Cell-Corrected RAS Method (CRAS) for Updating or Regionalizing a Matrix

R. Mínguez[1], J. Oosterhaven[2] and F. Escobedo[3]

Abstract

This paper presents a correction of the RAS method (CRAS) at the cell level of its solution. It incorporates data on the temporal or spatial variation in the cells of the known matrix that has to be updated or regionalized. Cell variation distributions are calculated from past input-output tables using simple formulas. After the solution of the regular RAS method is obtained, 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 to solve the optimization problem by means of an efficient iterative scheme. To test the behavior CRAS several simulations with a consistent time series of input-output tables for The Netherlands for 1968-1986 are made. They show that - in situations of structural change - applying CRAS improves the regular RAS estimate.

Key Words: RAS, biproportional adjustment, input-output analysis, structural change.

1 Introduction

The RAS method was designed by Stone (1961), Stone and Brown (1962) to solve the problem of updating a given old input-output (IO) matrix to a new required matrix for which only the row and column totals are given. Hewings (1969) showed that RAS could also be applied to the problem of regionalizing a known national matrix. Oosterhaven et al. (1986) combined both ideas to solve the problem of updating an interregional IO matrix such that it also satisfies the cell restrictions derived from a new national matrix. This is just a more systematic example of many other applications in which all kind of additional ad hoc data are available for the new matrix. All of these applications, however, stick to the principle of adding minimum information to the structure of one single known matrix. The variation only relates to adding different restrictions to the new unknown matrix, it does not relate to situations in which multiple old matrices are available, e.g.

[1]Department of Applied Mathematics, University of Castilla-La Mancha, Avenida de Camilo Jose Cela 2, 13071 Ciudad Real, Spain, email: roberto.minguez@uclm.es.

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

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

for many different years or for many different regions or countries. In this paper, we exclusively consider the situation of having multiple old IO matrices.

The somewhat related problem of having a single old matrix with a series of row and column totals for later years is studied by Omar (1967), who confirms the value of additional data for more future years. Tilanus (1966) treated a comparable problem using data from The Netherlands deriving the statistical correction method (SCM) for input-output predictions based on the observed structure of the predictions errors and the supplementary statistical hypotheses, which gave almost the same results as RAS. Johansen (1968), and Evans and Lindley (1973) argue in more general terms that individual elements contain more information than the row and column sums, and should therefore be taken into account for updating:

$$\log\left(\frac{z_{ij}}{z_{ij}^{\mathrm{o}}}\right) = \log(r_i) + \log(s_j) + \log(e_{ij}), \tag{1}$$

where e_{ij} is an stochastic term, and $\boldsymbol{r}$ and $\boldsymbol{s}$ are chosen to minimize $\sum_{ij}(\log e_{ij})^2$. Note that an advantage of this formulation is that it may readily be generalized to take account of complete matrices for a series of years.

The aim of this paper is to present an integrated approach that takes the historical behavior of the cells of the matrix that has to be updated into account, and that is easy to implement.

This paper is structured as follows. In Section 2 the proposed Cell-Corrected RAS (CRAS) method is presented. In 2.1 the problem is stated as a programming problem. Its solution is shown in 2.2, and the CRAS iterative algorithm to reach this solution is given in 2.3. In Section 3 a numerical comparison between CRAS and RAS using input-output tables for the Netherlands for 1969-1986 is made. Subsection 3.1 introduces the comparison measures and Subsection 3.2 compares 1-year, 3-year and 5-year forecasts using varying amounts of historical cell changes of comparable length. Section 4 concludes that CRAS performs better in cases of structural change.

2 The Cell-Corrected RAS Method

The goal of the conventional updating methods consists on obtaining an interindustry transactions matrix $\tilde{\boldsymbol{z}}$ of dimension m by n as close as possible to the original $\boldsymbol{z}^{\mathrm{o}}$ of the same dimension, knowing only the margins (the row and column sums) of the target matrix. The idea of minimizing the distance between a known matrix and the target, is logical as no further information is assumed to be available (see Miller (1998)). The solution of the old biproportional RAS algorithm of Stone is shown to be equivalent to minimizing Theil's well known information gain measure (Macgill (1977); Batten (1983), pp. 112-16). Other objective functions, which could be extended with a comparable cell-correction method, are possible, but are not considered here (see Jackson and Murray (2004), for a recent overview and comparison).

