International Input-Output Association

Working Papers in Input-Output Economics

WPIOX 09-008

Jan Oosterhaven and Fernando Escobedo-Cardeñoso

Cell-corrected ras (cras) as a spatial input-output projection technique

Working Papers in Input-Output Economics

The Working Papers in Input-Output Economics (WPIOX) archive has been set up under the auspices of the International Input-Output Association. The series aims at disseminating research output amongst those interested in input-output, both academicians and practitioners. The archive contains working papers in input-output economics as defined in its broadest sense. That is, studies that use data collections that are in the format of (or are somehow related to) input-output tables and/or employ input-output type of techniques as tools of analysis.

Code: WPIOX 09-008

Authors: **Jan Oosterhaven** and Fernando Escobedo-Cardeñoso

Title:

Cell-corrected ras (cras) as a spatial input-output projection technique

Abstract:

The RAS method is developed to extrapolate a single matrix such that it conforms to new row and column totals. This paper presents a cell-correction of RAS (CRAS) that uses the distribution of cell variations, calculated from a series of different RAS projections, to project the input-output table (IOT) of a specific region or country. The solution of CRAS is derived from an additional optimization problem, based on first order reliability methods, to obtain the most likely cell-corrections to the regular RAS solution. To test the performance of CRAS, cumulative simulations are made with eleven survey IOTs of Spanish regions for 1998-2005. The results show that CRAS outperforms RAS when a limited set of survey IOTs is used that are close to the target IOT. When more IOTs with more different IO structures are added CRAS gradually leads to results that become worse than applying RAS to the single IOT that is most similar in IO structure terms.

Keywords: Input-output analysis, RAS, spatial projection methods, Spanish regions.

Archives: Construction of input-output tables, Interregional studies, Methods and mathematics

Correspondence addresses:

Jan Oosterhaven Faculty of Economics and Business, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands. E-mail: j.oosterhaven@rug.nl.

Fernando Escobedo-Cardeñoso School of Civil Engineering and Planning, University of Castilla-La Mancha, Avenida de Camilo Jose Cela 2, 13071 Ciudad Real, Spain. E-mail: fernando.escobedo@uclm.es .

Date of submission: September 3, 2009

CELL-CORRECTED RAS (CRAS) AS A SPATIAL INPUT-OUTPUT PROJECTION TECHNIQUE

Jan Oosterhaven[1] and Fernando Escobedo-Cardeñoso[2]

Abstract

The RAS method is developed to extrapolate a single matrix such that it conforms to new row and column totals. This paper presents a cell-correction of RAS (CRAS) that uses the distribution of cell variations, calculated from a series of different RAS projections, to project the input-output table (IOT) of a specific region or country. The solution of CRAS is derived from an additional optimization problem, based on first order reliability methods, to obtain the most likely cell-corrections to the regular RAS solution. To test the performance of CRAS, cumulative simulations are made with eleven survey IOTs of Spanish regions for 1998-2005. The results show that CRAS outperforms RAS when a limited set of survey IOTs is used that are close to the target IOT. When more IOTs with more different IO structures are added CRAS gradually leads to results that become worse than applying RAS to the single IOT that is most similar in IO structure terms.

Keywords: input-output analysis, RAS, spatial projection methods, Spanish regions.

JEL Codes: C61, C67, D57, R15.

Acknowledgements: We thank Roberto Mínguez for his contribution to developing the CRAS method in the case of updating input-output tables (Mínguez, Oosterhaven and Escobedo, 2009).

1. INTRODUCTION

RAS is known as an iterative technique to update semi-positive input-output tables (IOTs), given an old table and new row and column totals (Stone, 1961; Bacharach, 1970).[3] It is also used to construct regional IOTs given a national IOT or given an IO table of a different region in combination with the row and column totals of the region at hand (Hewings, 1969, 1977). Both ideas are combined when an old interregional IO table has to be updated given new regional row and column totals, and new national cell totals (Oosterhaven, Piek and Stelder, 1986). In consultancy practice, it is quite often also used to construct national IOTs for countries that do not have an own IOT. All RAS applications have in common that one single (old) matrix is given that needs to satisfy some set of (new) constraints.

In view of the tremendous amount of national, regional, interregional and international IO tables now readily available on the internet, it is surprising that hardly any attention has been paid to the problem of constructing a new IO table using the information of as many of the existing IOTs that is relevant to the construction problem at hand. The one exception is the Cell-Corrected RAS method (CRAS) developed by Mínguez, Oosterhaven and Escobedo (2009). It is tested on the problem of updating Dutch IOTs over the period 1969-1986, using as many of the older tables that are available. In this setting, it is concluded that CRAS performs better than RAS when gradual changes need to be forecasted. Using many old tables leads to worse results than only using the single most recent table when sudden shocks, such as the oil price rises of 1973-74 and 1979-80, need to be covered.

Here we will test how CRAS performs as an estimation technique for regional or national IO tables where such tables do not exist. It appears that the temporal projection of IOTs is far simpler than the spatial projection of IOTs, which is our current topic. The reason for this is that time is one-dimensional and uni-directional (from past to future). Space, however, is at least two-dimensional and bi-directional. Moreover, distance may be defined in many ways, e.g. physical or socio-economic, whereas time essentially is simply time. Hence, when only one single IOT is used for a temporal RAS, clearly best choice is to take the most recent IOT available. When, however, a single IOT is used for a spatial RAS, the best choice is not so obvious. It is not simply the IOT of the region or country most close by physical space. In stead it is the IOT of the region or country most close by in terms of IO structure, but which region or country that is, is not clear beforehand, as will be shown in the application.

To test CRAS as a spatial projection method we need a set of identically defined, survey-based IOTs. The IOTs need to be survey-based, as non-survey IOTs are not suited for testing a non-survey construction technique, while they need to be identically defined across regions or countries for the obvious reasons. To test CRAS, we will use the set of survey-based symmetric IOTs for 11 Spanish regions as collected and harmonized for the construction of a seven region interregional semi-survey IOT for Spain for 2005 (Escobedo and Oosterhaven, 2009).[4] As Spain has 17 regions, it follows that 6 Spanish regions do not have a survey-based IO table yet. If the test on the eleven existing IOTs is successful, the obvious first application of CRAS at the regional level would be the non-survey construction of the six yet non-existent Spanish regional IOTs.[5]

The setup of this paper is as follows. Section 2 will briefly summarize the nature of the Cell-Corrected RAS method and the nature of its use as a spatial non-survey IOT construction method. Section 3 will discuss the setup of the test on the existing 11 Spanish regional IOTs. The core of the problem is twofold. First, the test has to be set up such that it comes as close as possible to its potential use as a non-survey technique. Second, a solution has to be found for defining the structural IO distance between the regions at hand. Section 4 discusses the results of the comparison of the 11 survey IOTs with their non-survey estimates, each based either on RAS or on CRAS applied to increasing amounts more and more different survey IOTs. Section 5 concludes that CRAS outperforms RAS when a limited set of survey IOTs is used that are close to the IOT that has to be projected. When more IOTs with more different IO structures are added CRAS gradually leads to results that are worse than using RAS on the single IOT that is most close by in IO structure terms.

2. THE CELL-CORRECTED RAS METHOD (CRAS)

The goal of a conventional spatial RAS projection consists of obtaining an input-output transactions matrix Z^R for region or country R of dimension m by n as close as possible to the input-output transactions matrix Z^S of region or country S of the same dimension, knowing only the margins (the row and column sums) of the target Z^R.[6]

Statement of the programming model

The proposed new spatial projection method CRAS has two stages.

