UNIVERSITAT ROVIRA I VIRGILI

WORKING PAPERS

Col·lecció "DOCUMENTS DE TREBALL DEL DEPARTAMENT D'ECONOMIA"

"Linear Aggregation In The Social Accounting Matrix Framework"

Maria Llop LLop
Antonio Manresa

Document de treball nº -12- 2010

DEPARTAMENT D'ECONOMIA
Facultat de Ciències Econòmiques i Empresarials

Edita:

Departament d'Economia
http://www.fcee.urv.es/departaments/economia/public_html/index.html
Universitat Rovira i Virgili
Facultat de Ciències Econòmiques i Empresarials
Avgda. de la Universitat, 1
432004 Reus
Tel. +34 977 759 811
Fax +34 977 300 661

Dirigir comentaris al Departament d'Economia.

Dipòsit Legal: T - 1841 - 2010

ISSN 1988 - 0812

DEPARTAMENT D'ECONOMIA
Facultat de Ciències Econòmiques i Empresarials

LINEAR AGGREGATION IN THE SOCIAL ACCOUNTING MATRIX FRAMEWORK*

Maria Llop
Departament d'Economia
Universitat Rovira i Virgili,
Facultat de Ciències Econòmiques i Empresarials
Avgda. Universitat nº 1,
43204 Reus, Spain.
Tel. 34 977 759898, Fax. 34 977 758907
E-Mail: maria.llop@urv.cat

Antonio Manresa
Departament de Teoria Econòmica and Centre de Recerca en Economia del Benestar
Universitat de Barcelona,
Avgda. Diagonal nº 690,
08034 Barcelona, Spain.
Tel. 34 93 4021867, Fax. 34 934021937
E-Mail: manresa@ub.edu

Abstract

In economic literature, information deficiencies and computational complexities have traditionally been solved through the aggregation of agents and institutions. In input-output modelling, researchers have been interested in the aggregation problem since the beginning of 1950s. Extending the conventional input-output aggregation approach to the social accounting matrix (SAM) models may help to identify the effects caused by the information problems and data deficiencies that usually appear in the SAM framework. This paper develops the theory of aggregation and applies it to the social accounting matrix model of multipliers. First, we define the concept of linear aggregation in a SAM database context. Second, we define the aggregated partitioned matrices of multipliers which are characteristic of the SAM approach. Third, we extend the analysis to other related concepts, such as aggregation bias and consistency in aggregation. Finally, we provide an illustrative example that shows the effects of aggregating a social accounting matrix model.

* The first author acknowledges the institutional support of the *Ministerio de Educación y Cultura* (grant SEJ2007-66318) and of the Generalitat de Catalunya (grant 2009SGR-322). The second author acknowledges the support of the *Ministerio de Educación y Cultura* (grant SEJ2006-07884).

1. INTRODUCTION

For different reasons, the aggregation of information is a common practice in economic analysis. As the data requirements and the data availabilities are not always reconcilable, some level of aggregation is necessary to solve the deficiencies that usually characterise the empirical sources. Aggregation is also used because the national accounts, which are prepared and compiled on a detailed basis, are published in an aggregated form by the national statistics institutes. Additionally, researchers use aggregated data when their interest lies in analysing the total effects of economic relationships. Finally, aggregation is a way of ensuring the confidentiality of economic data when there are restrictions on showing the individual information of agents and institutions.

In the input-output literature, the study of aggregation started in the early 1950s. In this field, a complete set of concepts has been developed, and nowadays it is possible to evaluate the properties of any input-output aggregation procedure. In fact, concepts such as aggregation bias, consistency in aggregation, and information requirements are perfectly defined and commonly used in input-output analysis. Because the theory of aggregation has also suggested specific criteria to reduce a given detailed input-output table, it minimises the errors due to the loss of information under aggregation.

The general question of input-output aggregation was first studied by Hatanaka (1952), Fei (1956), Theil (1957), Fisher (1958), and Ara (1959). Ijiri (1971) surveyed the fundamental queries that had been raised in the aggregation literature. Kimura (1985) developed the concept of consistency in aggregation within a dynamic input-output model. Howe and Johnson (1989) analysed the properties of an aggregated input-output model in relation with the original one. Kymn's paper (1990) reviewed all the theory of aggregation developed since then, and Olsen (1990) added some remarks and comments to the Kymn's paper. Oksanen and Williams (1992) used the factor-analytic approach to find similarities between industries to group them into aggregated sets. Olsen (1993) presented a joint theory of aggregation in input-output quantity-oriented models and price-oriented models. De Mesnard and Dietzenbacher (1995) considered the effects of aggregation over regions within an interregional input-output model, and focused on the interpretation of coefficients and multipliers under aggregation. Murray (1998) developed an optimisation-based approach for minimising the resulting error or the information loss in aggregated input-output models. Olsen (2000) presented an indicator of aggregation bias that allows the aggregation problems of a reduced input-output model to be identified. Dietzenbacher and Hoen (2000) analysed the effects of aggregation in an input-output table estimated in constant prices by means of double deflation. More recently, Lahr and Stevens (2002) discussed the implications of aggregation in regional input-output modelling, when the aggregated national input-output data are regionalised with trade-adjustments procedures.

In recent decades, Social Accounting Matrices (SAMs) have become a common instrument in economic analysis. SAM modelling can be defined as an extension of the input-output model because it reflects a greater set of income relations than those of the input-output approach. The SAM model completes the circular flow of income by capturing not only the intermediate demand relations, but also the relations between factor income distribution and private consumption. Since the pioneering contributions of Stone (1978) and Pyatt and Round (1979), social accounting techniques have been used to analyse the income generation process providing details about the sources and destinations of transactions between economic institutions.

