PAPER

Title: A methodological proposal for disaggregating accounting frameworks based on a Three-Steps RAS. An application to the Supply and Use tables of Galicia (NW Spain).

Authors and e-mails of them:

1. Fernando de la Torre Cuevas: fernando.delatorre@usc.es

2. Xesús Pereira López: xesus.pereira@usc.es

Department: Applied Economics and Quantitative Economics

University: University of Santiago de Compostela

Subject area: Input-output models in regional analysis. Theory and applications

Abstract: *(maximum 300 words)*

When using input-output analysis to elaborate comparisons between different economies a usual problem is the absence of accounting frameworks that follow a homogeneous classification. This happens due to arbitrary decisions made by statistical institutes or because of changes in statistical regulations and methodologies. A simple solution for this kind of problems is to aggregate industries and products till a common classification is achieved. However, aggregation can lead to substantial information losses that can difficult the study of economic structures. In this work a methodological proposal for product/industry disaggregation in Supply and Use tables is presented. This methodology shares common ground with other disaggregation algorithms found in literature. As a novelty, the Three-Steps RAS algorithm by Gilchrist & St. Louis is used as a way for incorporating exogenous information, respecting basic equilibria, and introducing other possible constraints. The methodology proposed can be easily implemented using non-specialised software. As an empirical application, the Supply and Use tables of Galicia (NW Spain) for the years 2008 and 2016 were estimated using the algorithm developed. By contrasting the estimated matrices with the officially published ones, the accuracy of disaggregation is evaluated.

Keywords: Input-Output; Disaggregation; RAS Method

JEL codes: C67; D57; R15

0. Introduction

For regional economies input-output (IO) statistics are sometimes not available or published with a low level of detail (Miller & Blair, 2009, pp. 160–161). This fact presents several problems. Firstly, if different input-output tables are to be compared it is important that they present the same product/industry classification. The most common way of achieving this common classification is to aggregate the available information. In input-output analysis, the mixture of different industries and products is a source of inaccuracy. By operating this way, relevant information about the economic structures that are to be analysed can be lost. In second place, environmental analysis normally requires for more disaggregated economic data (Chang, Ries, Man, & Wang, 2014; S. Lindner, Legault, & Guan, 2013; Liu, Lenzen, & Murray, 2012). While aggregation of environmental data in order to meet the economic product/industry classification has been the path usually taken in literature, it has been proved that disaggregation even if based on few real data has a superior outcome (Lenzen, 2011). These and other necessities justify the idea of having a relatively efficient and simple disaggregation scheme on hand. This is the purpose of the present work.

The investigation is structured as follows. In first place, the main contributions regarding disaggregation techniques in input-output contexts are summarised. It will be shown how these methodologies evolved and how they gradually incorporated RAS-type algorithms to adjust the matrices and ensure the disaggregated table's coherence. In section 2, a three-stage algorithm for disaggregating matrices will be presented. Section 3 provides an example of how the disaggregation technique operates through an empirical application. The experiment consists of comparing two sets of "true" disaggregated Supply and Use tables (SUTs) with the disaggregated estimation obtained using the present methodology. In addition, the algorithm´s performance will be tested against projection methods. Section 4 discusses the methodology presented both in analytical and in empirical terms. Section 5 summarises the conclusions achieved in the course of this investigation and outlines possible future lines of research.

1. Literature review

Traditionally, disaggregation problems occupied a secondary place in input-output literature. Scholars normally directed their attention to aggregation bias analysis. The literature review here presented does not intend to be an exhaustive comment about everything that has been written about this issue. The purpose of these lines is to give necessary credit to those contributions that inspired the methodology proposed afterwards. The linkages between these contributions and the present investigation will be addressed in section 4.

In Fei (1956) we can find an early contribution concerning the disaggregation of an industry in an input-output framework. Fei pointed out that when a matrix is aggregated, it is impossible to go back and find its disaggregated counterpart. This is because aggregation suppresses part of the information about the true matrix. In order to disaggregate an industry, Fei suggested that some of this information must be recovered. In an extreme case, at least the original number of sectors that were aggregated has to be known. With this information, a system of weights is elaborated and used to split the values of the aggregated matrix. The disaggregated matrix is, as said, different from the true one due to information losses. An adjustment process is needed. In this case, it consists in finding a matrix that corrects the errors included in the disaggregated estimation. The less information recovered, the harder the harder the task to be assumed in this adjustment stage.

In Miller & Blair (1985, 2009) a specific section devoted to disaggregation cannot be found. Despite that, these authors addressed a similar problem: the inclusion of a new sector in an already elaborated IO table. Two approaches are presented in their work: the “Final-Demand” and the "Complete Inclusion in the Technical Coefficients Matrix". The “Final-Demand” approach was first presented by Isard and Kuenne (1953) and Miller (1957). This methodology (Miller & Blair, 2009, pp. 634–635) is focused on impact analysis rather than the construction of more detailed input-output tables. An IO table close enough to the one that is to be disaggregated is considered as a reference. The technical coefficients associated with the inputs supplied by the pre-existing sectors to the new sector disaggregated are taken from there. Given a certain gross output for the new industry, its impact on intermediate demand can be calculated and included alongside the pre-existing final demands. In this case, the new sector does not hold any forward linkages with the old ones. The "Complete Inclusion in the Technical Coefficients Matrix" approach (Miller & Blair, 2009, pp. 636–637) fills this gap. However, in the absence of information about the new technical coefficients a unique result with economic meaning is not guaranteed. For a symmetric IO table Aislabie & Gordon (1990) assumed that the total output (and thus, the total input) of the disaggregated industry is known. In contrast with Miller & Blair, no information about the disaggregated technical coefficients is available. In the example provided, the known row (column) total is allocated considering different options. Thus, no unique solution is provided either.

A major advance in disaggregation methods was introduced by Wolsky (1984). His approach consists in a two-step procedure. An illustration of his method (Wolsky, 1984, p. 288) is reproduced in equation 1.

$$A = \begin{pmatrix} 0.02 & 0.05 & 0.05 & 0.05 \\ 0.35 & 0.10 & 0.45 & 0.45 \\ w_1 \times 0.08 & w_1 \times 0.25 & w_1 \times 0.15 & w_1 \times 0.15 \\ w_2 \times 0.08 & w_2 \times 0.25 & w_2 \times 0.15 & w_2 \times 0.15 \end{pmatrix} + \begin{pmatrix} 0 & 0 & w_2\delta_1 & -w_1\delta_1 \\ 0 & 0 & w_2\delta_2 & -w_1\delta_2 \\ \sigma_1 & \sigma_2 & \left(\left(\frac{1}{2}\right)\delta_3 + \xi\right)w_2 + \sigma_3 & -\left(\left(\frac{1}{2}\right)\delta_3 + \xi\right)w_2 + \sigma_3 \\ -\sigma_1 & -\sigma_2 & \left(\left(\frac{1}{2}\right)\delta_3 - \xi\right)w_2 - \sigma_3 & -\left(\left(\frac{1}{2}\right)\delta_3 - \xi\right)w_2 - \sigma_3 \end{pmatrix} \quad (1)$$