In the first one, data available for different regions are used in a standard RAS approach to estimate the parameters of the distributions of statistical deviations between the projected (RAS) regional IO tables and the true (survey) regional IO tables, $e \sim N(\mu^e, \sigma^e)$

(1) $$e_{ij}^{R(S)} = \frac{z_{ij}^R}{\tilde{z}_{ij}^{R(S)}}; \quad i = 1,...,m; \quad j = 1,...,n; \quad R, S = 1,...,T; \quad S \neq R$$

Where $e^{R(S)}$ is a stochastic term representing the unexplained deviation if we use Z^S as base matrix to estimate the target matrix Z^R by means of RAS; z_{ij}^R are the true values and $\tilde{z}_{ij}^{R(S)}$ are the values of the RAS projection; T is the number of regions or countries with an available IO table. Note that the problem is defined such that it also applies to rectangular IO matrices with $m \neq n$.

From (1), the first two distribution moment vectors μ^{e^R} (mean) and σ^{e^R} (standard deviation) of the stochastic deviations e can be calculated as follows:

(2) $$\mu_{ij}^{e^R} = \frac{\sum_{s=1, s\neq R}^{T} e_{ij}^{R(S)}}{T-1} \quad \text{and} \quad \sigma_{ij}^{e^R} = \sqrt{\frac{\sum_{s=1, s\neq R}^{T} \left(e_{ij}^{R(S)} - \mu_{ij}^{e^R}\right)^2}{T-2}} \quad ;\forall i, j, R$$

That is, we will have T values for μ_{ij}^e and σ_{ij}^e, one per target region *R*.

The second stage of the model uses the data of (2) to correct the RAS projection for region R ($\tilde{z}^{R(S)}$) by means of solving the following optimization problem:

(3) $$\underset{e}{Minimize} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{e_{ij}^{R(S)} - \mu_{ij}^{e^R}}{\sigma_{ij}^{e^R}} \right)^2$$

Subject to:

(4) $$\sum_{j=1}^{n} e_{ij}^{R(S)} \tilde{z}_{ij}^{R(S)} = u_i^R; \quad i = 1,...,m$$

(5) $$\sum_{i=1}^{m} e_{ij}^{R(S)} \tilde{z}_{ij}^{R(S)} = v_j^R; \quad j = 1,...,n$$

(6) $$e_{ij}^{R(S)} \geq 0; \quad i = 1,...,m; \quad j = 1,...,n$$

Where u_i^R equal the known row sums of the target matrix Z^R, and v_j^R equal the known column sums of Z^R. Equation (6) assures that the solution is semi-positive, although this last constraint is inoperational because all e values are centered around 1.

Once the optimization problem (3)-(6) is solved and the optimal values $e_{ij}^{R(S)^*}$ are available, the solution of CRAS, i.e. the values of the transaction matrix $\hat{Z}^{R(S)}$, is obtained as follows:

(7) $$\hat{z}_{ij}^{R(S)} = e_{ij}^{R(S)^*} \tilde{z}_{ij}^{R(S)}; \quad i = 1,\dots,m; \quad j = 1,\dots,n$$

Where (*) refers to the optimal values of $e^{R(S)}$.

Solution of the programming model[7]

It is instructive and handy to derive an explicit solution to better understand the behavior of the model. Consider the Lagrange function associated with problem (3)-(6):

(8) $$L(e,\lambda,\gamma) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{e_{ij}-\mu_{ij}^{e}}{\sigma_{ij}^{e}}\right)^2 + \sum_{i=1}^{m}\lambda_i\left(\sum_{j=1}^{n} e_{ij}\tilde{z}_{ij} - u_i\right) + \sum_{j=1}^{n}\gamma_j\left(\sum_{i=1}^{m} e_{ij}\tilde{z}_{ij} - v_j\right)$$

where λ_i and γ_j are the Lagrange multipliers.

The derivatives of the Lagrange function with respect to e, λ and γ are:

(9) $$\frac{\partial L}{\partial e_{ij}} = 2\frac{e_{ij}-\mu_{ij}^{e}}{\sigma_{ij}^{e^2}} + \tilde{z}_{ij}\left(\lambda_i + \gamma_j\right) = 0; \quad i = 1,\dots,m; \quad j = 1,\dots,n$$

(10) $$\sum_{j=1}^{n} e_{ij}\tilde{z}_{ij} - u_i = 0; \quad i = 1,\dots,m$$

(11) $$\sum_{i=1}^{m} e_{ij}\tilde{z}_{ij} - v_j = 0; \quad j = 1,\dots,n$$

Note that (9)-(11) represents a linear system with the following structure:

(12) $$\begin{bmatrix} A_{(mxn,mxn)} & B_{(mxn,m+n)} \\ B^T_{(m+n,mxn)} & O_{(m+n,m+n)} \end{bmatrix} \begin{bmatrix} e_{(mxn)} \\ \lambda_{(m+n)} \\ \gamma_{(m+n)} \end{bmatrix} = \begin{bmatrix} c_{(mxn)} \\ u_{(m+n)} \\ v_{(m+n)} \end{bmatrix}$$

where the dimensions of the corresponding matrices are in parentheses. Matrix A is a diagonal matrix with $a_{ij} = 2/\left(e_{ij}^{e}\right)^2$, matrix B contains the RAS solution $\tilde{z}_{ij}$, and O is an zero matrix. Note that, for convenience, the deviation matrix e has been reorganized in a column vector. The elements of the vector c are $c_{ij} = 2\mu_{ij}^e / \left(e_{ij}^e\right)^2$, and u and v are the vectors with the row sums and column sums of the target matrix, respectively.

For the system (12) to have a guaranteed unique solution, the rank of the coefficient matrix must be equal to its dimension $(mxn + m + n)$. The first column block $\begin{bmatrix} A \\ B^T \end{bmatrix}$ has rank mxn if $\sigma_{ij} \neq 0$, $\forall ij$, and the standard deviations are finite, because in that case A is a full diagonal matrix. However, the rank of matrix B is $m + n - 1$ if $\tilde{z}_{ij} \neq 0$, $\forall ij$, because in (10)-(11) there is a redundant constraint due to the compatibility condition that the sum of the row totals should equal the sum of the column totals, $u_{\bullet} = v_{\bullet}$. This redundancy must be removed, and therefore (12) must be generated eliminating one constraint in (10)-(11), no matter which. Once this condition holds, the system of linear equations is easily solved using sparse-oriented algorithms (LU, Gauss-elimination, etc.)

3. TESTING CRAS AS A SPATIAL PROJECTION METHOD

Next, we discuss how applying CRAS to the 11 Spanish survey-based regional IO tables has to be set up in order to test CRAS as a spatial IO projection method. The core of the problem is twofold. First, the test has to set up such that it comes as close as possible to its potential use as a projection technique. Second, a solution has to be found for defining the structural distance in terms of IOTs between the regions at hand, in order to determine how the performance of CRAS changes when the information of more and more IOTs is used.

Setting-up CRAS as a spatial IO projection technique

When transit trade is removed, the layout of the typical Spanish survey-based IOT for region *R* is shown in Table 1. The core problem is to use the 11 survey IOTs to simulate a situation that resembles the non-survey estimation of the lacking six IOTs, as much as possible. The solution of this problem should be based on the data that are indicated with “given” in Table 1. We claim that the data indicated with “estimation” in Table 1 can be estimated easily from the data that are “given” for the six Spanish regions for which there is not yet a regional IOT. The data indicated with “CRAS” then remain to be estimated by means of either RAS or CRAS.