As in the input-output model, the aggregation of information plays an important role within the SAM framework. One of the most important aspects in the social accounting context lies in choosing the aggregation level of agents and institutions considered in empirical analysis. Specifically, the division of consumers into socioeconomic groups and the number of foreign agents analysed is an important decision in applied research, since it determines the level of detail in the resulting description of the income generation process. Generally, the data deficiencies force us to consider aggregated regions and aggregated households, and this seriously impedes the identification of detailed income effects. Despite SAM aggregation being a common practice in applied research, as far as we know, the aggregation problem has not been analysed within the context of the social accounting matrix framework.

The objective of this paper is to adapt the conventional aggregation theory, typically applied to input-output modelisation, in such a way that it can be applied in the SAM model. Specifically, we extend the general context of aggregation to the social accounting database context, and we define the aggregation of the partitioned matrices of multipliers that are characteristic of the SAM framework. We also extend to the social accounting model other related concepts, such as aggregation bias of income and aggregation bias of multipliers. With this extension, therefore, we further analyse the aggregation problem and provide a conceptual framework that is useful for establishing the consequences of the information problems and the data deficiencies that commonly appear in the SAM database context.

The rest of the paper is organised as follows. Section 2 describes the social accounting matrix model of linear multipliers. In section 3 we apply the aggregation theory to the SAM framework, through the definition of aggregation functions, aggregation bias of income, aggregation bias of multipliers, and aggregation consistency. Section 4 shows an illustrative example. At the end of the paper we provide some concluding remarks.

2. THE SAM MODEL

The SAM model is based on the accounting identities reflected in a social accounting matrix. A SAM is a square matrix whose rows and columns add up to the same amount. This matrix contains the flows of income and expenditure related to all the economic agents by a temporal reference.[1] By convention, receipts of agents are entered in the rows, and expenditures are entered in the columns. Table 1 shows schematically the transactions that appear in a social accounting matrix.

[PLACE TABLE 1 HERE]

In the first row, T_{11} is a square matrix that contains the intermediate inputs. Matrix T_{13} shows the private consumption and has the same number of columns as the number of consumers in the SAM. Additionally, matrix T_{21} contains the factor income or value added, matrix T_{32} shows the factor income of consumers and matrix T_{33} shows the transferences between consumers. Finally, the last row and the last column in table 1 show the income relations with the remaining sectors, which include the government, the capital account and the foreign agents.

---

[1] See, for example, Pyatt (1988) for a detailed description of social accounting matrices.

To transform the representation of table 1 into a model, we assume that the structure of income and payments is constant. On the other hand, we must also divide the accounts of the SAM into two different categories: endogenous accounts and exogenous accounts. The standard representation of the SAM model can then be written as follows:

$$Y = A\,Y + X = [I - A]^{-1}\,X = M\,X, \tag{1}$$

where Y is the vector of column totals of the endogenous accounts, X is the vector of exogenous injections, I is the identity matrix and A is a square matrix of structural coefficients, calculated by dividing the transactions in the SAM by the corresponding column sum. In expression (1), $M = [I - A]^{-1}$ is the matrix of multipliers, and the element m_{ij} quantifies the increase in the income or receipts of account i caused by a unitary and exogenous injection received by account j. These elements thus show both the direct and the indirect effects on the endogenous accounts of the exogenous inflows received.

In the traditional endogeneity assumption of Stone (1978) and Pyatt and Round (1979), activities, factors of production and households are considered to be endogenous components. So, matrix A of structural coefficients has the following structure:

$$A = \begin{bmatrix} A_{11} & 0 & A_{13} \\ A_{21} & 0 & 0 \\ 0 & A_{32} & A_{33} \end{bmatrix},$$

where A_{11} contains the input-output coefficients, A_{13} contains the coefficients of households' sectorial consumption, A_{21} contains the factors of production coefficients, A_{32} contains the coefficients of factor income to consumers, and A_{33} contains the transactions between consumers. Note that expression (1) is an extension of the classical input-output model, since it includes a greater set of interdependencies than those of the input-output approach. The SAM model completes the circular flow by capturing not only the intermediate demand relations, but also the relations between factor income distribution and private consumption.

To provide a deeper insight into the analysis of SAM multipliers, Pyatt and Round (1979) divided matrix M into different circuits of interdependence. Specifically, it can be seen that:

$$
\begin{aligned}
Y &= A\,Y + X \\
&= (A - \bar{A})\,Y + \bar{A}\,Y + X \\
&= (I - \bar{A})^{-1}\,[(A - \bar{A})\,Y + X] \\
&= \dot{A}\,Y + (I - \bar{A})^{-1}\,X \\
&= \dot{A}^2\;Y + (I + \dot{A})\,(I - \bar{A})^{-1}\,X \\
&= \dot{A}^3\,Y + (I + \dot{A} + \dot{A}^2)\,(I - \bar{A})^{-1}\,X \\
&= (I - \dot{A}^3)^{-1}\,(I + \dot{A} + \dot{A}^2)\,(I - \bar{A})^{-1}\,X \\
&= M_3\,M_2\,M_1X, \qquad (2)
\end{aligned}
$$

where $\dot{A} = (I - \bar{A})^{-1}(A - \bar{A})$, $M_1 = (I - \bar{A})^{-1}$, $M_2 = (I + \dot{A} + \dot{A}^2)$, and $M_3 = (I - \dot{A}^3)^{-1}$. Finally, matrix $\bar{A}$ has the following structure:

$$
\bar{A} = \begin{bmatrix} A_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & A_{33} \end{bmatrix}.
$$

In the expression above, matrix M of the total SAM multipliers has been defined by three multiplicative components that convey different economic meanings.[2] After the corresponding matrix algebra is applied, it can be seen that the first block M_1 has the following elements:

$$M_1 = \begin{bmatrix} (I - A_{11})^{-1} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & (I - A_{33})^{-1} \end{bmatrix}.$$

Matrix M_1 contains the *own effects* explained by the connections between the accounts belonging to the same income relationships. Specifically, the perspective of income transmission reflected in M_1 responds to the effects of intersectorial linkages and the effects of transactions between consumers.