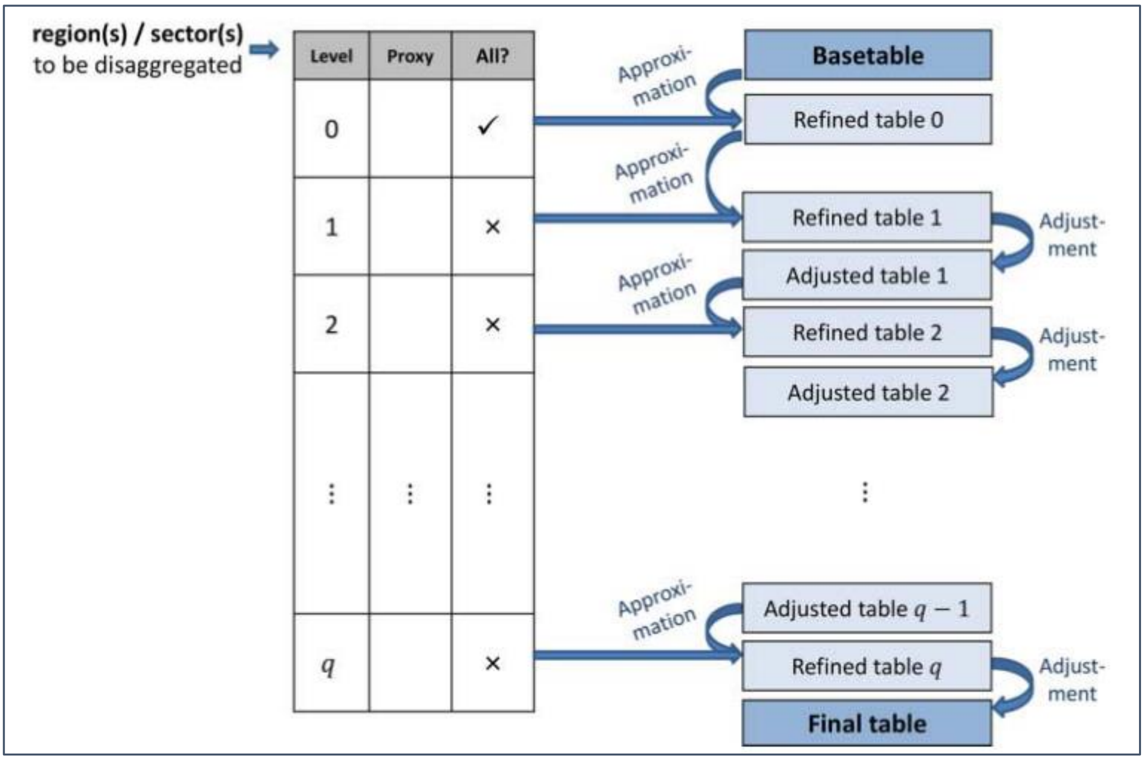
Firstly, an “augmented matrix” is built. In the example provided, it corresponds with the first matrix on the right hand side of the equation. The disaggregated sectors are introduced by splitting coefficients using a system of weights. This weights are stablished according to a known gross output distribution or other parameters available. In Wolsky´s case, the weights (w) are calculated as the proportion between the gross output of the disaggregated new sectors to the gross output of the aggregated one. In the example provided for the augmented matrix, only one new row is included. The UN Handbook of input-output table compilation and analysis (1999, p. 220), follows Wolsky´s approach, although it does not suggests further adjustments after obtaining the augmented matrix.

This model description is corrected by adding a second “distinguishing” matrix. In the example provided, it corresponds with the second matrix on the right hand side of the equation. If it´s true that Wolsky took Fei´s work as inspiration, this stage corresponds with the “adjustment process” described by de later. The distinguishing matrix embodies the differentiated weights for the purchases (δ) and sales (σ) of each new industry with the pre-existing ones. The distinguishing matrix also contains parameters (ξ) that inform about the relations between the disaggregated industries and themselves. The industries order is arranged so that the disaggregated sectors appear in the last rows and columns of the table, allowing the partition of the matrix. The new disaggregated matrix is obtained by summing the augmented and the distinguishing matrices. Wolsky´s method faces the problem of obtaining the δ, σ and ξ parameters in a context of limited information. To solve that issue, Wolsky itself provides a solution based on the estimation of their bounds considering the weights used for the augmented matrix. Gillen & Guccione (1990) criticised Wolsky´s work alleging the costs of estimating the distinguishing parameters. Their methodology required of information about total outputs, intermediate demand, and prices for the new industries to include in the model. This data might be available in some developed countries with some degree of aggregation, but it is rare to find them with the level of detail used in IO tables or SUTs. A similar proposal was made by Su et al. (Su, Huang, Ang, & Zhou, 2010) focused on environmental issues. However, the information used by these authors is also rarely available for many industries, specially at the regional level.

More recently, Lindner et al. (2012) contributed to Wolsky´s approach in to different directions. Firstly, they introduced an additional constraint related with the final demand. In order to ensure that all final demand elements stay positive, the authors imposed final demand-to-output ratios. While Wolsky focused only on disaggregating the coefficient matrix, this proposal allowed for the disaggregation of the final demand and de value added vectors at the same time too. A solution with economic meaning is also ensured. Secondly, the estimation of the distinguishing parameters is addressed in a more systematic way. This work generalises the calculation of lower and upper bounds for these parameters in the case of n new industries. Moreover, the range of possible values of these parameters is explored through a random walk algorithm in order to choose the most suitable values. In Barrera Lozano et al. (2015) a slight variation of Wolsky´s methodology is also developed. In this case, the augmented and distinguishing matrices are named *half-point* and *imaginary* matrices respectively. This work uses Wolsky´s scheme to disaggregate Social Accounting Matrices (SAMs).

The introduction of explicit adjustment rules which guarantee that the sum of the disaggregated sub-flows matches the original aggregated flow is another advance to be considered. In Wenz et al. (2015) a first approach to this issue can be found. In this contribution, the entries of a Multi-Regional Input-Output (MRIO) matrix are split following an iterative procedure with two steps (figure 2).

Figure 1. Schematic illustration of the refinement algorithm by Wenz et al.



Source: Wenz et al. (2015, p. 203).

Firstly, flows are divided into sub-flows according to different sources of information. This step is called *approximation*. These sources of information are hierarchically ranked according to its quality. The entries are split according to the lest accurate source of information. In second place, all new sub-flows are re-scaled to meet their correspondent aggregated flow total. The scaling process is delivered by difference. This step is called *adjustment*. In the first iteration, all entries are equally distributed into the new sub-flows. This is the least accurate proxy that can be considered. After this first iteration, more refined data is introduced. Information with a higher ranking might not be available for all parts of the matrix. In this case, the sub-flows remain as in the previous iteration and are only scaled if necessary to meet superordinate flow. This approximation-adjustment procedure is repeated until the most accurate source of information is considered.

