p_{i,j} = x_{i,j}/w$, $p_{\cdot,j} = h_j/w$, $p_{i,\cdot} = u_i/w$, and $w = \mathbf{u1}$ (i.e., the prescribed grand total of the elements of matrix **X**). Thus, $p_{i,j}$'s are considered to be elements of a two-dimensional joint probability distribution, and the target function is the 'additional' information contained by the probability distribution $p_{i,j}$ relative to the probability distribution $p^a_{i,j}$. The tasks and their solution techniques that can be formulated the way

[7] The mathematical formulation of information theory was developed by Shannon (1948) and was introduced to economics by Theil (1967).

above are called cross-entropy problems and methods.[8] Thus, the RAS method is a special case of cross-entropy methods.

By generalizing the deduction of Lemelin et al. (2013) it can be easily shown that the RAS method leads to the same result even if the reference matrix **A** is multiplied by any positive number. Multiplying by a γ scalar we obtain the objective function

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_{i,j} \ln(x_{i,j}/(\gamma \cdot a_{i,j})) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{i,j}\{\ln(x_{i,j}/a_{i,j}) - \ln\gamma\} = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{i,j} \ln(x_{i,j}/a_{i,j}) - w \cdot \ln\gamma,$$

which differs from the original objective in only one constant, thus, it has the same minimum point. Thus, from the mathematical point of view it does not matter whether the matrix **A** is multiplied by the $\gamma = w / \mathbf{1}^{\mathbf{T}}\mathbf{A}\mathbf{1}$ scalar to ensure that the grand sum of it is equal to the sum of the target margins. Of course, it is a very different question whether it is worthy to modify matrix **A** to $\mathbf{A}^*$ to satisfy the marginal constraints $\mathbf{A}^*\mathbf{1} = \mathbf{u}$, $\mathbf{1}^{\mathbf{T}}\mathbf{A}^* = \mathbf{v}$, that is to be a possible solution to the problem (1) at sight. Another question is, which of the lot of possible modifications should be chosen (the degree of freedom is considerable, since we have only n+m-1 independent conditions for the m x n elements of the reference matrix). Thus, we can reach the problem of the two-stage matrix estimation, in which case in the first stage matrix **A** is adjusted to the optimal $\mathbf{A}^*$, then using this as the reference matrix we generate the final estimate of the matrix **X** with a secondary adjustment model. It can be argued that the models used in the two stages should be "harmonized" (theoretically coherent) or deviated (using a completely identical model, there is clearly no point in breaking the process into two stages, since every

[8] The concept of cross-entropy was introduced and discussed first by Kullback and Leibler (1951).

optimization process starts computing by finding a possible solution).

2.3. Other matrix adjustment methods

Even in the case of nonnegative matrices, there are a variety of objective functions, different from the loss of information described above, that can be used and justified for the matrix adjustment problem. A very similar function is $\sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j}\ln(a_{i,j}/x_{i,j})$, which just revers the cast between the elements of the reference and the target matrix. For the comparative analysis of the results, see for example McNeil and Hendrickson (1985). The advantage of this objective is that each variable $x_{i,j}$ only appears once in the formula, so it is easier to calculate and its mathematical properties (monotony, non-negativeness, etc.) are easier to understand.

The basic versions of the so-called gravity models, mainly used for estimating the transport matrix,[9] see for example, Niedercorn and Bechdolt (1969) and Black (1972), can be considered equivalent to the RAS method, too (Mesnard 2011).

We do not mention here any possible, differently weighted objective functions that contain logarithm functions, but instead turn to the 'least squares' target functions.

According to Lahr and Mesnard (2004), Pearson's $\chi 2$ or the normalized quadratic deviation (the method of normalized least squares) was first used by Deming and Stephan (1940) and Friedlander (1961) to solve the matrix adjustment problem, and it was also suggested by Lecomber (1975) to update symmetric input-output tables (SIOTs). The

[9] The transportation problem is to deliver the quantity $x_{i,j}$ of the product x from the ith the starting point having a stock of u_i to the jth destination point demanding the quantity of v_j so that the transport cost is minimal.

objective function to minimize can be formalized by the two following and equivalent ways:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{i,j}-a_{i,j})^2/a_{i,j}=\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{i,j}/a_{i,j}-1)^2\cdot a_{i,j}. \qquad (4)$$

The second formula shows that this objective function represents the squared sum of the relative (%) difference of estimated and original matrix elements weighted with the original matrix elements. This means a compromise between Almon's (1968) simple sum of squared deviations[10]

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{i,j}-a_{i,j})^2 \qquad (5)$$

and the unweighted relative squared sum

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{i,j}/a_{i,j}-1)^2=\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{i,j}-a_{i,j})^2/a_{i,j}^2. \qquad (6)$$

It is easy to see that the simple square sum is likely to allow larger deviations at the small elements, while the unweighted relative square deviation is vice versa, the quantities to be distributed (the amount to be added or subtracted for the required row and column sums) are to be divided into larger elements, where the adjustment means a smaller percentage value.

---

[10] Although Almon used this formula for the coefficients of the input-output table, i.e. instead of the marginal constraints $\mathbf{X1} = \mathbf{u}$, $\mathbf{1}^{\mathrm{T}}\mathbf{X} = \mathbf{v}$, he minimizes the objective function subject to $\mathbf{X}\,\mathbf{g} = \mathbf{u}$, $\mathbf{1}^{\mathrm{T}}\mathbf{X}\,\hat{\mathbf{g}} = \mathbf{v}$, where $\mathbf{g}$ is the vector of known gross sectoral outputs.

As a generalization of the squared deviations and summarizing with weights 1, $1/a_{i,j}$ and $1/\ a_{i,j}^2$ (i.e. the objective functions (4)-(6)) Harthoorn and van Dalen (1987) introduce

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(\,x_{i,j}-a_{i,j})^2/g_{i,j}$$

where weights $1/g_{i,j}$ represents the 'relative confidence' of the elements $a_{i,j}$, regarded as 'first approximations' (Timurshoev et al. 2011).

It is also easy to see that the objective functions based on square variances, in contrast to the entropy (log) functions, allow the estimated $x_{i,j}$ to be different sign from the original $a_{i,j}$. Likewise, it is obvious that the original zero elements may also become non-zero. These issues will be discussed in the following Section.

3. Adjustment methods for matrices with negative and zero cells and margins

In this Section firstly, we present some modifications of methods for biproportional matrix adjustment discussed in the Section 2. These modifications are necessary to apply the methods in cases when there are negative entries in the reference matrix and/or there are negative or zero element in the prescribed row and column vectors.

The database of a multisectoral economic model often includes a cross-table (or contingency table) of such deaggregated categories. If these are to be estimated, in many cases, negative or zero values hinder the use of standard methods. For example, if the target value of a complete row or column margin (or some elements of it) is zero, the RAS estimation would set the entire row or column to zero in the first iteration, even if there are in fact both positive and negative values that are obviously and significantly different from zero, and their signs should be preserved. It is no better luck if one of the row or column sums of the reference matrix is zero (of course, then either all elements of the row or column must be zero or there must be positive and negative elements, as well), but the corresponding required margin value is zero. To highlight the weight of the problem, it is worth noting that the RAS algorithm that is otherwise not recommended in this case, would attempt to divide by zero when proportionally adjusting the given row or column.

As can be seen from their formulas in Section 2, the matrix correction methods discussed so far do not work or fail if some element of the reference matrix or one of the prescribed row or column sums is negative. Except for the banal ad hoc treatments of these problems, such as the replacement of the small negative $a_{i,j}$ elements to zero (Omar 1967)[11], or leaving them unchanged, Günlük-Şenesen and Bates (1988) and re-discovering their results, Junius and Oosterhaven (2003) first deals thoroughly with the treatment of the negative elements. The objective function of their 'generalized RAS' (GRAS) method

$$\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j}| \cdot x_{i,j} / a_{i,j} \cdot \ln(x_{i,j}/a_{i,j})$$

is distorted (Huang et al. 2008) and cannot be applied if not all columns and rows have a positive element (Temurshoev 2013), or as we will see, even in normal cases it does not always give the best results among the estimation methods available. Later on, Oosterhaven (2005) himself also points out that negative and positive differences can be eliminated in the originally proposed objective function (creating the illusion of perfect fit) and instead of it, he proposes the $\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j} \cdot x_{i,j} / a_{i,j} \cdot \ln(x_{i,j}/a_{i,j})|$ absolute information

[11] Quoted by Lahr and Mesnard (2004).

loss (AIL) function. Later again, the distortion of the GRAS objective function is corrected by Lenzen at al. (2007).[12] Huang et al. (2008) further changes the objective function to $\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{i,j}|\cdot(z_{i,j}\cdot\ln(z_{i,j}/e)+1)$, and name this ‘improved GRAS’ (IGRAS). They add a constant to the function, with which in the case of $z_{i,j} = 1$ it will be zero, but this does not affect the location of the optimum.

In any case, it can be seen from the logarithm in the objective function that if the model has any solution, for every *i, j* pairs $x_{i,j}/a_{i,j} = z_{i,j} \geq 0$ holds. Thus, the solution of the GRAS model guarantees that matrix elements preserve their sign (that is, the signs of the estimated matrix are the same as the reference matrix).

In addition to the entropy models based on some quantity of information, in the models containing quadratic objective functions can be formulated to preserve the signs of the reference matrix. Jackson and Murray (2004), for example, minimizes the following objective function (see their Model 10)

$$\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{i,j}|\cdot x_{i,j}/a_{i,j}\cdot\ln(x_{i,j}/a_{i,j}),$$

where $z_{i,j} = x_{i,j}/a_{i,j}$, subject to the usual marginal conditions and the $z_{i,j} \geq 0$ nonnegativity constraint. This so-called ‘sign-preserving squared differences’ model can be solved by the commercial mathematical programming softwares (for example the GAMS), but the inequality conditions do not allow to derive the optimal solutions using the method of

[12] The function $z\cdot\ln z$ has its minimum at the value $z = 1/e$, where e is the Euler-number (the base of the natural logarithm) and z denotes the ratio $x_{i,j}/a_{i,j}$. The minimum point should be $z = 1$. Function $z\cdot\ln(z/e)$ proposed by Lenzen at al. (2007) indeed has its minimum at $z = 1$.

