CELL-CORRECTED RAS (CRAS) AS A REGIONAL INPUT-OUTPUT CONSTRUCTION TECHNIQUE

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

Thursday, June 11, 2009

Abstract

The RAS method is used to update or regionalize a single matrix such that it conforms to new row and column totals. This paper presents a cell-correction of the RAS method (CRAS) that uses cell variation distributions calculated from input-output tables of different regions as a non-survey estimation technique for single region input-output tables. After solution of the regular RAS method, an additional optimization problem based on first order reliability methods (FORM) is solved, producing the most likely cell-corrections to the regular RAS solution. The advantage of the proposed formulation is its simplicity, which allows solving the optimization problem by means of an efficient iterative scheme. To test the behavior of the CRAS cumulative simulations are made with eleven symmetric input-output tables of Spanish regions for 1998-2005, harmonized to thirty economic sectors.

Keywords: Regional input-output analysis, Non-survey methods, RAS, Spanish regions.

1. Introduction

RAS is known as an iterative technique to update input-output (IO) tables, given an old table and new row and column totals (Stone, 1961, Bacharach, 1970). It is also used to construct regional input-output tables given a national IO table 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). All these applications have in common that one single (old) matrix is given to which minimal information is added such that it satisfies some set of (new) constraints.[3]

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 IO tables 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 IO tables 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 most recent single table when sudden shocks, such as the oil price rises of 1973-74 and 1979-80, need to be covered.

---

[1] School of Civil Engineering and Planning, University of Castilla-La Mancha, Avenida de Camilo Jose Cela 2, 13071 Ciudad Real, Spain. Email: fernando.escobedo@uclm.es.

[2] Faculty of Economics and Business, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands. Email: j.oosterhaven@rug.nl.

[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 from information theory (Bacharach, 1970, Snickars and Weibull, 1977, Bachem and Korte, 1979).

Here we will test how CRAS performs as a non-survey estimation technique for constructing regional input-output tables (IOTs). It should be noted that the temporal projection of IOTs is far simpler than the spatial projection of IOTs that 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, the 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 most close by physical space. In stead it is the IOT of the region most closed by in terms of IO structure, but which region that is not clear beforehand, as will be shown in the application.

To test CRAS as a spatial projection method we need a series of identically defined, survey-based IOTs of more or less comparable regions. 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 for the obvious reasons. This practically restricts the choice to a set of regions within one and the same country. We will use the set of single region survey-based symmetric IOTs for 11 Spanish regions as collected and harmonized for the construction of the seven region interregional semi-survey IOT for Spain for 2000 (Escobedo and Oosterhaven, 2009). As Spain has 17 regions, it follows that 6 Spanish regions do not have a survey-based symmetric IO table. 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.[4]

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 regional non-survey IOT construction method. Section 3 will discuss the setup of the test on the existing 11 regional symmetric 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 symmetric IOTs with their non-survey estimates, each based on application of CRAS to increasing amounts more and more different regional survey IOTs for the remaining regions. Section 5 concludes that CRAS performs better than RAS when a limited set of survey regional IOTs is used that are close to the IOT that has to be projected. When more IOTs are added of regions with more different IO structures CRAS leads to worse results than using RAS on the region that is most close by in IO structure terms.

2. The Cell-Corrected RAS method (CRAS)

The goal of conventional regionalizing methods consists on obtaining an input-output transactions matrix Z^R for region R of dimension m by n as close as possible to the input-output transactions matrix Z^S from region or nation S of the same dimension, knowing only the margins (the row and column sums) of the target matrix.

2.1 Statement of the programming model

The proposed method has two stages.

[4] 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 4 regions.

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, $e \approx N(\mu^e, \sigma^e)$:

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

Where $e_{ij}^{R(S)}$ is a stochastic term representing the unexplained deviation if we use S as base matrix to calculate matrix R of by means of RAS; $\tilde{z}^{R(S)}$ is the RAS projection of region R taking S as a base region; T is the number of regions with an available IO table.

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

(2a) $$\mu_{ij}^{e^R} = \frac{\sum_{s=1, s\neq R}^{T} e_{ij}^{R(s)}}{T-1}$$ and

(2b) $$\sigma_{ij}^{e^R} = \sqrt{\frac{\sum_{s=1, s\neq R}^{T} \left(e_{ij}^{R(s)} - \mu_{ij}^{e^R}\right)^2}{T-2}}$$

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

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

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

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

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

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

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

2.2 Solution of the programming model [5]

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

[5] In this paragraph we skip the superscript $R(S)$ to simplify the notation.