A similar procedure can be found in Avelino (2017). The Temporal EURO method proposed by this author is used to obtain quarterly or biannual IO tables. To ensure the annual coherence of the tables, an adjustment step is also implemented in the algorithm. In this case, scaling is done by multiplication like in the last step of Gilchrist & St. Louis Three-Steps RAS (1999, 2004). The advantage of scaling cells this way is that sign flips are avoided. Holý & Šafr (2020) went a step beyond and proposed the use of multidimensional RAS to disaggregate national IO tables into regional ones. In their contribution, disaggregated cells are iteratively scaled considering the national row and column totals. Other dimensions, such as domestic/imported flows or sectoral disaggregation are considered in their contribution. In addition, it must be noted that these authors only address the adjustment process of the disaggregated IO tables, not the procedure needed to generate their first approximation. Observations about the analytical formulation of their proposal can be found in De Mesnard (2020). A generalised multidimensional GRAS is also presented in Valderas Jaramillo & Rueda Cantuche (2021).

To conclude, Rueda Cantuche et al. (2020) proposal must be commented. In their work, these authors propose a method to convert SUT into different product/industry classifications. The RACE algorithm. This includes *ono-to-one*, *many-to-many, many-to-one* (aggregation) and *one-to-many* (disaggregation) conversion cases. Here we will focus on the last scenario. Rueda Cantuche et al. use two types of bridge matrices (R and C) to split the old product and industry classification of matrix A into the new one. In a first stage, columns are disaggregated considering bridge matrix C and scaled to ensure their coherence with the aggregated data. Matrix C is iteratively modified until a coherent reclassification is found. The same process is applied then to row re-classification. The main limitation of this approach, pointed out by their authors, is that bridge matrices are rarely available. In fact, for their empirical application, these author do not state these matrices were built. In addition, considering the empirical application presented in the article, RACE seems to perform worse than projection methods such as SUT-EURO.

2. Methodological proposal

Mathematically, the problem to be solved consist in transforming matrix $Z_{m\times n}$ into a new matrix $Z^*_{(m+m')\times(n+n')}$. In this case, let:

1. row m be disaggregated into rows $m, m+1, \cdots, m+p, \cdots, m+m'$ and
2. column n into columns $n, n+1, \cdots, n+q, \cdots, n+n'$.

To do so, a three-stages methodology is formulated. In the first stage, a system of weights in built taking as reference another matrix that has the desired dimensions. Next, an augmented matrix is calculated splitting the original entries of Z. Finally, this augmented matrix is adjusted to estimate the technological relationships hidden in matrix Z. An iterative proportional fitting algorithm is used for this purpose. Matrix Z^* is calculated at the end of these three stages.

a. Stage 1: calculation of the weight matrix *W*

A system of weights is needed to spread the cells of matrix Z obtaining an augmented matrix $Z^{(0)}_{(m+m')\times(n+n')}$. Let $Z^{W}_{(m+m')\times(n+n')}$ be the matrix used as reference for the weight calculation. Let W be the matrix that contains the system of weights. The interior elements of this matrix are defined as follows.

— For the non-disaggregated rows and columns $i = 1, \cdots, m-1$ and $j = 1, \cdots, n-1$:

$$w_{ij} = 1 \tag{2}$$

— For the generic new row $m+p$ and columns $j = 1, \cdots, n-1$:

$$w_{m+p,j} = \frac{z^{W}_{m+p,j}}{\sum_{i=m}^{m+m'} z^{W}_{ij}} \tag{3}$$

— For the generic new column $n+q$ and rows $i = 1, \cdots, m-1$:

$$w_{i,n+q} = \frac{z^{W}_{i,n+q}}{\sum_{j=n}^{n+n'} z^{W}_{ij}} \tag{4}$$

By definition:

$$\sum_{i=m}^{m+m\prime} w_{ij} = 1 \tag{5}$$

$$\sum_{j=n}^{n+n\prime} w_{ij} = 1 \tag{6}$$

— For the generic crossing between the new row $m + p$ and column $n + q$:

$$w_{m+p,n+q} = \frac{z_{m+p,n+q}^{w}}{\sum_{i=m}^{m+m\prime} \sum_{j=n}^{n+n\prime} z_{ij}^{w}} \tag{7}$$

Analogously, by definition:

$$\sum_{i=m}^{m+m\prime} \sum_{j=n}^{n+n\prime} w_{ij} = 1 \tag{8}$$

For the cases:

$$\sum_{i=m}^{m+m\prime} z_{ij}^{w} = 0 \tag{9}$$

$$\sum_{j=n}^{n+n\prime} z_{ij}^{w} = 0 \tag{10}$$

$$\sum_{i=m}^{m+m\prime} \sum_{j=n}^{n+n\prime} z_{ij}^{w} = 0 \tag{11}$$

The value of the w_{ij} entries cannot be defined this way. It may also happen that some entries of matrix Z^{w} are not available. In such situations[1]:

[1] This solution was first proposed by Fei (1956, p. 411) and it is similar to Wenz et al. (2015) first approximation step.

$$w_{ij} = 1/m' \quad \text{for} \quad j = 1, \cdots, n-1 \tag{12}$$

$$w_{ij} = 1/n' \quad \text{for} \quad i = 1, \cdots, m-1 \tag{13}$$

$$w_{ij} = 1/(m'n') \quad \text{for} \quad i = m, \cdots, m+m' \quad and \quad j = n, \cdots, n+n' \tag{14}$$

In addition, marginal weights are calculated. For clarity purposes in notation, let the subscript $\bullet$ denote the sum of a row or column.

— For the generic new row $m+p$:

$$w_{m+p,\bullet} = \frac{\sum_j z^w_{m+p,j}}{\sum_{i=m}^{m+m'} \sum_j z^w_{ij}} \tag{15}$$

— For the generic new column $n+q$:

$$w_{\bullet,n+q} = \frac{\sum_i z^w_{i,n+q}}{\sum_{j=n}^{n+n'} \sum_i z^w_{ij}} \tag{16}$$

b. Stage 2: the augmented matrix $\boldsymbol{Z}^{(0)}$

Once matrix W is calculated, $Z^{(0)}$ is derived as follows

— For the elements of rows and columns $m = 1, \cdots, m-1$ and $n = 1, \cdots, n-1$:

$$z^{(0)}_{ij} = z_{ij} \tag{17}$$

— For the generic new row $m+p$ and columns $j = 1, \cdots, n-1$:

$$z^{(0)}_{m+p,j} = z_{mj} w_{m+p,j} \tag{18}$$

— For the generic new column $n+q$ and rows $i = 1, \cdots, m-1$:

$$z^{(0)}_{i,n+q} = z_{in} w_{i,n+q} \tag{19}$$

— For the crossing between disaggregated rows and columns:

$$\begin{pmatrix} z^{(0)}_{mn} & \cdots & z^{(0)}_{m,n+n'} \\ \vdots & \ddots & \vdots \\ z^{((0))}_{m+m',n} & \cdots & z^{(0)}_{m+m',n+n'} \end{pmatrix} = z_{mn} \begin{pmatrix} w_{mn} & \cdots & w_{m,n+n'} \\ \vdots & \ddots & \vdots \\ w_{m+m',n} & \cdots & w_{m+m',n+n'} \end{pmatrix} \tag{20}$$

The entries of matrix $Z^{(0)}$ are obtained by combining the technological information of matrix Z^{w} with the aggregated information of matrix Z. The more complete and close this technological information is, the more accurate this first approximation will be.