Additionally, matrix M_2 is as follows:

$$M_2 = \begin{bmatrix} I & (I - A_{11})^{-1} A_{13} (I - A_{33})^{-1} A_{32} & (I - A_{11})^{-1} A_{13} \\ A_{21} & I & A_{21} (I - A_{11})^{-1} A_{13} \\ (I - A_{33})^{-1} A_{32} A_{21} & (I - A_{33})^{-1} A_{32} & I \end{bmatrix}.$$

This block contains the *open effects* caused by the accounts on the other parts of the circular flow of income, that is, from activities to households and factors, from factors to households and activities and, finally, from households to factors and activities. As it shows the effects of the accounts on the other income circuits of the system, the main diagonal in M_2 is unitary and the other elements are positive.

Finally, matrix M_3 has the following structure:

---

[2] Note that the decomposition in equation (2) is not unique. In consequence, the interpretation of the decomposed multipliers depends basically on the division of the matrix of expenditure share coefficients, that is, the structure of matrix $\overline{A}$.

$$M_3 = \begin{bmatrix} [I-(I-A_{11})^{-1}A_{13}(I-A_{33})^{-1}A_{32}A_{21}]^{-1} & 0 & 0 \\ 0 & [I-A_{21}(I-A_{11})^{-1}A_{13}(I-A_{33})^{-1}A_{32}]^{-1} & 0 \\ 0 & 0 & [I-(I-A_{33})^{-1}A_{32}A_{21}(I-A_{11})^{-1}A_{13}]^{-1} \end{bmatrix}.$$

The block M_3 contains the *circular effects* on the accounts that are activated because of the exogenous inflows received. The component M_3 is a block diagonal matrix, showing the closed-loop effects of the circular flow of income caused by the exogenous shocks received by the accounts. That is, this matrix shows the effects of any inflow starting from any part of the income circuit and coming back to its starting point, e.g. from activities to factors to households, and then back to activities in the form of consumption demand.

The decomposition of SAM multipliers identifies the different channels by which the income effects can be produced and transmitted throughout the economy. Logically, this kind of information is very useful for establishing the origin of the income shocks on the economic agents and institutions, providing deeper insights of the circular flow of income.

3. AGGREGATION IN THE SAM FRAMEWORK

3.1. Basic Concepts of Aggregation

This section describes the general notation and definitions used in the procedure of linear input-output aggregation. Similarly, as we will show in the next section, the concepts and definitions presented here can be applied to the social accounting matrix framework.

Let us assume that the *n* original accounts of the SAM model are aggregated into *N* accounts ($N < n$). We use the indexes I, $J = 1, \ldots, N$, to indicate the aggregated accounts, and the indexes $i, j = 1, \ldots, n$, to indicate the original disaggregated accounts. The aggregation of matrix A of structural coefficients of the SAM model is carried out by using an $N \times n$ aggregator matrix or *grouping matrix* G, which has the following structure:

$$G = \begin{bmatrix} 1\ldots1 & 0\ldots0 & \ldots & 0\ldots0 \\ 0\ldots0 & 1\ldots1 & \ldots & 0\ldots0 \\ \vdots & \vdots & \ddots & \vdots \\ 0\ldots0 & 0\ldots0 & \ldots & 1\ldots1 \end{bmatrix},$$

where a generic element in this matrix, g_{Ij} ($I = 1, \ldots, N$, and $j = 1, \ldots, n$), is defined as follows:

$$g_{Ij} = \begin{cases} 1, j \in I, \\ 0, j \notin I. \end{cases}$$

That is, each row of matrix G corresponds to an aggregated account, and each column corresponds to an original disaggregated account. We place a unitary element if the column of G corresponds to an account belonging to the aggregated set, and we place zero otherwise.[3]

Additionally, let H be an $N \times n$ matrix of *aggregation weights* with the same structure as G, that is:

$$H = \begin{bmatrix} \otimes\ldots\otimes & 0\ldots0 & \ldots & 0\ldots0 \\ 0\ldots0 & \otimes\ldots\otimes & \ldots & 0\ldots0 \\ \vdots & \vdots & \ddots & \vdots \\ 0\ldots0 & 0\ldots0 & \ldots & \otimes\ldots\otimes \end{bmatrix},$$

[3] Note that the definition of matrix G assumes that every disaggregated account belongs exactly to one aggregated account, and no partial aggregation of the original accounts can be made. In consequence, the columns in matrix G add up to one.

where the symbol $\otimes$ indicates non-zero values. The typical element in matrix *H*, h_{Ij} (I = 1, …., *N*, and j = 1, …, *n*), is defined as follows:

$$h_{Ij} = \begin{cases} \dfrac{Y_j}{\sum_{i \in I} Y_i}, j \in I, \\ 0, j \notin I. \end{cases}$$

Note that h_{Ij} are weights of the value of income in the original disaggregated account *j* (Y_j) with respect to the value of income in the aggregated account *I* ($\sum_{i \in I} Y_i$). Consequently, for each original account *j* the elements in matrix *H* add up to one, that is the row sum in this matrix is equal to one: $\sum_{j \in I} h_{Ij} = 1$.

It should be pointed out that, irrespective of the dimension of matrices *G* and *H*, the definitions above yield the following mathematical property: *HG*' = *I*; and the associated transpose to this property is: *GH*' = *I*.[4]

Finally, note that the grouping matrix *G* and the aggregation weights matrix *H* can be used for any level of aggregation applied to the original accounts of the SAM model, once these accounts have been renumbered and compacted into the aggregated ones. Thus, the definitions above are general enough to be applied in whatever aggregation required in empirical analysis, with the unique restriction that every disaggregated account has to belong completely to one aggregated set and no partial aggregations can be applied.