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

(10) $$\sum_{j=1}^{n}e_{ij}\tilde{z}_{ij}-u_{i}=0;i=1...m$$

(11) $$\sum_{i=1}^{m}e_{ij}\tilde{z}_{ij}-v_{j}=0;j=1...n$$

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 sum and column sums of the target matrix, respectively.

For the system of Equations (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 $\sigma_{ij}\neq 0$, $\forall ij$, while 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. So (12) must be generated eliminating one constraint in (10)-(11), no matter which.

3. Testing CRAS as a non-survey regional IOT method

Next, we discuss how applying CRAS to the 11 Spanish survey-based regional IOTs has to be set up in order to test CRAS as a non-survey regional IOT construction 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 non-survey technique (section 3.1). Second, a solution has to be found for defining the structural IO distance between the regions at hand, in order to determine how the performance of CRAS changes when the information of more and more regional IOTs is used.

3.1 Setting-up CRAS as a non-survey IO construction technique

When we assume that transit trade is zero, 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 set up a situation that resembles that of the non-survey estimation of the lacking 6 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 6 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.

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	$\begin{pmatrix} z_{11}^R & \dots & z_{1n}^R & z_{y_{11}}^R \dots \\ \dots & z_{ij}^R & \dots & ..z_{y_{iq}}^R .. \\ z_{m1}^R & \dots & z_{mn}^R & \dots z_{y_{mf}}^R \end{pmatrix}$ 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	$\begin{pmatrix} p_{11}^{R^E} & \dots & p_{1n}^{R^E} & p_{y_{11}}^{R^E} .. \\ \dots & p_{ij}^{R^E} & \dots & .p_{y_{iq}}^{R^E} . \\ p_{m1}^{R^E} & \dots & p_{mn}^{R^E} & ..p_{y_{mf}}^{R^E} \end{pmatrix}$ CRAS	0	0	
Rest of World sectors	$\begin{pmatrix} p_{11}^{R^M} & \dots & p_{1n}^{R^M} & p_{y_{11}}^{R^M} .. \\ \dots & p_{ij}^{R^M} & \dots & .p_{y_{iq}}^{R^M} . \\ p_{m1}^{R^M} & \dots & p_{mn}^{R^M} & ..p_{y_{mf}}^{R^M} \end{pmatrix}$ Estimation	0	0	
Value added	$\begin{pmatrix} g_1^R .. & ..g_j^R . & .g_n^R & ..g_{y_q}^R .. \end{pmatrix}$ Given	0	0	
Total output	$\begin{pmatrix} x_1^R .. & ..x_j^R . & .x_n^R & ..x_{y_q}^R .. \end{pmatrix}$ 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.

The arguments for making this selection are as follows. We, of course, 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, and 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 of the intra-regional purchases and sales per regional sector were known a priori. 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), that have been developed for the difficult estimation of precisely these intra-regional totals (see also Thuman, 1978).

For the 6 Spanish regions without an IOT 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. 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 6 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. Hence, it has to be assumed 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, for reasons to be given shortly.

For the eleventh 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 intermediate and local final demand satisfied by products from the whole of Spain can be calculated simply from the 'given' and the 'estimated' non-IO 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 own sectors intermediate and final sales to the whole of Spain are also calculated simply from the 'given' and the 'estimated' non-IO 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 row totals of the matrix of imports from RoS and the column total for the exports to RoS in Table 1, without using the info from the survey IOTs.[6] We see only one reasonable solution, namely to estimate only a *single row* of imports from RoS instead of a full matrix. When we estimate region *R* using the data of region *S*, we use region *S* its RoS import ratio (η^S) and region *S* its RoS export ratio (ε^S) to estimate the total intra-regional transactions for region *R(S)*. These two aggregate domestic trade coefficients are calculated from the survey IOT of region *S*, as follows:

(17) $$\eta^S = \frac{z_{\bullet\bullet}^{S^E}}{z_{\bullet\bullet}^{S} + z_{\bullet\bullet}^{S^E}} \text{ and } \varepsilon^S = \frac{e_{\bullet}^{S^E}}{z_{\bullet\bullet}^{S} + e_{\bullet}^{S^E}}$$