2.1 The Programming Model

The RAS method can be stated as follows (see Miller and Blair (1985)):

$$\underset{\boldsymbol{z}}{\text{Minimize}} \quad \sum_{\forall(i,j)\notin\Omega_0} z_{ij} \ln\left(\frac{z_{ij}}{z^{\mathrm{o}}_{ij}}\right), \tag{2}$$

subject to

$$\sum_{j=1}^{n} z_{ij} = u_i;\ i = 1, \ldots, m \tag{3}$$

$$\sum_{i=1}^{m} z_{ij} = v_j;\ j = 1, \ldots, n \tag{4}$$

$$\sum_{i=1}^{m} u_i = \sum_{j=1}^{n} v_j \tag{5}$$

$$z_{ij} \geq 0;\ i = 1, \ldots, m;\ j = 1, \ldots, n, \tag{6}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ contain the row sums and the column sums of the target IO matrix, and Ω_0 contains the pairs of indexes of the zero-elements of the IO matrix $\boldsymbol{z}^{\mathrm{o}}$. The use of this set is needed as the objective function is not defined for those specific values. Constraint (5) ensures the equality of sum total of the row sums and the column sums, and constraint (6) ensures positiveness of the cell values.

We add to this problem definition the availability of a set of T consecutive IO matrices $\boldsymbol{z}^{\theta};\ \theta = 1, \ldots, T$, all using the same classification of rows and columns. Under these conditions the RAS method can be applied between different consecutive years in order to obtain a series of biproportional relations between IO cell values:

$$\log\left(\frac{\tilde{z}^{\theta}_{ij}}{z^{\theta-t}_{ij}}\right) = \log(r^{\theta}_i) + \log(s^{\theta}_j);\ i = 1, \ldots, m;\ j = 1, \ldots, n;\ \theta = 1 + t, \ldots, T, \tag{7}$$

where $t < \theta$ is the length of the projection period and $\tilde{\boldsymbol{z}}^{\theta}$ is the updated RAS matrix, i.e. the solution to problem (2)-(6). Thus, it is possible to calculate $T - t$ observations of individual cell deviations for projection periods with length t that make equation (1) hold as follows:

$$e_{ij}^{\theta} = \frac{z_{ij}^{\theta}}{\tilde{z}_{ij}^{\theta}};\ i = 1, \dots, m;\ j = 1, \dots, n;\ \theta = 1 + t, \dots, T. \quad (8)$$

Using these data, the first two distribution moment vectors $\boldsymbol{\mu}^e$ (mean) and $\boldsymbol{\sigma}^e$ (standard deviation) of the stochastic deviations $\boldsymbol{e}$ can be calculated and used to improve the projection into the future.

The Cell-Correction method proposes a modification of the RAS solution $\tilde{\boldsymbol{z}}$ by means of solving the following optimization method:

$$\underset{\boldsymbol{e}}{\text{Minimize}} \quad \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{e_{ij} - \mu_{ij}^{e}}{\sigma_{ij}^{e}} \right)^2, \quad (9)$$

subject to

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

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

$$e_{ij} \geq 0;\ i = 1, \dots, m;\ j = 1, \dots, n \quad (12)$$

where the objective function (9) minimizes the sum of the squared differences between the deviations and their mean value from the historical observations, weighted by the inverse of the standard deviation. Constraints (10)-(11) ensure that the row and column sums of the corrected matrix are equal to the given row and column sums, and (12) ensures that the IO cell values remain semi-positive. Note that this last constraint will be inoperational because all the $\boldsymbol{e}$ values are around 1.

Observe also that, from a mathematical point of view, problem (9)-(12) consists of the minimization of a positive sum of continuously derivable convex functions defined on a compact set. Note that the solution of (10)-(12) is a linear space of dimension $(m-1) \times (n-1)$ (an arbitrary linear combination of $(m-1) \times (n-1)$ linearly independent vectors, see Castillo et al. (1999)). Hence, there exists only one unique solution provided that constraints (10)-(11) are mutually consistent, i.e.

$$\sum_{j=1}^{n} \tilde{z}_{ij} = u_i;\ i = 1, \dots, m; \quad \sum_{i=1}^{m} \tilde{z}_{ij} = v_j;\ j = 1, \dots, n, \quad (13)$$

which is always the case as $\tilde{\boldsymbol{z}}$ is the solution of problem (2)-(6). In fact, (13) directly follows from (10)-(11) if $e_{ij} = 1;\ i = 1, \ldots, m;\ j = 1, \ldots, n$, which is the solution to the RAS problem.