[4] The superscript ' denotes transposition of the corresponding matrix.

3.2. Definition of aggregation

The matrices G and H defined above allow the original SAM model to be transformed into a more compacted model, in which the number of accounts is lower than in the original one. Analogous to expression (1), the aggregated social accounting matrix model can be written as:[5]

$$Y^* = A^* Y^* + X^* = M^* X^*, \tag{3}$$

where Y^* represents the vector of income in the aggregated endogenous accounts, A^* represents the matrix of expenditure share coefficients of the aggregated accounts, and X^* represents the vector of aggregated exogenous injections.

By using the notations defined above, the equivalence between the compacted components of the SAM model and the original ones responds to the following calculations:

$$A^* = GAH^{'},$$

$$X^* = GX,$$

$$M^* = G(I - A)^{-1}H^{'},$$

which reflect the relation between the information of the original model, that is, matrices A and M and vector X, and the corresponding information of the aggregated model.

Taking into account the definition of matrix A^*, the reduced matrix of multipliers can be obtained as follows:

---

[5] In what follows, the superscript * accompanying a matrix or a vector denotes the aggregation of the corresponding matrix or vector.

$$M^{*} = [I - A^{*}]^{-1}$$
$$= [I - GAH']^{-1}$$
$$= [G(I - A)H']^{-1}$$
$$= [GM^{-1}H']^{-1}. \quad (4)$$

According to expression (2), we can also decompose the aggregated matrix of multipliers into different circuits of interdependence. Specifically, the first aggregated block of multipliers, M_1^*, is equal to:

$$M_1^{*} = [I - \overline{A}^{*}]^{-1}$$
$$= [I - G\overline{A}H']^{-1}$$
$$= [G(I - \overline{A})H']^{-1}$$
$$= [GM_1^{-1}H']^{-1}, \quad (5)$$

where $\overline{A}^{*} = G\overline{A}H'$ is the aggregated matrix of $\overline{A}$ (expression (2)). The block M_1^* shows the *aggregated own effects* of the multiplier decomposition, taking place between the accounts belonging to the same income group. Specifically, M_1^* contains the aggregated effects of intersectorial consumption and the aggregated effects of transfers between consumers.

The *aggregated open effects*, M_2^*, can be obtained as:

$$M_2^{*} = (\mathrm{I} + \dot{A}^{*} + \dot{A}^{*2})$$

$$= I + [(I - \overline{A}^{*})^{-1}(A^{*} - \overline{A}^{*})] + [(I - \overline{A}^{*})^{-1}(A^{*} - \overline{A}^{*})]^{2}$$

$$= I + [M_1^* G(A - \overline{A})H'] + [M_1^* G(A - \overline{A})H']^2, \quad (6)$$

where $\dot{A}^* = (I - \overline{A}^*)^{-1}(A^* - \overline{A}^*)$. This block shows the effects caused by the aggregated accounts on the other parts of the circular flow of income.

Finally, the *aggregated circular effects* of the multiplier decomposition, M_3^*, are equal to:

$$M_3^* = (I - \dot{A}^{*3})^{-1}$$

$$= (I - [(I - \overline{A}^*)^{-1}(A^* - \overline{A}^*)]^3)^{-1}$$

$$= (I - [M_1^*(G(A - \overline{A})H')]^3)^{-1}. \quad (7)$$

This expression contains the closed-loop effects on the aggregated accounts of the income shocks received.

Expressions (5) to (7) show the connection between the disaggregated matrices of coefficients and the aggregated matrices of decomposed multipliers. In other words, applying the equations above to the detailed original coefficients allows the multipliers of the aggregated SAM model and its decomposition into different income circuits to be directly obtained.

3.3. Aggregation bias of endogenous income

The most important question in any aggregation procedure is the resulting error when the aggregated model is used to make economic predictions. The theory of aggregation assumes implicitly that the disaggregated model is the correct one while the aggregated model suffers from errors. This standard assumption explains why the literature of input-output aggregation has turned attention towards the definition of optimal aggregation, consisting of those procedures that allow the error due to sectorial aggregation to be reduced. In this field, the contributions have provided guidance to researchers to ensure minimal discrepancies between the results of the aggregated input-output models and the results that would have been reported by the corresponding disaggregated models.

Typically, the starting point in the analysis of input-output aggregation bias is the comparison of the model prediction under aggregation and the theoretical equivalence of the aggregated system. According to Theil (1957), the input-output aggregation bias concerns the difference between the predicted values of total output, primary demand and intermediate consumption, with respect to the values for all these variables that would have been obtained through the mathematical definition of the compacted model.

In this section, we apply Theil's concept of income aggregation bias to the social accounting matrix framework. By taking into account the definition of the SAM model in equation (1) above, the prediction of the endogenous income after the aggregation is calculated as:

$$\overline{GY} = (I - A^{*})^{-1}\, GX, \tag{8}$$

where $\overline{GY}$ stands for the column vector of the endogenous income predicted after aggregation. On the other hand, the theoretical aggregation of the endogenous income is defined as:

$$GY = G(I - A)^{-1}X. \tag{9}$$

In the SAM database framework, the income aggregation bias or income aggregation error (ε^Y) can be obtained as the difference between the predicted endogenous income and the values obtained through the theoretical aggregation, that is, the difference between equations (8) and (9):

$$\varepsilon^Y = \overline{GY} - GY = (I - A^*)^{-1}GX - G(I - A)^{-1}X,$$

which is an $N \times 1$ vector that contains the discrepancy in the endogenous income of the aggregated model regarding the theoretical endogenous income. By applying the power series expansion, it follows that:

$$\begin{aligned} \varepsilon^Y &= (I - A^*)^{-1}GX - G(I - A)^{-1}X \\ &= [(I + A^* + A^{*2} + \ldots.)G - G(I + A + A^2 + \ldots.)]X \\ &= [(A^*G - GA) + (A^{*2}G - GA^2) + \ldots.]X. \end{aligned} \quad (10)$$