Finally, the total sum for new disaggregated rows and columns is to be estimated.

— For the new generic new row $m+p$:

$$z^{*}_{m+p,\bullet} = w_{m+p,\bullet} \sum_{j} z_{mj} \qquad (21)$$

— For the generic new column $n+q$:

$$z^{*}_{\bullet,n+q} = w_{\bullet,n+q} \sum_{i} z_{in} \qquad (22)$$

These marginal estimations will be considered as the correct values for row and column sums. In consequence, superscript $*$ will be associated with them.

c. Stage 3: obtention of the adjusted matrix Z^{*}

The augmented matrix $Z^{(0)}$ reflects the technological relations taken as reference from Z^{w}, which may differ from the real ones implicit in matrix Z. An adjustment process is needed in order to obtain a distinguished matrix Z^{*}. For this purpose, TRAS algorithm is used.

In the case of rows and columns, their marginal estimation will differ from the total sum of the disaggregated rows and columns in the augmented matrix. This is caused due to the lack of complete information. In mathematical terms:

$$z^{*}_{i\bullet} \neq z^{(0)}_{i\bullet} \quad \forall i = m, \cdots, m+p, \cdots, m+m' \qquad (23)$$

$$z^{*}_{\bullet j} \neq z^{(0)}_{\bullet j} \quad \forall j = n, \cdots, n+q, \cdots, n+n' \qquad (24)$$

For the remaining rows and columns if no other information is available:

$$z^{*}_{i\bullet} = z^{(0)}_{i\bullet} \quad \forall i = 1, \cdots, m-1 \qquad (25)$$

$$z_{\bullet j}^{*} = z_{\bullet j}^{(0)} \quad \forall j = 1, \cdots, n-1 \tag{26}$$

Despite this, additional data about the total sum of some rows and/or columns can be incorporated to the adjustment process.

The adjustment process has the form of a RAS-type problem. A benchmark matrix $Z^{(0)}$ is to be scaled in order to meet new row and column totals obtaining this way matrix Z^{*}. To avoid problems with possible negative entries, Generalised RAS (GRAS) is used (Günlük-Şenesen & Bates, 1988; Junius & Oosterhaven, 2003) for row and column scaling. Matrix $Z^{(0)}$ is divided into matrix $P^{(0)}$ that contains only its positive elements and $N^{(0)}$ with the absolute values of the negative elements. Thus, $Z^{(0)} = P^{(0)} - N^{(0)}$. Let:

- u be the vector that contains the row sum targets $z_{i\bullet}^{*} \quad \forall i = 1, \cdots, m + m'$.
- v be the vector that contains the column sum targets $z_{\bullet j}^{*} \quad \forall j = 1, \cdots, n + n'$
- e a sum vector of the appropriate dimensions containing ones.

The solution of the problem has the form:

$$\left(\hat{r}P^{(0)}\hat{s} - \hat{r}^{-1}N^{(0)}\hat{s}^{-1}\right)e = u \tag{27}$$

$$e\left(\hat{r}P^{(0)}\hat{s} - \hat{r}^{-1}N^{(0)}\hat{s}^{-1}\right) = v \tag{28}$$

GRAS is applied for $l = 0, 1, \cdots, \lambda, \cdots, \Lambda$ iterations. Resulting diagonal matrices $\hat{r}_{(m+m')\times(m+m')}$ and $\hat{s}_{(n+n')\times(n+n')}$ can be interpretated as a first estimation of the technological differences between the real matrix Z^{*} and the information taken as reference in matrix Z^{w}.

However, the obtention of the distinguished matrix is subjected to a third class of constraints related with matrix subtotals that are known. Matrix Z^{*} has to fulfil:

- For the disaggregated rows $i = m, \cdots, m + p, \cdots m + m'$:

$$\sum_{i=m}^{m+m\prime} z_{ij}^{*} = z_{mj} \quad \forall j = 1, \cdots, n-1 \tag{29}$$

— For the disaggregated columns $j = n, \cdots, n+q, \cdots, n+n'$:

$$\sum_{j=n}^{n+n\prime} z_{ij}^{*} = z_{in} \quad \forall i = 1, \cdots, m-1 \tag{30}$$

— For the crossing between disaggregated rows and columns:

$$\sum_{i=m}^{m+m\prime} \sum_{j=n}^{n+n\prime} z_{ij}^{*} = z_{mn} \tag{31}$$

To respect this third group of conditions, TRAS algorithm is used. A correcting matrix $K_{(m+m')\times(n+n')}$ is calculated.

— For the disaggregated rows m:

$$k_{ij} = \frac{z_{mj}}{\sum_{i=m}^{m+m'} z_{ij}^{*}} \quad \forall i = m, \cdots, m+m' \quad \forall j = 1, \cdots, n-1 \tag{32}$$

— For the disaggregated columns n:

$$k_{ij} = \frac{z_{in}}{\sum_{j=n}^{n+n'} z_{ij}^{*}} \quad \forall i = 1, \cdots, m-1 \quad \forall j = n, \cdots, n+n' \tag{33}$$

— For the crossing between disaggregated rows and columns:

$$k_{ij} = \frac{z_{mn}}{\sum_{i=m}^{m+m'} \sum_{j=n}^{n+n'} z_{ij}^{*}} \quad \forall i = m, \cdots, m+m' \quad \forall j = n, \cdots, n+n' \tag{34}$$

— For the cells not affected by the disaggregation process:

$$k_{ij} = 1 \quad \forall i = 1, \cdots, m-1 \quad \forall j = 1, \cdots, n-1 \tag{35}$$

Let $t = 0, 1, \cdots, \tau, \cdots, \mathrm{T}$ be the successive iterations including row, column, and correcting scaling. Firstly, the row and column scaling take place. The process stops at iteration $l = \Lambda$ when:

$$max\left|z_{i\bullet}^{(l)} - z_{i\bullet}^{*}\right| < \delta \quad \forall i = 1, \cdots, m+m' \tag{36}$$

$$max\left|z_{\bullet j}^{(l)} - z_{\bullet j}^{*}\right| < \delta \quad \forall j = 1, \cdots, n + n' \tag{37}$$

with δ being a small value.