The arguments for selecting the "estimation" part of Table 1 are as follows. We assume that nothing is known about both the intra-regional transactions and the exports and imports with regards to the Rest of Spain (RoS), because estimating them is the core of any non-survey estimation of regional IOTs. We should not assume that problem away by using the survey row and column totals of these matrices while comparing RAS or CRAS with a survey IOT. Unfortunately, in the past it has been assumed that the total intra-regional purchases and sales were known a priori for all sectors (Czamanski and Malizia, 1969; Morrison and Smith, 1974; Sawyer and Miller, 1983). As a consequence, it was unjustly concluded that RAS, as a non-survey technique, performed far better than competing non-survey techniques, such as e.g. the Location Quotient method (see Schafer and Chu, 1969). The latter, however, have been developed for the difficult estimation of precisely these intra-regional totals. So, they should not be assumed known a priori (see also Thuman, 1978).

For the six Spanish regions without an IOT, however, the exports to the Rest of the World (RoW) are given, as Eurostat requires National Statistical Offices (NSOs) to collect such data. The same holds for regional sectoral gross value added at market prices and its constituent components (net taxes on products, other net taxes on production, compensation of employees and gross operating surplus). Unknown total use and total output per regional sector may be estimated easily by using sector-specific ratios with gross value added at market prices as their base.[8] These ratios may be calculated either from the Spanish national IOT or from an appropriate average of the known regional IOTs. In order to separate the estimation error of these unknown totals from the estimation error of CRAS, we will use the actual information of each of the 11 regional survey-based IOTS while testing CRAS.

A more problematic decision is whether or not to assume that - for those six regions - the imports from the Rest of the World (RoW) can be estimated a priory or not, either as a full matrix or as a single row. The only regional foreign import data readily available in Spain are the detailed totals by products by region (Datacomex, 2009). Hence, it has to be assumed along each row of the foreign import matrix that all purchasing sectors and all categories of final demand have the same RoW import ratio. This will of course introduce an estimation error. In order not to pollute the estimation error of CRAS with the RoW import estimation error, we will use the RoW survey data while testing CRAS.

When the 'given' and the 'estimated' data are taken from the survey IOT of region *R*, RAS and CRAS are competing to estimate the remaining data. The IO data of the ten remaining regions S that form the database to estimate the remaining IO data for region *R* thus have the following structure:

(13) $$Z^S = \begin{pmatrix} z_{11}^S & \dots & z_{1n}^S & z_{y_{11}}^S \dots & e_1^{S^E} \\ \dots & z_{ij}^S & \dots & ..z_{y_{iq}}^S .. & e_i^{S^E} \\ z_{m1}^S & \dots & z_{mn}^S & \dots z_{y_{mf}}^S & e_m^{S^E} \\ p_{\bullet 1}^{S^E} & \dots & p_{\bullet n}^{S^E} & ..p_{y_{\bullet q}}^{S^E} .. & 0 \end{pmatrix}$$

Note that compared with Table 1, we have aggregated the RoS import table to a single RoS import row, as a full matrix is hardly ever required in IO applications, such as regional multiplier analysis.

For region R we only need the column and row sums of (13), i.e. we need to estimate:

(14) $$v^R = \begin{pmatrix} v_1^R & \dots & v_q^R & \dots & v_{n+f+1}^R \end{pmatrix} \text{ and } u^R = \begin{pmatrix} u_1^R \\ \dots \\ u_i^R \\ \dots \\ u_{m+1}^R \end{pmatrix},$$

to be substituted in (4)-(5).

The column sums of regional intermediate and final purchases from the whole of Spain can be calculated simply from the 'given' and the 'estimated' survey data in Table 1:

(15) $$1 \le q \le n+f \Rightarrow v_q^R = z_{\bullet q}^R + p_{\bullet q}^{R^E} = x_q^R - g_{\bullet q}^R - p_{\bullet q}^{R^M}$$

The row totals of the regional intermediate and final sales to the whole of Spain are also calculated simply from the 'given' and the 'estimated' survey data in Table 1:

(16) $$1 \le i \le m \Rightarrow u_i^R = z_{i\bullet}^R + e_i^{R^E} = x_i^R - e_i^{R^M}$$

The more difficult problem is how to estimate the totals of last row and last column of (13), without using the RoS trade data from the IOT of the target region *R*.[9] Hence, these totals have to be estimated by means of RoS trade data from the IOT of the base region *S*. To make maximum use of the known sector structure of region *R*, the sectoral trade ratios of region *S* are applied to region *R*'s 'estimated' sectoral purchases (15) and sectoral sales (16). This will lead to two, most certainly conflicting, residual estimates of region *R*'s total intra-regional transactions. Of the latter, we take the unweighted average to derive the required totals of the last row and last column of (13).

To clarify the need to take an average, the estimation will be formalized by means of the well known regional purchase coefficients (RPCs, Stevens and Trainer, 1980) and regional sales coefficients (RSCs, Boomsma and Oosterhaven, 1992). These equal one minus, respectively, the sectoral import ratio and the sectoral export ratio:

(17) $$RPC_q^S = 1 - \frac{z_{\bullet q}^{S^E}}{z_{\bullet q}^{S} + z_{\bullet q}^{S^E}} \; \forall q \text{ and } RSC_i^S = 1 - \frac{e_i^{S^E}}{z_{i\bullet}^{S} + e_{i\bullet}^{S^E}} \; \forall i$$

The intra-regional transaction total may then be estimated as the unweighted average of the estimates by means of the RPCs and the RSCs of region *S*:

(18) $$z_{\bullet\bullet}^{R(S)} = \frac{\sum_q RPC_q^S \left(z_{\bullet q}^R + z_{\bullet q}^{R^E} \right) + \sum_i RSC_i^S \left(z_{i\bullet}^R + e_i^{R^E} \right)}{2}$$

The exports' column total and the imports' row total with regard to RoS, then follow as the residual:

(19) $$q = n + f + 1 \Rightarrow v_q^R = e_{\bullet}^{R^E} = x_{\bullet}^R - e_{\bullet}^{R^M} - z_{\bullet\bullet}^{R(S)}$$

(20) $$i = m + 1 \Rightarrow u_i^R = z_{\bullet\bullet}^{R^E} = x_{\bullet}^R - va_{\bullet\bullet}^R - z_{\bullet\bullet}^{R^M} - z_{\bullet\bullet}^{R(S)}$$

The calculations (15)-(20) are made 110 times. For each of the 11 RAS projections we use the survey IOTs of the 10 remaining regions *S*.[10] Subsets of these ten RAS estimates for each *R* are then used to calculate the average deviation and the standard deviation of (2), which are used in the second stage of CRAS to produce the cell-corrected estimate of CRAS according to (7).

The next problem is which subsets of *S* to use. In temporal projections this choice is simple: the most recent table is the best choice. Temporal RAS and CRAS projections are then simply compared by adding more and more, less recent IOTs to the CRAS method, in order to compare the performance of both methods (see Mínguez, Oosterhaven and Escobedo, 2009). In spatial projections this is far more complicated, although theoretically it is still simple. The best choice for RAS is to take the survey IOT of the region *S* that resembles the projection region *R* best, and the best choice for CRAS is to add the second-best, the third-best, etc. regions *S*. Empirically, however, it is not known beforehand which region is first-best, second-best, etc.