Lagrangain multipliers.

Maybe that is why Huang et al. (2008) prevents the switch of the sign of the matrix elements in an alternative way. They use an additional $+M/2 \cdot \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j}| \cdot (\min(0, z_{i,j}))^2$ component in the Lagrangian function, where M is a sufficiently large positive number.

The numerical example of Junius and Oosterhaven (2003), constructed for adjusting a matrix with partly negative entries and marginal values, is used by several authors for the empirical analysis of the 'goodness of fit' of the competing methods, based on a variety of 'assessment' criteria. From these methods, Jackson and Murray (2004) have found the above-mentioned 'sign-preserving squared differences' model as the best.

Similarly, Huang et al. (2008) also concluded that the so-called 'improved normalized square deviation' (INSD) objective function

$$\sum_{i=1}^{m} \sum_{j=1}^{n} (x_{i,j}/a_{i,j} - 1)^2 \cdot |a_{i,j}| + M/2 \cdot \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j}| \cdot (\min(0, z_{i,j}))^2 \qquad (7)$$

shows the best performance (on average) using the example of Junius and Oosterhaven (2003). This, in addition to good estimation results, basically supports the justification for the similar (quadratic-type) target function chosen by the EU-GTAP project (Rueda–Revesz et al. (2016)).

As pointed out by Huang et al. (2008), and as Temurshoev et al. (2011)) has been deduced precisely, the INSD objective function is the first term of Taylor-series of the IGRAS objective function at $z_{i,j} = 1$, that is

$$|a_{i,j}| \cdot (z_{i,j} \cdot \ln(z_{i,j}/e) + 1) \approx |a_{i,j}| \cdot (0 + z_{i,j} \cdot (z_{i,j} - 2) + 1) \approx |a_{i,j}| \cdot (z_{i,j} - 1)^2.$$

Therefore, and since in GRAS estimates $z_{i,j} \geq 0$, Huang et al. (2008) claim that the INSD method is more prone to preserve the sign of elements than other non-biproportional methods.

In addition to the above and the empirical testing, Huang et al. (2008) gives the Lagrange functions associated with the constrained optimization problem corresponding to each method and derives the formulas of optimum solutions.[13] In the case of INSD, these are the following:

$$z_{i,j} = \begin{cases} 1 & if\ a_{i,j} = 0 \\ 1 + sgn(a_{i,j}) \cdot (\lambda_i + \tau_j) & if\ this\ is\ nonnegative\ or\ M = 0 \\ 0 & if\ 1 + sgn(a_{i,j}) \cdot (\lambda_i + \tau_j) < 0\ and\ M \to \infty \end{cases}, \quad (8)$$

$$\lambda_i = \{(u_i - \Sigma_j\, a_{i,j}) + \Sigma_j\, (M \cdot a_{i,j} \cdot \min(0, z_{i,j}) - \tau_j \cdot |a_{i,j}|)\} \,/\, \Sigma_j |a_{i,j}|, \text{ and} \quad (8a)$$

$$\tau_j = \{(v_j - \Sigma_i\, a_{i,j}) + \Sigma_i\, (M \cdot a_{i,j} \cdot \min(0, z_{i,j}) - \lambda_i \cdot |a_{i,j}|)\} \,/\, \Sigma_i |a_{i,j}|, \quad (8b)$$

where $z_{i,j} = x_{i,j} / a_{i,j}$, and λ_i and τ_j are the Lagrangian multipliers belonging to the row- and column sum deviations.

If $z_{i,j} \geq 0$, that is $a_{i,j}$ doesn't change its sign, the formulae for λ_i és τ_j simplify to the following forms:

$$\lambda_i = \{(u_i - \Sigma_j\, a_{i,j}) - \Sigma_j\, (\tau_j \cdot |a_{i,j}|)\} \,/\, \Sigma_j |a_{i,j}|, \text{ and} \quad (9)$$

[13] The errors in the first order conditions were corrected later by Temurshoev et al. (2011), but they also do not justify the division of the target function by 2 in the Lagrange function, which is equivalent to double-weighting the penalty function for deviations from margins.

$$\tau_j = \{(v_j - \Sigma_i\, a_{i,j}) - \Sigma_i\, (\lambda_i \cdot |a_{i,j}|)\} \,/\, \Sigma_i |a_{i,j}|. \quad (10)$$

It appears from the middle part of the formula (8) for the definition of $z_{i,j}$, that the Lagrange multipliers should be added in the case of positive, and should be subtracted in the case of negative $a_{i,j}$ elements.

Multiplying the middle part of (8) by $a_{i,j}$ yields

$$x_{i,j} = z_{i,j} \cdot a_{i,j} = a_{i,j} + a_{i,j} \cdot \mathrm{sgn}(a_{i,j}) \cdot (\lambda_i + \tau_j). \quad (11)$$

Introducing $d_{i,j} = x_{i,j} - a_{i,j}$, subtracting $a_{i,j}$ from each side of (10), and considering that $a_{i,j} \cdot \mathrm{sgn}(a_{i,j}) = |a_{i,j}|$, we obtain

$$d_{i,j} = x_{i,j} - a_{i,j} = |a_{i,j}| \cdot (\lambda_i + \tau_j) \quad (12)$$

expressing the relation between the Lagrangian multipliers and the (optimal) changes of the elements of the matrix.

From the equations (9) and (10), these multipliers depend on each other, and the row and column direction absolute value share of the element, as well.

It is also worth noting that similarly to the RAS method (where the Lagrangian multipliers of the rows and columns can change inversely in groups, and uniformly, but in any proportion within a group), it is apparent from formula (12) that it leads to the same estimate when choosing any φ value for λ_i replacing $\lambda_i + \varphi$ and replacing τ_j with $\tau_j - \varphi$, they still satisfy (8), (9) and (10). So in this sense, if there is a solution, there are infinite ones, but while in case of a 'multiplicative' RAS, the degree of freedom of the Lagrangian multipliers occurs in a proportionality factor, while for the INSD in an additive component.

Since in the general (sign-preserving) case discussed by Huang et al. (2008) equations (8)-(10) are simultaneous (λ_i and τ_j depend on each other and $z_{i,j}$, and vice versa), they suggest an iterative solution algorithm with $z_{i,j}^{(0)} = 1$, $\lambda_i^{(0)} = 0$, $\tau_j^{(0)} = 0$ initial values. In the not sign-preserving case discussed later in Section 3.2., however, λ_i and τ_j depend on each other only. In the first iteration suggested by Huang et al. (2008), according to equations (9) and (10) the Lagrangian multipliers are

$$\lambda_i^{(1)} = g_i \,/\, \Sigma_j\,|a_{i,j}| \qquad (13)$$

$$\tau_j^{(1)} = h_j \,/\, \Sigma_i\,|a_{i,j}| \qquad (14)$$

where $g_i = u_i - \Sigma_j\, a_{i,j}$ and $h_j = v_j - \Sigma_i\, a_{i,j}$ denote the differences of the prescribed row and column sums from those of the matrix **A**.

Now take $\mathbf{S} = |\mathbf{A}|$, where $|\mathbf{A}|$ denotes the matrix containing the absolute values of the elements of **A**, $\mathbf{w} = \mathbf{1}^{\mathbf{T}}\mathbf{S}$, $\mathbf{q} = \mathbf{S1}$, and $\mathbf{R} = \hat{\mathbf{q}}^{-1}\mathbf{S}$ and $\mathbf{C} = \mathbf{S}\hat{\mathbf{w}}^{-1}$, where **R** and **C** are matrices containing the row- and column-wise absolute value distribution of **S**. Substituting (13) and (14) into (12) we obtain that in the first iteration the elements of the matrix change by

$$d_{i,j}^{(1)} = |a_{i,j}|\cdot(\lambda_i^{(1)} + \tau_j^{(1)}) = g_i \cdot |a_{i,j}| \,/\, \Sigma_j\,|a_{i,j}| + h_j \cdot |a_{i,j}| / \Sigma_i\,|a_{i,j}| = g_i \cdot r_{i,j} + h_j \cdot c_{i,j}. \quad (15)$$

This means that the iteration adjusts the rows and columns of the matrix by the row- and column-wise absolute value share of the elements.

Note, that the first step of the iteration would add the discrepancy (do the adjustment to the target margins) 'twice', in the $\sum_{i=1}^{m}\sum_{j=1}^{n} d_{i,j}^{(1)} = \sum_{i=1}^{m} g_i + \sum_{j=1}^{n} h_j$ formula of the overall total of the resulting matrix both the $\sum_{i=1}^{m} g_i$, and the $\sum_{j=1}^{n} h_j$ components alone would eliminate the overall discrepancy. We will come back to this issue in section 3.2.

Using the recently introduced notations multiplying equations (9) and (10) (which determine the optimal values of λ_i and τ_j) by $q_i = \Sigma_j|a_{i,j}|$ and $w_j = \Sigma_i|a_{i,j}|$ respectively we get their

$$\lambda_i \cdot q_i = g_i - \Sigma_j (\tau_j \cdot s_{i,j}) \quad (16)$$

$$\tau_j \cdot w_j = h_j - \Sigma_i (\lambda_i \cdot s_{i,j}) \quad (17)$$

simpler versions. With matrixalgebraic notations (16) and (17) can be combined into the

$$\begin{bmatrix} \mathbf{q}^{\wedge} & \mathbf{S} \\ \mathbf{S}^{\mathrm{T}} & \mathbf{w}^{\wedge} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\tau} \end{bmatrix} = \begin{bmatrix} \mathbf{g} \\ \mathbf{h} \end{bmatrix} \quad (18)$$

system of inhomogenous linear equations, where $\boldsymbol{\lambda}$, $\boldsymbol{\tau}$, $\mathbf{g}$ and $\mathbf{h}$ are the column vectors containing the λ_i , τ_j, g_i and h_j elements respectively.