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

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

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

Where the intra-regional transaction total is calculated as the unweighted average of the estimates with the aggregate export coefficient and the aggregate import coefficient of region *S*:

(20) $$z_{\bullet\bullet}^{R(S)} = \frac{\left(1-\eta^S\right)\left(z_{\bullet\bullet}^{R} + z_{\bullet\bullet}^{R^E}\right) + \left(1-\varepsilon^S\right)\left(z_{\bullet\bullet}^{R} + e_{\bullet}^{R^E}\right)}{2}$$

The calculations (15)-(20) will be made 110 times, i.e. for each of the 11 applications of CRAS we have survey IOTs for the 10 remaining regions *S* available,[7] which will be used to make ten different non-survey RAS estimates for each of the 11 regions *R*.[8]

Subsets of these ten RAS estimates for each *R* will be used to calculate the average deviation and the standard deviation of (2), which will then be used in the second step of CRAS to produce the cell-corrected estimate of CRAS according to (7).

3.2 Accuracy of different RAS estimates for the Spanish regional IO tables

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 simply compared by adding more and more less recent IOTs, to calculate the performance of both methods (see Minguez et al. 2009). In spatial projections this is far more complicated. 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 regions 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

[6] 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 at all.

[7] We use the symmetric IO tables of 11 regions in current prices with 30 sectors, all of them from the period 1998-2005.

[8] Note that all RAS solutions are obtained within an error tolerance of 1×10^{-6}.

parts of the IOTs, thus excluding estimates of the trade with the RoS. We could also compare the RoS results separately, but the comparing the intra-regional part is far more important as its estimation errors determine the estimation errors of the regional multipliers for which regional IOTs are used most.

The comparisons are made by inspecting the distance between the projection z and the true value z^{true}, using different matrix distance measures (deMesnard and Miller, 2006). We focus on additive norms $\left\| z - z^{true} \right\|$, as using multiplicative norms $\left\| \ln(z) - \ln(z^{true}) \right\|$ does not change their basic properties (deMesnard, 2004). We will use the following norms:

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

(24)
$$MAPE = \frac{1}{m \times n} \sum_i \sum_j \frac{\left| z_{ij} - z_{ij}^{true} \right|}{\left| z_{ij}^{true} \right|} \times 100\%$$

- Weighted Absolute Percentage Error, given by:

(25)
$$WAPE = \frac{\sum_i \sum_j \left| z_{ij} - z_{ij}^{true} \right|}{\sum_k \sum_l z_{kl}^{true}} \times 100\%$$

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

(26)
$$NSE = \sum_i \sum_j \frac{\left(z_{ij} - z_{ij}^{true} \right)^2}{z_{ij}^{true}}$$

- Weighted Normalized Squared Error, given by:

(27)
$$WNSE = \frac{\sum_i \sum_j \left(z_{ij} - z_{ij}^{true} \right)^2}{\sum_k \sum_l z_{kl}^{true}}$$

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

(28)
$$MIG = \sum_i \sum_j \left| z_{ij}^{true} \ln\left(\frac{z_{ij}}{z_{ij}^{true}} \right) \right|$$

Once the norms are calculated for both methods, the comparison is made with the following formula:

(29)
$$c_p = \frac{\tilde{n} - \hat{n}}{\tilde{n}} \times 100\%$$

where $\tilde{n}$ and $\hat{n}$ are the norms obtained when using the RAS and CRAS method, respectively, and c_p is the performance comparison parameter that gives the percentage difference between the CRAS and the RAS method. Positive values of c_p imply a better performance of CRAS.

From the raw Spanish regional IO tables we calculate eleven harmonized IOTs and the 11x10 sets of u and v vectors (row and column totals of the RAS part). Then we estimate each of the 11 non-survey regional IOTs by using each of the remaining 10 regional IOTs as the base matrix for a regular RAS estimate. The results are ordered by the size of their estimation errors, as shown in Table 2 for the WAPE norm, which we consider to be the individually most relevant distance measure. Figure 2 shows the unweighted average of all five norms used. See Figure 1 for the location of the Spanish regions.