Following Theil (1957), the $N \times 1$ vector of first-order aggregation bias (ε_f^Y) is defined taking into account the first-order terms of equation (10):

$$\varepsilon_f^Y = (A^*G - GA)X. \quad (11)$$

As ε_f^Y is an approximation of the total aggregation error, it can lead to imprecise conclusions about the aggregation procedure. Olsen (2001) demonstrated that the first-order aggregation bias may or may not be null regardless of whether the total aggregation bias is null or not. In consequence, the value of first-order bias is not related to the value of total aggregation bias. For this reason, the use of ε_f^Y is not recommended as it does not detect perfect aggregations and, consequently, does not give information about the goodness of the aggregations.[6]

Expressions (10) and (11) are measures of the aggregation error in absolute terms, showing the difference in the endogenous income of the accounts. To obtain an indicator of the first-order aggregation bias in relative terms (R_f), we can apply the following calculation:

$$R_f = [\widehat{GY}]^{-1}\varepsilon_f^Y = [\widehat{Y^*}]^{-1}(A^*G - GA)X. \qquad (12)$$

In this expression, the symbol $\wedge$ represents a diagonal matrix containing the elements of vector GY, that is, the values of the endogenous income in the aggregated accounts (expression (9)). Note that R_f is an $N \times 1$ vector of the first-order bias regarding the total endogenous income in each compacted account.

Similarly, the $N \times 1$ vector of total bias in relative terms (R) can be obtained as:

$$R = [\widehat{GY}]^{-1}\varepsilon^Y = [\widehat{GY}]^{-1}[(A^*G - GA) + (A^{*2}G - GA^2) + \ldots]X. \qquad (13)$$

The vectors R_f and R allow an immediate interpretation of the aggregation bias of income, since they measure the discrepancies of the aggregated model as percentages of the total endogenous income.

Finally, we can calculate a total measure of the first-order bias of income (T_f) as follows:

$$T_f = [e'GY]^{-1}[e'\varepsilon_f^Y]$$

---

[6] For instance, section 4 shows an example of perfect aggregation (total aggregation bias null), in which the first-order aggregation bias is different from zero.

$$= [e'G(I-A)^{-1}X]^{-1}[e'(A^{*}G-GA)X], \qquad (14)$$

where e' is a unitary row vector, and T_f is a scalar showing the total first-order bias of the endogenous income.

According to expression (14), the total bias of endogenous income (T) can be obtained as:

$$T = [e'GY]^{-1}[e'\varepsilon^{Y}]$$

$$= [e'G(I-A)^{-1}X]^{-1}[e'[(A^{*}G-GA)+(A^{*2}G-GA^{2})+\ldots.]X], \qquad (15)$$

being a scalar of the total error in the endogenous income due to aggregation.

Note that expressions (14) and (15) allow a unique value to be obtained, which is measured as a percentage of the endogenous income of the aggregated sets that synthesize the error in the endogenous income due to aggregation. These measures may be very useful to illustrate the goodness of aggregation procedures, because they are easily interpretable.

3.4. Aggregation bias of multipliers

The literature of input-output aggregation has focused on analysing the resulting error in income when sectorial information concerns large accounts. However, another important question in any aggregation procedure is the resulting error in the multipliers, that is, the resulting error in the income effects of the accounts per unit of exogenous shocks. In other words, the bias of income captures the total error in the endogenous income of each account, and the bias of multipliers will capture the error in the income effects of the accounts when these accounts receive unitary and exogenous shocks of income. In this section, we propose a method to quantify the bias of the SAM multipliers due to aggregation. This extension allows the effects of using aggregated data and aggregated accounts in the SAM framework to be individually shown.

The starting point in the analysis of the multipliers aggregation bias is the comparison of the multipliers prediction under aggregation and the theoretical equivalence of the aggregated system. Taking this idea into account, the prediction of the SAM multipliers after aggregation is equal to $(I - A^{*})^{-1} = [GM^{-1}H']^{-1}$. On the other hand, the theoretical aggregation is equal to $G(I - A)^{-1}H'$. Thus, the multipliers aggregation bias or multipliers aggregation error (ε^{M}) can be obtained as the difference between the predicted values and the values obtained through the mathematical aggregation, that is:

$$\varepsilon^{M} = [GM^{-1}H']^{-1} - G(I - A)^{-1}H', \qquad (16)$$

which is an $N \times N$ matrix that contains the discrepancy between the multipliers of the aggregated SAM model regarding the theoretical multipliers that are obtained by aggregating the original detailed model.

Similarly, we can obtain a measurement of the aggregation error of the partitioned matrices of multipliers. Specifically, the bias in the first block of multipliers, ε^{M_1}, is calculated as:

$$\varepsilon^{M_1} = [I - \overline{A}^{*}]^{-1} - [G(I - \overline{A})H']^{-1}, \qquad (17)$$

being an $N \times N$ matrix, showing the difference between the predicted own effects and the corresponding theoretical equivalence of the aggregated own effects. Similarly, the $N \times N$ matrix of the bias in the open effects, ε^{M_2}, is equal to:

$$\varepsilon^{M_2} = (I + \dot{A}^{*} + \dot{A}^{*2}) -$$

$$[I + [(GM_1^{-1}H')^{-1}G(A - \overline{A})H'] + [(GM_1^{-1}H')^{-1}G(A - \overline{A})H']^2]. \quad (18)$$

Finally, the bias in the aggregated circular effects, ε^{M_3}, is obtained as:

$$\varepsilon^{M_3} = (I - \dot{A}^{*3})^{-1} - [(I - ([GM_1^{-1}H']^{-1}G(A - \overline{A})H')^3)^{-1}] \quad (19)$$

which is an $N \times N$ matrix containing the differences in the third block of the SAM multiplier decomposition.

