AGRODEP Technical Note 0017

December 2020

A Comparative Analysis of Updating and Balancing Methods for Social Accounting Matrices

Fousseini Traoré
Alhassane Camara
Kike Yra Fonton

Table of Contents

List of Tables

List of Figures

I. Introduction

Computable General Equilibrium (CGE) models, which are widely used in the literature to quantify potential impacts of economic public policies, are calibrated on data from Social Accounting Matrices (SAM). SAM belongs to national accounting systems and represents interactions between activities and institutions within an economy. According to Decaluwé & al. (2001) "on one side, a SAM shows a coherent representation of transactions which have taken place within an economy - country, region or a set of countries or regions – while on other side, it provides policy makers with an accounting basis for an analytical framework that can facilitate their choices." [1]

Building a SAM for recent years may, however, be difficult and time-consuming, since it requires accessing, gathering, and compiling data from different sources.[2] Furthermore, even if these data are available at the time of building the SAM, they may be obsolete and hence, useless. To overcome this issue, the literature suggests updating existing SAM using the latest information and data about the economy under study (Robinson et al., 2001). Several methods have been employed through the literature, thus paving the way to discussions about their strengths and weaknesses.

In a pioneering study work, Robinson et al. (2001) use Monte Carlo simulations to randomly disrupt a SAM that was initially balanced, in order to compare Cross-Entropy and RAS methods. The unbalanced SAM is then updated to make it coherent with new information using both methods. Results indicate that if the analyst focuses on the matrix of coefficients (ratios), the cross-entropy method is better than the RAS method. If the analyst, however, focuses on flows, then both methods are equivalent; the RAS method reporting slightly better results. In the same line, Ahmed and Preckel (2007) analyze the precision of both methods (CE and RAS). To do so, they first update an old Input-Output (I-O) table. Then they compare it with observed data. Results show that divergences between the updated I-O table and the observed one, are relatively small when using the CE method. Lee (2015) studies performances of RAS, CE, Ordinary Least Square (OLS) and Linear Programming. Like previous authors, Lee reveals that the CE approach is more precise than other considered methods.

Gunluk-Senesen and Bates (1988) present a more nuanced picture of the superiority of the CE method over other approaches. According to the authors, the attractiveness of methods based on minimizing distance function changes along with the level of aggregation. In fact, at the aggregate level, methods of minimizing distance function are not satisfactory, while they become more adequate at the disaggregate level. Thus, Gunluk-Senesen and Bates (1988) support that an improvement of the performance of those methods can be reached with little disaggregation, i.e., when it is possible to select sectors to be disaggregated. In

[1] For more details, see Fofana (2007), Decaluwé, Martens and Savard (2001), Pyatt and Round (1985)
[2] TRE:Supply and Use Table, TCEI : Table of Integrated Economic Accounts, TOFE : Government Financial Operations Table, BP : Balance of payments, a household survey.

addition, the (difference) criteria that one can use may be biased in favor of one of those methods due to the nature of the objective function. Finally, it is worth mentioning that the RAS method is a specific case of the more global CE approach.

When updating a SAM, it is critical to deal with structural changes that take place between the initial SAM and the final SAM. These changes include: occurrence of new transactions (type 1) or changes in the sign of net transactions (type 2). Type 1 changes may concern any cell[3] in the SAM; the big challenge for the analyst is to (i) identify them before updating the SAM and (ii) find suitable methods to include them in the updated SAM instead of using a purely statistical technique. Type 2 changes usually concern savings from government, the rest of world or a certain category of households. When they occur (for example, government's savings may change from negative to positive value), analysts must find ways to consider them as new constraints to be included in the updated SAM. Another major issue faced by analysts when rebalancing a SAM is negative values (subsidies or in some cases savings), as the objective functions in some optimization problems (like cross-entropy) do not allow negative values. The challenge for analysts is to employ techniques that can maintain signs for transactions such as subsidies, and are flexible enough to replicate potential sign reversal in transactions such as savings.

In the same vein[4] of previous studies, the objective of this technical note is to compare four methods (CE, OLS, Huber and Hellinger's methods) when updating and balancing a SAM. Unlike previous studies, the contributions of this note are:

i. This note goes beyond the traditional comparison of CE and RAS methods: it presents an application of Huber and Hellinger methods. To the best of our knowledge, those methods have not been implemented in the literature on updating and balancing a SAM.

ii. On the contrary of studies based on a Monte Carlo simulation framework, this note makes a comparison of the above methods using an analysis of divergence between structures of updated SAM and observed one[5]. In addition, unlike Ahmed and Preckel (2007), this note compares results of considered methods following different criteria: a) comparison between updated ratios and observed one, b) the ability of methods to reproduce observed structural changes, c) number of accounts presenting an extreme value, d) analysis of the structure of production and factor intensity between sectors.

This note is structured as: section 1 presents a description of methods, sections 2 presents data and results and section 3 concludes.

[3] Which have an economic interpretation.
[4] With the exception of the study by Lee (2014).
[5] Ahmed and Preckel (2007) when using input and output table.

II. Description of methods

In this section, we present the conceptual framework of each above-mentioned method.

a. Cross-Entropy (CE)

The cross-entropy method is a flexible and powerful technique to deal with inconsistent and limited data[6]. Robinson et al. (2001) describe cross-entropy as a method to solve ill-posed problems; in the case of balancing or updating a SAM, the problem is ill-posted (under identified) because one must determine n^2 positive parameters while it is only possible to use 2n-1 independent restrictions (sums of columns and rows). To find a unique solution without losing degrees of freedom, we must impose restrictions, using only all available and reliable information. Information can include: measurement errors on observed variables, knowledge about some sub-matrices of SAM, availability of previous SAM, knowledge of sums of columns and rows of the new SAM, macroeconomic aggregates including GDP, value added per sector, public consumption, private consumption, investment or public savings. When a prior SAM is available, the optimization problem consists in estimating new coefficients sufficiently close to those of previous SAM under multiple constraints reflecting new information on the structure of the economy. Constraints can be formulated using linear relationships between cells of SAM or using inequalities.