To test RAS versus CRAS, our choice is to compare the best choice of regions in both cases. Hence, we have to determine the rank order of the 10 non-survey RAS projections of each of the 11 survey IOTs. To determine this rank order and to evaluate the performance of CRAS we will only compare the intra-regional parts of the IOTs, thus excluding estimates of the trade with the RoS. We could also compare the RoS results separately, but comparing the intra-regional part is far more important as its estimation errors determine the estimation errors of the regional or national multipliers for which regional or national IOTs are used most.

Accuracy of different RAS estimates for the Spanish regional IO tables

The comparisons are made by inspecting the distance between a projection z and the true value z^{true}, using different matrix distance measures (deMesnard and Miller, 2006). We only use additive measures, as their multiplicative equivalents have the same basic properties (deMesnard, 2004). Moreover, we only use measures that weigh the error made by the size of the IO cell at hand, as their unweighted equivalents provide inferior information. Thus, we only use the following matrix distance measures:

- Weighted Absolute Percentage Error (Butterfield and Mules, 1980):

(21) $$WAPE = \frac{\sum_i \sum_j \left| z_{ij} - z_{ij}^{true} \right|}{\sum_k \sum_l z_{kl}^{true}}$$

- Weighted Normalized Squared Error (Deming and Stephan, 1940):

(22) $$WNSE = \frac{\sum_i \sum_j \left(z_{ij} - z_{ij}^{true} \right)^2}{\sum_k \sum_l z_{kl}^{true}}$$

- Minimum Information Gain (Tilanus and Theil, 1965):

(23) $$MIG = \frac{\sum_i \sum_j \left| z_{ij}^{true} \ln\left(\frac{z_{ij}}{z_{ij}^{true}} \right) \right|}{\sum_k \sum_l z_{kl}^{true}}$$

Table 2 shows the result of (21)-(23) for the single best RAS projection per target IOT. Of course, as might be expected from the above definitions, *MIG* shows the least variation in performance of RAS per target IOT, whereas *WSNE* shows the largest variation. The *WAPE* nicely summarizes that the best RAS projection needs to be improved upon, as the weighted average cell-error runs from 25% to 50% per target IOT, which should be considered high. Next, especially when one looks at *WSNE*, it appears that projecting the IOT of the Madrid region is most difficult. This is not too surprising, as the capital region has an economic structure unlike any other Spanish region. Looking also at *WAPE* and *MIG*, the Baleares Islands and the Comunidad Valenciana also appear to be difficult to project. The case of Baleares is clear as it is an island economy with a much stronger tourism sector than any of the other ten Spanish regions with an IOT. Finally, it appears that only for Asturias, one needs a different base IOT to get the best score with each separate distance measure. In all other cases, the same base IOT is needed with either two or with all three of the distance measures.

The rank order of the performance of all possible ten base IOTs, of which Table 2 only shows the best projection, does differ, but not much. Therefore, from here on we only present the results for the average of the three measures.[11]

To get an unweighted average, all three measures need to be normalized. To get results that are easy to understand, we normalize each with the minimal value of its single best RAS projection for target region *R*, i.e. we normalize with the values of Table 2. This implies that the normalization is dependent upon target region *R* and the best base region *S* Consequently, for that R(S)-combination, all RAS values of other base IOTs will be larger than one, whereas CRAS values may either be smaller or larger than one, depending on whether that CRAS projection performs better or worse than the best RAS projection at hand. The theoretical minimum is zero, indicating a CRAS projection that produces the true target IOT, while the corresponding best RAS projection does not.

Thus, we only show results for our so-called Average Normalized distance Measure:

$$(24)\qquad ANM^{(C)RAS,R(S)} = \frac{1}{3}\left[\frac{WAPE^{(C)RAS,R(S)}}{\min\limits_{S\neq R} WAPE^{RAS,R(S)}} + \frac{WNSE^{(C)RAS,R(S)}}{\min\limits_{S\neq R} WNSE^{RAS,R(S)}} + \frac{MIG^{(C)RAS,R(S)}}{\min\limits_{S\neq R} MIG^{RAS,R(S)}}\right];$$

The rank-order of the results of (24) for the ten RAS-projections of each of the eleven target IOTs are shown in Table 3. Figure 1 shows the location and the size of the regions at hand. When the information of Table 3 and Figure 1 is combined, three types of regions may be distinguished.

First, large regions with long coasts, like Andalucía, Comunidad Valenciana, and Galicia. As we can see in Table 3, Andalucía appears only twice as one of the best four regions to predict the regional IOT of another region, and Comunidad Valenciana and Galicia do not appear one single time, so this type of regions is not a good choice as base matrix to make a projection. This is most likely due to the fact that these large regions with long coasts have different specific main economic sectors (agriculture, manufacturing, and fishing, respectively).

Second, large inland regions, like Aragón, Castilla-La Mancha and Castilla y León. They offer a good choice as base IOT to project each others' IOTs, as they appear five, four and six times, respectively, in the group of the best four base IOTs. This outcome may be due to the fact that these regions all lack singularities, such as small size, specific economic structure, high value added sectors, and having a coast. Castilla y León, for example, is a large region with a balanced economic structure, not strongly specialized in specific sectors, while it has a relatively central position.

Third, small regions, like Asturias, Navarra and País Vasco. These offer the best choice as base matrix. Asturias, for example, appears always as one of the best three base IOTs with all target regions. Navarra appears eight times and País Vasco seven times. The economic structure of these regions has a predominance of the industrial sector, especially in Navarra and País Vasco, with a service sector less important than in national level (MPT, 2009).

Outside these three groups stand the Balearic Islands and the Madrid region that offer the first-worst or the second-worst base IOT for most of the other nine regions. This is undoubtly due to the fact that, although quite different, both are small regions with a very strong service sector, specialized in tourism and central government, respectively. They only serve as a good base region once, namely while projecting each other's IOT.

This discussion gives a first indication of the type of characteristics that are likely to define which regions have a comparable IO structure. Size and location are important, but sectoral structure seems to be the single most important factor that determines which IOTs present the best base IOT for the spatial projection of an unknown IOT.

Figures 2 and 3 contain the same information as Table 3, but with the numerically differences added. Figure 2 emphasizes the importance of choosing the right base matrix. In the case of Andalucía and Navarra, for example, there are only two regions that are more of less suitable for a conventional RAS projection, and in the case of Galicia, there is only one. In many cases, picking the third, fourth or fifth best base IOT already leads to estimation errors that are up to 50% larger than that of the best base matrix, which already was not to good to start with (see Table 2). In addition, Figure 3 shows that picking a really wrong base IOT for a conventional RAS projection easily leads to errors that are 50% to 150% larger than those of the best choice.

To conclude, this section indicates that the performance of using RAS as a non-survey technique in the old fashioned way is weak, even when the best choice of base IOT is made. Moreover, not choosing the best base matrix may lead to errors that could be up to three times larger. Finally, it shows that the non-similar part, so-called the "non-fundamental part" of the economy is unfortunately quite large, at least in the case of the Spanish regions studied (see Jensen, Hewings and West (1987) for the concept of fundamental economic structure, and Thakur (2008) for a recent temporal application). Next, we consider if using CRAS can improve the quality of RAS as a spatial projection method.

4. COMPARING RAS, AND CRAS WITH MORE AND MORE REGIONS

By choosing the first-best region in each single projection we give RAS a head start to CRAS, but to not disadvantage CRAS unduly we successively add the second-best, the third-best etc. when comparing CRAS with RAS. To make this comparison we estimate the parameters of the distributions of statistical deviations between the projected and the true IOTs for each of the Spanish regional tables, as indicated in (1). Then we calculate the average and standard deviation corresponding to those statistical deviations, as in (2), and apply that data in (3).