Note that problem (9)-(11) has been widely used in structural reliability analysis (First Order Reliability Methods, FORM, see Stewart and Melchers (1997)) as an invariant quantitative measure of risk. In fact, the square root of the optimal objective function value is known as the "reliability index" (see Hasofer and Lind (1974)) and the optimal solution $\boldsymbol{e}^*$ is known as "point of maximum likelihood", which in this particular case corresponds to the values of the unexplained deviations that satisfy constraints (10)-(12) and whose value of the joint probability density function is maximum (minimum distance from the expected value, i.e., the most likely value).

The most general expression of the objective function in order to get the reliability index is $\sqrt{(\boldsymbol{e}-\boldsymbol{\mu})^T\boldsymbol{V}^{-1}(\boldsymbol{e}-\boldsymbol{\mu})^T}$ (see Ditlevsen (1981), Veneciano (1974) and Low and Tang (1994)), where $\boldsymbol{V}$ is the covariance matrix, which becomes (9) if $\boldsymbol{e}$ variables are independent (diagonal covariance matrix). Hence, the proposed approach assumes the statistical independence between deviations.

2.2 Solution to the Programming Model

To derive the solution to the programming model consider the Lagrange function associated with problem (9)-(12):

$$\mathcal{L}(\boldsymbol{e}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{e_{ij}-\mu^e_{ij}}{\sigma^e_{ij}}\right)^2 + \sum_{i=1}^{m}\lambda_i\left(\sum_{j=1}^{n} e_{ij}\tilde{z}_{ij} - u_i\right) + \sum_{j=1}^{n}\mu_j\left(\sum_{i=1}^{m} e_{ij}\tilde{z}_{ij} - v_j\right), \tag{14}$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are the Lagrange multipliers.

The derivatives of the Lagrange function with respect to $\boldsymbol{e}, \boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are:

$$\frac{\partial \mathcal{L}}{\partial e_{ij}} = 2\frac{e_{ij}-\mu^e_{ij}}{\sigma^{e^2}_{ij}} + \tilde{z}_{ij}(\lambda_i + \mu_j) \;=\; 0;\ i = 1, \ldots, m;\ j = 1, \ldots, n \tag{15}$$

$$\sum_{j=1}^{n} e_{ij}\tilde{z}_{ij} - u_i \;=\; 0;\ i = 1, \ldots, m \tag{16}$$

$$\sum_{i=1}^{m} e_{ij}\tilde{z}_{ij} - v_j \;=\; 0;\ j = 1, \ldots, n. \tag{17}$$

From (15) deviations are written in terms of the dual variables as:

$$e_{ij} = \mu^e_{ij} - \tilde{z}_{ij}(\lambda_i + \mu_j)\sigma^{e^2}_{ij}/2, \tag{18}$$

Substituting (18) in (16) and (17) gives the following relationships between the Lagrange multipliers:

$$\lambda_i^{(\nu+1)} = \frac{-u_i + \sum_j \tilde{z}_{ij}(\mu_{ij}^e - \mu_j^{(\nu)}\tilde{z}_{ij}\sigma_{ij}^{e^2}/2)}{\sum_j (\sigma_{ij}^e \tilde{z}_{ij})^2/2};\ i = 1, \ldots, m \tag{19}$$

$$\mu_j^{(\nu+1)} = \frac{-v_j + \sum_i \tilde{z}_{ij}(\mu_{ij}^e - \lambda_i^{(\nu)}\tilde{z}_{ij}\sigma_{ij}^{e^2}/2)}{\sum_i (\sigma_{ij}^e \tilde{z}_{ij})^2/2};\ j = 1, \ldots, n, \tag{20}$$

where ν is the iteration counter.

Equations (19) and (20) can be solve iteratively until the norms $||\boldsymbol{\lambda}^{(\nu+1)} - \boldsymbol{\lambda}^{(\nu)}||$ and $||\boldsymbol{\mu}^{(\nu+1)} - \boldsymbol{\mu}^{(\nu)}||$ are lower than a pre-specified tolerance ϵ. Once the values of the dual variables are given, the most likely deviation $\boldsymbol{e}^*$ is calculated using (18), and the optimal values of CRAS ($\hat{\boldsymbol{z}}$) are obtained as:

$$\hat{z}_{ij} = e_{ij}^* \tilde{z}_{ij};\ i = 1, \ldots, m;\ j = 1, \ldots, n, \tag{21}$$

where $(*)$ refers to optimal values.