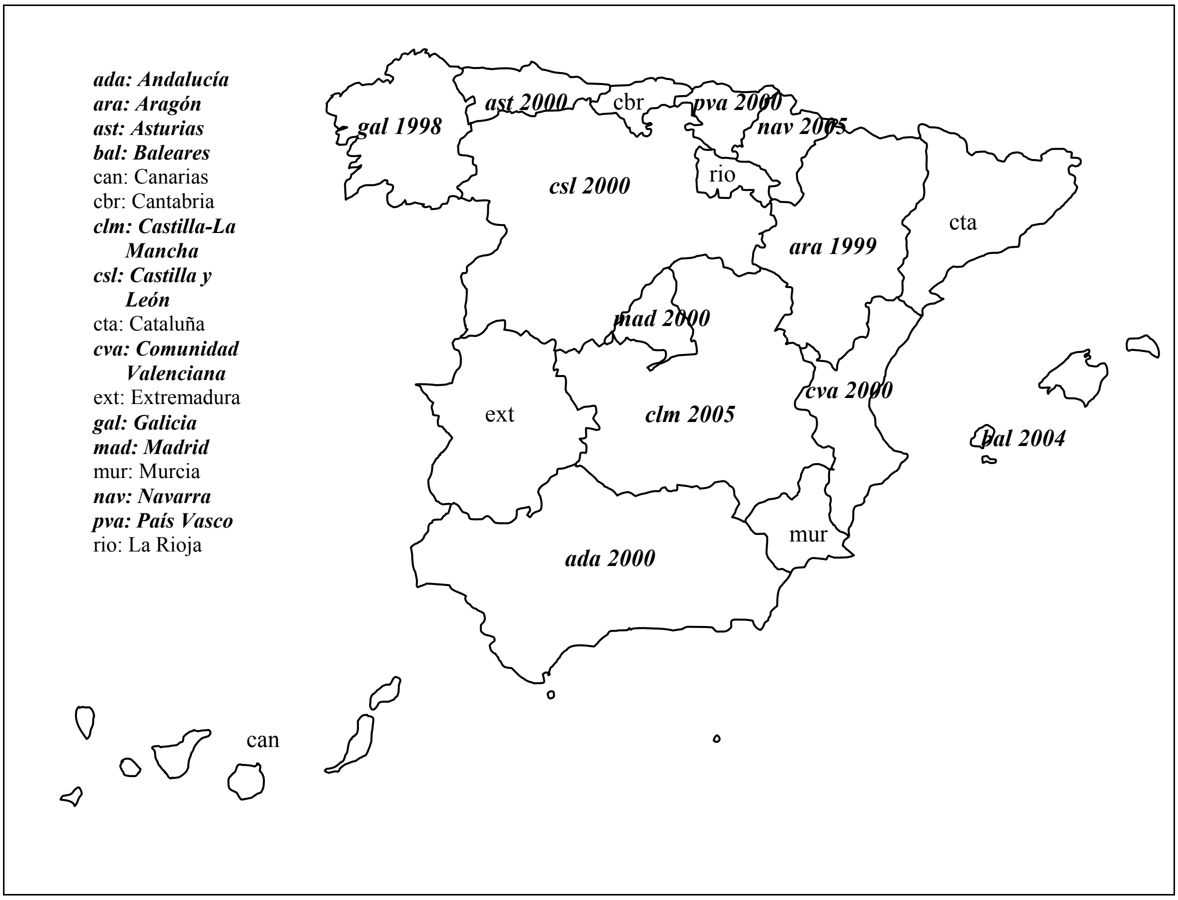

Figure 1. Spanish regions with an regional symmetric IO table are bold and in italics.

In both Table 2 and Figure 2 we note that Aragón and Castilla y León are the two regions that in general give the most accurate RAS-estimate for the other ten regions. This may be due to their economic structure, which is not strongly specialized in specific sectors. On the other extreme we see that Balearic Islands in general offers the worst base matrix, due to the fact that this inland economy is very different from the rest of the Spanish regional economies. The capital region of Madrid also proves to be bad starting point for spatial extrapolation. This clearly indicates that the performance of using RAS as a non-survey technique in the old fashioned way very strongly depends on the right choice of the base matrix.