Since $\mathbf{1}^{\mathrm{T}}\mathbf{g} = \mathbf{1}^{\mathrm{T}}\mathbf{h}$, $\hat{\mathbf{q}}\,\mathbf{1} = \mathbf{q} = \mathbf{S1}$ and $\hat{\mathbf{w}}\,\mathbf{1} = \mathbf{w} = \mathbf{S}^{\mathrm{T}}\mathbf{1}$, therefore the following holds:

$$\begin{bmatrix} \mathbf{q}^{\wedge} & \mathbf{S} \\ \mathbf{S}^{\mathrm{T}} & \mathbf{w}^{\wedge} \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ -\mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad (19)$$

Since (19) can be interpreted as a solution of a homogenous system of linear equations, its $\begin{bmatrix} \mathbf{q}^{\wedge} & \mathbf{S} \\ \mathbf{S}^{\mathrm{T}} & \mathbf{w}^{\wedge} \end{bmatrix}$ (symmetric) coefficient matrix (denoted subsequently by $\mathbf{S}^*$) is singular (i.e. its rows/columns are linearly interdependent). Therefore the (18) system of linear equations cannot be solved by multiplying it from the left by the (non-existent) inverse of the $\mathbf{S}^*$ matrix. Instead, one must express one variable by the rest and must be dropped along with the corresponding equation. Finally, the reduced set of linear equations (which contains (m+n-1) equations and the same number of variables) can be solved by multiplying it from the left by the (non-existent) inverse of the reduced coefficient matrix.

3.1. Non-sign-preserving methods

Estimates of the matrices with negative entries can easily be a sign-switching. Previously, the authors tried to avoid this with using sign-preserving algorithms. It was needed especially in entropy models because the logarithmic objective function is not defined in case of negative $z_{i,j}$. However, there may also be cases where the required marginal values are such that the signs of the elements $a_{i,j}$ and $x_{i,j}$, must be different.

Lemelin (2009) also presents such a case when attempting to extend and test the GRAS, ant the Kullback and Leibler (1951) cross-entropy methods for zero-margin matrices. In his study the numerical example of Junius and Oosterhaven is first interpreted as a net world trade matrix, where the element $a_{i,j}$ of the inconsistent initial matrix **A** shows the net exports of the *i*th country from the *i*th product. The row and column sums of the estimated **X** matrix must be zero. Then, the example changes to the matrix of net investment positions, where $a_{i,j}$ represents the net claim of the *j*th country from *i*th asset. In the latter case, the row sums of **X** must also be zero, but in the column sums there may be negative values, as well. That is why Lemelin determined the prescribes a negative target sum for column 2 with originally positive elements to force the sign-switch of some elements, and to test Kullback and Leibler's cross-entropy, and Junius and Oosterhaven's GRAS methods, somewhat modified to these cases. Table 1 shows the initial matrix and required row and column sums.

Table 1. Initial matrix of net international investment positions and prescribed margins

Financial assets	Country# 1	Country# 2	Country# 3	Country# 4	**Initial row sums**	**Prescribed row sums**
Financial asset#1	7	3	5	-3	**12**	0
Financial asset#2	2	9	8	1	**20**	0
Financial asset#3	-2	0	2	1	**1**	0
Initial column sums	**7**	**12**	**15**	**-1**		
Prescribed column sums	**9**	**-16**	**17**	**-10**		

Table 2 shows the matrix **X** estimated with cross-entropy method.

Table 2. Adjusted matrix of net international investment positions using cross-entropy model

Financial assets	Country#1	Country#2	Country#3	Country#4	Estimated row sums	Prescribed row sums
Financial asset#1	27495.08	-36579.17	24854.35	-15770.26	**0**	**0**
Financial asset#2	-11049.53	36563.17	-34936.82	9423.19	**0**	**0**
Financial asset#3	-16436.55	0	10099.48	6337.07	**0**	**0**
Estimated column sums	**9**	**-16**	**17**	**-10**		
Prescribed column sums	**9**	**-16**	**17**	**-10**		

Lemelin explains the apparently unrealistic results of the cross-entropy method by that the method seeks to preserve the proportions of the elements of the matrix. Thus, even if small column sums are to be corrected at a higher rate, the relating column entries change at the same large scale (like RAS).

Using the GRAS method, Lemelin obtained the following results (see Table 3 here, and Table 8 in Lemelin 2009).

Table 3. Adjusted matrix of net international investment positions using GRAS model

Financial assets	Country# 1	Country# 2	Country# 3	Country# 4	Estimated row sums	Prescribed row sums
Financial asset#1	17.07	-23.44	18.65	-12.28	**0**	**0**
Financial asset#2	-2.49	7.44	-6.52	1.58	**0**	**0**
Financial asset#3	-5.57	0	4.87	0.71	**0**	**0**
Estimated column sums	**9**	**-16**	**17**	**-10**		
Prescribed column sums	**9**	**-16**	**17**	**-10**		

Comparing the results obtained with the two methods, Lemelin states that the modified GRAS method proved to be better than the cross-entropy method. Unfortunately, he did not investigate the solutions available with quadratic-type targeting functions, although they obviously allow for sign-switching. We will address this issue in next section.

After discussing reasons for changes in inventories (which products and how often change in the related column of the input-output tables), Lenzen (2014) reverses the general negative judgement of the sign-switching and looks as an advantage that, if necessary, an adjustment process can change the sign of the elements of the reference matrix.

3.2. The additive RAS method

In the turbulent early years of economic transition of the Hungarian economy (the early 90s) while trying to update the Hungarian input-output tables, its auxiliary matrices and other macroeconomic matrix categories I developed my "additive-RAS" algorithm (Révész, 2001) and used it instead of the RAS in the case of zero (or close to zero) known (target) margins or negative reference matrix elements of unpleasant magnitude and appearing in unlucky locations (in which cases the RAS is unusable or at least unreliable). In the first step this additive-RAS algorithm for each row distributes the difference of the target row total and the corresponding row-total of the reference matrix proportionately to their row-wise absolute value share (as matrix R was defined above) according to the

$$x_{i,j}^{(1)(\mathrm{r})} = a_{i,j} + g_i^{(1)} \cdot r_{i,j} \qquad (20)$$

formula, where $g_i^{(1)} = g_i$. Then a similar adjustment has to be done column-wise according to the

$$x_{i,j}^{(1)} = x_{i,j}^{(1)(\mathrm{r})} + h_j^{(1)} \cdot c_{i,j} \qquad (21)$$

formula, where $h_j^{(1)} = \mathrm{v_j} - \Sigma_\mathrm{i}\, x_{i,j}^{(1)(\mathrm{r})}$.

In general, the n-th iteration (i.e. which contains the n-th row-wise and n-th column-wise adjustment) can be described by the

$$x_{i,j}^{(n)(r)} = x_{i,j}^{(n-1)} + g_i^{(n)} \cdot r_{i,j} \quad (22)$$

(where $g_i^{(n)} = u_i - \Sigma_j x_{i,j}^{(n-1)}$) and

$$x_{i,j}^{(n)} = x_{i,j}^{(n)(r)} + h_j^{(n)} \cdot c_{i,j} \quad (23)$$

formulas, where $h_j^{(n)} = v_j - \Sigma_i x_{i,j}^{(n)(r)}$. Based on this the total change in the individual elements, caused by the first n iteration ($d_{i,j}^{(n)} = x_{i,j}^{(n)} - a_{i,j}$) can be described as

$$d_{i,j}^{(n)} = \sum_{k=1}^{n} (g_i^{(k)} \cdot r_{i,j} + h_j^{(k)} \cdot c_{i,j}) = r_{i,j} \cdot \sum_{k=1}^{n} g_i^{(k)} + c_{i,j} \cdot \sum_{k=1}^{n} h_j^{(k)}. \quad (24)$$

If the process converges then obviously its $d_{i,j}^{(\Sigma)}$ limit value can be computed as

$$d_{i,j}^{(\Sigma)} = r_{i,j} \cdot g_i^{(\Sigma)} + c_{i,j} \cdot h_j^{(\Sigma)} \quad (25)$$

where $g_i^{(\Sigma)} = \lim_{n \to \infty} g_i^{(\Sigma)}$ and $h_j^{(\Sigma)} = \lim_{n \to \infty} h_j^{(\Sigma)}$.

Since the $d_{i,j}^{(\Sigma)}$ elements of the 'final' matrix should satisfy the row-total and column-total requirements (otherwise the adjustment process would continue by distributing the remaining discrepancy), summing the equations of (25) by j we get the following:

$$g_i = \Sigma_j d_{i,j}^{(\Sigma)} = \Sigma_j (r_{i,j} \cdot g_i^{(\Sigma)} + c_{i,j} \cdot h_j^{(\Sigma)}) = g_i^{(\Sigma)} \cdot \Sigma_j (r_{i,j} + c_{i,j} \cdot h_j^{(\Sigma)}) = g_i^{(\Sigma)} + \Sigma_j (c_{i,j} \cdot h_j^{(\Sigma)}) \quad (26)$$

Similarly summing the equations of (25) by j we get the

$$h_j = \Sigma_i d_{i,j}^{(\Sigma)} = \Sigma_i (r_{i,j} \cdot g_i^{(\Sigma)} + c_{i,j} \cdot h_j^{(\Sigma)}) = \Sigma_j (r_{i,j} \cdot g_i^{(\Sigma)}) + h_j^{(\Sigma)} \cdot \Sigma_i c_{i,j} = \Sigma_j (r_{i,j} \cdot g_i^{(\Sigma)}) + h_j^{(\Sigma)} \quad (27)$$

conditions for the so far unknown $g_i^{(\Sigma)}$ and $h_j^{(\Sigma)}$ values. Equations (26) and (27) can be described in matrixalgebraic notations as

$$\mathbf{g} = \mathbf{g}^{(\Sigma)} + \mathbf{C}\, \mathbf{h}^{(\Sigma)} = \hat{\mathbf{q}}\, \hat{\mathbf{q}}^{-1} \mathbf{g}^{(\Sigma)} + \mathbf{S}\, \hat{\mathbf{w}}^{-1} \mathbf{h}^{(\Sigma)} \quad (28)$$

$$\mathbf{h} = \mathbf{R}^{T} \mathbf{g}^{(\Sigma)} + \mathbf{h}^{(\Sigma)} = \mathbf{S}^{T} \hat{\mathbf{q}}^{-1} \mathbf{g}^{(\Sigma)} + \hat{\mathbf{w}}\, \hat{\mathbf{w}}^{-1} \mathbf{h}^{(\Sigma)} \quad (29)$$

respectively, where $\mathbf{g}^{(\Sigma)}$ and $\mathbf{h}^{(\Sigma)}$ mean the column vectors containing the elements of $g_i^{(\Sigma)}$ and $h_j^{(\Sigma)}$ respectively. Equations (28) and (29) can be combined in the

$$\begin{bmatrix} \hat{\mathbf{q}} & \mathbf{S} \\ \mathbf{S}^{\mathbf{T}} & \hat{\mathbf{w}} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{q}}^{-1}\mathbf{g}^{(\Sigma)} \\ \hat{\mathbf{w}}^{-1}\mathbf{h}^{(\Sigma)} \end{bmatrix} = \begin{bmatrix} \mathbf{g} \\ \mathbf{h} \end{bmatrix} \tag{30}$$

system of inhomogenous linear equations.