As we observe in Table 4 and Figure 4, for all the regions there are three to nine CRAS combinations that produce errors that are up to 70% smaller than the best RAS projection. The CRAS method with the two, three or four most similar IOTs produces *always* a better projection than the best RAS projection, while CRAS method with the five most similar IOTs is better than the best RAS in nine out of eleven cases, and where CRAS does not improve RAS, it is only 3% and 1% worse.

On the other hand, when still more, increasingly dissimilar IOTs are added, the performance of CRAS deteriorates quite systematically. This is especially the case when the last one or two IOTs are added, which are quite often those of Madrid and the Baleares. Still, with each number of base IOTs there are cases to be found in which CRAS outperforms RAS. The problem of course is how the analyst, not knowing which CRAS combinations perform better, decides on the number and the specific base IOTs to include in a CRAS projection.

The target IOTs of the Baleares and Madrid, and to a lesser extend also those of Castilla-La Mancha and Galicia, are of special interest, as in those four cases almost any CRAS approach, regardless of the number of IOTs used, is better than the best RAS. There is little that these four regions have in common in terms of IO structure, but Figure 4 shows that the spread in their CRAS performance measure is relatively small, while Table 2 shows that the best RAS projection, especially of the Baleares and Madrid, tends to be worse than that of the other regions. So, when one has a bad conventional RAS projection it is easier to improve it by adopting of CRAS, but again when one has a bad conventional RAS projection will not be known beforehand.

Finally, Table 5 summarizes the improvement of the CRAS projection with the best two, three, four and five base IOTs compared to the best RAS projection. We see that there are very relevant improvements to be made in all cases, with a minimum of 51% for Andalucía and a maximum of 72% in the case of Galicia. In all cases, except for Asturias and the Baleares, the largest improvement is reached when the two most similar IOTs are combined in CRAS 2. The main practical problem is of course how to determine these two best regions.

Hence, in conclusion, in our opinion, choosing CRAS with the three to four most similar IOTs probably is the best strategy, as trying to choose the single best IOT for a conventional RAS projection almost certainly leads to a worse result as one easily picks the second or third best instead of the first best, whereas choosing CRAS with the three to four best IOTs, including one or two wrong choices, most certainly produces a better result.

5. CONCLUSIONS

The availability of many different regional or national input-output tables provides the researcher with extra information that may be used to improve the accuracy of any spatial IO projection method. We show that the CRAS method precisely does that by adding cell-specific corrections to RAS, which only uses one single known matrix. The cell corrections of CRAS are determined by minimizing the sum of the squared mean deviations of RAS projections between the multiple known tables, weighted by the inverse of their standard deviation.

In the test of the performance of CRAS relative to RAS with eleven Spanish regional survey IOTs for 1998-2005, it is shown that it is crucial to choose the right IOT table as a base matrix to get a good performance of the RAS method as a spatial IO projection technique. When the wrong IOT is chosen estimation errors may easily be up to three times as large compared to using the right IOT, which already suffers from estimation errors between 25% and 50%.

This sensitivity for the right choice of base table is considerably reduced when CRAS is used with three to four IOTs of regions with an economic structure that is considered to be most similar to the IOT that has to be projected. In that case CRAS may give a reduction in the error of 50% to 70% compared to the RAS projection. Based on our Spanish test, however, the analyst is advised against adding more IOTs to a CRAS projection, as those IOTs may become too different from the IO structure of the target region or country. Only when there are no base IOTs that seem similar to the target IOT, as in our test with Madrid and the Baleares, it might be advisable to use as many distantly comparable IOTs as possible.

REFERENCES

Bacharach, Michael. 1970. *Biproportional Matrices and Input-Output Change*. Cambridge, England: Cambridge University Press.

Bachem, Achim, and Bernhard Korte. 1979. "On the RAS-Algorithm," *Computing*, 23(2), 189–198.

Boomsma, Piet, and Jan Oosterhaven. 1992. "A double-entry method for the construction of bi-regional input-output tables," *Journal of Regional Science,* 32(3), 269–84.

Butterfield, Martin, and Trevor Mules. 1980. "A testing routine for evaluating cell by cell accuracy in short-cut regional input-output tables," *Journal of Regional Science*, 20(3), 293–310.

Czamanski, Stanislaw, and Emil E. Malizia. 1969. "Applicability and Limitations in the Use of National Input-Output Tables for Regional Studies," *Papers of the Regional Science Association* 23(1), 65–78.

Datacomex. 2009. *Spanish Foreign Trade Statistics*. Madrid, Spain: Ministry of Industry, Tourism and Trade (http://datacomex.comercio.es/).

deMesnard, Louis. 2004. "Biproportional methods of structural change analysis: a typological survey," *Economic Systems Research*, 16(2), 205–230.

deMesnard, Louis, and Ronald E. Miller. 2006. "A note on added information in the RAS procedure: Reexamination of some evidence," *Journal of Regional Science*, 46(3), 517–528.

Deming, W. Edwards, and Frederick F. Stephan. 1940. "On a least-squares adjustment of a sampled frequency table when the expected marginal totals are known," *Annals of Mathematical Statistics*, 11(4), 427–444.

Escobedo-Cardeñoso, Fernando, and Jan Oosterhaven. Forthcoming. *Hybrid Interregional Input-Output Construction Methods: Applied to the Seven Region Spanish Input-Output Table*. The Netherlands: University of Groningen.

Hewings, Geoffrey J. D. 1969. "Regional input-output models using national data: The structure of the West Midlands economy," *The Annals of Regional Science*, 3(1), 179–191.

——. 1977. "Evaluating the possibilities for exchanging regional input-output coefficients," *Environment and Planning A*, 9(8), 927–944.

Jensen, Rodney C., Geoffrey J. D. Hewings and Guy R. West. 1987. "On a taxonomy of economies," *Australian Journal of Regional Studies*, 2, 3–24.

Junius, Theo, and Jan Oosterhaven. 2003. "The Solution of Updating or Regionalizing a Matrix with both Positive and Negative Entries," *Economic Systems Research* 15(1), 87–96.

Miller, Ronald E. 1998. "Regional and interregional input-output analysis," in a W. Isard, I. J. Azis, M. P. Drennan, R. E. Miller, S. Saltzman, and E. Thorbecke (eds.), *Methods of Interregional and Regional Analysis*. Brookfield, VE, USA: Ashgate, pp. 41–133.

Mínguez, Roberto, Jan Oosterhaven, and Fernando Escobedo. 2009. "Cell-Corrected RAS Method (CRAS) for Updating or Regionalizing an Input-Output Matrix," *Journal of Regional Science* 49(2), 329–348.

Morrison, William I., and Peter Smith. 1974. "Nonsurvey Input-Output Techniques at the Small Area Level: An Evaluation," *Journal of Regional Science* 14(1), 1–14.

MPT. 2009. *Perfil económico-financiero de las Comunidades Autónomas 2008*. Madrid: Ministerio de Política Territorial, Gobierno de España.

Oosterhaven, Jan, Gerrit Piek, and Dirk Stelder. 1986. "Theory and practice of updating regional versus interregional interindustry tables," *Papers of the Regional Science Association*, 59(1), 57–72.

Sawyer, Charles, and Ronald E. Miller. 1983. "Experiments in Regionalization of a National Input-Output Table," *Environment and Planning A* 15(11), 1501–20.