Table 2. Rank order of WAPE difference between original tables and RAS estimates

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

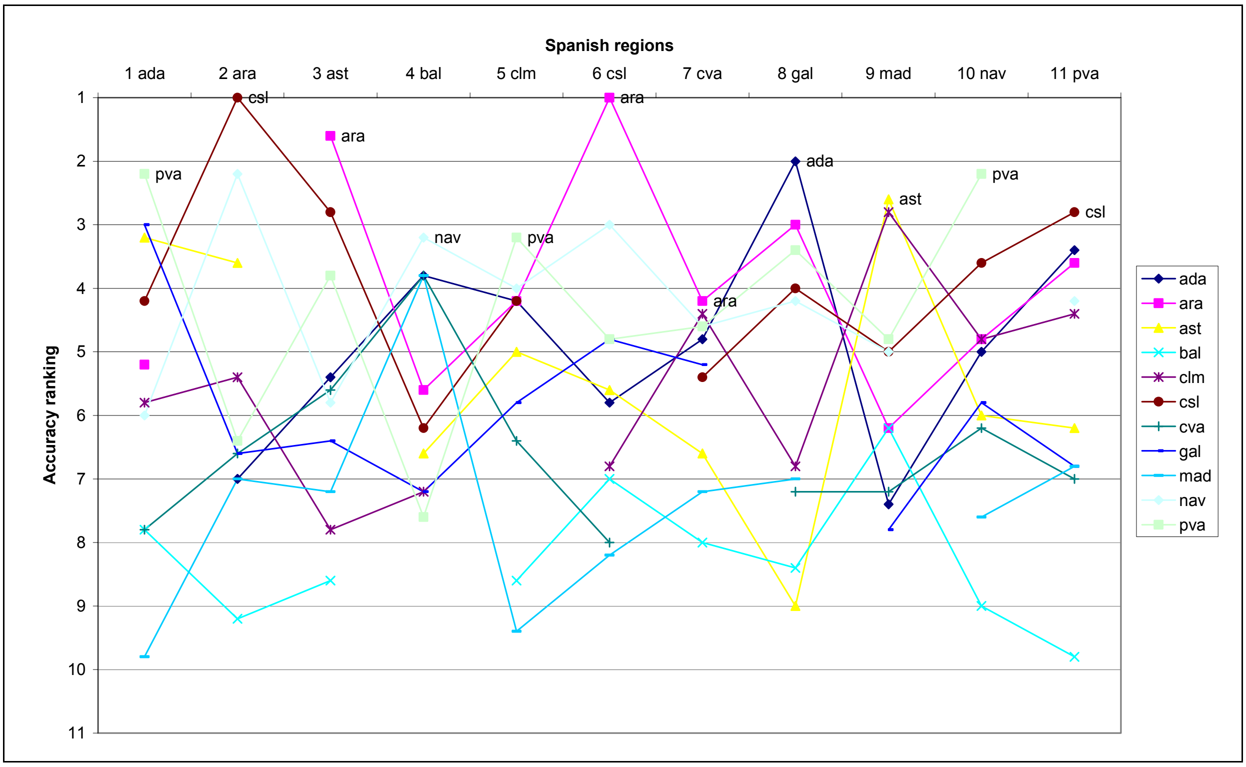

FIGURE 2. Average of five norm ranking of the regional IOTs as base matrices for RAS.

By choosing the first-best region in each single projection we will give RAS a head start to CRAS, but to not disadvantage CRAS unduly we will successively add the second-best, the third-best etc. when comparing CRAS with RAS.

4. Results of comparing RAS with CRAS with only the best five regions

Below follow preliminary results, which compare only the best five regional IOT when applying CRAs with the single best regional IOT in case of RAS. To make this comparison we estimate the parameters of the distributions of statistical deviations between the projected and the true IO tables for each of the regional Spanish tables when they are used as goal matrix in the RAS projection, as it is indicated in Equation 1. Then we calculate the average and standard deviation corresponding to those statistical deviations as in Equations (2a) and (2b).

After this we compare the performance of RAS method when we use the five closest regions to calculate the goal region, where we have 5x11 estimates, with the performance of CRAS method when we do just the same routine but with the difference that when we use a certain regional IO table as a base matrix we modify the RAS solution with the statistical deviations associated to the objective table.

If we look in detail at the minimum information gain ratio (MIG) ratio in Table 4, based on Equation (29), we see that in the 55 cases only five present a worse CRAS result than the RAS one. We can also see that regions as Aragón and Castilla y León sometimes perform worse with CRAS than with RAS, because they do well with RAS. On the other hand a region as Madrid, and specially Islas Baleares, improves around a 50% as an average when CRAS is used to calculate those regional IO tables.