Let's assume T is a matrix with T_{ij} elements[7], each element (cells) represents a payment from a column j to a row i. For each account, total receipts (y_i) must equal total expenditures (y_j). In other words, we have the following identity:

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

To obtain the coefficients of $a_{i,j}$, we divide each cell by the total sum:

$$a_{i,j} = \frac{T_{i,j}}{\sum_{i,j} T_{i,j}} \qquad (2)$$

[6] See Golan (1996). In fact, the method can use a priori information about some sub-matrices of SAM and inequality constraints; it can also deal with measurement errors on variables by including stochastic terms in the optimization problem.
[7] We use the same notation as Robinson et al. (2001)

When a prior matrix $a^0_{i,j}$ exists, the cross-entropy method finds a new set of coefficients $a_{i,j}$, which minimizes the distance of Kullback-Leibler (1951)[8] between $a^0_{i,j}$ and $a_{i,j}$:

$$Min\ CE_{\{a_{i,j}\}} = \sum_{i,j} a_{i,j} ln\left(\frac{a_{i,j}}{a^0_{i,j}}\right) \quad (3)$$

$$S.C:$$

$$\sum_{i,j} a_{i,j} = 1 \quad (4)$$

$$\sum_{i,j} G^k_{i,j} a_{i,j} = \gamma^k \quad (5)$$

Equation 3 is the objective function to be minimized; equation 4 is the additivity constraint and finally, equation 5 is a set of linear constraints reflecting macroeconomic conditions. G is a matrix (n X n) that aggregates cells and γ is the observed value of macroeconomic aggregate.

Cross-entropy values reflect the way in which the quantity of information introduced in the program changes the (updated) balanced SAM compared to the initial one. The more restrictive the constraints introduced, the greater the distance between the updated SAM and the initial one. In contrast, when introduced information is less binding, the cross-entropy measure of distance is closer to zero. This method can be extended to include stochastic terms into constraints (see Golan, 1996).

b. Ordinary Least Square (OLS)

The OLS method consists in finding a new set of $a_{i,j}$ which minimizes the square of the distance between $a_{i,j}$ *and* $a^0_{i,j}$

$$Min\ MCO_{\{a_{i,j}\}} = \sum_{i,j} \left(a_{i,j} - a^0_{i,j}\right)^2 \quad (6)$$

$$S.C:$$

$$\sum_{i,j} a_{i,j} = 1 \quad (7)$$

$$\sum_{i,j} G^k_{i,j} a_{i,j} = \gamma^k \quad (8)$$

[8] From a formal perspective, we would rather use the term “divergence” of Kullback-Leibler since the value isn’t symmetrical and doesn’t satisfy the triangular inequality.

This approach assumes that there exists a linear relationship between $a_{i,j}$ and $a^0_{i,j}$. Equations 7 and 8 are the same as in the cross-entropy optimization problem. The quadratic form in the objective function indicates that higher distances are negatively more weighted than lower ones. The main drawback of this method is its lack of robustness against outliers. As previously mentioned, due to the rapid expansion of the quadratic curve, outliers tend to drive the solution away from the true minimum.

c. Minimization of the cost function of Huber (1964)

This method consists in minimizing a cost function of errors. This function has a quadratic form when error terms are lower than a certain threshold while it becomes linear when error terms are beyond this threshold. Due to the nature of cost function, the solution (minimum) is relatively more robust against the effect of outliers.

$$Min\ HU_{\{a_{i,j}\}} = \sum_{i,j} C\left(a^0_{i,j} - a_{i,j}\right) \quad (9)$$

$$S.C:$$

$$\sum_{i,j} a_{i,j} = 1 \quad (10)$$

$$\sum_{i,j} G^k_{i,j} a_{i,j} = \gamma^k \quad (11)$$

$$C(\delta) = \begin{cases} \delta^2\ si\ |\delta| < b & (12) \\ \quad 2b|\delta| - b^2\ \ otherwise \end{cases}$$

b is the threshold. The cost function C is continuous, convex and these properties make the convergence to a global minimum more reliable.

To make the above problem feasible, the following function is used:

$$C(\delta) = 2b^2\left(\sqrt{1 + \left(\frac{\delta}{b}\right)^2} - 1\right) \quad (13)$$

This function is closer to that of Huber; however, its derivatives are continuous. The function tends to be δ^2 when δ is lower while it tends to be linear when δ is higher. To determine the threshold b, we analyze the distribution of absolute errors from OLS method using a Box-plot. The value of b is the upper bound of Box-plot and is determined as follow:

$$min\ [max(\delta_{ij}),\ Q3 + 1.5 * (Q3 - Q1)]$$

Where δ_{ij} is an absolute error resulting from the OLS method, Q3 and Q1 are the first and the third quartiles of δ_{ij}.

d. Minimization of Hellinger distance

The objective of this approach is to minimize the Hellinger distance HE ($a_{i,j}^0$, $a_{i,j}$) where $a_{i,j}^0$,$et\ a_{i,j}$ are the same as in previous optimization problems

$$Min\ HE_{\{a_{i,j}\}} = \sum_{i,j} \left(\sqrt{a_{i,j}^0} - \sqrt{a_{i,j}} \right)^2 \quad (14)$$

$$S.C:$$

$$\sum_{i,j} a_{i,j} = 1 \qquad (15)$$

$$\sum_{i,j} G_{i,j}^k a_{i,j} = \gamma^k \qquad (16)$$

Equations 15 and 16 are the same as in previous problems. Like the cross-entropy method, the assumption is that coefficients can be considered as probabilities. However, unlike the "Kullback-Leibler distance," the Hellinger distance is symmetric. Both methods belong to the same paradigm (f-divergence, see Csiszar and Shields, 2004)

III. Data and Applications

a. Data

We use both a prior SAM (2006) and a final or observed SAM (2011) from Rwanda[9]. The objective is to update the initial SAM (2006) using above-mentioned methods and compare each updated SAM with the final or observed one (2011). In the optimization problem, we use macroeconomic constraints[10] that shape the new structure of the considered economy in 2011. Before updating the SAM, we however check the coherence between the initial SAM (2006) and the final SAM (2011).

