Entropy Theory and RAS are Friends

Robert A. McDougall

May 14, 1999

Abstract

Recent research in applications of entropy theory to matrix balancing problems in economics has put powerful new tools in the hands of data base developers, but overshadowed some previous findings. We recall earlier findings that the RAS is an entropy-theoretic model. Investigating the properties of a more recently proposed entropy-theoretic model, we find that in general the RAS remains preferable. We show further that the RAS can be obtained also as a generalised cross-entropy model. Finally, we present examples illustrating how entropy-theoretic techniques can extend the RAS to handle a wider range of problems.

1 Introduction

Recent years have seen new interest in applications of entropy theory in economics, including applications to data balancing problems such as those encountered in input-output (*IO*) table construction. Works by Golan, Judge, and Robinson [11] (henceforward GJR) and Golan, Judge, and Miller [10] (henceforward GJM) have been effective in promoting this new interest. This interest however has not been without its disadvantages. The following mistaken views for example may be encountered:

- that the application of entropy theory to IO data construction is a new development,

- that the RAS model widely used in such tasks is not an entropy-theoretic method,

- that the RAS has been superseded by more recently developed entropy-theoretic methods.

This paper is intended to recall to notice earlier findings on the relationship of entropy theory and the RAS, to investigate the properties of some more recently proposed entropy-theoretic

methods, to assess their merits relative to the RAS, and to offer some suggestions on how entropy theory may most fruitfully be applied to data balancing problems.

Following a brief historical overview (section 2), we examine first the matrix filling problem and maximum entropy methods (section 3). This introduces ideas for section 4, on matrix balancing methods and cross entropy methods, including the RAS. We then undertake a preliminary examination of generalised cross entropy methods (section 5), provide examples of constructive applications to matrix balancing problems of entropy-theoretic techiques (section 6), and conclude (section 7).

2   Historical overview

The RAS or biproportionate adjustment model has been invented several times in several different disciplines. Bregman [3] reports "This method was proposed in the 1930s by the Leningrad architect G.V. Sheleikhovskii for calculating traffic flow." Bacharach's [1] monograph relates that "In 1940, Deming and Stephan had treated as a biproportional constrained matrix problem the statistical problem of estimating an unknown contingency matrix from known marginals and an initial estimate of the matrix [7]"; and that in 1941, Leontief [16] "proposed a biproportional form for the relationship between the values taken by an input-output matrix at different points of time." But the impetus to its use in IO table construction came from Stone [8, 21], who in 1962 proposed its use by the Cambridge Growth Project in constructing use matrices for IO tables for the United Kingdom, and gave it the name *RAS* it commonly bears in this field of application.

Besides use matrices for IO tables, it has been applied in constructing other economic data arrays such as make matrices for IO tables (e.g., Cambridge Growth Project [9], 1963) and bilateral trade matrices (e.g., Bénard [2], 1963).

The iterative scaling method commonly used in solving the RAS was proposed by Deming and Stephan [7] in 1940 in the contingency table context, and independently by Stone in 1962 in the IO table context.

While the RAS model was being discovered and rediscovered, Shannon [19] in 1948 initiated the field of information theory, appropriating for one of its key concepts the name *entropy* from thermodynamics. Shannon's interest was in communication engineering (where applications include the use of error correcting codes on compact disks, and the Lempel-Ziv compression scheme widely used for file compression on computers), but connections were established to other fields such as computer science (Kolmogorov complexity, [14] 1968), finance theory (log optimal portfolios, Kelly [13] 1956), spectral analysis in geophysics and elsewhere (Burg [5] 1975), and statistics (contingency tables, Fisher's information measure; see, e.g., Kullback [15]).