Comparing this with (18) we can see that both the coefficient matrices and the right-hand-side constant vectors are the same as their counterpart in (18) and (30). Therefore the solutions of the (18) and (30) set of linear equations are the same too. This means that if $\boldsymbol{\lambda}$, $\boldsymbol{\tau}$ are the solution of (18), then those $\mathbf{g}^{(\Sigma)}$ and $\mathbf{h}^{(\Sigma)}$ vectors which satisfy the $\hat{\mathbf{q}}^{-1}\mathbf{g}^{(\Sigma)} = \boldsymbol{\lambda}$ and $\hat{\mathbf{w}}^{-1}\mathbf{h}^{(\Sigma)} = \boldsymbol{\tau}$ equations, hence which can be computed as

$$\mathbf{g}^{(\Sigma)} = \hat{\mathbf{q}}\,\boldsymbol{\lambda} \tag{31}$$

$$\mathbf{h}^{(\Sigma)} = \hat{\mathbf{w}}\boldsymbol{\lambda} \tag{32}$$

are the solutions of the (30) set of linear equations. By substituting (31) and (32) into (25) we get the

$$d_{i,j}^{(\Sigma)} = r_{i,j} \cdot q_i \cdot \lambda_i + c_{i,j} \cdot w_j \cdot \tau_j = s_{i,j} \cdot \lambda_i + s_{i,j} \cdot \tau_j = |a_{i,j}| \cdot (\lambda_i + \tau_j) \tag{33}$$

formula for the resulting total changes (in the individual matrix elements) of the additive-RAS algorithm. This is just the same as (12), i.e. what for this case Huang et al (2008) derived as the optimal solution of the INSD-model.

Therefore, we proved that the result of the additive-RAS algorithm is identical to that of the INSD-model if the sign of the matrix elements do not change. Fortunately, sign flips occur only if the ratio of the target- and actual margins is extremely high. For example, (since the shares in the absolute values are smaller than the value shares) unless this ratio falls below -100 per cent, the iteration certainly does not cause sign flips. This is true in the case of even more extreme margin adjustment ratios. In any case, extreme margin adjustment ratios raise concern about the applicability of the reference matrix, i.e. about whether the structure of the searched (target) matrix may preserve the similarity to the structure of the reference matrix.

The above 'shares in the absolute values are smaller than the value shares' statement requires certain qualifications. This is true only if they are computed from the same matrix. In the above presented algorithm the absolute value shares are computed from the $a_{i,j}$ elements of the reference matrix, while the 'actual' value shares (i.e. in the n-th iteration) are computed from the already adjusted $x_{i,j}^{(n)(r)}$ and $x_{i,j}^{(n)}$ matrices. Therefore, if for some reasons the structure of the $x_{i,j}^{(n)(r)}$ and $x_{i,j}^{(n)}$ matrices differ considerably from the structure of the reference matrix then the additive RAS algorithm may cause sign flips. Although the algorithm still may converge and may produce apparently reasonable results, it can not be guaranteed that these results are the best estimates according to some usual optimum criteria (distance measure).

Hence if the additive-RAS algorithm produces sign flips and consequently its mathematical characteristics become opaque (unclear) then it is worth modifying the algorithm appropriately. Concretely – similarly to what practically the multiplicative RAS algorithm does with the value shares, - we may compute the absolute value shares from the n-th iteration's ('current') $x_{i,j}^{(n)(r)}$ and $x_{i,j}^{(n)}$ matrices (more precisely we denote these by $\tilde{x}_{i,j}^{(n)(r)}$ and $\tilde{x}_{i,j}^{(n)}$ respectively, since these differ from their counterparts in the original additive-RAS algorithm) and distribute the discrepancies proportionately to these modified absolute value shares. Therefore the (22)-(23) adjustment-formulas of the n-th iteration will be replaced by the following:

$$\tilde{x}_{i,j}^{(n)(r)} = \tilde{x}_{i,j}^{(n-1)} + g_i^{(n)} \cdot r_{i,j}^{(n)} \qquad (34)$$

where $r_{i,j}^{(n)} = |\tilde{x}_{i,j}^{(n-1)}| / \Sigma_j |\tilde{x}_{i,j}^{(n-1)}|$, and

$\tilde{x}_{i,j}^{(n)} = \tilde{x}_{i,j}^{(n)(r)} + h_j^{(n)} \cdot c_{i,j}^{(n)}$ (35)

where $c_{i,j}^{(n)} = |\tilde{x}_{i,j}^{(n)(r)}| / \Sigma_i |\tilde{x}_{i,j}^{(n)(r)}|$.

In general, since by definition $\Sigma_j r_{i,j} = \Sigma_j r_{i,j}^{(n)} = \Sigma_i c_{i,j} = \Sigma_i c_{i,j}^{(n)} = 1$ therefore of the general equations of the additive RAS method (see the (22)-(23) and (34)-(35) equations) one can see that

$\Sigma_j x_{i,j}^{(n)(r)} = \Sigma_j \tilde{x}_{i,j}^{(n)(r)} = u_i$, $\Sigma_i x_{i,j}^{(n)} = \Sigma_i \tilde{x}_{i,j}^{(n)} = v_j$,

i.e. the marginal conditions hold.

Naturally, in the case of non-negative elements both the RAS- and the modified-RAS algorithms the solution and the iteration steps are the same as those of the traditional RAS.

Based on the above introduction of the modified additive-RAS method it is still a rhetorical question what are the further mathematical characteristics of the resulting $\tilde{x}_{i,j}$ matrix, how far it fits to the reference matrix, or to some pre-adjusted reference matrix. Apart from the few mathematical characteristics described above we can say that the modified additive-RAS algorithm is similar to some naturopath drugs, which apparently works well, but its biological effect-mechanisms and the conditions of its applicability are not properly known. Possibly because of this unclear nature the method has not caught the attention of mathematicians. In any case, the precise mathematical discussion of the modified RAS-algorithm remains to be accomplished and may reveal quite a few interesting properties.

Fortunately, in our more than 25-year experience, we found that the additive-RAS method and the modified additive-RAS method mostly converge fast and usually the resulting matrix fits well to the reference matrix. To illustrate this, we present not one of our exercises but we test it with the numerical example of Huang et al (2008) instead.

The numerical test confirmed that the additive-RAS algorithm produces the same result as what Huang et al (2008) published as optimal solution of the INSD-model, which they had found to be the estimation method with the best fit in terms of the AIL (average information loss) measure (which was computed to be 11,28).

In addition, if we follow Temurshoev et al's (2011) interpretation (or clarification) of the iteration method suggested by Huang et al (2008) it is easy to prove that in the case of the *not enforced sign-preserving case*, our and their iteration algorithms are the same too. Concretely, while Huang et al (2008) only say that "By initializing $\boldsymbol{\alpha}$, $\boldsymbol{\lambda}$, $\boldsymbol{\tau}$ as **I**, **0**, **0** respectively and calculating them with equations (24), (28) and (29) iteratively, we obtain the final solution" where the $\boldsymbol{\alpha}$ stands for the matrix of our $z_{i,j}$ 'cell indices' (i.e is the matrix of the ratios of the corresponding elements of the resulting and reference matrices) and where their equations (24), (28) and (29) correspond to our equations (8), (8a) and (8b) respectively, Temurshoev et al (2011) not only correct this by saying that $\boldsymbol{\alpha}$ is (*not square* but) the *m x n* matrix of ones, but also say that *within* each iteration steps $\boldsymbol{\lambda}$ has to be computed *first* and only then $\boldsymbol{\tau}$ is computed already using the just computed values of $\boldsymbol{\lambda}$, while $\boldsymbol{\alpha}$ has to be computed *at last*. This we call the *recursive* interpretation of the algorithm suggested by Huang et al (2008) as opposed to the other legitimate interpretation which may be called the *contraction-like* algorithm in which the iterating variables (vectorised and grouped together in the say **w** vector) change simultaneously according to the $\mathbf{w}^{(n+1)} = f(\mathbf{w}^{(n)})$ symbolic scheme, where f is the operator of the iteration steps.

In the case of the *not enforced* (but still) *sign-preserving case* the first step of the Huang et al (2008) suggested iteration algorithm interpreted as contraction-like algorithm are identical to equations (13) and (14). In the *recursive interpretation* – using equation (10) – equation (14) is replaced by the

$\tau_j^{(1)} = h_j / \Sigma_i |a_{i,j}|$ $\tau_j = \{ h_j - \Sigma_i (g_i / (\Sigma_j |a_{i,j}|) \cdot |a_{i,j}|)\} / \Sigma_i|a_{i,j}| = \{ h_j - \Sigma_i (g_i r_{i,j})\} / \Sigma_i|a_{i,j}|$ (14')

formula, where the numerator is just the *residual column discrepancy* remaining after the first additive-RAS row-wise adjustment. Therefore $|a_{i,j}| \cdot \tau_j^{(1)} = \{ h_j - \Sigma_i (g_i r_{i,j})\} \cdot c_{i,j}$ represents the changes made by the first column-wise additive-RAS adjustment.