The initial SAM (2006) has 117 accounts including 53 activities, 53 products, 3 factors, and 8 institutions, while the final SAM (2011) has 126 accounts including 54 branches, 54 products, 9 factors, and 9 institutions. We aggregated some accounts to have a perfect correspondence between the two SAMs and

[9] The Rwanda SAMs come from the International Food Policy Research Institute (IFPRI).
[10] Data on macroeconomic constraints come from World Development Indicators (WDI) from World Bank.

we finally obtained 24 accounts in both matrices (see tables 13 and 14 in the Annex). Macroeconomic constraints are ratios reflecting the structure of GDP (See table 1).

Table 1 shows that the structure of Rwanda's economy observed in the SAMs is very similar to that resulting from World Bank data. Therefore, it is reasonable to use those data as external sources to update the initial SAM (2006).

Table 1. Macroeconomic indicators

	SAM		WDI		Ratio WDI/SAM	
Indicators	**2006**	**2011**	**2006**	**2011**	**2006**	**2011**
Share of primary sector[1] in GDP	0.36	0.26	0.36	0.28	0.99	1.09
Share of industry sector in GDP	0.12	0.14	0.13	0.17	1.10	1.19
Share of services in GDP	0.44	0.54	0.45	0.47	1.01	0.88
Share of finale expenditure in GDP[2]	0.94	0.93	0.96	0.92	1.02	0.99
Share of investment in GDP	0.17	0.20	0.17	0.23	0.97	1.13

Source: Authors
Note: 1 : agriculture and livestock ; 2 including households final expenditure and government public expenditure

b. Results

Before comparing the updated SAMs using above-mentioned methods with the observed SAM (2011), we first analyze differences between the SAM (2006) and the final or observed SAM (2011) in order to identify changes in the structure of the economy (apparition of new transactions, changes in signs). This comparison shows that some changes have taken place in Rwanda's economy between 2006 and 2011. In fact, the final SAM (2011) has relatively 5 new transactions that were absent in the initial SAM (2006): imports of livestock products, imports of private services, labor payment to the rest of world, capital payment to the government and exports of services (cells highlighted in yellow in table 2). Thus, the question is: do above-mentioned methods reproduce those changes when updating the initial SAM (2006) using changes in macroeconomic conditions over the 2006-2011 period?

First, we solve optimization problems without imposing any constraints on the stability of several zeros nor on signs in the initial SAM.

Table 2. Differences between initial SAM (2006) and the observed SAM (2011)

		a1	**a2**	**a3**	**a4**	**a5**	**a6**	**a7**	**a8**	**a9**	**a10**	**a11**	**a12**	**a13**	**a14**	**a15**	**a16**	**a17**	**a18**	**a19**	**a20**	**a21**	**a22**	**a23**	**a24**
Agriculture	a1																								
Livestock, Fishery, silviculture	a2																								
Mining	a3																								
Food processing	a4																								
Other Industries	a5																								
Private services	a6																								
Public administration	a7																								
Agriculture	a8																								
Livestock, Fishery, Forestry	a9																								
Mining	a10																								
Food processing	a11																								
Other Industries	a12																								
Private services	a13																								
Public administration	a14																								
Labor	a15																								
Capital	a16																								
Margins	a17																								
Households	a18																								
Direct tax	a19																								
Tariffs	a20																								
Sales tax	a21																								
Government	a22																								
Saving/Investment	a23																								
Rest of the World	a24																								

Source: Authors' computation

We compare performances of above-mentioned methods using multiple criteria: i) analyzing the divergence between updated coefficients and observed coefficients, ii) identifying new accounts, stability of zeros and signs, iii) counting the number of accounts presenting an extreme value and finally iv) comparing the structure of the production of the updated SAM with that of the observed SAM.

i) Analysis of divergence between updated coefficients and observed ones

We present two measures of divergences between updated coefficients and observed ones: Mean Absolute Error (MAE) and Mean Relative Error (MRE). Results presented in table 3 show that MAEs are relatively lower with the CE and Hellinger's methods compared to OLS and Huber's methods (about 1.17 times higher).

Table 3. Mean Absolute Error (absolute value)

Absolute Errors	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
CE	576	0.0004785	0.0020273	0	0.0185584
HELLINGER	576	0.000477	0.0020267	0	0.0185584
OLS	576	0.0005574	0.0022177	0	0.0223009
HUBER	576	0.0005563	0.0022159	0	0.0222394
OLS^b	576	0.000561	0.0021345	0	0.019688
$HUBER^b$	576	0.0005291	0.0020303	0	0.0184323

^{b} Relative errors are introduced into the objective function to be minimized.
Source: Authors' computation

OLS^b and $Huber^b$ methods are more precise than OLS and Huber, although they remain relatively less efficient than CE and Hellinger's method. When using the MRE criteria[11], OLS^b and $Huber^b$ methods become relatively more precise than CE and Hellinger's method (table 4).

Table 4. Average proportional deviations (absolute value)

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
CE	91	1.600031	7.680123	0.0030538	72.4061
HE	91	1.534106	7.482143	0.0058156	70.58877
OLS^b	91	1.290636	5.836826	0.0100438	54.37582
$HUBER^b$	91	1.2906	5.835788	0.0021862	54.37077

Source: Authors' computation
^{b}Relative errors are introduced into the objective function to be minimized.

[11] To avoid dividing by zero, we only use the sub-sample of cells that are not null.

ii) Analysis of changes in the structure of SAM

As a reminder, we have observed in the previous section that the observed SAM (2011) presents 5 new transactions in comparison with the initial SAM (2006). [12] In this section, we analyze the efficiency of methods in term of their abilities to replicate observed structural changes in Rwanda's economy over the period 2006-2011. Results in table 5 show that all four methods fail to reproduce correctly those changes. In addition, Huber's method and OLS, are relatively less efficient than CE and Hellinger's method. In fact, those methods assign zeros to transactions that have lower shares (weights) in the initial SAM. Therefore, they present a higher number of cells with null values in the updated SAMs, although their equivalents in the observed SAM report positive values. In other words, Huber's method and OLS overestimate the number of zeros in the updated SAM. However, when we use Huber[b] method and OLS[b], we observe an increase in the efficiency of Huber's method and OLS, although they remain less efficient than CE and Hellinger's method.