Theil [22] in 1967 collected and extended work on the application of entropy theory to economic topics including the construction of economic data matrices (he does not seem to have been aware of the finance theory work mentioned above). Following Uribe, de Leeuw and Theil [23], he identified some relationship between entropy theory and the RAS; but the final step of identifying the RAS model as an entropy-theoretic method was made by Bacharach [1] in 1970. Unknown to him, other researchers had already obtained similar results (e.g., Sinkhorn [20] in 1964, Bregman [3] in 1967); those researchers in turn, though familiar with applications of the RAS model to transport economics and statistics, were not aware of its employment in IO economics.

Further information on the history of the RAS may be obtained from Bacharach [1] and Schneider and Zenios [18].

It will be apparent to the reader that the early history of the RAS method in IO economics was marked by a lack of awareness of relevant entropy-theoretic work in other fields. This is unsurprising, since that work tends to be specialised and technical. More recently, GJR in 1994 and GJM in 1996 have done much to increase awareness of relevant work outside economics, offset unfortunately by neglect of relevant work inside. Thus for instance GJR write that "in contrast to [the RAS method], we make use of the entropy principle and consider a method based only on the information that is available," neglecting the literature that establishes for the RAS precisely the connections to information and entropy theory that they deny.

3 Entropy optimization and proportional allocation

Consider the matrix filling problem: find matrix elements v_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J$ consistent with row target totals $v_{i\bullet}$ and column target totals $v_{\bullet j}$. Equivalently, find column coefficients $b_{ij} = v_{ij}/v_{\bullet j}$ consistent with the row and column totals.

Such a problem might arise, for example, in disaggregating an industry in an IO table, knowing only the industry-wide input values $v_{i\bullet}$ and disaggregated industry output values $v_{\bullet j}$; or in constructing a bilateral trade matrix, knowing only country export values $v_{i\bullet}$ and import values $v_{\bullet j}$. We shall refer to the column coefficients b_{ij}, $i = 1, \ldots, I$ as the *column structure* of column j. Similarly we shall refer to the row coefficients $v_{ij}/v_{i\bullet}, j = 1, \ldots, J$ as the row structure of row i.

This problem has a simple and obvious solution, namely, to allocate values equiproportionally across columns and across rows:

$$v_{ij} = \frac{v_{i\bullet}}{v_{\bullet\bullet}} \frac{v_{\bullet j}}{v_{\bullet\bullet}} \, v_{\bullet\bullet} = \frac{v_{i\bullet} v_{\bullet j}}{v_{\bullet\bullet}}, \tag{1}$$

3

or, in terms of column coefficients,

$$b_{ij} = \frac{v_{i\bullet}}{v_{\bullet\bullet}}, \tag{2}$$

where $v_{\bullet\bullet}$ is the common sum $\sum_j v_{\bullet j} = \sum_i v_{i\bullet}$ (required to exist, for the problem to be feasible). We call this the *proportional allocation* method.

The proportional allocation method might seem too obvious to require justification. Nevertheless, a reasoned justification can uncover some less obvious considerations.

Proposition 1 *For the matrix filling problem, the proportional allocation method should be preferred, in the absence of information about variation in column structure or variation in row structure.*

Since we have no basis for differentiating the column structure across columns, we should not differentiate it, but assume a uniform structure. For example, in disaggregating an input-output table, if we know nothing about how cost structures vary across the disaggregate industries, we should give them all the same cost structure. Equivalently, having no basis for differentiating the row structure, we should assume a uniform structure.

Again, this appears intuitively obvious, but further justification is available if needed. If we arbitrarily differentiate the column structures or row strucures, we are liable to create spurious mechanisms in any model built on the data set. If for example we differentiate cost structures across industries, then in a model using these data, a fall in the price of a given input leads to a relative cost reduction for the industries that use that input most intensively; and that is liable to have further effects, for example, a shift in the export pattern toward those industries' products. But if the differentiation in cost structures in the data base is arbitrary, then those effects are most likely spurious.

In brief, we should use the proportional allocation method because it avoids importing spurious information into the matrix. ∎

In justifying the proportional allocation method, we do not rely on the argument that it yields the most accurate estimates. For reasonable assumptions about probability distributions and choices of accuracy measures, the proportional allocation method may well be the most accurate. But any estimation method for this problem is bound to be highly inaccurate, so any use of the estimates should not rely on their accuracy.