Table 4. MIG ratio (%) for regional forecasts with information of 5 closest IO tables

		Target regional table										
		ada	ara	ast	bal	clm	csl	cva	gal	mad	nav	pva
Base matrix	ada		-1%			-9%	3%	19%				
	ara	19%				54%	-11%	48%		63%	31%	
	ast				54%				27%		39%	11%
	bal											
	clm	1%	18%				28%	35%		29%		
	csl	14%	-31%	16%		31%		47%	23%	62%	11%	
	cva	17%		47%	36%	26%	35%		41%	32%		8%
	gal			35%	60%						40%	8%
	mad		49%		44%							26%
	nav	22%	-1%	37%		36%	19%		37%	56%		2%
	pva			14%	52%			17%	24%		9%	

Finally we discuss the position of the CRAS estimates in the rank order with respect to the five competing RAS estimates, with special attention to the question whether an a priori choice of the best RAS-base region could have been made in case CRAS does not give the single best result. As we can see in Table 5 this is clearly the case in the relation between the regions of Aragón and Castilla y León, where the RAS approach is the best. This may be due to the fact that Castilla y León and Aragón are similar regions as they are inland, relatively big in area and with similar economic structure. In fact, Aragón is the closest economic structure out of the ten regions to Castilla y León.

Table 5. Best base matrix according to the different performance indicators.

	Objective regional table										
	ada	*ara*	*ast*	*bal*	*clm*	*csl*	*cva*	*gal*	*mad*	*nav*	*pva*
MAPE	ara*	ada*	nav*	gal*	ara*	ara	ara*	pva*	ara*	ara*	ast
WAPE	csl	csl	csl	ast*	ara*	ara	clm	nav*	ara*	ara*	nav
NSD	nav	csl	gal*	cva	ara	ara	ara*	pva	ara*	pva	gal
WNSD	csl	csl	csl	ast*	ara*	ara	clm	csl*	ara*	ara	nav
MIC	csl*	csl	gal*	gal*	ara*	ara	csl*	pva*	csl*	ara*	nav*

Note: without * = RAS method; with * = CRAS method.

A special comment deserves the Figure 3 where we can see that in eight out of eleven occasions a CRAS estimate was ranked first as an average. There were three exceptions; two are relative to the above depicted relation between Castilla y León and Aragón, where each RAS estimate is the best approach for each other. The third and last exception deals with the case of País Vasco, where the best estimate is the RAS approach with base matrix as Navarra. This may be due to País Vasco is a very specific region in Spain and the closest economic structure to this region is Navarra.

In the eight cases out of eleven where one of the CRAS estimates was ranked first as an average the base matrix was a region which has a close economic structure to the region we want to approximate. In five occasions, Andalucía, Castilla-La Mancha, Comunidad Valenciana, Madrid and Navarra, the CRAS estimate with Aragón as base matrix was the best approach, what involves that a relatively medium-large region like Aragón, with a balanced economic structure which has no weak sectors and geographically central and with good transport infrastructures with the main regions, may be used as a base matrix with CRAS method to calculate other regions which are large (Andalucía, Castilla-La Mancha, Comunidad Valenciana) or small and developed (Madrid and Navarra). The other three remaining regions (Asturias, Galicia and Baleares), which are coastal and small had a best estimate with a CRAS approach based on a similar region.