After that, the third step is implemented. Let $t = 0, 1, \cdots, \tau, \cdots, \mathrm{T}$ be the iterations that included row, column, and correcting scaling processes. For the generic iteration τ:

$$Z^{(\tau)} = K^{(\tau-1)} \circ [\hat{r}^{(\tau-1)} Z^{(\tau-1)} \hat{s}^{(\tau-1)}] \tag{38}$$

Where $\circ$ stands for the Hadamard product, i.e., the element-by-element multiplication of the two matrices. After $t = \mathrm{T}$ iterations, the algorithm converges towards the solution Z^* which respects the three types of constraints considered in the distinguishing process.

3. Empirical application

a. Data preparation

To test the methodology proposed we have aggregated the 2008 and 2016 SUTs of Galicia (NW Spain). All data used can be easily found in the Galician Statistics Institute´s website[2]. The matrices considered were the "Supply table at basic prices and transformation into purchaser´s prices" and the "Use table at purchaser´s prices". The original matrices with 110 products and 72 industries were aggregated into 72 products and 25 industries. Imports, trade margins, final demand and gross value added were also aggregated considering this classification. The disaggregated classification corresponds to the one published for 2011 and 2016 SUTs. The aggregated classification was taken from the 2013 tables. The 2008 SUT published has 114 products and 78 industries. To achieve a common classification, straightforward aggregation was used[3]. Galicia´s SUT tables for the year 2011 were taken as the system of reference (Z^w).

[2] All matrices can be downloaded here: https://www.ige.eu/web/mostrar_actividade_estatistica.jsp?idioma=gl&codigo=0307007003

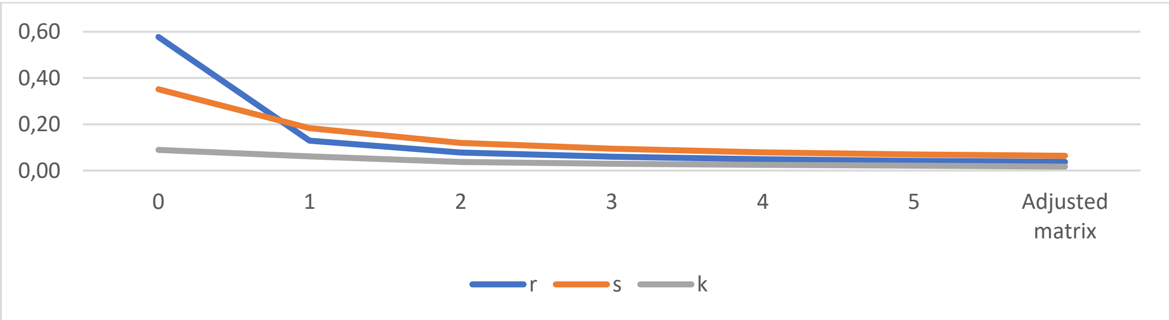
[3] In the Annex a table of correspondence is presented.

To preserve SUT´s coherence, the Supply tables were adjusted considering their marginal estimations. These were calculated according to equations (21) and (22). To overcome possible issues due to conflicting information, a modified version of KRAS (Lenzen, Gallego, & Wood, 2009) was implemented[4]. The row and column sums obtained for the Supply matrix were used to adjust the Use matrix. Here again, if conflicting information turns the adjustment process infeasible, targets were slightly modified using the modified KRAS. Row (r) and column (s) were scaled until conditions (35) and (36) were met with $\delta < 0.5$. After that, conditions (28), (29) and (30) were ensured using the correspondent correcting matrix K. SUTs were iteratively adjusted to obtain a coherent system. The process stopped after iteratively adjusting 5 times each table.

b. Results obtained

As for the uniqueness of the estimation obtained using the methodology proposed, figures 2, 3, 4 and 5 show how the process gradually converges towards a unique solution in all matrices. Scaling coefficients gradually diverged less from unity in each iteration. This means that smaller adjustments were needed as the algorithm was implemented. The third scaling step, operated by matrix K proved not to distort the unique character of the solutions obtained in the row and column adjustments (GRAS).

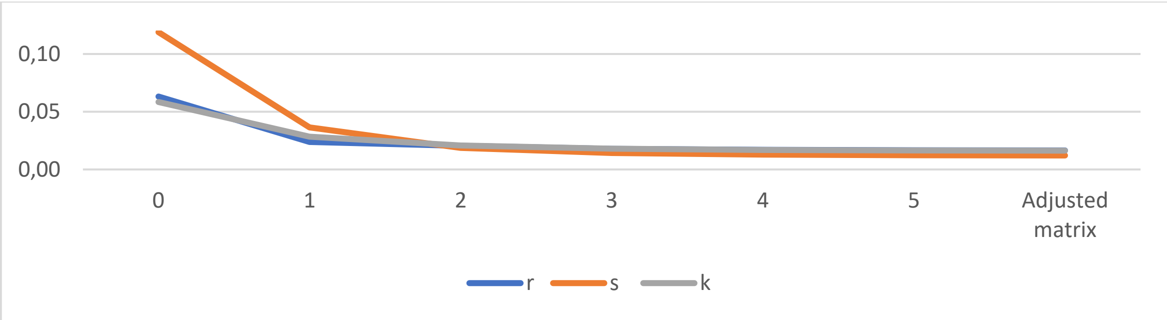
Figure 2. Mean Absolute Deviation from 1 of the row (r), column (s) and submatrix (k) scaling coefficients. Supply table, 2008.



Source: Own elaboration.

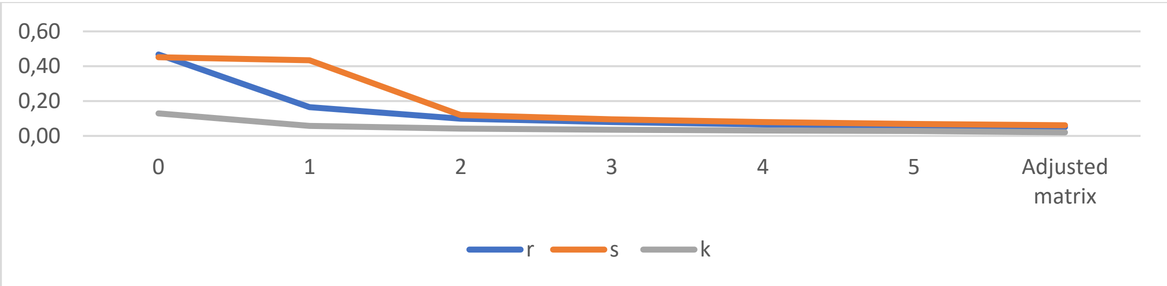
[4] Detailed formulation of this algorithm will be presented at the VIII SHAIO Permanent Workshop.

Figure 3. Mean Absolute Deviation from 1 of the row (r), column (s) and submatrix (k) scaling coefficients. Use table, 2008.