Finally, all four methods also fail to replicate changes in signs within cells. This is due to the nature of the functions used in optimization problems (logarithm, square root). In fact, public saving is negative in the initial SAM (2006) while it is positive in the observed SAM (2011). Results show that CE and Hellinger's method maintain the signs of cells in the initial SAM while OLS and Huber's method assign zeros to public saving in the updated SAM, because of its lower share (weight) in the initial SAM.

Table 5. Comparison of methods using structural change criteria

	Rwanda					
	CE	**Hellinger**	**Huber**	**OLS**	**Huber[b]**	**OLS[b]**
Cells with no values (zeros) in the updated SAM while their equivalents in the observed SAM have values (non-zeros)	5[13]	5[14]	18[15]	18[16]	3[17]	5[18]
Cells with values (non-zeros) in the updated SAM while their equivalents in the observed SAM have no value (zeros)	0	0	82	76	23	32
Cells with negative values in the updated SAM while their equivalents in the observed SAM have positive values	1=public saving	1=public saving	0	0	1=public saving	1=public saving
Cells with positive values in the updated SAM while their equivalents in the observed SAM have negative values.	0	0	0	0	0	0

Source: Authors' computation

[b]Relative errors are introduced into the objective function to be minimized.

[12] Imports of livestock products, imports of private services, labor payment to the rest of world, capital payment to the government and exports of services.

[13] These accounts come from the difference between initial SAM and observed SAM. They are presented above.

[14] These accounts come from the difference between initial SAM and observed SAM.

[15] Including 4 out of 5 accounts coming from the difference between initial SAM and observed SAM.

[16] Including 4 out of 5 accounts coming from the difference between initial SAM and observed SAM

[17] Salaries paid to the rest of world, capital income paid to the government, services exports.

[18] These accounts come from the difference between initial SAM and observed SAM.

To sum up, these results reveal that more attention must be paid (i) to accounts that have changed sign; and ii) to the stability of the number of cells with no value (zeros). First, the analyst must fix values of accounts that have changed sign to the observed values; second, he/she also must keep constant the number of cells with no value (zeros) in the initial matrix in order to avoid the creation of transactions that are entirely meaningless in the updated SAM (mainly when we use OLS or Huber's method).

Consequently, we re-run the updating process, including new constraints: (i) we keep constant cells with no value (zeros) in the initial SAM, (ii) we fix the value of public saving (% of GDP) to the value observed in 2011.

Results in table 6 show that the introduction of these constraints help reduce absolute divergences between updated SAMs and the observed one. The reduction is relatively higher with OLS and Huber's method and their modified versions.

Table 6. Analysis of absolute errors when constraints are included in the optimization problem.

Absolute errors	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
CE	576	0.0004354	0.0018493	0	0.0185584
HELLINGER	576	0.0004342	0.0018453	0	0.0185584
OLS	576	0.0004961	0.0020928	0	0.0209747
HUBER	576	0.0004941	0.0020875	0	0.0211208
OLS^b	576	0.0004278	0.0018495	0	0.0185584
$HUBER^b$	576	0.0004457	0.0019369	0	0.0185584

Source: Authors' computation

[b]Relative errors are introduced into the objective function to be minimized.

Reported values of MRE show that all four methods use additional information from new constraints efficiently. MRE values are relatively lower than in the previous case; the performance gain is relatively higher with the modified versions of OLS and Huber's method. (OLS^b and $Huber^b$).

Table 7. Analysis of relative errors when constraints are included in the optimization problem

Variable Obs	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
CE	91	0.7735783	1.671276	0.0052806	9.846386
HE	91	0.7580043	1.630112	0.0000259	9.900541
OLS^b	91	0.6888942	1.549918	0.0068873	10.87809
$HUBER^b$	91	0.6921274	1.551873	0.0095737	10.8891

[b]Relative errors are introduced into the objective function to be minimized.

Source: Authors' computation

iii) Number of accounts presenting an extreme value

It is also important to compare the number of accounts presenting an extreme value when we use the different methods presented above. In this section, we still use the MRE (absolute) values, and outliers that present a relative divergence above 75%.

Results of all four methods are presented in Tables 8-11; it is obvious that the number of accounts presenting an extreme value in OLS or Huber's method is two times higher than that in CE or Hellinger's method.

Table 8: Deviations observed with the cross entropy method

		a1	**a2**	**a3**	**a4**	**a5**	**a6**	**a7**	**a8**	**a9**	**a10**	**a11**	**a12**	**a13**	**a14**	**a15**	**a16**	**a17**	**a18**	**a19**	**a20**	**a21**	**a22**	**a23**	**a24**
Agriculture	a1																								
Livestock, Fishery, forestry	a2																								
Mining	a3																								
Food processing	a4																								
Other Industries	a5																								
Private services	a6																								
Public administration	a7																								
Agriculture	a8																								
Livestock, Fishery, Forestry	a9																								
Mining	a10																								
Food processing	a11																								
Other Industries	a12																								
Private services	a13																								
Public administration	a14																								
Labor	a15																								
Capital	a16																								
Margins	a17																								
Households	a18																								
Direct tax	a19																								
Tariffs	a20																								
Sales tax	a21																								
Government	a22																								
Saving/Investment	a23																								
Rest of the World	a24																								

Source: Authors' computation
Note: Accounts highlighted in color present a deviation of 75% (absolute value)

Table 9: Deviations observed with the Hellinger method

		a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18	a19	a20	a21	a22	a23	a24
Agriculture	a1																								
Livestock, Fishery, Forestry	a2																								
Mining	a3																								
Food processing	a4																								
Other Industries	a5																								
Private services	a6																								
Public administration	a7																								
Agriculture	a8																								
Livestock, Fishery, forestry	a9																								
Mining	a10																								
Food processing	a11																								
Other Industries	a12																								
Private services	a13																								
Public administration	a14																								
Labor	a15																								
Capital	a16																								
Margins	a17																								
Households	a18																								
Direct tax	a19																								
Tariffs	a20																								
Sales tax	a21																								
Government	a22																								
Saving/Investment	a23																								
Rest of the World	a24																								

Source: Authors' computation
Note: Accounts highlighted in color present a deviation of 75% (absolute value)