Typically however accuracy is not the only consideration. Suppose for example we developed some random method for filling in the matrix. It would likely be more inaccurate on average than the proportional allocation method; but it might not be much more inaccurate. But it would be far inferior methodologically, because it would create spurious mechanisms in any model built on top of the data, that is, mechanisms we had no justification for putting in the model.

4

The argument for proportional allocation is not of course an argument against using further information where available. This further information might bear directly or indirectly on the problem in hand. As an example of direct information, in disaggregating the food processing industry, one would know better than to assume that the sugar industry uses raw sugar in the same proportions as other food processing industries. As an example of indirect information, in constructing a bilateral trade matrix, one might infer from other trade pattern studies that the gravity model would provide better results than proportional allocation.

Golan, Judge and Robinson [11] (*GJR*) however use maximum entropy to derive a different method. Maximizing an objective function

$$H = -\sum_i \sum_j b_{ij} \log b_{ij} \tag{3}$$

subject to constraints

$$\sum_j b_{ij} v_{\bullet j} = v_{i \bullet}, \tag{4}$$

$$\sum_i b_{ij} = 1, \tag{5}$$

they obtain

$$b_{ij} = \frac{1}{\Omega_j(\lambda_i)} \exp[-\lambda_i v_{\bullet j}], \tag{6}$$

where

$$\Omega_j(\lambda_i) = \sum_{i=1}^m \exp[-\lambda_i v_{\bullet j}]. \tag{7}$$

We note that the objective function (3) is not an entropy but a sum of entropies. Specifically, we may write

$$H = \sum_j H_j, \tag{8}$$

where each

$$H_j = \sum_i b_{ij} \log b_{ij}$$

H_j is an entropy, since it is a sum of terms involving variables b_{ij}, $i = 1, \ldots, m$ subject to an adding-up constraint $\sum_i b_{ij} = 1$. But H itself is not an entropy.

Entropy theory in other words supplies the H_j, but the method of combining them to form a single objective function, formula (8), was chosen at its authors' [11] discretion. The simple

sum (8) is a natural and obvious way of combining the entropies, but if on examination we find some problem with the estimators it yields, we are free to define a new objective function combining them in some other way. The only obvious restriction on such an objective function is that it should be a strictly increasing function of the individual entropies. As we shall see, we can also chose a different set of variables to include in the entropy measure. We can even change from an entropy maximization to a cross-entropy minimization method. In short, the model of equations (3, 4), does not represent the entropy optimization approach, but one particular entropy approach. We call this particular approach the *maximum sum of entropies* or *MSE* approach.

We now examine the MSE solution (6, 7). We note first that comparing coefficients within a column, we have

$$b_{i_2 j} > b_{i_1 j} \text{ if and only if } \lambda_{i_2} < \lambda_{i_1}.$$

Since the λ_i are the same for all columns, this shows that the size order of coefficients is the same within all columns.

Next, note that the within-column variation in the coefficients b_{ij} itself varies with the column sum $v_{\bullet j}$. In the columns with small column sums, the coefficients tend to be clustered together, close to their average value $1/m$, while in the columns with large column sums the coefficients are more widely dispersed. To see this, observe that for any two rows i_1, i_2, the ratio of coefficients

$$\frac{b_{i_2 j}}{b_{i_1 j}} = \left[\frac{\exp(-\lambda_{i2})}{\exp(-\lambda_{i1})} \right]^{v_{\bullet j}}$$

is close to one for small $v_{\bullet j}$, and distant from one for large $v_{\bullet j}$; provided only that λ_{i_1} and λ_{i_2} are unequal.