Schafer, William A., and Kong Chu. 1969. "Non-survey techniques for constructing regional interindustry models," *Papers of the Regional Science Association*, 23(1), 83–104.

Snickars, Folke, and Jörgen W. Weibull. 1977. "A minimum information principle theory and practice," *Regional Science and Urban Economics*, 7(1-2), 137–68.

Stevens, Benjamin H. and Glynnis A. Trainer. 1980. "Error generation in regional input-output analysis and its implications for non-survey models," in a S. Pleeter (ed.) *Economic impact analysis: methodology and applications*. Boston, MA, USA: Martinus Nijhoff, pp. 68–84.

Stone, Richard. 1961. *Input-Output and National Accounts*. Paris: Office of European Economic Cooperation.

Thakur, Sudhir K. 2008. "Identification of temporal fundamental economic structure (FES) of India: An input–output and cross-entropy analysis," *Structural Change and Economic Dynamics,* 19(2), 132–151.

Thumann, R. G. 1978. "A comment on Evaluating the possibility for exchanging regional input-output coefficients", *Environment and Planning A*, 10(3), 321–5.

Tilanus, Christian B. and Henri Theil. 1965. "The information approach to the evaluation of input-output forecasts," *Econometrica*, 33(4), 847–862.

TABLE 1. Layout of the standardized Spanish regional input-output table*

	Intermediate demand and local final demand	Exports to RoS	Exports to RoW	Total output
Own region sectors	$\begin{pmatrix} z_{11}^{R} & \cdots & z_{1n}^{R} & z_{y_{11}}^{R} \cdots \\ \cdots & z_{ij}^{R} & \cdots & ..z_{y_{iq}}^{R} .. \\ z_{m1}^{R} & \cdots & z_{mn}^{R} & \ldots z_{y_{mf}}^{R} \end{pmatrix}$ CRAS	$\begin{pmatrix} e_{1}^{R^E} \\ e_{i}^{R^E} \\ e_{m}^{R^E} \end{pmatrix}$ CRAS	$\begin{pmatrix} e_{1}^{R^M} \\ e_{i}^{R^M} \\ e_{m}^{R^M} \end{pmatrix}$ Given	$\begin{pmatrix} x_{1}^{R} \\ x_{i}^{R} \\ x_{m}^{R} \end{pmatrix}$ Estimation
Rest of Spain sectors	$\begin{pmatrix} p_{11}^{R^E} & \cdots & p_{1n}^{R^E} & p_{y_{11}}^{R^E} .. \\ \cdots & p_{ij}^{R^E} & \cdots & .p_{y_{iq}}^{R^E} . \\ p_{m1}^{R^E} & \cdots & p_{mn}^{R^E} & ..p_{y_{mf}}^{R^E} \end{pmatrix}$ CRAS	0	0	
Rest of World sectors	$\begin{pmatrix} p_{11}^{R^M} & \cdots & p_{1n}^{R^M} & p_{y_{11}}^{R^M} .. \\ \cdots & p_{ij}^{R^M} & \cdots & .p_{y_{iq}}^{R^M} . \\ p_{m1}^{R^M} & \cdots & p_{mn}^{R^M} & ..p_{y_{mf}}^{R^M} \end{pmatrix}$ Estimation	0	0	
Value added	$\left(g_{1}^{R} .. \quad ..g_{j}^{R} . \quad .g_{n}^{R} \quad ..g_{y_q}^{R} ..\right)$ Given	0	0	
Total output	$\left(x_{1}^{R} .. \quad ..x_{j}^{R} . \quad .x_{n}^{R} \quad ..x_{y_q}^{R} ..\right)$ Estimation			

* The meaning of the symbols is: x = total output and total use; z = intra-regional intermediate and final demand; R, S = Spanish regions; m = number of supplying sectors; n = number of purchasing sectors; y = indicator of final demand category, f = number of final demand categories; e = exports; E = Rest of Spain (RoS); M = Rest of the world (RoW); p = imports; g = value added.

TABLE 2. Matrix distance measures of best RAS projection per target IO table

	Target input-output table										
	ada	*ara*	*ast*	*bal*	*clm*	*csl*	*cva*	*gal*	*mad*	*nav*	*pva*
WAPE in %	26.9	29.9	33.8	48.1	42.3	26.0	46.7	33.9	50.3	32.3	36.5
Base IOT	ast	csl	ara	nav	csl	ara	ast	ast	clm	csl	nav
WNSE/1000	166	16	32	212	117	32	176	45	723	19	111
Base IOT	ast	csl	nav	nav	csl	ara	ast	ast	clm	ara	nav
*MIG*1000*	334	304	359	553	405	326	536	373	496	349	386
Base IOT	pva	csl	ada	nav	ada	nav	pva	ast	pva	csl	ada

TABLE 3. Rank order of RAS projections, average of three matrix distance measures

		Target input-output table										
		ada	*ara*	*ast*	*bal*	*clm*	*csl*	*cva*	*gal*	*mad*	*nav*	*pva*
Base matrix accuracy ranking	*1st*	ast	csl	nav	nav	csl	ara	ast	ast	clm	csl	nav
	2nd	pva	nav	csl	mad	nav	nav	clm	pva	ast	ara	ada
	3rd	nav	ast	ara	ast	ast	ast	pva	nav	pva	ast	ast
	4th	csl	pva	pva	clm	ada	clm	ara	csl	bal	pva	ara
	5th	ara	clm	ada	ada	pva	gal	nav	ada	ara	clm	csl
	6th	clm	gal	gal	ara	bal	ada	ada	ara	nav	gal	clm
	7th	gal	cva	cva	cva	ara	pva	gal	clm	csl	cva	cva
	8th	cva	ada	clm	csl	gal	cva	csl	cva	ada	ada	gal
	9th	bal	bal	mad	gal	cva	bal	bal	bal	cva	mad	mad
	10th	mad	mad	bal	pva	mad	mad	mad	mad	gal	bal	bal

TABLE 4. Normalized performance of best RAS, and best CRAS by number of regions

	Target input-output table										
	ada	*ara*	*ast*	*bal*	*clm*	*csl*	*cva*	*gal*	*mad*	*nav*	*pva*
*Best RAS**	1.02	1.00	1.13	1.00	1.07	1.00	1.03	1.00	1.03	1.02	1.09
CRAS 2	0.50	0.39	0.45	0.51	0.31	0.35	0.44	0.28	0.38	0.30	0.43
CRAS 3	0.78	0.69	0.42	0.44	0.43	0.78	0.54	0.43	0.40	0.61	0.52
CRAS 4	0.87	0.73	0.54	0.53	0.66	0.96	0.97	0.42	0.79	0.81	0.70
CRAS 5	1.01	1.03	0.76	0.46	0.70	1.01	0.99	0.43	0.81	1.00	0.82
CRAS 6	1.17	1.16	0.73	0.76	0.79	2.31	1.24	0.56	0.85	1.24	1.14
CRAS 7	1.32	1.20	0.99	0.80	0.81	1.95	1.34	0.63	0.81	2.69	1.32
CRAS 8	1.36	1.33	1.53	0.72	0.82	2.58	1.33	0.83	0.86	3.35	1.48
CRAS 9	1.32	1.37	1.75	0.70	0.84	2.46	1.38	0.93	0.95	5.66	1.89
CRAS 10	1.93	2.21	1.76	0.78	1.51	3.74	1.34	1.47	1.09	5.91	1.89

* The best RAS may not be equal to 1, as it is the average over three matrix distance measures, which are not always normalized with regard to the same base region (see Table 2).
Best overall approach. CRAS outperforms RAS.