Table 10: Deviations observed with the Huber method

		a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18	a19	a20	a21	a22	a23	a24
Agriculture	a1																								
Livestock, Fishery, Forestry	a2																								
Mining	a3																								
Food processing	a4																								
Other Industries	a5																								
Private services	a6																								
Public administration	a7																								
Agriculture	a8																								
Livestock, Fishery, Forestry	a9																								
Mining	a10																								
Food processing	a11																								
Other Industries	a12																								
Private services	a13																								
Public administration	a14																								
Labor	a15																								
Capital	a16																								
Margins	a17																								
Households	a18																								
Direct tax	a19																								
Tariffs	a20																								
Sales tax	a21																								
Government	a22																								
Saving/Investment	a23																								
Rest of the World	a24																								

Source: Authors' computation

Note: Accounts highlighted in color present a deviation of 75% (absolute value)

Table 11: Deviations observed with the OLS method

		a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18	a19	a20	a21	a22	a23	a24
Agriculture	a1																								
Livestock, Fishery, Forestry	a2																								
Mining	a3																								
Food processing	a4																								
Other Industries	a5																								
Private services	a6																								
Public administration	a7																								
Agriculture	a8																								
Livestock, Fishery, Forestry	a9																								
Mining	a10																								
Food processing	a11																								
Other Industries	a12																								
Private services	a13																								
Public administration	a14																								
Labor	a15																								
Capital	a16																								
Margins	a17																								
Households	a18																								
Direct tax	a19																								
Tariffs	a20																								
Sales tax	a21																								
Government	a22																								
Saving/Investment	a23																								
Rest of the World	a24																								

Source: Authors' computation

Note: Accounts highlighted in color present a deviation of 75% (absolute value)

Table 12: Deviation observed with the OLS^b and Huber^b methods

		a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18	a19	a20	a21	a22	a23	a24
Agriculture	a1																								
Livestock, Fishery, Forestry	a2																								
Mining	a3																								
Food processing	a4																								
Other Industries	a5																								
Private services	a6																								
Public administration	a7																								
Agriculture	a8																								
Livestock, Fishery, Forestry	a9																								
Mining	a10																								
Food processing	a11																								
Other Industries	a12																								
Private services	a13																								
Public administration	a14																								
Labor	a15																								
Capital	a16																								
Margins	a17																								
Households	a18																								
Direct tax	a19																								
Tariffs	a20																								
Sales tax	a21																								
Government	a22																								
Saving/Investment	a23																								
Rest of the World	a24																								

Source: Authors' computation
Note: Accounts highlighted in color present a deviation of 75% (absolute value

)

iv) Analysis of the structure of production and factor intensity (without any modification in OLS and Huber's method)

In the industry sectors (A3-A5) and service sectors (A6 and A7) where we respectively observe an increase in the share of value added (VA) (% of GDP) over the period 2006-2011, all four methods overestimate the share of VA in production. It is noticeable that overestimations are relatively higher with OLS[19] and Huber's method (see figure 1). When we turn to the analysis of induced changes in factor intensity, the following observations arise:

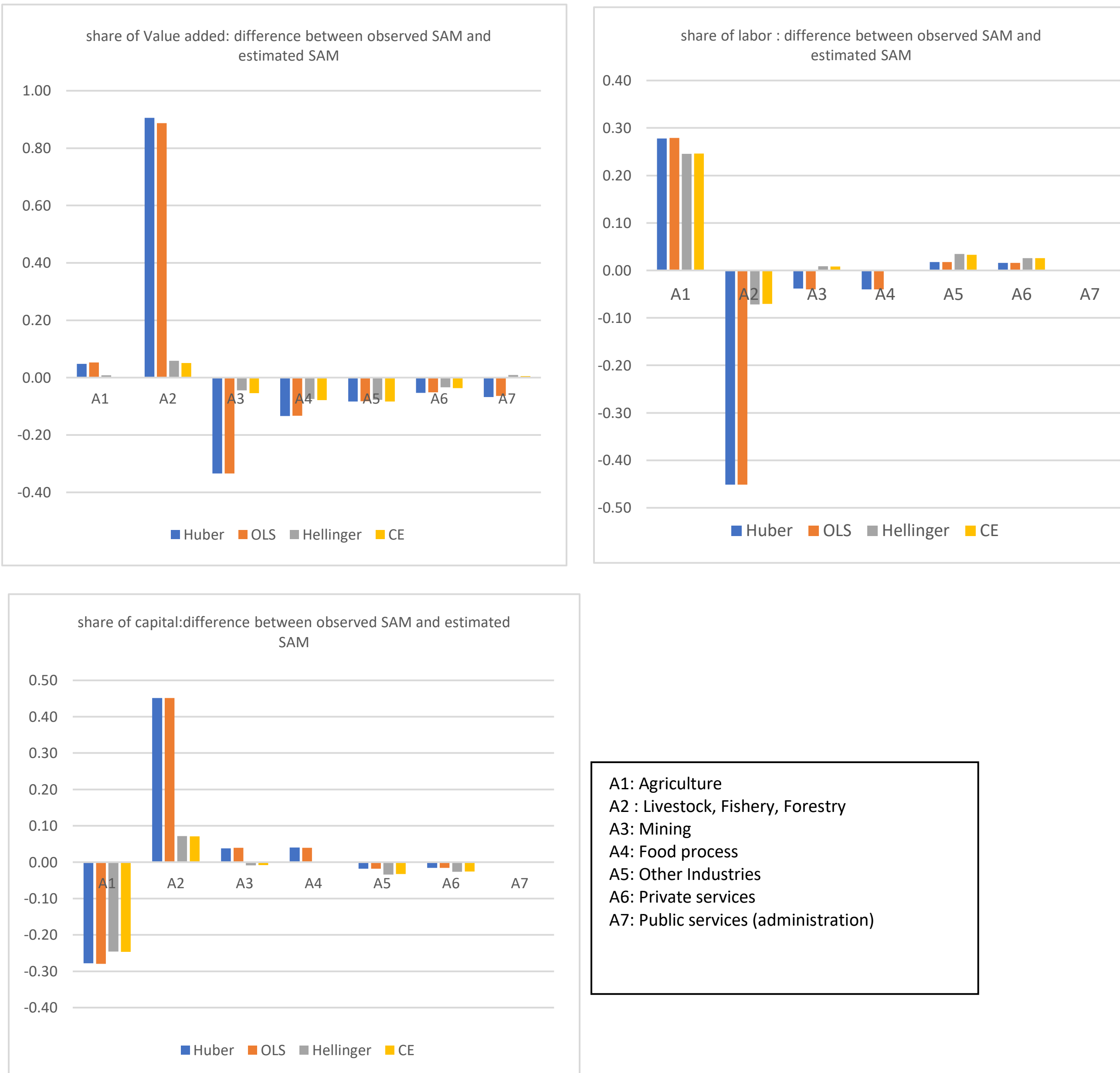
(i) Within the sub-sectors, A3 and A4, which are capital intensive (based on initial SAM data), OLS and Huber's method underestimate the share of capital (at the expense of labor share) while CE and Hellinger's method exhibit better estimates. However, in the sub-sector A5, which is also capital intensive, but relatively less capital intensive than the sub-sectors A3 and A4, all four methods overestimate the share of capital[20].