2.3 The CRAS-Algorithm

The programming problem (9)-(12) can, of course, be solved by any non-linear optimization routine implemented in frameworks such as, GAMS (Brooke et al., 1998), AIMMS (Bisschop and Roelofs, 1999), AMPL (Fourer et al., 1993), LINDO and What's Best (Schrage, 1991). It is, however, instructive and handy to derive an efficient iterative solution that is comparable to the RAS-algorithm. For that reason, we present the CRAS-algorithm for correcting a given RAS solution matrix:

Algorithm 1 (Cell-Corrected RAS method).

Input: *An updated RAS solution matrix ($\tilde{\boldsymbol{z}}$) , the mean ($\boldsymbol{\mu}^e$) and standard deviation ($\boldsymbol{\sigma}^e$) from historical observations, and an pre-specified tolerance ϵ.*

Output: *The corrected RAS matrix ($\hat{\boldsymbol{z}}$) within tolerance ϵ.*

- **Step 1:** *Initialize the iteration counter $\nu = 1$, and start from given Lagrange multipliers vectors $\boldsymbol{\lambda}^{(\nu)} = \mathbf{0}$ and $\boldsymbol{\mu}^{(\nu)} = \mathbf{0}$.*

- **Step 2:** *Update the $\boldsymbol{\lambda}^{(\nu+1)}$ vector using (19) and the $\boldsymbol{\mu}^{(\nu)}$-values.*

- **Step 3:** *Update the $\boldsymbol{\mu}^{(\nu+1)}$ vector using (20) and the $\boldsymbol{\lambda}^{(\nu+1)}$-values.*

- **Step 4:** *Check for convergence, if* $||\boldsymbol{\lambda}^{(\nu+1)} - \boldsymbol{\lambda}^{(\nu)}|| \leq \epsilon$ *and* $||\boldsymbol{\mu}^{(\nu+1)} - \boldsymbol{\mu}^{(\nu)}|| \leq \epsilon$ *the optimal values of the Lagrange multipliers have been found, go to* **Step 5***, otherwise, update the iteration counter* $\nu \rightarrow \nu + 1$ *and continue with* **Step 2***.*

- **Step 5:** *Calculate the optimal values of the deviations* $\boldsymbol{e}^*$ *using (18) and the values of the final Cell-Corrected matrix* $\hat{\boldsymbol{z}}$ *using (21).*

∎

3 Numerical Comparison of CRAS with RAS

To test the performance of CRAS relative to RAS we use a consistent series of Dutch national IO tables in current prices with 24 sectors (which excludes the 'sector' non-classified) for the period 1969-1986, compiled by the Netherlands Central Bureau of Statistics (CBS)[4]. First, we discuss the performance indicators used. Then, we present the results of a series of comparisons of 1-year, 3-year and 5-year projections of the intermediate 24x24 part of these tables.

3.1 Performance Indicators

The comparison of CRAS with RAS approach is made by inspecting the distance between the corresponding solutions $\tilde{\boldsymbol{z}}$ and $\hat{\boldsymbol{z}}$ from (21) and the true value $\boldsymbol{z}^{\text{true}}$, using different matrix distance measures (deMesnard and Miller (2006)), which hereafter will be called norms. The use of different norms is motivated by the fact that they can lead to different results. We focus on additive norms $||\tilde{\boldsymbol{z}} - \boldsymbol{z}^{\text{true}}||$ and $||\hat{\boldsymbol{z}} - \boldsymbol{z}^{\text{true}}||$, as using multiplicative norms $||\log(\tilde{\boldsymbol{z}}) - \log(\boldsymbol{z}^{\text{true}})||$ and $||\log(\hat{\boldsymbol{z}}) - \log(\boldsymbol{z}^{\text{true}})||$ does not change the basic properties of the norms (deMesnard (2004)).

The norms used for the matrix $\boldsymbol{M} = \boldsymbol{Z} - \boldsymbol{Z}^{\text{true}}$, with the differences between the forecasted and true elements, are the following:

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

$$\text{MAPE} = \frac{1}{n^2} \sum_i \sum_j \frac{|m_{ij}|}{|z_{ij}^{\text{true}}|} \times 100\%. \tag{22}$$

[4]Other available, internally consistent series of Dutch IO tables relate to 87 sectors for the period 1987-1995 and to 104 sectors for the period 1995-2003. For our purpose of comparing CRAS with RAS using the longest series is the most suited. This 1969-1986 series also has a consistent 91 sector classification, which may be used for further testing.