TABLE 5. Performance of the CRAS approach in relation to the best RAS approach*

Region	ada	ara	ast	bal	clm	csl	cva	gal	mad	nav	pva
CRAS 2	51%	61%	60%	49%	71%	65%	57%	72%	63%	70%	60%
CRAS 3	24%	31%	63%	56%	59%	22%	47%	57%	61%	40%	52%
CRAS 4	14%	27%	52%	47%	38%	5%	6%	58%	23%	20%	36%
CRAS 5	1%	-3%	23%	54%	35%	-1%	4%	57%	21%	2%	25%

* $(1 – ANM^{CRAS} / ANM^{RAS})$*100%.

FIGURE 1. Spanish regions with a regional survey IO table (italics, year in legend).

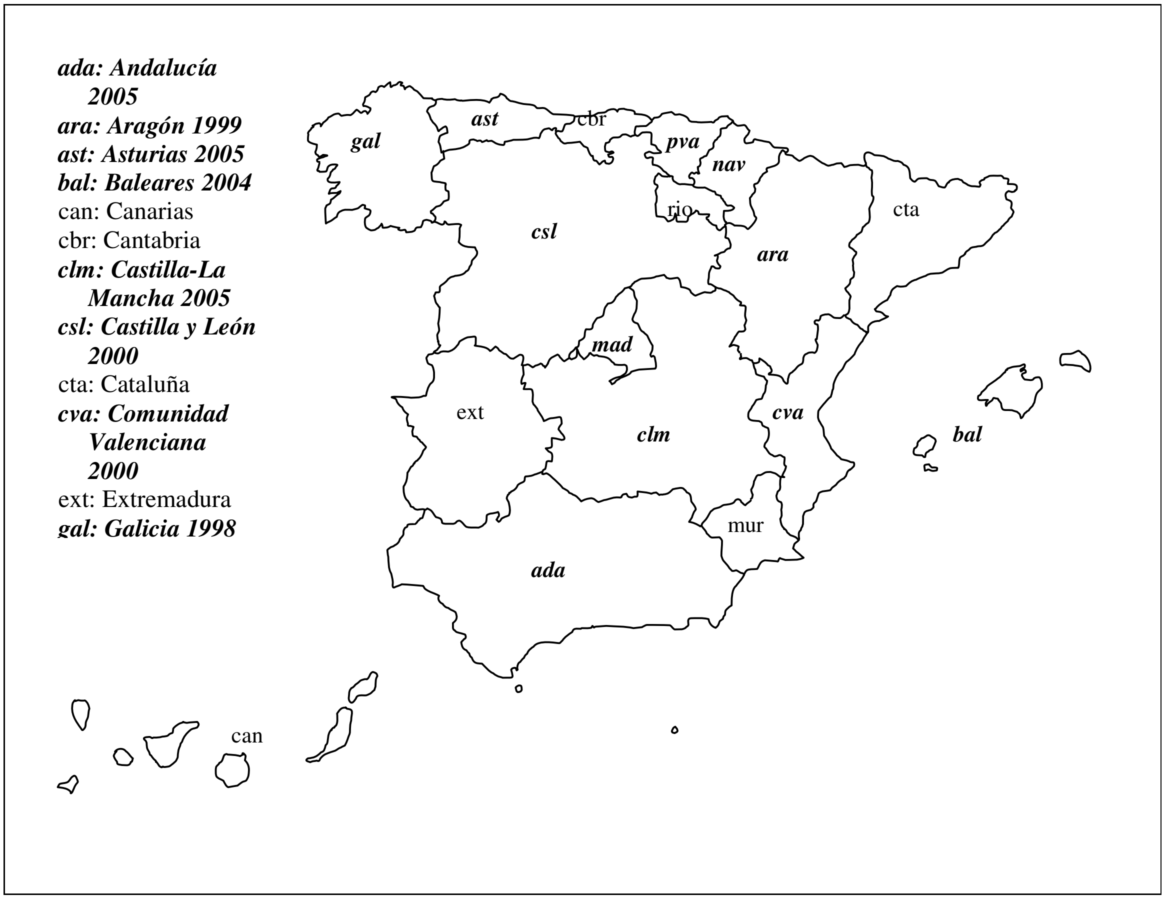


FIGURE 2. Best five RAS projections of the eleven Spanish regional survey IO tables.

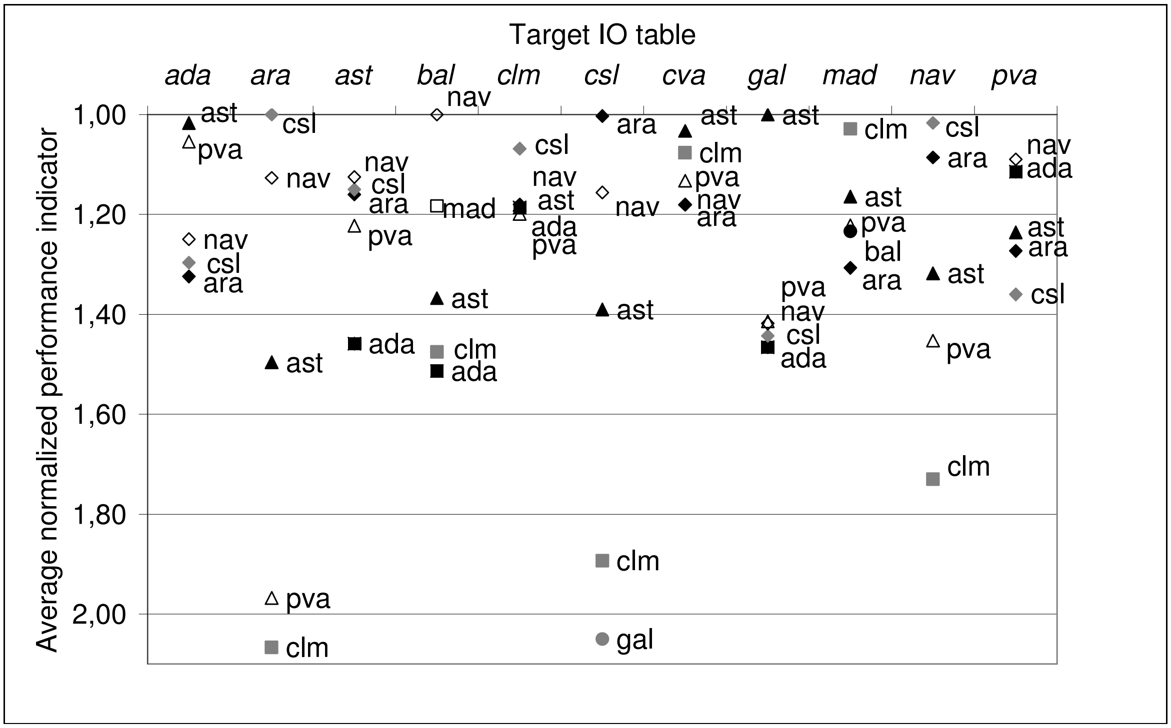


FIGURE 3. Worst five RAS projections of the eleven Spanish regional survey IO tables.