(ii) In the sub-sector A6, which is labor intensive, all four methods underestimate the share of labor, and hence overestimate share of capital. They show perfect estimates of factor intensity in the sub-sector A7 (public services).

In the primary sector[21], where we observe a decrease in the share of VA (%GDP) over the period 2006-2011, OLS and Huber's method underestimate the share of VA in production within all sub-sectors, while CE and Hellinger's method show better estimates. In the sub-sector A1, which is capital intensive, all four methods overestimate share of capital, while in the sub-sector A2, which is labor intensive, they overestimate the labor share. In both cases, OLS and Huber's method exhibit remarkably higher overestimations.

[19] With the exception of sector A5 where OLS presents lower errors.
[20] Overestimations are higher with CE and Hellinger's method
[21] Where more than 90% of production comes from value added (based on data from the initial SAM)

Figure 1. Structure of the production and factor intensity, Rwanda



v) *Analysis of the structure of production and factor intensity (with OLS^b and Huber^b method)*

In the industry sub-sectors of A3 and A4, OLS^b and Huber^b' method underestimate the share of VA in production whereas CE and Hellinger's method overestimate it; with the underestimation higher than the overestimation. In the sub-sector of A5, all four methods overestimate the value-added share.

In the service sub-sector A6, all four methods overestimate the share of value added in the production[22] while they underestimate it in the sub-sector of A7[23] (see figure 2). The analysis of changes in factor intensity show the following points:

i. In the sub-sectors A3 and A4, which are capital intensive, all methods show better estimates of capital share[24]. However, in the sub-sector A5, which is also capital intensive, but relatively less capital intensive than sub-sectors A3 and A4, those methods overestimate the capital share, and the Huber^b' method shows the highest overestimation.

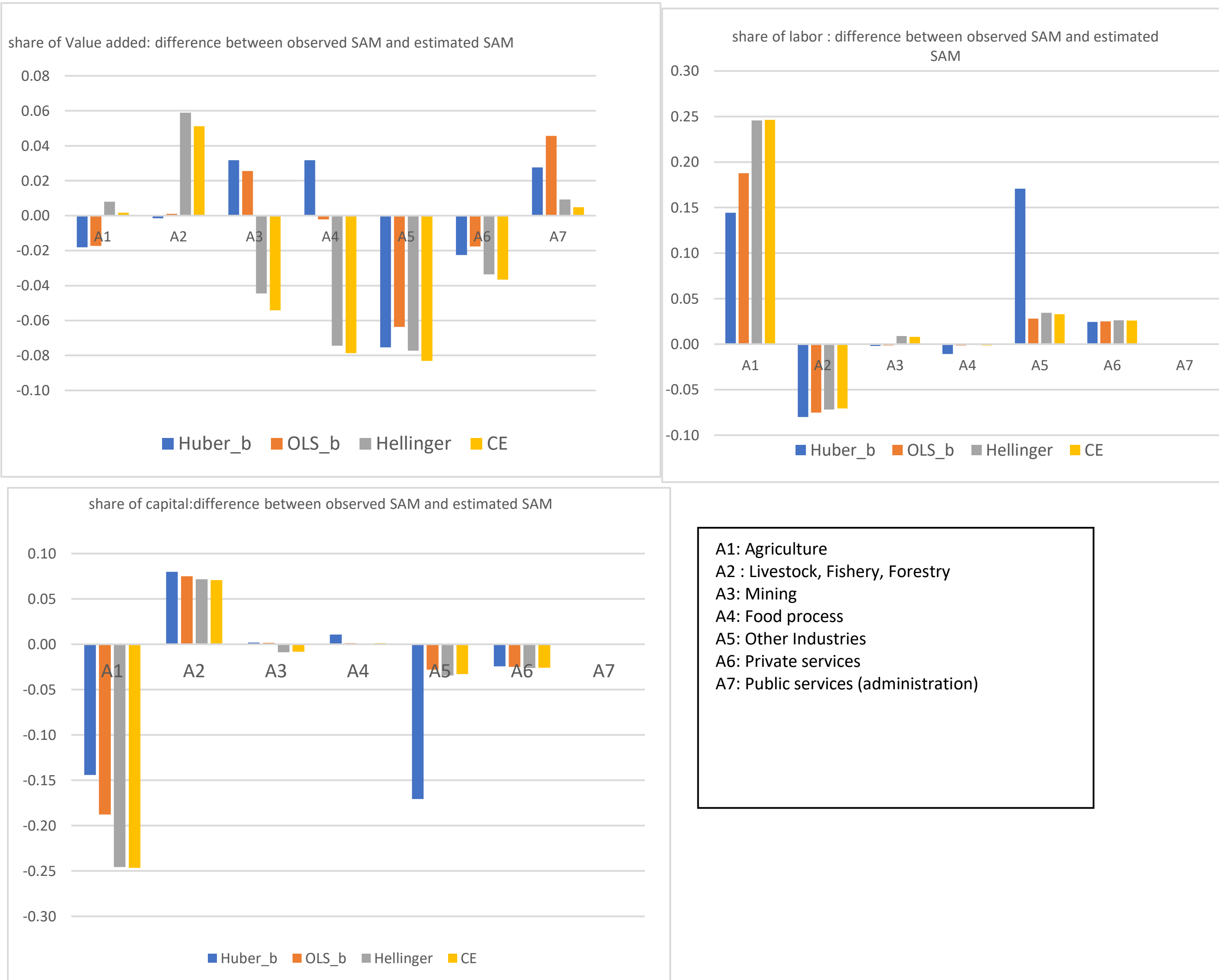
ii. In the sub-sector A6 which is labor-intensive, all methods underestimate the labor share (overestimate capital share). They exhibit perfect estimates of factor intensity in the sub-sector A7 (public services).