- Weighted Average Percentage Error, given by:

$$\text{WAPE} = \sum_i \sum_j \left(\frac{z_{ij}^{\text{true}}}{\sum_k \sum_l z_{kl}^{\text{true}}} \right) \frac{|m_{ij}|}{z_{ij}^{\text{true}}} \times 100\%. \tag{23}$$

- Normalized Square Error (Deming and Stephan (1940)):

$$\text{NSE} = \sum_i \sum_j \frac{(z_{ij} - z_{ij}^{\text{true}})^2}{z_{ij}^{\text{true}}}. \tag{24}$$

- Weighted Normalized Square Error, given by:

$$\text{WSE} = \sum_i \sum_j \left(\frac{z_{ij}^{\text{true}}}{\sum_k \sum_l z_{kl}^{\text{true}}} \right) \frac{(z_{ij} - z_{ij}^{\text{true}})^2}{z_{ij}^{\text{true}}}. \tag{25}$$

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

$$\text{IG} = \sum_i \sum_j \left| z_{ij}^{\text{true}} \log \left(\frac{z_{ij}}{z_{ij}^{\text{true}}} \right) \right|. \tag{26}$$

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

$$c_p = \frac{\tilde{n} - \hat{n}}{\hat{n}} \times 100\%, \tag{27}$$

where $\hat{n}$ and $\tilde{n}$ are the norms (22)-(26) obtained when using the CRAS and RAS 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.

3.2 Comparisons for 1-, 3-, and 5-Year Projections

The 5-year projections of Figure 1 represent the most interesting and most relevant results. In Figure 1 (and in Figures 3 and 4), the CRAS method corrects the RAS projection for the same year using the maximal information provided by all possible, historic 1-year projections. The 1976-values in Figure 1 thus compare the CRAS projection for the period 1971-76 with the comparable RAS projection, where CRAS uses the additional information on the historic projections for the periods 1969-70 and 1970-71, while the 1981-86 CRAS projection also uses the information all other 1-year projections up till 1980-81. Note that 1-year projections provide the maximum amount of information and we assume that the deviations follow an ergodic process which does not depend on the

Figure 1: Performance ratio (%) of CRAS to RAS for 5-year projections.

Figure 2: Oil price development from 1969 to 1986 (in dollars, Forbes (2007)).

prediction year, i.e., each new data point adds the same amount of new information (see Walters (1982)). Other statistical assumptions will of course require different type of comparisons.

Figure 1 shows that CRAS systematically outperforms RAS for 5-year projections, independent of the norm used, except for the last projection for 1986. Looking for an explanation for the exception, the oil price rises of 1973-74 and 1978-79, and the subsequent oil price fall of 1985-86 come to mind (see Figure 2). The CRAS projection for 1986 uses information on all 1-year projections of 1969-70 up to 1980-81, which includes the information on the two oil price rises. This information is used to improve the projection for 1981-86 that had to predict the IO consequences of the oil price fall of 1985-86. Naturally, the added information pointed into an entirely different direction than was needed to predict 1986. Considering this, the partially better performance of RAS over CRAS for this particular year is disappointingly small.

Comparing the outcomes for the different norms, the largest out-performance of CRAS is found for the squared errors (WSE and NSE). When the weighted errors (WSE and WAPE) are compared with their unweighted equivalents (NSE and MAPE), in almost all cases, the out-performance is larger for the weighted errors. Both results indicate that CRAS especially outperforms RAS in projecting the larger IO cells with the larger changes, which are the more important ones from an economic point of view.

The weighted squared errors, furthermore, indicate that the out-performance of CRAS is smallest for 1981 and 1982. Again the oil price shocks provide part of the explanation. Note that all previous projections had to predict 5-year changes with oil price rises, but the projections for 1981 and 1982 were the first that had to cover oil price declines. This underscores the conclusion that may be derived from the explanation of the 1986 exception, namely that CRAS outperforms RAS in situations of systematic structural change, but that RAS may outperform CRAS when unprecedented asymmetric (price) shocks have to be projected.[5]

[5]The oil price explanation of the exceptions, where RAS does partially better than CRAS, is confirmed

	5-year forecasts										
t	**76**	**77**	**78**	**79**	**80**	**81**	**82**	**83**	**84**	**85**	**86**
2	8.8	5.5	-3.7	-7.9	-11.4	6.5	-2.7	-3.2	2.1	-9.2	-6.9
3		8.9	1.6	-2.8	-1.5	2.4	3.0	1.4	3.5	-2.2	-8.4
4			4.3	3.1	2.6	6.2	3.1	4.6	4.9	1.2	-2.1
5				5.0	5.7	5.7	5.4	4.4	9.3	1.3	0.6
6					5.4	8.2	4.6	5.9	7.6	8.5	-0.2
7						7.5	6.8	5.5	8.9	6.9	2.4
8							6.9	7.3	7.3	4.7	0.4
9								8.0	9.3	4.4	-0.2
10									9.8	6.9	-0.5
11										7.5	-0.1
12											0.3

Table 1: WAPE ratio (%) for 5-year forecasts with information of t previous IO tables.