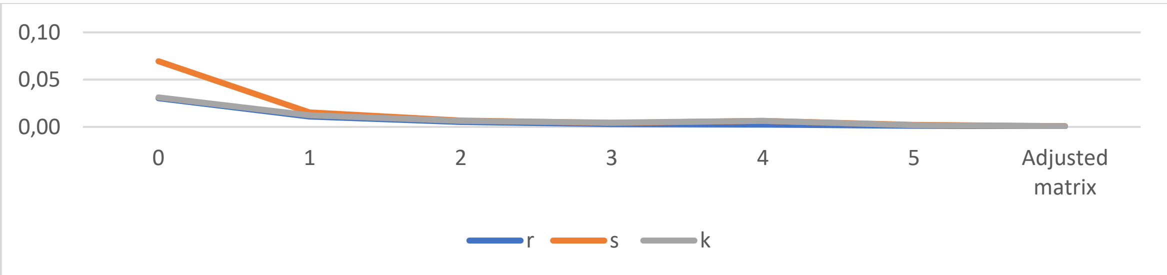
Source: Own elaboration.

Figure 4. Mean Absolute Deviation from 1 of the row (r), column (s) and submatrix (k) scaling coefficients. Supply table, 2016.



Source: Own elaboration.

Figure 5. Mean Absolute Deviation from 1 of the row (r), column (s) and submatrix (k) scaling coefficients. Use table, 2016.



Source: Own elaboration.

In order to measure the accuracy of the estimations, the same approach proposed by Rueda Cantuche et al. (2020) is followed. The results obtained are compared to the ones that can be calculated by *projection*, i.e., by using an updating algorithm with benchmark matrices that have the desired product/industry classification. In this case, the 2011 SUTs for Galicia are used as benchmark matrices. The 2008/2016 production-by-industry and supply/use-by-product vectors were considered as targets. The 2011 SUT was projected forward and backwards using GRAS algorithm (Günlük-Şenesen & Bates, 1988; Junius & Oosterhaven, 2003). GRAS outperforms SUT-EURO method used in Rueda Cantuche et al. (2020) according to Temursho et al. empirical findings (2011). When problems due to conflicting information appeared, the already mentioned version of KRAS was used to introduced small modifications on the target vectors.

The differences between the present disaggregation methodology and the projection approach were evaluated through the Wage Average Percentage Error (WAPE) formulated in Mínguez et al. (2009) as in Rueda Cantuche et al. (2020, p. 160).

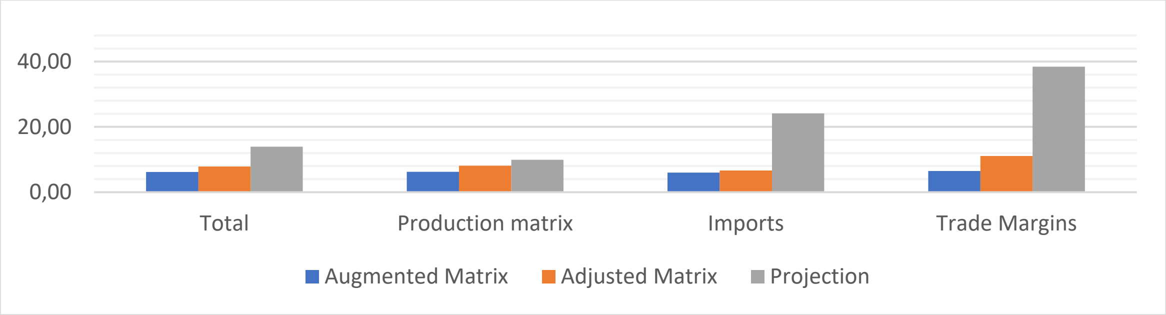
$$WAPE = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{\left| x_{ij}^{true} \right|}{\sum_i \sum_j x_{ij}^{true}} \right) \frac{\left| x_{ij} - x_{ij}^{true} \right|}{\left| x_{ij}^{true} \right|} \cdot 100 \qquad (39)$$

Where superscript “true” refers to the published matrices (2016, and the slightly aggregate version of 2008). This way, each percentage deviation is measured and weighed considering its importance in the original matrix.

Results are presented in figures 6, 7, 8, and 9. These empirical application suggests:

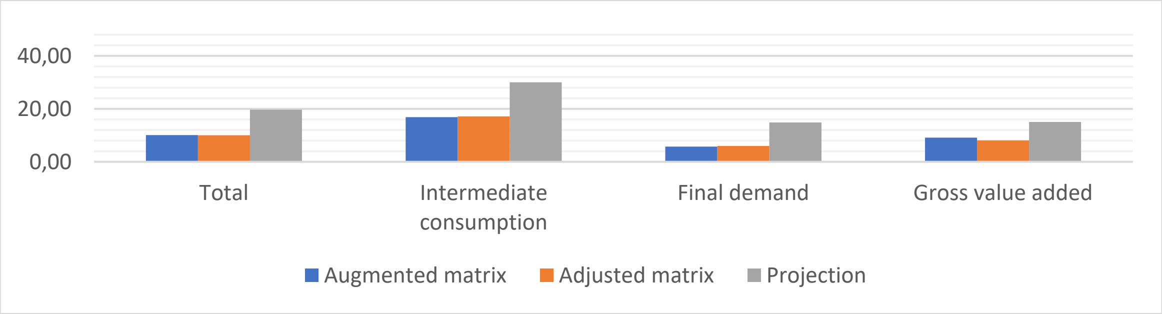
1. The disaggregation methodology here presented generally performs better than projection methods even when full information about row and column totals is available. The only case where projections outperformed disaggregation was found in the 2016 production matrix.
2. The closer matrix Z^w is to the “true” matrices, the better estimations are obtained. This seems to occur for al parts of the SUT system including imports, trade margins, final demand and gross value added submatrices. However, a five-year distance between the reference and the true matrices yielded estimations with total WAPE below 12%.
3. Adjusted matrices ensure coherent SUTs. This is achieved at the expense of some losses in the estimation’s accuracy. Nevertheless, the biggest distortion introduced by this stage (in the 2008 trade margins) adds less than a 5% error.

Figure 6. Wage Average Percentage Error. Augmented, Adjusted and Projected matrices. Supply table 2008.



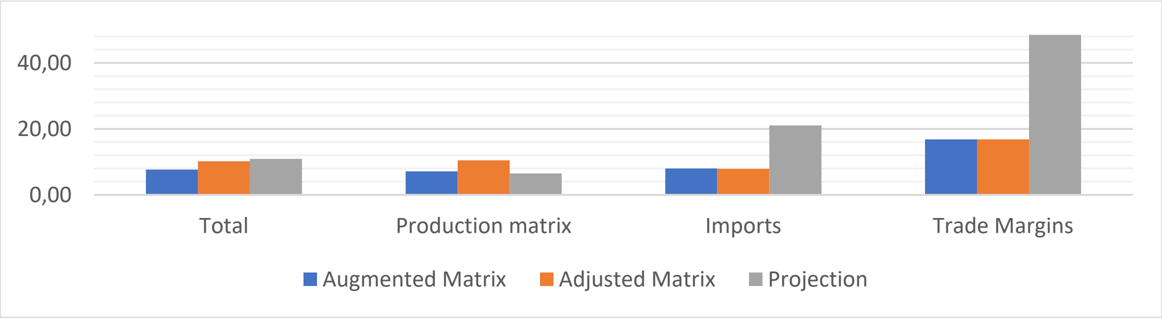
Source: Own elaboration.