Expressions (16) to (19) allow measurements of the multipliers bias due to aggregation of accounts in the SAM framework to be obtained. These calculations complete the information about the differences in the income effects of the individual accounts due to the exogenous shocks received, and this provides helpful knowledge of the goodness of the aggregation procedures applied to a social accounting model of multipliers.

3.5. Consistency in aggregation

The concept of consistency in input-output aggregation, which has been widely analysed in the literature, suggests reasonable criteria to ensure a "good" aggregation of a given input-output table. The idea of consistent aggregation refers to the degree of relation between the disaggregated model and the aggregated one. Ideally, the aggregated model should give the same results as those obtained through the aggregation of the original disaggregated model. This property of aggregation is known in the input-output literature by different terms, such as "perfect aggregation", "consistent aggregation", "exact aggregation", "acceptable aggregation", and "intrinsic aggregation".

Taking into account the definition of aggregation bias described above, a perfect aggregation of a SAM model should imply a null error in equation (10). This can be accomplished in different ways:

- First, as it is stated in the theorem of Hatanaka (1952), the total aggregation bias will be zero for any vector X of exogenous income, if and only if $A^{*}G = GA$. Note that $(A^{*2}G - GA^{2}) = A^{*}A^{*}G - GAA = A^{*}(GA) - (A^{*}G)A = 0$, and, similarly, the higher order terms of equation (10) will be zero as well. This property suggests that if two or more accounts have the same column coefficients in matrix A, the aggregation of these accounts will result in a null total aggregation bias.

- Second, de Mesnard and Dietzenbacher (1995) demonstrated that the condition $A^{*}G = GA$ is equivalent to $M^{*}G = GM$, that is $(I - A^{*})^{-1}G = G(I - A)^{-1}$. Then, the aggregation biais in equation (10) will be zero. This result suggest that if the direct influences (captured in matrices A and A^{*}) show a null aggregation biais, then the global influences (captured in matrices M and M^{*}) will show a null aggregation biais as well.

- Third, the aggregation bias also depends on the characteristics of vector X of exogenous income. If some accounts are not aggregated, and vector X only has positive entries for these unaggregated accounts, being zero the exogenous income for the aggregated ones, the first-order aggregation bias will be zero.[7] However, it has to be taken into account that a null first-order aggregation bias does not ensure a null total aggregation bias.[8]

---

[7] See Miller and Blair (1985).

[8] Olsen (2001).

To sum up, the concept of consistency in aggregation refers to the characteristics that must be present in the economy being analysed to ensure a null aggregation error. At empirical level, however, consistency may not be accomplished as it depends on the specific data and the specific features of the economy. Keeping this limitation in mind, the input-output literature has provided methods to ensure a minimum error in the aggregation procedures applied to a given input-output model.[9] This knowledge is very useful from a practical point of view, mainly when the data restrictions prevent a perfect aggregation.

4. AN ILLUSTRATIVE EXAMPLE

This section shows an example of a social accounting matrix model of multipliers, consisting in reducing the number of sectors of production, the number of factors, and the number of consumers.[10] Specifically, we compare two different representations for the same hypothetical economy: a model with two activities of production, two factors and two households as endogenous components, regarding a model with one activity, one factor, and one household as endogenous components. Given that the data requirements for a detailed SAM model are often difficult to obtain and that the statistical agencies does not publish annually the complete SAM databases, our example could be very useful for understanding the implications of the usual practice of using reduced information in empirical SAM modelisation.

[PLACE TABLE 2 HERE]

[9] See, for instance, Fisher (1966).

[10] Our example follows the conventional endogeneity assumption of Stone (1978) and Pyatt and Round (1979).

Table 2 shows a disaggregated SAM database, in which we distinguish seven accounts: two activities of production, two factors, two consumers and an exogenous account (that contains the income relations of government, capital account, and foreign agents). Table 3 shows the income flows for the same hypothetical economy, in which sectors, factors and households have been aggregated each one into a compacted account.

[PLACE TABLE 3 HERE]

By using the information from the two SAMs, the matrices of structural coefficients are given as:

$$A = \begin{bmatrix} 0.25 & 0.45 & 0 & 0 & 0.3 & 0.066 \\ 0.15 & 0.1 & 0 & 0 & 0.7 & 0.066 \\ 0.2 & 0.3 & 0 & 0 & 0 & 0 \\ 0.15 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.417 & 0.5 & 0 & 0 \\ 0 & 0 & 0.583 & 0.5 & 0 & 0 \end{bmatrix}, \text{ and } A^* = \begin{bmatrix} 0.475 & 0 & 0.480 \\ 0.375 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

The aggregation of sectors, factors and consumers is carried out by using the following aggregation matrices:

$$G = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \text{ and } H' = \begin{bmatrix} 0.5 & 0 & 0 \\ 0.5 & 0 & 0 \\ 0 & 0.6 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0.4 \\ 0 & 0 & 0.6 \end{bmatrix},$$

which allow the aggregated matrix of coefficients to be obtained through the following calculation:

$$A^{*} = GAH' = \begin{bmatrix} 0.475 & 0 & 0.480 \\ 0.375 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