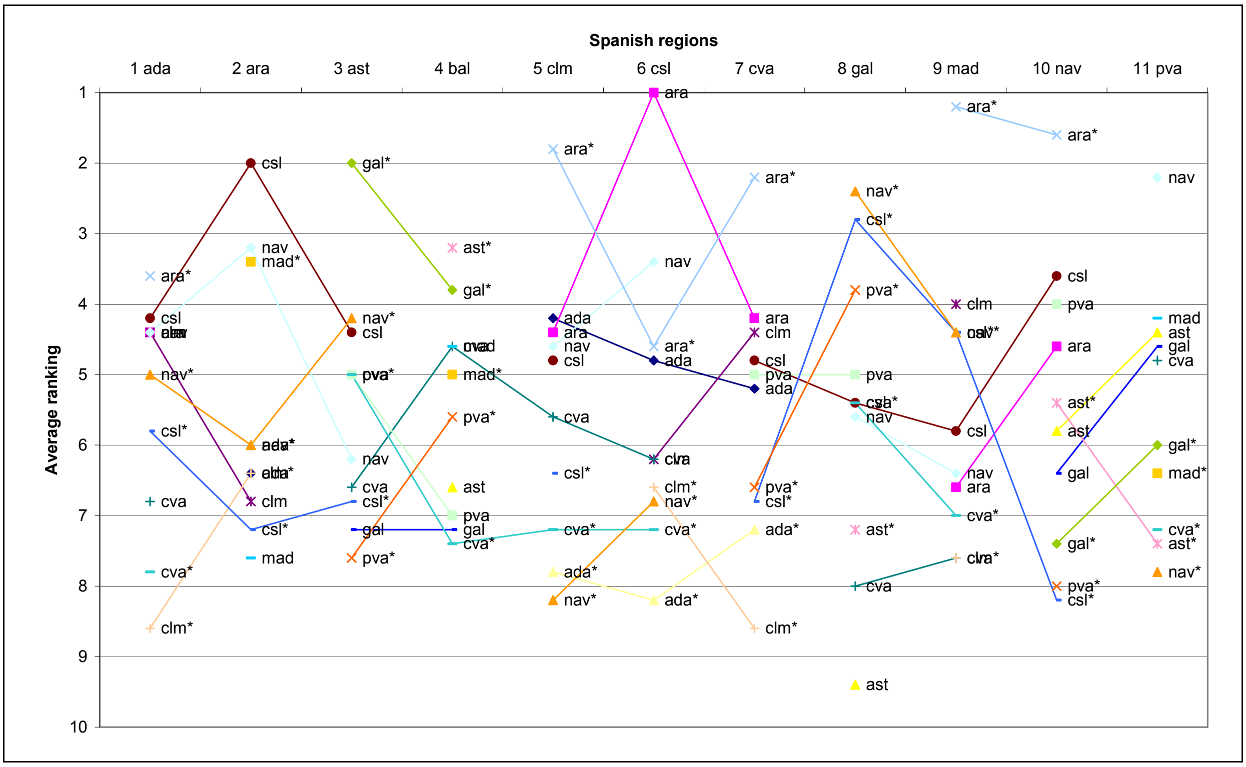

FIGURE 3. Average five norm ranking of the regional IOTs as base matrices for RAS and CRAS.

4. Preliminary conclusions

The availability of many different regional input-output tables provides the researcher with extra information that may be used to improve the accuracy of the RAS method to construct regional IO tables. We use that information by means of the CRAS method, that adds cell-specific corrections to RAS, a biproportional method which uses only one single known matrix. CRAS, however, uses all matrices that are considered relevant for the projection problem at hand, in this case, the five closest, in terms of economic structure, regional tables of a country. The cell corrections of CRAS are determined by minimizing the sum of the squared mean deviations of RAS projections between the multiple known tables, in time or space, weighted by the inverse of their standard deviation.

The RAS method is the best option when we want to obtain a regional table and we have another regional table very close in terms of economic structure. Nevertheless, if we do not have a single clearly close regional table but we have a set of regional tables more or less close to the regional table we want to approximate, CRAS outperforms RAS as the best estimate, as a CRAS-obtained matrix is based on the closest regional economic structure out of that set of five regions. Specifically in the case of regions with general characteristics and no singularities (small size, specific main economic sectors, coast, etc.) a CRAS estimate based on a medium-large region with a balanced economic structure and a relatively central position with good transport infrastructures may be the best approach.

Acknowledgements

We thank Roberto Mínguez for his contribution to developing the CRAS method (Mínguez et al. 2009).

References

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

Bachem, Achim, and Bernhard Korte. 1979. "On the RAS-Algorith," *Computing*, 23, 189–198.

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.

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, Fernando and Jan Oosterhaven. 2009. *Hybrid Interregional Input-Output Construction Methods: Applied to the Seven Region Spanish Input-Output Table*. Mimeo, University of Groningen, The Netherlands.

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.

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

Jackson, Randall and Alan Murray. 2004. "Alternative input-output matrix updating formulations," *Economic Systems Research*, 16(2), 135–148.

Miller, Ronald E. 1998. "Regional and interregional input-output analysis," in 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.

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.

Schafer, W.A. and K. Chu, 1969. "Non-survey techniques for constructing regional interindustry models," *Papers of the Regional Science Association*, 23, 83-101.

Snickars, Folke, and Jorgen W. Weibull. 1977. "A minimum information principle theory and practice," *Regional Science and Urban Economics*, 7, 137–68.

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

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

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