It is easy to see from equations (9) and (10) and to prove by mathematical induction that in all further iteration steps **λ** and **τ** also represent the percentage row-wise and column-wise residual additive-RAS adjustment requirements respectively (remaining after the previous adjustments or in other words applying the (13) and (14') formulas of the 'first' iteration but after '*reinitializing*' the $a_{i,j}$, $c_{i,j}$, $r_{i,j}$, g_i and h_j parameters). Therefore, the recursive interpretation of the INSD model's iteration algorithm suggested by Huang et al. is absolutely the same as the additive-RAS algorithm.

By reconsidering the meaning of equations (8), (9) and (10) in the light of the just analysed iteration algorithm we can say that in the *n*-th iteration for each pair of (i,j) indices the percentage change in the corresponding matrix element ($z_{i,j}$) is the sum of the percentage change in the corresponding row- and column-totals still required after the first *n-1* iterations, minus the weighted average of these required row-total changes *weighted by the reinitialized* $c_{i,j}$ shares.

Although in the *case of sign-flips* we know little more of the mathematical characteristics of the additive-RAS and INSD algorithms than what is said in Huang et al (2008) both the additive-RAS and modified additive-RAS algorithms yielded quite reasonable estimates for the (somewhat extreme) numerical example given by Lemelin (2009). The results of the additive RAS-algorithm can be seen in Table 4.

Table 4: Additive-RAS estimates for the matrix of international investment positions

	Country 1	Country 2	Country 3	Country 4	**Totals**	**Given totals**
Asset 1	7,89	-4,42	5,10	-8,58	**0**	**0**
Asset 2	2,62	-11,58	9,64	-0,67	**0**	**0**
Asset 3	-1,52	0,00	2,27	-0,75	**0**	**0**
Totals	**9**	**-16**	**17**	**-10**		
Given totals:	**9**	**-16**	**17**	**-10**		

Comparing the above table with the reference matrix presented in Table 1 one can see that almost all elements have changed in the right direction (i.e. to eliminate the discrepancy between the target and actual margins) and the magnitudes of the individual (cell) changes are also reasonable. Note, that our method does not use any arbitrary normalisation.

Comparing our results with those of Lemelin (see Table 3 above or in Table 8 in Lemelin (2009)) it can be seen clearly that our additive-RAS solution is superior to his solution based on the Kullback-Leibler cross-entropy minimizing criterion (minimand), especially regarding the $a_{2,1}$ element and the whole 2nd and 4th columns (where in his solution it is mysterious why two entries have increased while for both of them both the corresponding row-total and column total should have decreased).

To compare the fit of our additive-RAS and Lemelin's estimates we computed the MAD (mean average deviation) statistics. This again showed the superiority of the additive-RAS method over Lemelin's estimates: for Lemelin's solution this error measure proved to be more than twice as high than for our additive-RAS estimates (7.28 versus 3.42). Interestingly, for this numerical matrix adjustment problem the modified additive-RAS algorithm produced different results depending on whether the adjustment started row-wise or column-wise. In the latter case the MAD value was 5.42 (still better than that of Lemelin's model) while when the column-wise adjustment is done first, the results and hence the MAD-value is the same as in the case of the additive-RAS solution (3.42).

Interestingly, when in equation (35) we (re)defined $c_{i,j}^{(n)}$ as $c_{i,j}^{(n)} = | \tilde{x}_{i,j}^{(n)} | / \Sigma_i | \tilde{x}_{i,j}^{(n)} |$, or in other words, when we recalculate the absolute vale shares only *after each iterations* (but not after each adjustment steps), then the results of this 'less frequently'-modified additive-RAS algorithm produced almost the same results as the additive-RAS algorithm (or INSD model) even when starting the adjustments row-wise (more precisely the MAD value in this latter case was 3.47).

3.3. The model estimating and transforming the 2010 EU I-O tables to GTAP format

To illustrate the practical application and generalization possibilities we present a recent research project in which the matrix adjustment problem had to be formulated in a more general way and where in the solution process we had to apply various tricks by considering the macroeconomic statistical and economic aspects of the problem.

In the so-called EU-GTAP project (see Rueda et al (2016) or in the **EU-GTAP project - final report-161005.pdf** file) at the request of the European Commision's General Directorate for Trade (DG Trade) and with the methodological support and supervision of the Eurostat and the GTAP-consortium[14] the project team of the EC Joint Research Center compiled the EU-countries' Input-Output tables, Tax and Subsidy matrices in GTAP format (in the 57 sectors of the GTAP-database and both in basic and producers prices) and according to the SNA2008 (see Eurostat (2008)) and ESA2010

[14] More information about this organisation van be found on their homepage (www.gtap.org)

methodology. The transformation from the 64 (Nace 2 classification based) sectors of the Eurostat Input-Output tables to the 57 GTAP sectors required the disaggregation of the mining sector, the textile and clothing sector, the metallurgy sector, the food-beverages-tobacco sector, and the electricity-gas-heat supply sector. This disaggregation process, i.e. to elaborate country and category specific share matrices (which show the shares of the disaggregated sectors within the aggregate sector) required the acquisition, processing and reconciliation of many auxiliary data. Missing and confidential Input-Output tables and Tax and Subsidy matrices were estimated by the team using sound methodology. The results were built into the 9.2. version of the GTAP database.

The additive-RAS method was also used in this project. For Spain allegedly exists the 'Taxes less subsidies' matrix but it is confidential. Therefore, it was estimated by the additive-RAS method using the so-called 2010 'Use-table' as the reference matrix (practically a proxy for the tax base). This estimated negative or zero values in each row where the prescribed row-total (borrowed from the 'Supply table') was negative (for example, in the case of the products of agriculture, mining and land transport). Similarly, the additive-RAS method estimated negative or zero values in each column where the prescribed column-total (borrowed from the Input-Output table) was negative (concretely for the food-beverages-tobacco industry). As opposed to this the RAS-method estimated positive values at the intersection of those rows and columns where the prescribed margins were negative. Clearly, this is absolutely unacceptable. In general, the additive-RAS method estimated a much more plausible distribution of the row- and column-totals across the elements of the corresponding rows and columns. All these confirmed the superiority of the additive-RAS method over the RAS-method in the estimation of such matrices.

In the EU-GTAP project the two-directional matrix adjustment problem appeared as a part of a more complex estimation problem. Since the team had to estimate both the domestic input-output matrix (commodity flows) and the import matrix and simultaneously, so that for the row-totals, column-totals and cell-specific (i.e. corresponding to a given row and column) upper or lower bounds may be given only for their sum (domestic+import). Therefore, we may call this problem a *two-matrix adjustment problem*.

In addition, the team had to prescribe the *non-negativity* of most of the elements of the matrices to be estimated. Many (but relatively few) *exceptions* were also introduced into the estimating model (mainly due to the errors and inconsistencies of the statistical data, and due to some curious accounting techniques used by some national statistical institutes which led to negative elements in unusual locations like exports, investments and consumption). Finally, to ensure the add-up consistency between the estimated disaggregated matrix elements and their exogenously given aggregate counterparts, *block-total constraint*s were also introduced into the model.

The *core of the model* developed for the solution of this complex problem (i.e. *without* the mentioned exceptions and absolute or relative upper and lower bounds on the input coefficients, exports and stock accumulations) can be formulated as follows:

Sets:

I GTAP sectors (the general element of the set is denoted by i)

V final demand categories (the general element of the set is denoted by v)

B sectors of the common aggregation of the GTAP and the 2010 Eurostat Input-Output tables (the general element of the set is denoted by b)

M(b,i) mapping of sets B and I, i.e. the set of those (b,i) pairs where GTAP sector i is included in the common aggregation sector b

Variables:

$\mathbf{D^p}(i,j)$ as the intermediate (production) demand block of the domestic IOT;

$\mathbf{D^f}(i,v)$ as the final demand block of the domestic IOT;

$\mathbf{M^p}(i,j)$ as the intermediate (production) demand block of the import IOT; and

$\mathbf{M^f}(i,v)$ as the final demand block of the import IOT.

Parameters:

$\mathbf{x}(i)$ gross output of the i-th GTAP-sector

$\mathbf{m}(i)$ total imports of the i-th GTAP-sector

$\mathbf{v}(i)$ gross value added of the i-th GTAP-sector

ε arbitrary small scalar value (0.1 in the GAMS code)

λ arbitrary big scalar value (10 in the GAMS code)

$\mathbf{D_0^p}(i,j)$ reference (prior) matrix for $\mathbf{D^p}(i,j)$

$\mathbf{D_0^f}(i,v)$ reference (prior) matrix for $\mathbf{D^f}(i,v)$

$\mathbf{M_0^p}(i,j)$ reference (prior) matrix for $\mathbf{M^p}(i,j)$

$\mathbf{M_0^f}(i,v)$ reference (prior) matrix for $\mathbf{M^f}(i,v)$

$\mathbf{D}_a^\mathbf{p}(b,b')$ block-totals (at the common aggregation level) of the intermediate demand block of the GTAP-profile-cleaned domestic IOT

$\mathbf{M}_a^\mathbf{p}(b,b')$ block-totals (at the common aggregation level) of the intermediate demand block of the GTAP-profile-cleaned import IOT

$\mathbf{D}_a^\mathbf{f}(b,v)$ block-totals (at the common aggregation level) of the final demand block of the GTAP-profile-cleaned domestic IOT

$\mathbf{M}^{\mathbf{f}}_{a}$ (b,v) block-totals (at the common aggregation level) of the final demand block of the GTAP-profile-cleaned import IOT.