In the disaggregated SAM model, the matrix of total multipliers is equal to:

$$M = \begin{bmatrix} 1.984 & 1.347 & 0.770 & 0.880 & 1.538 & 0.222 \\ 0.712 & 1.795 & 0.710 & 0.818 & 1.470 & 0.167 \\ 0.610 & 0.808 & 1.367 & 0.422 & 0.749 & 0.095 \\ 0.369 & 0.381 & 0.187 & 1.214 & 0.378 & 0.050 \\ 0.439 & 0.527 & 0.663 & 0.783 & 1.501 & 0.064 \\ 0.540 & 0.662 & 0.891 & 0.853 & 0.625 & 1.080 \end{bmatrix},$$

and the decomposition into different income channels yields:

$$M_1{}^{11} = \begin{bmatrix} 1.481 & 0.741 & 0 & 0 & 0 & 0 \\ 0.247 & 1.235 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$M_2 = \begin{bmatrix} 1 & 0 & 0.488 & 0.556 & 0.963 & 0.148 \\ 0 & 1 & 0.499 & 0.519 & 0.938 & 0.099 \\ 0.200 & 0.300 & 1 & 0 & 0.474 & 0.059 \\ 0.150 & 0.100 & 0 & 1 & 0.238 & 0.032 \\ 0.158 & 0.175 & 0.417 & 0.500 & 1 & 0 \\ 0.192 & 0.225 & 0.583 & 0.500 & 0 & 1 \end{bmatrix},$$

[11] Note that as Table 3 does not reflect transfers between households, there are no own multiplier' effects in the consumers' accounts.

$$M_3 = \begin{bmatrix} 1.286 & 0.319 & 0 & 0 & 0 & 0 \\ 0.265 & 1.295 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.367 & 0.422 & 0 & 0 \\ 0 & 0 & 0.187 & 1.214 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.501 & 0.064 \\ 0 & 0 & 0 & 0 & 0.625 & 1.080 \end{bmatrix}.$$

In the aggregated model, the matrices of multipliers are equal to:

$$M^* = \begin{bmatrix} 2.899 & 1.391 & 1.391 \\ 1.087 & 1.522 & 0.522 \\ 1.087 & 1.522 & 1.522 \end{bmatrix},$$

$$M_1^* = \begin{bmatrix} 1.905 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, M_2^* = \begin{bmatrix} 1 & 0.914 & 0.914 \\ 0.375 & 1 & 0.343 \\ 0.375 & 1 & 1 \end{bmatrix}, M_3^* = \begin{bmatrix} 1.522 & 0 & 0 \\ 0 & 1.522 & 0 \\ 0 & 0 & 1.522 \end{bmatrix}.$$

The total income aggregation bias is null, given that we have complete information of the income relations of the accounts, that is, we have the entire detailed database. On the other hand, the first-order bias shows values different from zero in two accounts:

$$\varepsilon^Y = \overline{GY} - GY = \begin{bmatrix} 200 \\ 100 \\ 125 \end{bmatrix} - \begin{bmatrix} 200 \\ 100 \\ 125 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

$$\varepsilon_f^Y = (A^*G - GA)X = \begin{bmatrix} 2.458 \\ -0.625 \\ 0 \end{bmatrix},$$

and the relative first-order bias being equal to:

$$R_f = \begin{bmatrix} 0.006 \\ -0.001 \\ 0 \end{bmatrix}.$$

The interpretation of R_f is as follows. The first-order error in sectors of production is about 0.6 per cent of the endogenous income, in factors of production is about -0.1 per cent, and in consumers is about 0 per cent.

In order to synthesize the bias of income into a unique value, we can calculate the total measurements of expressions (14) and (15). Specifically, the total first-order bias of aggregation is $T_f = 0.004$, meaning that the first-order error is about 0.4 per cent of the total endogenous income in the aggregated SAM model, and the total bias T is equal to zero.

Additionally, we can calculate the aggregation bias of multipliers due to the aggregation of accounts as follows:

$$\varepsilon^M = (I - A^*)^{-1} - [G(I - A)H']^{-1} =$$

$$= \begin{bmatrix} 2.899 & 1.391 & 1.391 \\ 1.087 & 1.522 & 0.522 \\ 1.087 & 1.522 & 1.522 \end{bmatrix} - \begin{bmatrix} 2.899 & 1.391 & 1.391 \\ 1.087 & 1.522 & 0.522 \\ 1.087 & 1.522 & 1.522 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Given that we have complete information of the detailed database, the multipliers error is equal to zero. Similarly, it can be checked that the bias of the partitioned matrices of multipliers are equal to null matrices, meaning that the error in all the block multipliers is zero.

Let us now assume that the data of vector X does not coincide with the values that appear in the original information. This situation is very common in applied research, given that the complete SAM databases are not annually published and only partial information is known for the periods not covered by the official statistics. Then, the values of X could correspond to other periods different from the base year. For instance, if the values of X are now 10% greater than those corresponding to the base year, the initial vector X and the new vector (X') are given by:

$$X = \begin{bmatrix} 10 \\ 35 \\ 10 \\ 15 \\ 5 \\ 20 \end{bmatrix}, \text{ and } X' = \begin{bmatrix} 11 \\ 38.5 \\ 11 \\ 16.5 \\ 5.5 \\ 22 \end{bmatrix}.$$

By using the new values of exogenous income, it can be checked that the total aggregation biais is now different from zero:

$$\varepsilon^Y = \overline{GY} - GY = \begin{bmatrix} 220 \\ 110 \\ 137.5 \end{bmatrix} - \begin{bmatrix} 200 \\ 100 \\ 125 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \\ 12.5 \end{bmatrix},$$

and the relative total bias is equal to:

$$R = \begin{bmatrix} 0.047 \\ 0.024 \\ 0.029 \end{bmatrix}.$$

These values show that the total error in sectors of production is about 4.7 per cent of the endogenous income, in factors of production is about 2.4%, and in consumers is about 2.9 per cent of endogenous income. The total measure of the income biais is equal to $T = 0.100$, meaning that the error is about 10 per cent of the total endogenous income.