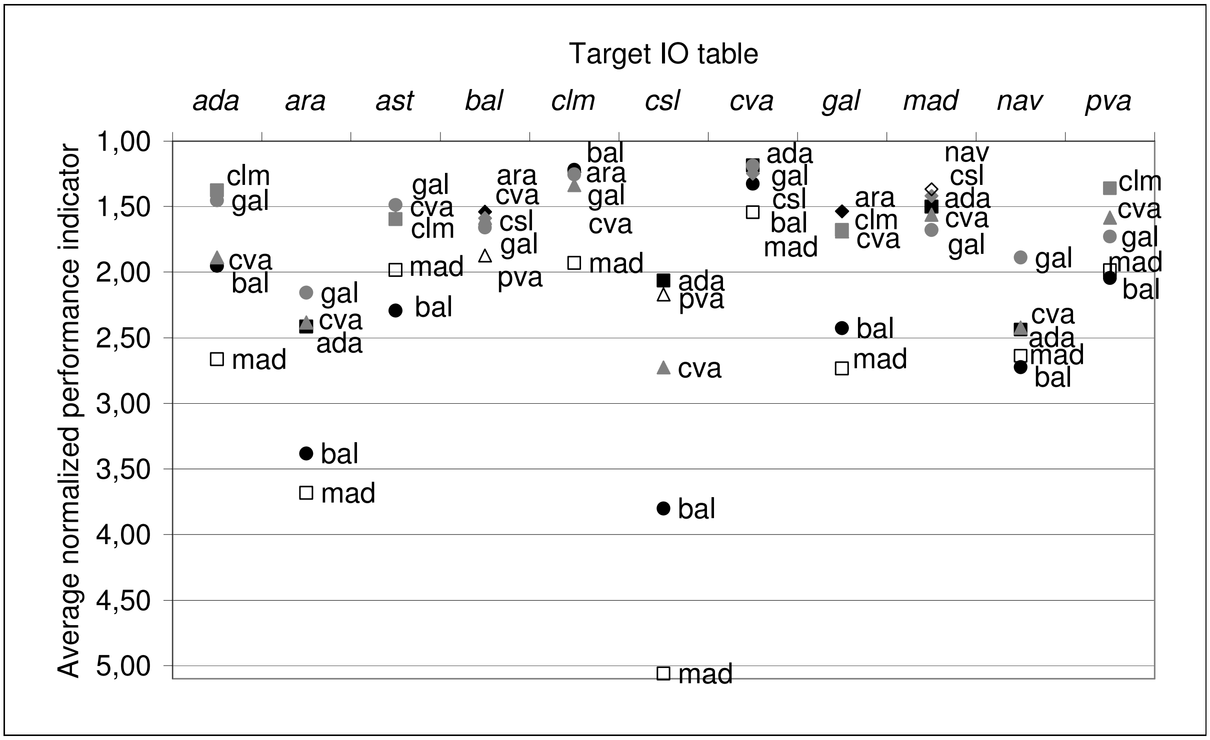
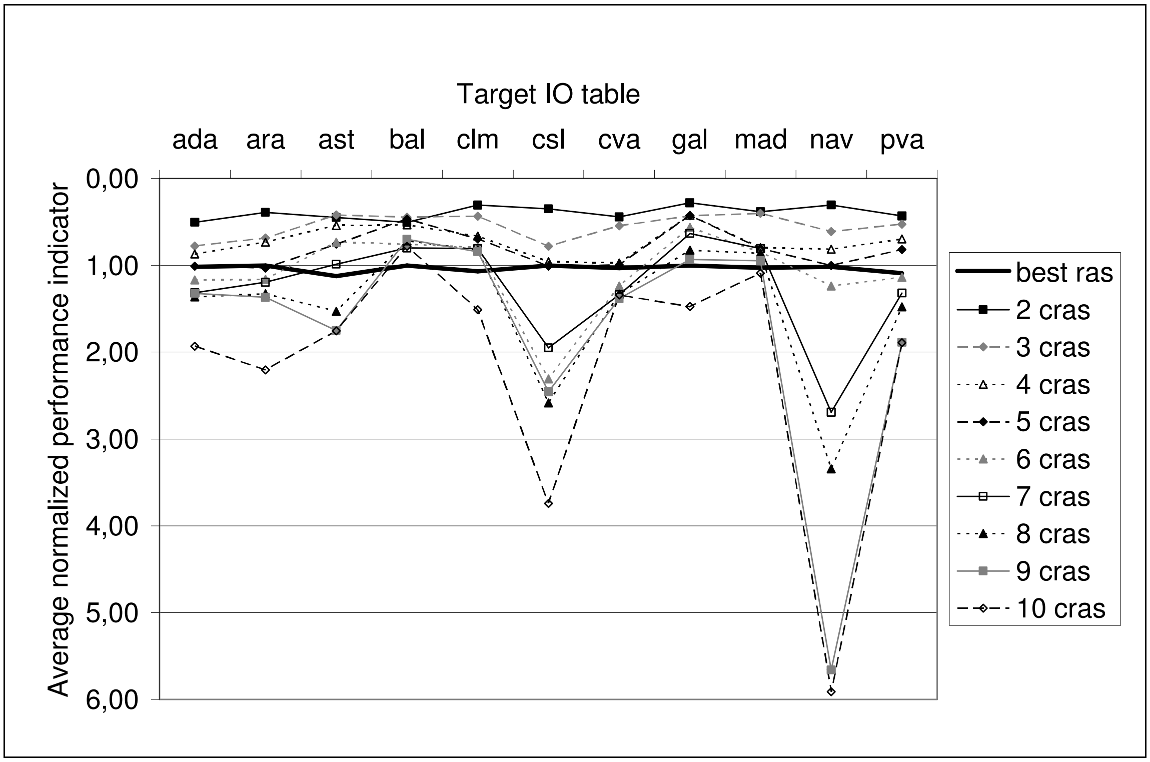


FIGURE 4. Projection error of RAS versus CRAS with more and more, less alike regions.



Footnotes:

[1] Faculty of Economics and Business, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands. E-mail: j.oosterhaven@rug.nl.
[2] School of Civil Engineering and Planning, University of Castilla-La Mancha, Avenida de Camilo Jose Cela 2, 13071 Ciudad Real, Spain. E-mail: fernando.escobedo@uclm.es.
[3] Meanwhile it has been proved that the old iterative solution to the RAS updating problem (Stone, 1961) is equivalent to solving the non-linear minimization of information gain (*MIG*) (Bacharach, 1970, Snickars and Weibull, 1977, Bachem and Korte, 1979). See Junius and Oosterhaven (2003) for a Generalized RAS (GRAS) algorithm, which also covers cases with negative cells and negative constraints.
[4] As a spatial projection technique, CRAS may also be tested on e.g. the harmonized set of European IOTs (Eurostat, 2002).
[5] There are two more Spanish regions with IO data, Cataluña and Canarias, but they only have a use table. The application of CRAS to Cataluña and Canarias may therefore be more accurate than that for the other four regions.
[6] The idea of minimizing the distance between a known matrix and the target is logical as no further information is assumed to be available (Miller, 1998).
[7] This section is copied from Mínguez, Oosterhaven and Escobedo (2009). We have deleted the superscript $R(S)$ to simplify the notation.
[8] In practice, this should be done at the most disaggregate level possible.
[9] This is precisely the reason why using a national IOT to construct a regional IOT can not be done with any measure of accuracy, as the national IOT does not contain any information on interregional trade.

[10] We use the symmetric IO tables of 11 regions in current prices with 30 sectors, all of them from the period 1998-2005. Note that all RAS solutions are obtained within this absolute error tolerance: $\frac{\sum_i \left| z_{i\bullet}^R - u_i \right| + \sum_j \left| z_{\bullet j}^R - v_j \right|}{2} < 0.0001$.

[11] These separate results are available upon request with the second author.