Then, we defined the minimisation problem as:

$$Min\sum_{i,j}\left(\left[\frac{\mathbf{D}^{\mathbf{p}}(i,j)+\varepsilon}{\mathbf{D}^{\mathbf{p}}_{0}(i,j)+\varepsilon}-1\right]^2+\left[\frac{\mathbf{D}^{\mathbf{p}}_{0}(i,j)+\varepsilon}{\mathbf{D}^{\mathbf{p}}(i,j)+\varepsilon}-1\right]^2+\left[\frac{\mathbf{M}^{\mathbf{p}}(i,j)+\varepsilon}{\mathbf{M}^{\mathbf{p}}_{0}(i,j)+\varepsilon}-1\right]^2+\left[\frac{\mathbf{M}^{\mathbf{p}}_{0}(i,j)+\varepsilon}{\mathbf{M}^{\mathbf{p}}(i,j)+\varepsilon}-1\right]^2\right)+$$

$$+\lambda\sum_{i,v}\left(\left[\frac{\mathbf{D}^{\mathbf{f}}(i,v)+\varepsilon}{\mathbf{D}^{\mathbf{f}}_{0}(i,v)+\varepsilon}-1\right]^2+\left[\frac{\mathbf{M}^{\mathbf{f}}(i,v)+\varepsilon}{\mathbf{M}^{\mathbf{f}}_{0}(i,v)+\varepsilon}-1\right]^2\right)$$

subject to[15]:

$$\mathbf{x}(i)=\sum_{j}\mathbf{D}^{\mathbf{p}}(i,j)+\sum_{v}\mathbf{D}^{\mathbf{f}}(i,v)$$

$$\mathbf{m}(i)=\sum_{j}\mathbf{M}^{\mathbf{p}}(i,j)+\sum_{v}\mathbf{M}^{\mathbf{f}}(i,v)$$

$$\mathbf{v}(j)=\mathbf{x}(j)-\sum_{i}\left(\mathbf{D}^{\mathbf{p}}(i,j)+\mathbf{M}^{\mathbf{p}}(i,j)\right)$$

$$\mathbf{D}^{\mathbf{p}}_{a}(b,b')=\sum_{i\,|\,(b,i)\in M}\sum_{j\,|\,(b',j)\in M}\mathbf{D}^{\mathbf{p}}(i,j)$$

$$\mathbf{M}^{\mathbf{p}}_{a}(b,b')=\sum_{i\,|\,(b,i)\in M}\sum_{j\,|\,(b',j)\in M}\mathbf{M}^{\mathbf{p}}(i,j)$$

$$\mathbf{D}^{\mathbf{f}}_{a}(b,v)=\sum_{i\,|\,(b,i)\in M}\mathbf{D}^{\mathbf{f}}(i,v)$$

$$\mathbf{M}^{\mathbf{f}}_{a}(b,v)=\sum_{i\,|\,(b,i)\in M}\mathbf{M}^{\mathbf{f}}(i,v)$$

Note the following features of the objective function:

[15] The | vertical bar in the following constraints represents the 'if', meaning that the summation is restricted to those elements of the set which meet the condition on the right hand side of the bar.

- The ‘augmentation’ of the value of the matrix elements by the arbitrarily chosen scalar ε was introduced by Möhr et al. (1987) and used also recently e.g. by Lemelin et al. (2013) to eliminate the (frequently hidden) inconsistency of the constraints so that if the given row- and column-totals require the increase of some elements there be enough non-zero elements in the reference matrix which can be modified accordingly[16].
- Besides, we have avoided cases where big initial values may turn into very small values by computing relative errors with both variables and their reference values in the denominators of the objective function.
- A weighting scalar λ was introduced (first by Byron (1978) as the degree of reliability of the elements of the reference matrix) to make sure that the estimates for the final demand be closer to the reference matrix or in other words to counterbalance the fact that the number of final demand elements are much less than those of the intermediate demand.

4. The most important matrices to be adjusted in multisectoral macroeconomic analyses

The best method for a specific biproportional matrix adjustment problem also depends on the economic content of the matrix. This section discusses the previous statement in detail. In addition to reviewing the types of matrices to be adjusted, we highlight the specialities that influence the choice of the appropriate mathematical process and its expected effectiveness. We also outline the methods by which standard mathematical

[16] and ’to avoid having to take the log of zero in the cross-entropy method’ (Lemelin et al,, 2013)

procedures should be complemented.

4.1. Input-output tables "type A"

If we do not distinguish between domestic and imported products, the balances can be written using equation $\mathbf{x} + \mathbf{u} = \mathbf{T1} + \mathbf{y}^h + \mathbf{z}$, where $\mathbf{x}$ is the vector of the gross production of each sector (or product), $\mathbf{u}$, $\mathbf{z}$, and $\mathbf{y}^h$ are the vectors representing the import, export and final use by product (industry), respectively, and a t_{ij} element of the $\mathbf{T}$ matrix stands for intermediate use of the *i*th (industry's) product(s) in the production of *j*th (industry's) product(s). The so-called input-output table "type A"[17] also shows the allocation imported and domestic products together in the left-upper quadrant but sets column and row sum equalities by subtracting imports from the final use (with the equation $\mathbf{x} = \mathbf{T1} + \mathbf{y}^h + \mathbf{z} - \mathbf{u}$ for net product balances, where $\mathbf{z} - \mathbf{u}$ is the net export).

When adjusting the table of net product balances (i.e., the matrix of ($\mathbf{T}$, $\mathbf{y}^h$, $\mathbf{z}$, -$\mathbf{u}$), the sum of which is the gross production values $\mathbf{x}$) to a different (new) gross production vector and column sums one faces several negative entries in the reference matrix due to the -$\mathbf{u}$ component. Therefore, standard RAS method cannot be used, more specifically, there is no guarantee that it will work well. Or, as Jackson and Murray (2004) described, because of the negative elements the behaviour of RAS will be "erratic", i.e., the results of iterations can be unpredictable.

The problem of negative elements can occur in the open static input-output models in other categories, at other locations in the matrix, too, above all in the column of the

---

[17] Input-output tables consist of two overlapping accounts. Thus they show, on one hand the use of each products (row-wise), and on the other hand the distribution of the value of production on expenditures and incomes (column-wise). For more see Zalai (2012)).

change in inventories (which shows changes by products or industries). If the entire input-output table is to be estimated, i.e. the elements of added value are unknown, then rarely (if current expenditures are greater than the production value), the added value itself may be negative. It follows naturally that one of its components (either the operating surplus or the net taxes on production) is negative.

Other negative elements in the input-output tables are due to the specific accounting of some exceptional events. In the product-by-product input-output table for 2010 in the Eurostat database, for example, these can be found in the column of fixed capital formation, consumption and export (mainly due to the net margin-based settlement of re-export items). The scope and word limits of this article does not allow the description of the possible causes of negative values. We only note here that although it is often difficult to trace the causes, in some important cases, this must be tried before simply applying the general and "blind" methods of adjustment.

Biproportional adjustment of input-output tables can be problematic because of zero row and column sum, as well. For example, for a "type-A" input-output table, some of the row sums that express gross production values can easily be zero if there is no domestic production of the given product. For the column of changes in inventories, it may easily occur that its sum is zero or very close to zero, too. When using the input-output tables as a reference matrix, it is important to realize that the values of inventory changes in some previous years or in other regions (especially because the statistical error is commonly reported here) are suffering from high eventuality (see Lenzen et al (2014), for example), and thus they cannot serve as a good starting point for the estimation. In the reference matrix, inventory changes must be given in some other reasonable way (for example, as a long run average in proportion to gross output).

4.2. Matrices of taxes less subsidies on products

When there are more subsidies then taxes on a given product, negative cells occur in the matrix of taxes less subsidies on products (which is included in the background tables or matrices subtracted of input-output tables). Negative elements may also appear in net product taxes if they are listed only in a single row (see the input-output tables at basic prices where they are summed up in the row below the basic price intermediate and final uses of the sectors) or in a single column (such as in the supply table where in addition to the basic price values, all taxes and subsidies on the total use of the products are indicated in a separate column).

Matrices of net taxes on products are rare sparse matrices, with disadvantages and advantages, as well. The latter can be exploited with a well-designed estimation method. Disaggregating data of net taxes to matrices of certain tax and subsidy types one can utilize (as we did in the EU-GTAP project described in subsection 3.3) that most of the elements contain only a single tax or subsidy, thus the estimation procedure "pulls" the sums to be distributed to these cells from the prescribed margins.

The same sparse feature can be used to adjust the matrix of net product taxes as follows. Let **T** be the reference matrix of taxes on products, and **S** is the reference matrix of the subsidies on products (containing negative or 0 entries, mostly the latter). Thus, **T**+**S** represents the matrix of taxes less subsidies on products. For a period (or region) other than of the reference matrix, in many cases (at most) only the margins are known from the data published by the statistical offices (row sums, i.e. product taxes less subsidies by product groups, can be gained usually from the supply table, and column sum from the input-output table). Let's denote these rows and column sums by **r** and **c**, respectively. The schema for the adjustment task can then be specified in the table below.

Table 5. The schema for the adjustment task of matrices of taxes less subsidies on products

T	**S**	**r**
S		?
c	**?**	

This task cannot be solved by the RAS method because some of the margins (identified by a question mark in the table) are unknown. In addition, when using RAS or similar methods, nothing ensures that the matrices to be placed in **S** in the estimated matrix will also be identical. To handle this task with the RAS or additive-RAS method, one can write the task in the following way (spreading out the rows and columns of the matrix **S**).

Table 6. Arrangement of matrices of taxes less subsidies on products for the use of additive RAS

T	$<\mathbf{s}_{.1}>$	$<\mathbf{s}_{.2}>$...	$<\mathbf{s}_{.n}>$	**r**
$<\mathbf{s}_{1.}>$	$-\mathbf{S}_{11}$	$-\mathbf{S}_{12}$ ...	$-\mathbf{S}_{1n}$	**0**
$<\mathbf{s}_{2.}>$	$-\mathbf{S}_{21}$	$-\mathbf{S}_{22}$ ...	$-\mathbf{S}_{2n}$	**0**
:	:	:	:	:
$<\mathbf{s}_{n.}>$	$-\mathbf{S}_{n1}$	$-\mathbf{S}_{n2}$ ...	$-\mathbf{S}_{nn}$	**0**
c	**0**	**0** ...	**0**	

In table $\mathbf{s}_{.j}$ is *j*th column, $\mathbf{s}_{i.}$ the *i*th row of **S,** < > is the sign of the diagonalization, and $\mathbf{S}_{ij}$ is the transpose of the matrix containing s_{ij} in the appropriate cell and zeros everywhere else.

Because of the zero margins RAS cannot solve this equally redesigned task in an economically meaningful way, but the additive RAS does work even under these circumstances. In the solution, row and column elements with the same index in the matrix **T** and **S** will be the same.