In the sub-sector A1, OLS^b and Huber^b's method overestimate the share of value added (VA) in production whereas CE and Hellinger's method show better estimates. On the contrary of sub-sector 1, CE and Hellinger's method underestimate the value-added share in subsector A2 when OLS^b and Huber^b's method present better results.

In the sub-sector A1 which is capital intensive, the four methods overestimate the capital share when in sub-sector A2 which is labor intensive, they overestimate the labor share. OLS^b and Huber^b's method exhibit higher overestimations in sub-sector A2 while they present lower overestimations in subsector A1.

[22] CE and Hellinger's method show higher overestimations.
[23] OLS^b, Huber^b method show higher underestimations.
[24] Although we observe a lower overestimation with CE and Hellinger methods in A3, and a lower underestimation of Huber in A4.

Figure 2. Structure of the production and factor intensity, Rwanda



Furthermore, it is of interest to analyze the ability of the above-mentioned methods to reproduce the evolution of factor intensity over time. Results presented in table 13 show that CE, Hellinger's method, modified versions of OLS, and Huber's method replicate the main features of the evolution of factor intensity between 2006 and 2011.[25] In addition, those methods closely reproduce the direction of changes in factor intensity, when OLS and Huber's method (without any modifications) do not. In fact, OLS and Huber's method highly overestimate the share of labor in subsector A2, which is the most labor-intensive sub-sector[26].

Table 13. Capital intensity in sectors

	SAM 2006	SAM 2011	CE 2011	Hellinger 2011	OLS 2011	Huber 2011	OLS^b 2011	$Huber^b$ 2011
A1	0.76	0.521	0.768	0.767	0.801	0.799	0.709	0.665
A2	0.371	0.451	0.381	0.38	0	0	0.376	0.372
A3	0.516	0.518	0.527	0.527	0.479	0.48	0.517	0.516
A4	0.56	0.571	0.57	0.572	0.531	0.531	0.570	0.560
A5	0.511	0.488	0.521	0.523	0.506	0.506	0.516	0.659
A6	0.479	0.463	0.489	0.489	0.479	0.479	0.488	0.488
A7	0	0	0	0	0	0	0.000	0.000

Source: Authors' computation

In summary, the above analysis reveals the relative superiority of CE and Hellinger methods over OLS and Huber methods. The main results of that comparison can be synthesized as follows:

- CE and Hellinger method exhibit lower mean errors regarding the estimation of ratios.
- When we introduce the relative error measure in the objective function, OLS and Huber's method become more efficient; and in some cases, they are better than CE Hellinger's method.
- All methods fail to reproduce observed structural changes (occurrence of new transactions, changes in sign). However, CE and Hellinger's method perform better than other methods. In fact, when we compare the updated SAM[27] to the observed one, OLS and Huber's method create new transactions in the updated SAM that should not appear, even though we introduce a modification in the objective function. Additionally, OLS and Huber's method assign zeros to the transactions with lower shares (weights) in the initial SAM.
- OLS and Huber's method show a larger number of accounts presenting an extreme value of relative mean error (> 75%). However, introducing a modification in the objective function eliminate those differences between OLS, Huber's method, and CE, Hellinger's method.

[25] However, in subsector A5, they do not replicate those features correctly.
[26] After subsector A7 (public services) that uses mainly labor.
[27] Without imposing constraints about maintaining zeros from the initial SAM.

- All methods overestimate (underestimate) the share of value added (% of the production) within sectors where we observe an increase (decrease) in the contribution to GDP. The over (under) estimations are larger with OLS and Huber's method.
- When the contribution of the sector to GDP increases (decreases), all methods underestimate (overestimate) the share of a factor in which the considered sector is more (less) intensive. Cross-entropy, Hellinger's method and the modified versions of OLS and Huber's method present better estimates[28].
- Furthermore, cross entropy, Hellinger's method and the modified version of OLS and Huber's method are more suited to reproduce the direction and the extent of changes in the factor intensity over time.

IV. Conclusion

Social Accounting Matrices are part of the larger group of national accounting systems and are mainly used as data sources for Computable General Equilibrium Models (the latter being widely used to assess ex-ante impacts of public policies). Building a recent disaggregated SAM might, however, be difficult and time-consuming because of issues linked to data availability and access, data processing and compilation from various sources[29]. The literature presents different methods to update an old SAM i.e. built in a previous year, using available new information on the structure of the economy under study, thus paving the way to discussions about their strengths and weaknesses.

This technical note goes beyond the traditional comparison between CE and RAS methods and present applications of Huber and Hellinger's methods, which are absent from the literature. Furthermore, it uses several criteria to identify strengths and weaknesses of those methods: a) comparison between updated ratios and observed ones, b) the ability of methods to replicate structural changes, c) number of accounts presenting an extreme value, d) analysis of the structure of production and factor intensity between sectors.

Results show that CE and Hellinger's method are relatively superior to OLS and Huber's method in that they minimize mean errors concerning the estimation of ratios; they do not significantly modify the structure of accounts; and they preserve factor intensity within sectors.

These results must, however, be considered with caution. In this note, we use only one numerical case study while additional insight could be gained by analyzing several examples. Therefore, an avenue for the future research is to perform the same exercise using different structures (reaggregation) of the same SAM or SAMs from different countries.

[28] With exception of sector A5.
[29] Input-output table, Economic Account table, Balance of payment, households survey.