The above conclusion suggests that CRAS will perform better when more historic information is available. To investigate this hypothesis, Table 1 compares the 5-year projection performance ratios of CRAS for all possible 1-year projections that CRAS could use. To minimize the amount of outcomes, we only show the results for our favorite norm, the weighted absolute percentage error (WAPE). Thus, the diagonal of Table 1 shows the same information as the WAPE-line in Figure 1. Note that the WAPE norm takes a middle position in Figure 1 and runs almost parallel to the minimum information gain (MIG), which is the second reason to select WAPE as the most representative norm.

The first row of Table 1 shows all 5-year CRAS projections that only use the two most recent 1-year projections to calculate $\boldsymbol{\mu}^e$ and $\boldsymbol{\sigma}^e$. When using this minimal amount of extra information, CRAS outperforms RAS in only 4 of the 11 cases on the first row. Going down the columns of Table 1, i.e. adding information on older and older projections to calculate $\boldsymbol{\mu}^e$ and $\boldsymbol{\sigma}^e$, we see that the CRAS projection improves in 38 of the 55 cases of adding one more observation. Thus, it may indeed be concluded the CRAS performs better the more historic IO data is used.

Figure 3 and Table 2 provide the same information as Figure 1 and Table 1, but for 3-year projections. In Figure 3, CRAS again outperforms RAS, but with this shorter

by an inspection of the cells with the largest differences between the two projections and the real 5-year development of the 24x24 IO cell values for the 11 target years. In the case of RAS these are found in the construction to construction cells (4x), in crude oil and natural gas exploitation to public utilities (2x), in petroleum industry to chemicals, in utilities to utilities, and in three more cells. In the case of CRAS these are found in petroleum industry to chemicals (5x), in construction to construction (2x), in transport equipment production to construction, and in the same three more cells. When the CRAS and RAS projections are directly compared which each other, the largest differences are almost exclusively found in the crude oil and natural gas exploitation to public utilities (8x).

Figure 3: Performance ratio (%) of CRAS to RAS for 3-year projections.

	3-year forecasts												
t	**74**	**75**	**76**	**77**	**78**	**79**	**80**	**81**	**82**	**83**	**84**	**85**	**86**
2	-0.5	2.5	-5.2	-16.4	-22.0	-10.3	-3.8	-7.3	2.0	-2.8	-11.9	-15.2	-11.0
3		7.5	0.9	-9.5	-10.0	-11.6	-0.9	0.0	2.7	0.9	-11.0	-12.0	-11.1
4			4.9	-2.9	-3.3	-4.6	-3.1	4.8	3.1	1.1	-4.1	-15.3	-9.3
5				0.3	0.4	-2.6	-0.8	4.4	9.2	4.1	-1.2	-12.1	-10.1
6					0.9	-0.0	0.2	6.4	5.8	3.2	0.7	-7.1	-7.2
7						-0.9	1.8	5.2	7.2	3.0	-3.0	-7.0	-4.5
8							1.0	6.8	5.2	5.2	-1.7	-7.4	-5.1
9								6.8	7.3	5.5	-0.4	-6.7	-3.0
10									7.6	7.1	0.4	-7.7	-4.5
11										7.6	1.4	-6.9	-4.7
12											2.3	-6.7	-4.2
13												-6.8	-3.8
14													-3.4

Table 2: WAPE ratio (%) for 3-year forecasts with information of t previous IO tables.

projection period there are more exceptions. The 1971-74 projection is worse than that for later years. Most likely because this CRAS projection only uses the 1-year projections of 1969-70 and 1970-71, during which oil price remained almost constant (see Figure 2), whereas it had to predict the IO consequences of the oil price rise of 1973-74. Also, the CRAS projections for 1979 and 1980 are worse or only little better than RAS, again most likely because the IO consequences of the oil price increases of those two years had to be projected. For the, even worse CRAS-projections of 1985 and 1986, again oil price changes provide for at least part of the answer. For the 1982-85 and 1983-86 projections CRAS needed to project the IO consequences of oil price declines with only information on years with either rising or more or less constant oil prices, which sent CRAS in the wrong direction.