5. CONCLUSIONS

In this paper we extended the conceptual tool of linear input-output aggregation to the social accounting matrix framework. After defining the aggregation in a SAM database context, we showed the aggregation of the partitioned matrices of multipliers that are characteristic of the social accounting matrix model. We also extended the analysis to other related concepts, such as aggregation bias of income and aggregation bias of multipliers. Finally, we presented a numerical example that illustrates the usefulness of the conceptual set developed in the paper. Our example, which is very frequent in applied research, concerns the aggregation of sectors of production, factors of production, and households in a SAM model of multipliers.

The study of aggregation is very important for empirical work. The aggregation of agents and institutions into large accounts is a usual practice in the social accounting matrix framework, as it allows the information deficiencies that commonly characterise the empirical sources to be overcome. This paper extends the theory of linear aggregation which is extremely valuable for establishing the consequences of the information problems and the data deficiencies that frequently characterise the SAM database context. This extension can help to clarify the consequences of not using detail in the description of the income generation process of the economy.

REFERENCES

Ara K, 1959, "The Aggregation Problem in Input-Output Analysis" *Econometrica* **27** 257-262.

De Mesnard L, Dietzenbacher E, 1995, "On the Interpretation of Fixed Input Coefficients under Aggregation" *Journal of Regional Science* **35** 233-243.

Dietzenbacher E, Hoen A R, 2000, "Double Deflation and Aggregation" *Environment and Planning A* **31** 1695-1704.

Fei J C H, 1956, "A Fundamental Theorem for the Aggregation Problem in Input-Output Analysis" *Econometrica* **24** 400-412.

Fisher W D, 1958, "Criteria for Aggregation in Input-Output Analysis" *Review of Economics and Statistics* **40** 250-260.

Fisher W D, 1966, "Simplification of Economic Models" *Econometrica* **34** 563-584.

Hatanaka M, 1952, "Note on Consolidation within a Leontief System" *Econometrica* **20** 301-303.

Howe E C, Johnson C R, 1989, "Linear Aggregation in Input-Output Models" *SIAM Journal on Matrix Analysis and Applications* **10** 65-79.

Ijiri Y, 1971, "Fundamental Queries in Aggregation Theory" *Journal of the American Statistical Association* **66** 766-782.

Kimura Y, 1985, "On the Aggregation Problem in Input-Output Models" *Papers of the Regional Science Association* **56** 167-176.

Kymn K O, 1990, "Aggregation in Input-Output Models: a Comprehensive Review, 1946-1971" *Economic Systems Research* **2** 65-93.

Lahr K O, Stevens B H, 2002, "A Study of the Role of Regionalisation in the Generation of Aggregation Error in Regional Input-Output Models" *Journal of Regional Science* **42** 477-507.

Miller R E, Blair P D, 2009, *Input-Output Analysis: Foundations and Extensions*, (Cambridge University Press).

Murray A T, 1998, “Minimizing Aggregation Error in Input-Output Models”, *Environment and Planning A* **30** 1125-1128.

Oksanen E H, Williams J R, 1992, “An Alternative Factor-Analytic Approach to Aggregation of Input-Output Tables” *Economic Systems Research* **4** 245-256.

Olsen J A, 1985, “Adaptation of Detailed Input-Output Information: Restructuring and Aggregation” *Review of Income and Wealth* **31** 397-411.

Olsen J A, 1990, “Aggregation in Input-Output Models: A Comment” *Economic Systems Research* **2** 436-438.

Olsen J A, 1993, “Aggregation in Input-Output Models: Prices and Quantities” *Economic Systems Research* **5** 253-275.

Olsen J A, 2000, “Aggregation in Macroeconomic Models: An Empirical Input-Output Approach” *Economic Modelling* **17** 545-558.

Olsen J A, 2001, “Perfect Aggregation of Industries in Input-Output Models”, in Lahr, M. L. and Dietzenbacher E. (editors), *Input-Output Analysis: Frontiers and Extensions* (Palgrave, New York).

Pyatt G, 1988, “A SAM Approach to Modeling”, *Journal of Policy Modeling* **10** 327-352.

Pyatt G, Round J, 1979, “Accounting and Fixed Price Multipliers in a Social Accounting Matrix Framework” *Economic Journal* **89** 850-873.

Stone R, 1978, *The Disaggregation of the Household Sector in the National Accounts*. World Bank Conference on Social Accounting Methods in Development Planning, Cambridge.

Theil H, 1957, “Linear Aggregation in Input-Output Analysis” *Econometrica* **25** 111-122.

Table 1. Structure of a Social Accounting Matrix

	Activities	Factors	Households	Rest of Accounts	Total
Activities	T_{11}	0	T_{13}	T_{14}	Y_1
Factors	T_{21}	0	0	T_{24}	Y_2
Households	0	T_{32}	T_{33}	T_{34}	Y_3
Rest of accounts	T_{41}	T_{42}	T_{43}	T_{44}	Y_4
Total	Y_1	Y_2	Y_3	Y_4	

Table 2. A Disaggregated Social Accounting Matrix

		Activities		Factors		Households			
		Activity 1	Activity 2	Factor 1	Factor 2	Household 1	Household 2	Exogenous Account	Total
Activities	Activity 1	25	45	0	0	15	5	10	100
	Activity 2	15	10	0	0	35	5	35	100
Factors	Factor 1	20	30	0	0	0	0	10	60
	Factor 2	15	10	0	0	0	0	15	40
Households	Household 1	0	0	25	20	0	0	5	50
	Household 2	0	0	35	20	0	0	20	75
	Exogenous Account	25	5	0	0	0	65	0	95
	Total	100	100	60	40	50	75	95	520

Table 3. An Aggregated Social Accounting Matrix

	Activities	Factors	Households	Exogenous Account	Total
Activities	95	0	60	45	200
Factors	75	0	0	25	100
Households	0	100	0	25	125
Exogenous Account	30	0	65	0	95
Total	200	100	125	95	520