This gives us:

Proposition 2 *The MSE method leads to more uniform cost structures for smaller industries, and more differentiated cost structures for larger industries, in this sense: that for columns j_1, j_2 such that $v_{\bullet j_2} > v_{\bullet j_1}$,*

$$if \quad \frac{v_{i_2 j_1}}{v_{i_1 j_1}} < 1 \quad then \quad \frac{v_{i_2 j_2}}{v_{i_1 j_2}} < \frac{v_{i_2 j_1}}{v_{i_1 j_1}}, \tag{9}$$

$$if \quad \frac{v_{i_2 j_1}}{v_{i_1 j_1}} > 1 \quad then \quad \frac{v_{i_2 j_2}}{v_{i_1 j_2}} > \frac{v_{i_2 j_1}}{v_{i_1 j_1}}. \tag{10}$$

Equivalently, since the coefficients sum to one within columns, the MSE solution has this property: in the columns with small column sums, the coefficients for rows with small row sums are relatively large, and the coefficients for rows with large row sums, relatively small; while in the columns with large column sums, the coefficients for rows with small row sums are relatively

small, and the coefficients for rows with large row sums, relatively large. For an IO table, this means that the large industries are relatively intensive users of the large commodities, and the small industries, relatively intensive users of the small commodities. But there is no reason to expect the larger industries' cost structures to differ in any particular way from the smaller industries'. So the differentiation in their cost structures in the MSE solution is undesirable.

We conclude therefore that the MSE solution is inferior to proportional allocation, in that it differentiates arbitrarily between column structures (and likewise row structures) on the basis of irrelevant information, namely, column target totals (row target totals).

An example is provided by GJR [11]. Filling in a matrix with row totals $\begin{bmatrix} 80 & 110 & 140 & 145 \end{bmatrix}'$ and column totals $\begin{bmatrix} 56 & 62 & 91 & 266 \end{bmatrix}$, the proportional allocation method yields the same column share vector

$$\begin{bmatrix} 0.168 \\ 0.232 \\ 0.295 \\ 0.305 \end{bmatrix},$$

for each column, while the MSE method yields

$$\begin{bmatrix} 0.221 & 0.218 & 0.204 & 0.133 \\ 0.246 & 0.246 & 0.243 & 0.221 \\ 0.265 & 0.267 & 0.274 & 0.315 \\ 0.268 & 0.270 & 0.279 & 0.331 \end{bmatrix}$$

where the cost structure becomes increasingly differentiated as we pass from the first to the last column (for convenience of display, we have reordered rows and columns in increasing order). In the first column (the one with the smallest column total), the ratio of the largest column share element to the smallest is $0.268/0.221$, or 1.22; but in the last (largest) column it is $0.331/0.133$, or 2.49, more than twice as large. Thus in this example the MSE method yields very different cost structures in the smallest and largest industries, even though there is absolutely no relevant information on which to discriminate the cost structures.

Why does the MSE solution have this behavior? We can explain this in terms of the objective function, which is, as will be recalled, the simple sum of the entropies of each column's coefficients. Roughly speaking, to maximize the objective function, we try to maximize each column's entropy, subject to other considerations such as satisfying the constraints. Now for each column, the maximum entropy solution is that which puts all coefficients equal,

$$b_{ij} = \frac{1}{m}, \ i = 1, \ldots, m.$$

But we cannot adopt this solution for all columns j, since that would breach the row sum conditions (4). A solution that satisfies those conditions is the proportional allocation solution, but this does not maximize the objective function. Starting with the proportional allocation solution, we can increase the entropy of the column with the smallest total, by moving the

coefficients closer to the uniform value $1/m$. To maintain the row sum conditions, we make offsetting adjustments in the column with the largest total. Now in the row sum conditions, the columns are weighted by their totals $v_{i\bullet}$. So we can conserve the row sum conditions by making only large changes in the small-column coefficients but only small changes in the large coefficients. This gives us a large increase in the entropy of the small-total column, and only a small decrease in the entropy of the large-total column. Since both those entropies are given equal weight in the objective function, this leads to a net increase in the objective function. In short, the use of unweighted entropies in the objective function favors a solution in which the small-column coefficients are relatively concentrated, while the large-column coefficients are relatively dispersed.