The above results confirm that CRAS performs well under conditions of systematic structural change, and that RAS may do better when unprecedented price shocks have to be covered. This conclusion is further substantiated by the worst 3-year WAPE-ratio of -22.0% for the 1978 projection in Table 2. That specific CRAS-projection only uses the projections of 1973-74 and 1974-75, with the IO consequences of the first price oil hike, whereas it had to make a projection for a new period without further price increases. Under such conditions the CRAS method more or less assumes that the oil price will

Figure 4: Performance ratio (%) of CRAS to RAS for 1-year projections.

continue to grow, which gives a wrong projection.

Table 2 also confirms the conclusion of Table 1 that adding more IO projections improves the performance of CRAS relative to RAS. Inspecting the columns of Table 2 shows that adding one more historic IO projection improves the CRAS projection in 58 out of the 78 cases. A comparable conclusion holds for 1-year projections that improve the CRAS projection in 70 out of 105 cases in Table 3. However, especially the last two columns of Table 2 also show that adding more information does not work when it is more information on the same old trend. In case of projecting the IO consequences of something entirely new, like the sudden oil price decline of 1985-86, adding more information on periods in which the oil price rises or remains constant, does not help.

Table 2 also shows that CRAS outperforms RAS for the shorter 3-year projections in only 37 of the total of 91 simulations, whereas it outperformed RAS in 51 of the 66 simulations with 5-year projections. The 3-year score becomes even worse if we move to 1-year projections in Table 3. There, CRAS only outperforms RAS in 10 of the 120 cases. This bad performance for shorter periods not only holds for the WAPEs of Table 3, but also for the other norms as is shown in Figure 4, which reproduces the WAPEs from the diagonal of Table 3.

So, our final conclusion is that CRAS only outperforms RAS in longer term projections. The most likely explanation is that shorter term changes in the values of IO cell are far more stochastic than the 5-year changes. This qualifies the earlier conclusion that CRAS outperforms RAS when systematic structural changes need to be covered. To this we should add "for longer term projections". For practical situations, however, this restriction is not a very strong one. As most national statistical offices produce IO tables with delays of four and more years, being able to project them better for 1-year periods is not a very interesting proposition. Hence, the significantly better performance of CRAS over RAS for 5-year projections is most relevant and most promising.

4 Conclusions

This paper presents a new method that adds cell-specific corrections to the well known biproportional RAS method, which only takes into account the row and column totals of the unknown target matrix. The cell corrections of CRAS are determined by minimizing

	1-year forecasts														
t	**72**	**73**	**74**	**75**	**76**	**77**	**78**	**79**	**80**	**81**	**82**	**83**	**84**	**85**	**86**
2	-31.8	-8.0	-10.1	-23.6	-36.7	-27.9	-18.0	-25.0	-5.0	-22.0	-15.5	-23.1	-25.8	-16.4	-14.6
3		0.0	-5.3	-16.7	-16.3	-31.7	-10.1	-12.1	2.3	-15.0	-14.4	-18.9	-15.8	-11.7	-7.4
4			-3.2	-8.6	-0.8	-26.7	-12.8	-16.5	-0.1	-12.1	-5.2	-16.3	-17.7	-11.9	-5.0
5				-4.8	0.5	-22.6	-15.5	-6.4	-0.3	-5.9	-1.6	-16.3	-17.8	-9.4	-5.4
6					0.5	-22.7	-13.6	-2.4	-4.9	1.0	-3.9	-12.1	-16.5	-13.0	-4.2
7						-21.7	-11.9	0.9	-3.1	1.2	-5.5	-9.0	-15.0	-11.2	-3.2
8							-12.5	-0.7	-4.1	0.8	-12.0	-15.3	-12.5	-6.8	-2.9
9								-2.3	-4.5	3.2	-10.6	-14.2	-11.9	-8.1	-1.9
10									-5.7	4.9	-7.9	-10.4	-5.4	-5.7	-3.3
11										3.9	-7.6	-9.5	-6.0	-8.7	-1.0
12											-7.7	-11.2	-5.6	-9.8	-1.9
13												-10.5	-6.1	-8.4	-1.9
14													-5.8	-8.7	-2.6
15														-9.1	-2.4
16															-1.2

Table 3: WAPE ratio (%) for 1-year forecasts with information of t previous IO tables.

the sum of the squared mean deviations of RAS projections from known IO tables, in time or space, weighted by the inverse of their standard deviation.