The size of the matrix can still be a problem. For 64 sectors of EU input-output tables we have $64 \cdot 65 = 4160$ rows and columns inside the table. However, for the zero elements **S**, there is no need to write the balance. Since **S** matrix is sparse, it allows to omit a large part of rows and columns, which, however, make it easier to handle the problem. The scheme described above can be used to address other problems (e.g. with additive RAS), in which the components are estimated based on the only available marginal data (e.g., estimating regional input-output tables estimation).

4.3. Transformation matrices of consumption and investment

Some more sophisticated CGE models derive investment demand from the demand of investing sectors and the investment product structure typical of the industry ("material-technical" composition, import rates). These structures are described by the investment matrix (more specifically the matrix for gross capital accumulation), the rows of which contain the capital goods (or the suppliers of them), and the columns represent the investor industries and sectors (the buyers of fixed capital goods). The investment matrix is also a sparse matrix. Besides the elements of the construction and manufacture of machinery and possibly the diagonal of industry by industry input-output tables (own investments) one can hardly find entries different from zero. Thus, biproportional estimation methods have a small room to maneuver for distributing deviations of the prescribed and actual margins, and the margins (or one group of them) can easily be inconsistent, i.e. the problem proves infeasible.

Of course, if product flows are also calculated at the basic price in this investment transformation matrix, but the column sums are measured at purchasers' prices, then the difference between the puchasers' and base price expenditures, i.e. the net value of product taxes and subsidies, should be accounted for in a separate row, as in the basic price input-output tables. Thus, the transformation matrix may also have negative values in this row.

Likewise, consumption is often estimated and assigned to the producing sectors (or product groups) based on the COICOP consumption categories, using the so-called Lancaster transformation matrix. There may also be inconsistent margins here (though mostly for statistical reasons). In practice, however, it is not the biggest problem for estimation. It is the separation of trade margins and product taxes from the COICOP categories at consumer prices. In the case of basic input-output tables, the trade margins and the product taxes are usually separated, so this is also the case with transformation. Generally, the consumption transformation matrices produced by the statistical offices are usually made at purchasers' price. Thus, this can serve as a reference matrix when transforming to the consumption column of the basic price input-output table, if only the row of margins and net product taxes is imputed first (based on some estimate but must keep in mind that there may also be negative as indicated in the investment matrix). Alternatively, the transformation can be performed at a purchasers' price and the margins and product taxes then be separated afterwards by the estimated rates per product (by supplier sector), but in accordance with the product taxes and margins given for consumption. Of course, both paths are pretty bumpy (for example, to ensure that the "percentage" values of the margins or product taxes estimated with RAS or other adjustment method for consumer spending remain reasonable), but the discussion of these problems is not possible in this article.

Another problem is that, although in the consumption and investment transformation matrix and in the so-called bilateral trade matrix, the elements of the sectors, in principle, may not include negative entries, but if they are considered as the breakdown of the consumption, investment and export columns of the input-output table, they can inherit the above-mentioned negative values in these columns of the input-output table. Moreover, also the sale of used fixed assets appear in the accumulation of capital in the investment matrix. This can cause negative entries, too.

4.4. Further matrices to be adjusted in the national economy statistics

If a matrix displays both the incomes and the expenditures (including the savings) of some economic agents[18] (for example, the household groups) so that each column represent the **budget of an agent**, and where the expenditures are accounted as negative amounts, then the column-totals become zeros. If we adjust this matrix (used as a reference matrix) by the above discussed methods then the incomes and expenditures will be estimated simultaneously so that total incomes (or total expenditures) are not known ex ante, but can be computed as the sum of the positive elements (while the sum of the negative elements constitute the total expenditures).

However, the RAS-method does not work in this case, where the column-totalsa are zeros: in the first iteration the column-wise adjustment would turn each elements to zero irreversibly. This is clearly an unacceptable result in the case of a budget-matrix.

Similarly, let us consider the **matrix of assets and liabilities**. Its rows represent the individual financial instruments (cash, deposits, bonds, loans, etc.) while the columns represent the economic agents. If we account the liabilities of the given instruments (debts) as negative entries, then obviously the matrix will contain many such negative elements and its row-totals must be zeros (for each instrument the total claims are equal to the total liabilities (see for example the above discussed numerical example of Lemelin (2009)).

---

[18] This combined accounting may be necessary if the total incomes (and hence expenditures) of the given agents are not known.

Since the Input-Output tables form a block of the (branch-accounts containing) **Social Accounting Matrix** (SAM) these also contain negative elements. Nevertheless other cells of the SAM also may contain negative elements. One reason for this is the following:

Compilation of the SAMs are usually done for calibrating a SAM-multiplier-model, which uses expenditure share coefficients calibrated by dividing the columns (belonging to the endogenous accounts) by the corresponding column-total. Therefore, it is essential that every transaction must be accounted in the column of that category, with which it is (or at least can be assumed to be approximately) proportional. Even, total incomes (= total expenditures) must be defined accordingly, so that they must be economic theoretically meaningful categories, and within which the expenditure shares can be assumed to exist. For this purpose it happens frequently that the *j*-th account's (agent's, category's) t_{ij} expeditures on the *i*-th account is accounted (transposed or mirrored on the main diagonal) with negative sign as part of the t_{ji} transactions (see for example, in Robinson et al (1998)). So if the original expenditure of the i-th account on the j-th account was lower than this t_{ij} (does not counterbalance it) then the resulting value of t_{ji} will become negative. For example, if we want to determine an exogenous account's (usually the expenditures of the government and of the rest of the world) transaction to a certain other account (for example, the housing investment subsidies) endogenously (proportionately to the total income of this account[19]) then we can account this as the (negative) transaction (payment) of this account to the given exogenous account.

---

[19] The row-totals (which represent the total income of the account) are equal to the corresponding column-totals (total expenditures) by construction, i.e. since the "expenditures" include the savings (not spent amount).

The above mentioned cases (occasionally supplemented with further tricks used to handle them) are summarized in the following table:

Table 7: Summary table of the estimation problems of the most important macroeconomical matrix categories

Category name	negative and other problematic elements	possible methods for solving the problems
Input-output tables	changes in inventories, net taxes, (re)exports, consumption, investment, value added (and its components), imports (as negative column of the final demand of the usual “A”-type I-O table), zero output (A-típusú ÁKM-ben)	creating mirror accounts*, estimating the changes in inventories separately (e.g. proportionately to the output or total uses), grossing up the reexports, applying the additive-RAS or two-stages RAS-methods
matrix of taxes less subsidies	negative elements (if subsidies surpass the taxes), sparse matrix	estimating the subsidy matrix separately (see table 6), applying the additive-RAS method
Consumption (transformation) matrix	negative elements due to the (re)sales of consumption goods produced in previous years	replacing the corresponding negative elements of the reference matrix with zeros or estimating them separately, applying the additive-RAS method
Investment (transformation) matrix	negative elements due to the (re)sales of machines produced in previous years	see at the consumption matrix
Budget (Income-Expenditure) matrices	expenditures accounted as negative elements	estimating the incomes and expenditures separately, applying the additive-RAS method
Matrix of assets and liabilities	liabilities accounted as negative elements	similarly to the previous
Social Accounting Matrix	certain transfers accounted as negative entries	accounting them as positive amounts in the transposed cell of the matrix, applying the additive-RAS method

* i.e. if there are negative elements in a column (row) then reallocate them into an additional row (column) labelled 'decreases in …". For example, decreases in inventories can be displayed in an additional row (as a source), while negative net taxes (subsidies) may be displayed in an additional column (as a demand) as suggested for example, by Lenzen [2014], p.205. However, for these modified matrices the row- and column-totals may not be available in the statistics.

5. Combined and sequential applications of biproportional matrix adjustment methods

As in Section 3.3 (The model estimating and transforming the 2010 EU I-O tables to GTAP format) we have already noticed, the matrix adjustment task can be defined as a more complex conditional optimum problem. In this, biproportional matrix adjustment methods can only have role in solving some of the subtasks, especially in the steps of producing the reference matrix (a prior). Biproportional methods therefore not only compete but may also be complementary to each other. Type A input-output tables representing the consolidated balances of domestic and import products can be estimated using the additive RAS method without knowing the sectoral product structure of the import (Révész, 2009). In doing so, we first estimate the type A IOT by the additive RAS method, thus obtaining the breakdown of the import by producing sector. Then, we estimate the so-called "type B" input-output table (which contains the user breakdown only for the domestic products in the upper matrix block) by a similar method in which an estimate is obtained for the row of imports, that is, the breakdown by user. Finally, we estimate the product by user import matrix using the import row referred to above as column sums and the column vector of import by product groups (gained from the type A IOT estimation in the first step) as row sums, and with using the reference import matrix, we can estimate the new import matrix by RAS technique. Interestingly, with the addition of domestic and import product balances, we can get different numbers from the estimated A-type IOT in the first step, which can be overwritten.

Sequential application of bi-directional matrix alignment methods is proposed by the so-called two- or three-stage RAS method (see, for example, Bacharach, 1970, pp. 93-99, and Gilchrist and St. Louis, 1999). In this at first stage, the reference matrix or some of its blocks are adjusted with a RAS method to the required margins, and then the matrix that now satisfies the marginal conditions is to be used as a reference matrix for a new RAS estimation or for solving an entropy model using a more complex target function.

6. Summary

From the naive, heuristic applications to sophisticated procedures of biproportional matrix adjustment, science has gone a long way. These methods can be applied in a growing number of areas, due to the precise formulation of the problem, the explored mathematical properties of the proposed methods, the improved statistical data (which make it possible to produce a better reference matrix), the development of computing (more efficient solving software) and the accumulated international experience. Of course, many mathematical properties and relationships must be clarified. In our article we have also covered them. We have demonstrated that the specific knowledge of the economic phenomena and the characteristics of the reference matrix is a prerequisite for a successful application. In the case of the transformation of EU IOTs into the GTAP sector breakdown we have also shown by practical examples that the good estimation results are mainly due to the good reference matrix (obtained from the initial matrix using a complex, 6 step pre-adjustment). Others, including McNeil and Hendrickson (1985) and Round (2003), also found that if the reference matrix is close to of the target matrix, various models with common target functions leads to very similar estimation results.