Figure 7. Wage Average Percentage Error. Augmented, Adjusted and Projected matrices. Use table 2008.



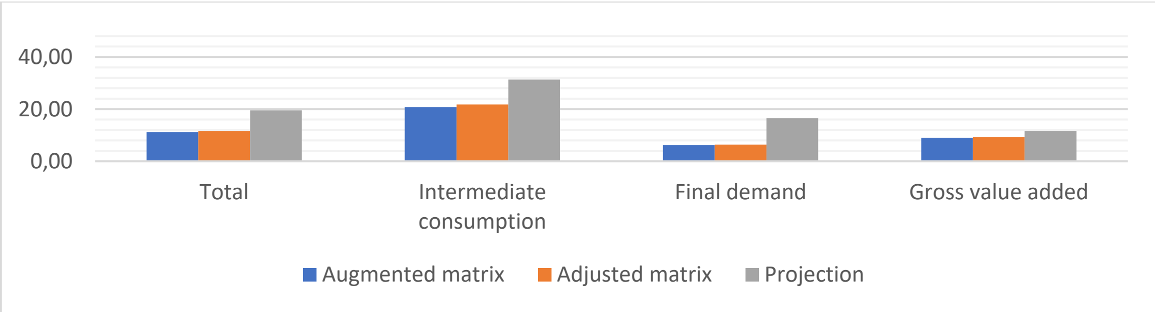
Source: Own elaboration.

Figure 8. Wage Average Percentage Error. Augmented, Adjusted and Projected matrices. Supply table 2016.



Source: Own elaboration.

Figure 9. Wage Average Percentage Error. Augmented, Adjusted and Projected matrices. Use table 2016.



Source: Own elaboration.

4. Discussion

In analytical terms, the methodology presented in this investigation incorporates much of the advances found in literature. Like in Fei (1956) and Wolsky (1984), an algorithm based in augmentation and adjustment operations is given. Augmentation stage follows a similar path by stablishing a weight system based on external information. This also relates with Miller and Blair´s (1985, 2009) proposal, who take other matrices as systems of reference. The adjustment process in this present investigation can be considered more transparent and less demanding in terms of information. Wolsky´s δ, σ and ξ parameters are here substituted by the r, s, and k coefficients. The solution for this coefficients, given a system of reference Z^w and an aggregated matrix Z is unique and has a straightforward calculation. The adjustment process is delivered considering as marginal totals as a target. This approach can be useful in contexts of limited information. Different from Wenz et al. (2015) the scaling adjustments to meet the original aggregated sums are done by multiplication (37) and not by difference. This ensures the absence of undesirable sign flips. Regarding this third adjustment step (k) the present methodology makes use of TRAS algorithm (Gilchrist & St. Louis, 1999, 2004). To conclude, this investigation follows Lindner et al. (2012), allowing for the introduction of constraints regarding final demand, gross value added, imports, etc. Thus, a complete Supply-Use system, and not only a symmetric matrix, can be disaggregated.

In empirical terms, the algorithm proved to present a unique solution for the disaggregation problem. This fact constitutes an improvement in comparison with Miller

& Blair (1985, 2009) and Aislabie & Gordon (1990). This disaggregation methodology also presented a better performance than projections via GRAS algorithm. This results contrasts with the one obtained by Rueda Cantuche et al. (2020, pp. 161–163). In this contribution, disaggregation performed considerably worse than projections made with the SUT-EURO method within a three-years period. In the present case, WAPE results obtained by disaggregation were still better even when there was a five-year gap between the system of reference and the true matrices. In a direct comparison, the present methodology appears to yield more accurate estimations than RACE does. This algorithm seems to perform better with unknown import/export data than RACE with known import/export data. The highest WAPE value obtained in the present case is 21.8% (intermediate consumptions, 2016) while the lowest WAPE obtained by Rueda Cantuche et al. considering known import/export data is 22.5% (final demand at purchaser´s prices, Czech Republic, 2007). Moreover, this proposal overcomes the absence of bridge matrices to re-classify products and industries by using other published matrices as a system of reference (Z^w). Here, matrices of a same region have been used, but it must be noted that national tables or others can be employed as long as they have the desired product/industry classification. As we have seen, the closer the system of reference is to the true matrices, the better estimations will be obtained.

5. Conclusions

The absence of survey-based detailed tables for regional economies is a problem to overcome when studying their structure, environmental issues, or other questions. Crucial information can be lost due to high aggregation. Moreover, disaggregation is sometimes needed in order to compare matrices over time and also between regions and countries. To address this problems, a methodology for disaggregating SUTs is here presented. This methodology can be used to disaggregate symmetric tables too. The stages and steps formulated in section 2 provide a systematic and transparent way for disaggregating product lines and industries. It has also been proved that information requirements for the proposed method are more realistic. As long as there is a matrix with the desired classification that can be taken as reference, it can be implemented. If information about the distribution of production by industries and supply/uses by products is available, missing information in the reference matrix is not necessarily an issue. The estimations obtained and summarised in section 3 suggest that this way of disaggregating matrices outperforms projection methods. These promising results, however, must be considered with caution due to the experiment´s limited scope.

Further investigation lines have also been identified in the course of the present work. Firstly, rectifications concerning aggregated matrix subtotals could be considered in more dimensions following Holý & Šafr (2020) and Valderas Jaramillo & Rueda Cantuche (2021) multidimensional RAS approaches. In second place, sign flips in certain parts of the SUTs could be allowed by implementing Lenzen et al. (2014) RAS variant. In third place, information requirements could be reduced by applying Path-RAS approach developed by Pereira López & Rueda Cantuche (2013). To conclude, it would be advisable to keep testing this methodology in different scenarios and evaluating its performance. In particular, the reproduction of Rueda Cantuche et al. (2020) experiment would be needed to clarify which proposal yields more accurate estimations. However, all these tasks go beyond the purpose of the present investigation.

6. References

Aislabie, C., & Gordon, M. (1990): "Input-output in practice: Sector disaggregation in regional input-output tables", *Economic Systems Research*, vol. 2, nº4, p. 407–420.

Avelino, A. F. T. (2017): "Disaggregating input–output tables in time: the temporal input–output framework", *Economic Systems Research*, vol. 29, nº3, p. 313–334.

Barrera Lozano, M., Mainar, A. J., & Vallés, J. (2015): "Disaggregation of sectors in social accounting matrices using a customized Wolsky method", *Applied Economics Letters*, vol. 22, nº13, p. 1020–1024.

Chang, Y., Ries, R. J., Man, Q., & Wang, Y. (2014): "Disaggregated I-O LCA model for building product chain energy quantification: A case from China", *Energy and Buildings*, vol. 72, p. 212–221.