An advantage of CRAS is that it can as easily be solved as RAS itself, namely by means of an iterative algorithm with good convergence properties.[6] The most important advantage is that it clearly outperforms RAS when making longer term projections of five and more years, which are the most relevant projection periods in practical statistical work. This conclusion, however, has to be qualified in the sense that CRAS does not outperform RAS when historically or spatially unprecedented structural change or spatial differences have to be simulated.

In the empirical tests for the Netherlands for the period 1969-86 this proved to be the case when one of the two oil price hikes of 1973-74 or 1979-80 had to be covered, or the sharp oil price decline of 1985-86. Further empirical testing with constant price IO tables will have to prove whether CRAS outperforms RAS in longer term constant price projections without any qualification, as we expect those IO tables to only reflect systematic structural change without unprecedented shocks.[7]

[6]These follow from the empirical simulations for the Netherlands and are available upon request with the first author.

[7]Further testing may also be useful to see whether using longer term historic projections, instead of the 1-year historic projections of this paper, further improves the out-performance of CRAS over RAS in longer term projections.

References

Batten, M. (1983). *Spatial Analysis of Interacting Economies*. Boston, The Hague, London: Kluwer-Nijhoff.

Bisschop, J. and M. Roelofs (1999). *AIMMS, The User's guide*. Paragon Decision Technology.

Brooke, A., D. Kendrick, A. Meeraus, and R. Raman (1998). *GAMS: A user's guide*. Washington: GAMS Development Corporation.

Butterfield, M. and T. 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.

Castillo, E., A. Cobo, F. Jubete, and R. E. Pruneda (1999). *Orthogonal Sets and Polar Methods in Linear Algebra: Applications to Matrix Calculations, Systems of Equations and Inequalities, and Linear Programming*. New York: John Wiley & Sons.

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

deMesnard, L. and R. E. Miller (2006). A note on added information in the RAS procedure: Reexamination of some evidence. *Journal of Regional Science 46*(3), 517528.

Deming, W. E. and F. 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.

Ditlevsen, O. (1981). *Uncertainty modeling with applications to multidimensional engineering systems*. New York: McGraw-Hill Book Co.

Evans, G. J. and R. M. Lindley (1973). The use of ras and related models in manpower forecasting. *Economics of Planning 13*(1-2), 53–73.

Forbes (2007). http://www.forbes.com/.

Fourer, R., D. M. Gay, and B. W. Kernighan (1993). *AMPL: A Modelling Language for Mathematical Programming*. Massachusetts: Boyd & Fraser publishing company.

Hasofer, A. M. and N. C. Lind (1974). Exact and invariant second moment code format. *J. Engrg. Mech. 100*(EM1), 111–121.

Hewings, G. 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.

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

Johansen, L. (1968). Explorations in long-term projections for the Norwegian economy. *Economics of Planning 8*(1-2), 70–117.

Low, B. K. and W. H. Tang (1994). Efficient reliability evaluation using spreadsheet. *Journal of Engineering Mechanics, ASCE 123*(7), 749–752.

Macgill, S. M. (1977). Theoretical properties of biproportional matrix adjustments. *Environment and Planning A 9*(6), 687–701.

Miller, R. E. (1998). *Methods of Interregional and Regional Analysis*, Chapter Regional and interregional input-output analysis, pp. 41–133. Brookfield, VE, USA: Ashgate.

Miller, R. E. and P. D. Blair (Eds.) (1985). *Input-output analysis : Foundations and extensions*. Englewood Cliffs, NJ, USA: Prentice-Hall.

Omar, F. H. (1967). *The projection of input-output coefficients with application to the United Kingdom*. Ph. D. thesis, Department of Economics, University of Nottingham.

Oosterhaven, J., G. J. Piek, and T. M. Stelder (1986). Theory and practice of updating regional versus interregional interindustry tables. *Papers of the Regional Science Association 59*(1), 57–72.

Schrage, L. (1991). *LINDO: An Optimization Modeling System*. London: The Scientific Press.

Stewart, M. G. and R. E. Melchers (1997). *Probabilistic Risk Assessment of Engineering Systems*. London: Chapman & Hall.

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

Stone, R. and A. Brown (1962). *A Programme for Growth, 1. A Computable Model of Economic Growth*. London: Chapman and Hall.

Tilanus, C. (1966). *Input-output experiments : The Netherlands 1948-1961*. Rotterdam: Rotterdam University Press.

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

Veneciano, D. (1974). Contributions to second moment reliability. Res. Rep. R74-33, Dept. of Civil Engineering., MIT, Mass.

Walters, P. (1982). *An introduction to ergodic theory.* New York: Springer.