Biproportional matrix adjustment methods can be applied not only in isolation, but also sequentially (see, for example, the two-stage RAS method). In addition, the methods and professional tricks presented in this paper can be utilized in more complex mathematical programming problems.

References

Almon, C. (1968): **Recent methodological advances in input–output in the United States and Canada**. – Előadás a 4. Nemzetközi ÁKM konferencián (Fourth International Conference on Input–Output Techniques), Genf.

Bacharach, M. (1970): **Biproportional Matrices and Input-Output Change** (Cambridge, UK: Cambridge University Press)

Black, William R. (1972): **Interregional commodity flows: Some experiments with the gravity model**. Journal of Regional Science, 12 (1): 107-.118.

Bregman, L. M. (1967): **Proof of the Convergence of Sheleikhovskii's Method for a Problem With Transportation Constraints**, USSR Computational Math. and Mathem. Phys. 1(1), 191-204.

Byron, R.P. (1978): **The Estimation of Large Social Account Matrices,** Journal of the Royal Statistical Society, Series A, 141, Part 3, pp. 359-367

Deming, W. E. és Stephan, F. F. (1940): **On a least-squares adjustment of a sampled frequency table when the expected marginal totals are known**, Annals of Mathematical Statistics, 11, pp. 427–444.

Eurostat (2008): " **Eurostat Manual of Supply, Use and Input-Output Tables**", Luxembourg: European Commission, Eurostat

Friedlander, D. (1961): **A technique for estimating contingency tables, given marginal totals and some supplemental data**, Journal of the Royal Statistical Society, Series A, 124, pp. 412–420.

Furness, K. P. (1965): **Time function iteration**, Traffic Engineering and Control, 7, pp. 458–460.

Gilchrist, D. – St. Louis, L. (1999): **Completing input–output tables using partial information with an application to Canadian data**, Economic Systems Research, 11, pp. 185–193.

Günlük-Şenesen, G. – Bates, J. M. (1988): **Some Experiments with Methods of Adjusting Unbalanced Data Matrices**, Journal of the Royal Statistical Society. Series A (Statistics in Society) Vol. 151, No. 3, pp. 473-490

Harrigan, F. and Buchanan, I. (1984): **A quadratic programming approach to input-output estimation and simulation**, Journal of Regional Science, 24: 339–358. doi:10.1111/j.1467-9787.1984.tb00807.x

Harthoorn, R. és J. van Dalen (1987) : **On the adjustment of tables with Lagrange multipliers,** NA-024. Central Bureau of Statistics, The Netherlands, National Accounts Research Division.

Jackson, R. W. – Murray, A. T. (2004): **Alternative Input–Output Matrix Updating Formulations**, Economic Systems Research, Vol. 16, No. 2, June 2004, pp. 135-148.

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

Koppány Krisztián (2016): **Növekedési hozzájárulások számítása input-output táblák strukturális felbontása alapján**, Statisztikai Szemle, 94. évfolyam 8–9. szám, 881-914. oldal, http://www.ksh.hu/statszemle_archive/2016/2016_08-09/2016_08-09_881.pdf (Estimating Growth Contributions By Structural Decomposition Of Input-Output Tables)

Kullback, S. – Leibler, R. A. (1951), "**On Information and Sufficiency**" Ann. Math. Stat. 4, pp. 99-111.

Lahr, M. L.: (2001): '**A strategy for producing hibrid regional input-output tables**'. In: M. L. Lahr and E. Dietzenbacher (eds.): Input-Output Analysis: Frontiers and Extensions. New York: Palgrave, pp. 211–242.

Lahr, M. – Mesnard, L. (2004): **Biproportional Techniques in Input–Output Analysis: Table Updating and Structural Analysis**, Economic Systems Research, Vol. 16, No. 2, June 2004, pp. 115-134.

Lecomber, J. R. C. (1975). **A critique of methods of adjusting, updating and projecting matrices.** In: **Estimating and Projecting Input-Output Coefficients.** R. I. G. Allen and W. F. Gossling. London, UK, Input-Output Publishing Company: pp. 1-25.

Lemelin, A. (2009): **A GRAS variant solving for minimum information loss**, Economic Systems Research, Vol. 21, No. 4, pp. 399–408.

Lemelin, A. – Fofana, I. – Cockburn, J. (2013): **Balancing a Social Accounting Matrix: Theory and application** (revised edition), Partnership for Economic Policy working paper, http://ssrn.com/abstract=2439868

Lenzen, Manfred – Moran, Daniel D. – Geschke, Arne - Keiichiro Kanemoto (2014): **A non-sign preserving GRAS-variant**, Economic Systems Research, Vol. 26, No. 2, 197–208.

Lenzen, Manfred – Wood, Richard – Gallego, Blanca (2007): **Some Comments on the GRAS Method**, Economic Systems Research, 19:4, 461-465, DOI:10.1080/09535310701698613

MacGill, S. M. (1977): "**Theoretical properties of biproportional matrix adjustments**", Environment and Planning A, 9: 687-701.

McNeil, S. and Hendrickson, C. (1985): "**A note on alternative matrix entry estimation techniques**", Transportation Research: Vol. 19B, No. 6, pp. 509-519, 1985, Pergamon Press Ltd.

Mesnard, L. (2011): **Six matrix adjustment problems solved by some fundamental theorems on biproportion**, working paper, University of Burgundy and CNRS, http://ssrn.com/abstract=1692512

Ming-Chang Lee (2014): **Social accounting matrix balanced based on mathematical optimization method and general algebraic modeling system**, British Journal of Economics, Management & Trade 4(8): 1174-1190, 2014

Möhr, M., – Crown, W.H. – Polenske, K.R. (1987): **A Linear Programming Approach to Solving Infeasible RAS Problems**. Journal of Regional Science, 27, 587–603.

Niedercorn, J. H. – Bechdolt, B.V. Jr. (1969): **An economic derivation of the 'gravity law' of spatial interaction**, Journal of Regional Science, 9, 2: 273-82.

Omar, F. H. (1967): **The Projection of Input–Output Coefficients with Application to the United Kingdom**. Unpublished PhD dissertation, University of Nottingham.

Oosterhaven, J. (2005): **GRAS versus minimizing absolute and squared differences: a comment**, Economic Systems Research, 17, pp. 327–331.

Ortúzar, J.D. and L.G. Willumsen (2011): **Modelling Transport**. Fourth Edition. UK, John Wiley and Sons

Polenske, K.R. (1997): **Current uses of the RAS Technique:A Critical Review**. In: A. Simonovits and A.E. Steenge (eds.) Prices, Growth and Cycles. London, MacMillan, 58–88.

Révész, T. (2001): **Költségvetési és környezetpolitikák elemzése általános egyensúlyi modellekkel**, Budapesti Közgazdaság–tudományi Egyetem, Ph.D. értekezés, 2001. március (Fiscal and Environmental Policy Analyses Using General Equilibrium Models)

Révész, T. (2009): **Negyedéves adatokon alapuló ágazati bontású előrejelző és hatás-elemző modell** – Az áfa-bevallási adatbázisnak a legfrissebb hazai ÁKM-mel integrált újszerű alkalmazása – A Kockázatkutató Intézet részére készített tanulmány (TAM-REP (3).DOC file) (Multisectoral forecasting and impact analysis model based on quarterly VAT data – Innovative integrated use of VAT database and Hungarian IOT)

Robinson, S. – Cattaneo, A. - El-Said, M. (1998): **Estimating a Social Accounting Matrix Using Cross Entropy Methods**, TMD discussion paper No. 33, International Food Policy Research Institute

Round, J. I. (2003). "**Constructing SAMs for development policy analysis. Lessons learned and challenges ahead**", Economic Systems Research, 15(2), 161-183.

Rueda-Cantuche, José - Revesz, Tamas - Amores, Antonio F. - Velázquez, Agustín - Mraz, Marian - Ferrari, Emanuele - Mainar, Alfredo - Montinari, Letizia - Saveyn, Bert (2016): **Improving the European Input-Output Database for Global Trade Analysis** (EU-GTAP), Final report June 30, 2016, European Commission JRC Nº33705-2014-11 and DG TRADE 2014/G2/G10

Schneider, M. H. – Zenios, S. A. (1990): **A comparative study of algorithms for matrix balancing**, Operations Research, 38, 3: 439-455

Shannon, C. E. (1948): "**A Mathematical theory of communication**", Bell System Technical Journal 27, 379-423.

SNA (2009): **System of National Accounts, 2008** (SNA2008.pdf), az Európai Bizottság, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and World Bank közös kiadványa

Stone, R. (1961): **Input–Output and National Accounts** (Organization for European Economic Cooperation, Párizs).

Stone, R. – Brown, A. (1962): **A Computable Model of Economic Growth** (Chapman and Hill, London).

Stone, R. (1981): **Balancing the National Accounts; The Adjustment of Initial Estimates: a Neglected Stage in Measurement**, paper presented at the Ivor Pearce Conference, University of Southampton, 5-7 January 1982.

Temurshoev, Umed – Webb, Colin – Yamano, Norihiko (2011): **Projection of Supply and Use tables: methods and their empirical assessment**, Economic Systems Research, 23:1, 91-123, DOI: 10.1080/09535314.2010.534978

Temurshoev, U. – Miller, R. E. – Bouwmeester, M. C. (2013): **A note on the GRAS method**, Economic Systems Research, 25:3, 361-367, Theil, H. (1967): **Economics and information theory**, Rand McNally & Company, Chicago, Studies in mathematical and managerial economics, 7, 488 oldal

Yule, G. Udny (1912): "**On the Methods of Measuring Association Between Two Attributes**". Journal of the Royal Statistical Society. 75 (6): 579–652. doi:10.2307/2340126 (http://www.jstor.org/stable/2340126?origin=crossref)

Zalai, E. (2012): **Matematikai közgazdaságtan I.** Általános egyensúlyi modellek és mikroökonómiai elemzések – **II.** Többszektoros modellek és makrogazdasági elemzések, (**Mathematical Economics I.** General Equilibrium Models and Microeconomic Analyses – **II.** Multisectoral models and macroeconomic analyses) Közgazdasági es Jogi Könykiadó, Budapest