De Mesnard, L. (2020): *Holy-Safr 's multidimensional RAS method is not unique*.

Fei, J. C.-H. (1956): "A Fundamental Theorem for the Aggregation Problem of Input-Output Analysis", *Econometrica*, vol. 24, nº4, p. 400–412.

Gilchrist, D. A., & St. Louis, L. V. (1999): "Completing Input–Output Tables using Partial Information, with an Application to Canadian Data", *Economic Systems Research*, vol. 11, nº2, p. 185–194.

Gilchrist, D. A., & St. Louis, L. V. (2004): "An algorithm for the consistent inclusion of partial information in the revision of input-output tables", *Economic Systems Research*, vol. 16, nº2, p. 149–156.

Gillen, W. J., & Guccione, A. (1990): "Disaggregating Input-Output Models: an Alternative to Wolsky's Method", *Economic Systems Research*, vol. 2, nº1, p. 39–42.

Günlük-Şenesen, G., & Bates, J. M. (1988): "Some Experiments with Methods of Adjusting Unbalanced Data Matrices", *Journal of the Royal Statistical Society*, vol. 151, nº3, p. 473–490.

Holý, V., & Šafr, K. (2020): *Disaggregating Input-Output Tables by the Multidimensional RAS Method*, p. 1–17.

Isard, W., & Kuenne, R. E. (1953): "The Impact of Steel Upon the Greater New York-Philadelphia Industrial Region", *The Review of Economics and Statistics*, vol. 35, nº4, p. 289–301.

Junius, T., & Oosterhaven, J. (2003): "The solution of updating or regionalizing a matrix with both positive and negative entries", *Economic Systems Research*, vol. 15, nº1, p. 87–96.

Lenzen, M. (2011): "Aggregation versus disaggregation in input-output analysis of the environment", *Economic Systems Research*, vol. 23, nº1, p. 73–89.

Lenzen, M., Gallego, B., & Wood, R. (2009): "Matrix balancing under conflicting information", *Economic Systems Research*, vol. 21, nº1, p. 23–44.

Lenzen, M., Moran, D. D., Geschke, A., & Kanemoto, K. (2014): "A Non-Sign-Preserving RAS Variant", *Economic Systems Research*, vol. 26, nº2, p. 197–208.

Lindner, S., Legault, J., & Guan, D. (2013): "Disaggregating the Electricity Sector of China'S Input-Output Table for Improved Environmental Life-Cycle Assessment", *Economic Systems Research*, vol. 25, nº3, p. 300–320.

Lindner, Sören, Legault, J., & Guan, D., (2012): "Disaggregating Input-Output Models With Incomplete Information", *Economic Systems Research*, vol. 24, nº4, p. 329–347.

Liu, C. H., Lenzen, M., & Murray, J. (2012): "A disaggregated emissions inventory for Taiwan with uses in hybrid input-output life cycle analysis (IO-LCA)", *Natural Resources Forum*, vol. 36, nº2, p. 123–141.

Miller, R. E. (1957): "The Impact of the Aluminum Industry on the Pacific Northwest: A Regional Input-Output Analysis", *The Review of Economics and Statistics*, vol. 39, nº2, p. 200–209.

Miller, R. E., & Blair, P. D. (1985): *Input-Output Analysis. Foundations and Extensions*, Prentice-Hall, Englewood Cliffs.

Miller, R. E., & Blair, P. D. (2009): *Input – Output Analysis. Foundations and Extensions* (2ª ed.), Cambridge University Press, Cambridge.

Mínguez, R., Oosterhaven, J., & Escobedo Cardeñoso, F. (2009): "Cell-corrected RAS method (CRAS) for updating or regionalizing an input-output matrix", *Journal of Regional Science*, vol. 49, nº2, p. 329–348.

Pereira López, X., & Rueda Cantuche, J. M. (2013): "Métodos de actualización global de tablas input-output con información limitada: su importancia para el análisis", *XXXIX Reunión de Estudios Regionales*.

Rueda Cantuche, J. M., Amores, A. F., & Remond Tiedrez, I. (2020): "Can supply, use and input–output tables be converted to a different classification with aggregate information?", *Economic Systems Research*, vol. 32, nº1, p. 145–165.

Su, B., Huang, H. C., Ang, B. W., & Zhou, P. (2010): "Input-output analysis of CO2 emissions embodied in trade: The effects of sector aggregation", *Energy Economics*, vol. 32, nº1, p. 166–175.

Temursho, U., Webb, C., & Yamano, N. (2011): "Projection of supply and use tables: Methods and their empirical assessment", *Economic Systems Research*, vol. 23, nº1, p. 91–123.

UN Department of Economic and Social Affairs (1999): *UN Handbook of input-output table compilation and analysis*, United Nations, New York.

Valderas Jaramillo, J. M., & Rueda Cantuche, J. M. (2021): "The multidimensional nD-GRAS method: applications for the projection of multiregional input-output frameworks and valuation matrices", *Papers in Regional Science*.

Wenz, L., Willner, S. N., Radebach, A., Bierkandt, R., Steckel, J. C., & Levermann, A. (2015): "Regional and Sectoral Disaggregation of Multi-Regional Input–Output Tables – a Flexible Algorithm", *Economic Systems Research*, vol. 27, nº2, p. 194–212.

Wolsky, A. M. (1984): "Disaggregating Input-Output Models", *The Review of Economics and Statistics*, vol. 66, nº2, p. 283–291.

7. Annex

Table 1. Product correspondence between 2008 (MIOGAL-08) and 2011 (MIOGAL-11) for Supply and Use Tables.

	MIOGAL-08		MIOGAL-11
R47A	Comercio polo miúdo agás de vehículos de motor e de combustible para vehículos de motor	R47	Comercio polo miúdo, salvo de vehículos de motor e motocicletas
R47B	Comercio polo miúdo de combustible para vehículos de motor		
R49A	Transporte por ferrocarril	R49	Transporte terrestre e por tubaxe
R49B	Outro transporte terrestre		
R50	Transporte marítimo e por vías navegables interiores	R50_51	Transporte marítimo e por vías navegables interiores; transporte aéreo
R51	Transporte aéreo		
R59_60M	Actividades cinematográficas de vídeo e televisión, gravación de son e edición musical; actividades de programación e emisión de radio e televisión de mercado	R59_60	Actividades cinematográficas de vídeo e televisión, gravación de son e edición musical; actividades de programación e emisión de radio e televisión
R60NM	Actividades de programación e emisión de radio e televisión de non mercado		
R74_75M	Outras actividades profesionais, científicas, técnicas e veterinarias de mercado	R74_75	Outras actividades profesionais, científicas, técnicas e veterinarias
R75NM	Actividades veterinarias de non mercado		
R94M	Actividades asociativas de mercado	R94	Actividades asociativas
R94NM	Actividades asociativas de non mercado		

Source: Own elaboration.