So far, we have argued that the naïve-seeming proportional allocation solution is actually superior to the sophisticated-seeming entropy optimization solution—or, more accurately, to its MSE variant. For the proportional allocation model is itself an entropy optimization model.

Proposition 3 *Proportional allocation is a maximum entropy model.*

For want of space, we omit proof, but indicate briefly three routes by which the result may be obtained:

- Instead of the simple sum of column entropies in the MSE, use a weighted sum, where the weights are the column target totals.

- Instead of defining the objective function as a function of several column-specific entropies, use a single, whole-matrix entropy measure.

- Instead of maximizing an entropy measure, minimise inter-column cross entropy.

4 Entropy optimization and the RAS method

Consider now the matrix balancing problem: given initial estimates $u_{ij}, i = 1, \ldots, m, j = 1, \ldots, J,$ find new estimates v_{ij} as like as can be to the original v, but consistent with row target totals $v_{i\bullet}$ and column target totals $v_{\bullet j}$.

Here a traditional approach is the RAS method; find row scaling factors r_i and column scaling factors s_j, such that revised estimates

$$v_{ij} = r_i u_{ij} s_j \tag{11}$$

have the required target totals. In place of this approach GJR [11] propose a cross-entropy procedure: minimise

$$CE = \sum_i \sum_j b_{ij} \log \frac{b_{ij}}{a_{ij}},$$

where a_{ij} denotes the column share $u_{ij}/u_{\bullet j}$; subject to the constraints

$$\sum_j a_{ij} v_{\bullet j} = v_{i \bullet},$$

$$\sum_i a_{ij} = 1.$$

We note that the objective function CE is not itself a conventional cross entropy measure, but rather a sum of cross entropies,

$$CE = \sum_j \mathrm{CE}_j = \sum_i b_{ij} \log \frac{b_{ij}}{a_{ij}}, \tag{12}$$

where CE_j is the cross entropy between the initial and final estimates for column j. Combining these separate cross entropies into a single measure by simple summation is a natural choice, but not the only one. Recognizing that the GJR procedure is just one of many possible cross entropies procedures, we refer to it as the minimum sum of cross entropies, or *MSCE*, approach.

As GJR [11] show, the MSCE problem has solution

$$b_{ij} = \Omega_j^{-1} a_{ij} e^{-\lambda_i v_{\bullet j}},$$

where

$$\Omega_j = \sum_i a_{ij} e^{-\lambda_i v_{\bullet j}}.$$

From the solution we see immediately:

Proposition 4 *the MSCE revises cost structures more drastically for large industries than for small, in that, for any two inputs i_1, i_2, for any two industries j_1, j_2 such that $v_{\bullet j_1} < v_{\bullet j_2}$,*

$$\frac{v_{i_2 j_2}}{v_{i_1 j_2}} \Big/ \frac{u_{i_2 j_2}}{u_{i_1 j_2}} \begin{cases} < \frac{v_{i_2 j_1}}{v_{i_1 j_1}} \Big/ \frac{u_{i_2 j_1}}{u_{i_1 j_1}} & \textit{if} & \frac{v_{i_2 j_1}}{v_{i_1 j_1}} \Big/ \frac{u_{i_2 j_1}}{u_{i_1 j_1}} < 1 \\ = \frac{v_{i_2 j_1}}{v_{i_1 j_1}} \Big/ \frac{u_{i_2 j_1}}{u_{i_1 j_1}} & \textit{if} & \frac{v_{i_2 j_1}}{v_{i_1 j_1}} \Big/ \frac{u_{i_2 j_1}}{u_{i_1 j_1}} = 1 \\ > \frac{v_{i_2 j_1}}{v_{i_1 j_1}} \Big/ \frac{u_{i_2 j_1}}{u_{i_1 j_1}} & \textit{if} & \frac{v_{i_2 j_1}}{v_{i_1 j_1}} \Big/ \frac{u_{i_2 j_1}}{u_{i_1 j_1}} > 1 \end{cases}$$

In words: if the input ratio decreases between the initial and final estimates for industry j_1, it decreases more for the (larger) industry j_2; if the input ratio increases for industry j_1, it increases more for industry j_2; if it is constant for j_1, it is constant also for j_2.