Annex

Table 12 Correspondence between the sectors in the matrices of 2006 and 2011

2011		2006	
awhea	Wheat	awhea	Wheat
amaiz	Maize	amaiz	Maize
arice	Paddy rice	arice	Paddy rice
asorg	Sorghum	asorg	Sorghum
apota	Irish potatoes	apota	Irish potatoes
aspot	Sweet potatoes	aspot	Sweet potatoes
acass	Cassava	acass	Cassava
aroot	Other roots	aroot	Other roots
apuls	Pulses	apuls	Pulses
avege	Other vegetables	avege	Vegetables
abana	Bananas	abana	Bananas/plantains
afrui	Other fruits	afrui	Fruits
aoils	Oil seed	aoils	Oil seeds
acoff	Coffee	acoff	Coffee
agtea	Green tea	agtea	Green tea
apyre	Pyrethrum	aotex	Other export crops
aotex	Other export crops		
Agriculture in SAM			
aboli	Bovine cattle, live	aboli	Bovine cattle, live
ashli	Sheep and goats, live	ashli	Sheep and goats, live
aswli	Swine, live	aswli	Swine, live
apoli	Poultry, live	apoli	Poultry, live
amilk	Raw milk	amilk	Raw milk
aegg	Eggs	aegg	Eggs
aoliv	Other livestock products	aoliv	Other livestock products
afore	Forestry	afore	Forestry
afish	Fishing	afish	Fishing
Livestock, fishing and forestry in SAM			
amini	Mining	amini	Mining
Mining			
amfdp	Meat, fish and dairy products	amfdp	Meat, fish and dairy products
acere	Processed cereals	acere	Processed cereals
acomg	Processed coffee	acomg	Processed coffee
atemg	Processed tea	atemg	Processed tea
abake	Bakery, processed sugar	abake	Bakery & processed sugar
atbev	Traditional beverages	atbev	Traditional beverages
antbe	Modern beverages	antbe	Modern beverages
atoba	Tobacco	atoba	Tobacco
Food processing n SAM			
atext	Textile and clothing	atext	Textile and clothing
awood	Wood, paper and printing	awood	Wood, paper and printing
achem	Chemicals	achem	Chemicals
amine	Non-metallic minerals	amine	Non-metallic minerals
afurn	Furniture and other manufactured products	afurn	Furniture and other manufactured products
aelec	Electricity, gas and water	aelec	Electricity, gas and water
acons	Construction	acons	Construction
Other industry			
areta	Wholesale and retail trade	areta	Wholesale and retail trade

ahote	Hotels and restaurants	ahote	Hotels and restaurants
atran	Transports	atran	Transports
acom m	Communication	acom m	Communication
afina	Finance and insurance	afina	Finance and insurance
arsta	Real estate	arsta	Real estate
aserv	Business services	aserv	Business services
arepa	Repair	arepa	Repair
aeduc	Education	aeduc	Education
aheal	Health	aheal	Health
aoser	Other personal services	aoser	Other personal services
Private service			
agov	Public administration	agov	Public administration
Public service			

Table 13. Correspondence between the factor and institution accounts of the 2006 matrices and 2011

2011		**2006**	
flab-ag	Labor - Agriculture	flab	Labor
flab-us	Labor - Unskilled		
flab-ls	Labor - Low Skilled		
flab-hs	Labor - High Skilled		
Labor			
flnd	Crop land	fcap	Capital
fliv	Livestock	flnd	Land
fcap-ag	Capital - Agriculture		
fcap-na	Capital - Non-Agriculture		
fcap-ss	Capital - Sector Specific		
Capital			
hhd-rur	Rural households	hhd	Households
hhd-urb	Urban households		
Households			
trc	Transaction costs	trc	Transaction costs
gov	Government	gov	Government account
dtax	Direct taxes	dtax	Direct Taxes
mtax	Import tariffs	mtax	Import duties & tariffs
stax	Sales taxes	stax	Sales taxes
s-i	Savings/investment	s-i	Savings-investment
row	Rest of world	row	Rest of the world
Institutions			

References

Ahmed, S. A., & Preckel, P. V. (2007, July). A comparison of RAS and entropy methods in updating IO tables. In *American Agricultural Economics Association, 2007 Annual Meeting* (pp. 1-20).

Csiszár, I.; Shields, P. (2004). *"Information Theory and Statistics: A Tutorial" (PDF)*. Foundations and Trends in Communications and Information Theory.

Decaluwé, B., Martens, A., & Savard, L. (2001). *La politique économique du développement et les modèles d'équilibre général calculable: une introduction à l'application de l'analyse mésoéconomique aux pays en développement*. PUM.

Fofana, I. (2007). Elaborer une matrice de comptabilité sociale pour l'analyse d'impacts des chocs et politiques macroéconomiques. *Centre Interuniversitaire sur le Risque, les Politiques Economiques et l'Emploi (CIRPEE)*.

Golan, A., & Judge, G. (1996). Recovering information in the case of underdetermined problems and incomplete economic data. *Journal of statistical planning and inference*, *49*(1), 127-136.

Golan, A., Judge, G. G., & Miller, D. (1996). *Maximum entropy econometrics: Robust estimation with limited data* (pp. 6-38). New York: Wiley.

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)*, 473-490.

International Food Policy Research Institute (IFPRI) A 2006 Social Accounting Matrix for Rwanda - Social Accounting Matrix (SAM) 2006. Washington, DC: International Food Policy Research Institute (IFPRI)

International Food Policy Research Institute (IFPRI). 2014. Rwanda Social Accounting Matrix (SAM), 2011. Washington, DC: International Food Policy Research Institute (IFPRI) [dataset].http://dx.doi.org/10.7910/DVN/28532

Lee, M. C., & Su, L. E. (2015). Social accounting matrix balanced based on mathematical optimization method and general algebraic modeling system. *Oxford Journal of Scientific Research*, 75.

Lemelin, A., Fofana, I., & Cockburn, J. (2013). Balancing a Social Accounting Matrix: Theory and application (revised edition*). Available at SSRN 2439868*.

Robinson, S., Cattaneo, A., & El-Said, M. (2001). Updating and estimating a social accounting matrix using cross entropy methods. *Economic Systems Research*, *13*(1), 47-64.