9

Another way to put this is that the MSCE model tends to stay closer to the initial estimates for small industries (that is, for columns with small target totals), and to deviate from them more for large industries (large target totals). This would be appropriate if one considered the initial IO estimates more reliable for small industries than for large. In practice, it would seem rarely appropriate.

In comparing the two methods, we note that for work with non-negative data, the RAS method has several attractive properties which have no doubt contributed to its popularity and durability. It is easy to solve; indeed in small applications it is commonly solved using a simple iterative procedure without resort to optimization methods. Provided the initial estimates are not too sparse, it always provides a solution; if it does provide a solution, that solution is unique (see, e.g., Bacharach [1]). It preserves sign, without the need for side conditions; here it contrasts with alternative methods such as least squares, which in general require side conditions to avoid negative-valued solutions. And being a rescaling method, it is transparent, in that there is a simple relation (11) between the final and initial estimates.

The MSCE method has some of the same attractive properties. Provided the data are not too sparse, it can provide a solution, and the solution if it exists is unique. Like the RAS, it preserves sign. On the other hand, it is less transparent than the RAS. It is also less readily solved; though with solution software such as GAMS (Brooke, Kendrick, Meeraus, and Raman [4]), solution presents little difficulty even to the inexpert user.

There is however one important and desirable property that the RAS method has and MSCE does not:

Proposition 5 *For any pair of inputs i_1, i_2, the RAS preserves the ordering of input intensities across industries, in that for any pair of industries j_1, j_2,*

$$
\frac{v_{i_2 j_2}}{v_{i_1 j_2}}
\begin{cases}
\leq \frac{v_{i_2 j_1}}{v_{i_1 j_1}} & if \quad \frac{u_{i_2 j_2}}{u_{i_1 j_2}} \leq \frac{u_{i_2 j_1}}{u_{i_1 j_1}} \\
= \frac{v_{i_2 j_1}}{v_{i_1 j_1}} & if \quad \frac{u_{i_2 j_2}}{u_{i_1 j_2}} = \frac{u_{i_2 j_1}}{u_{i_1 j_1}} \\
\geq \frac{v_{i_2 j_1}}{v_{i_1 j_1}} & if \quad \frac{u_{i_2 j_2}}{u_{i_1 j_2}} \geq \frac{u_{i_2 j_1}}{u_{i_1 j_1}}
\end{cases}
$$

In words: if the ratio of usage of input i_2 to usage of input i_1 is smaller in industry j_2 than in industry j_1 in the initial estimates, then it is smaller also in the final estimates; if the ratio is greater in industry j_2 than in industry j_1 in the initial estimates, then it is greater in the final estimates; if the ratios are equal in the initial estimates, then they are equal also in the final estimates. In general, the MSCE estimates do not preserve the intensity ordering.

Before proving the proposition, we discuss why this property matters. Suppose for example that you have to update an IO table, for an economy in which the capital-labor ratio is increasing through time. In rebalancing the estimates, you expect then to increase the capital-labor ratios in individual industries. However you would normally do this in a way that preserves the industry ordering: the industry with the highest capital-labor ratio in the original data set would have

10

the highest capital-labor ratio in the rebalanced data; the industry with the lowest capital-labor ratio would be the same; and likewise with the intermediate rankings. This is because there is nothing in the new data to support any reversal in relative input intensities; there cannot be, since the new data contain no industry-specific information about cost structures. If you use the RAS to rebalance the data, it is guaranteed that the industry ordering will be preserved; if you use the MSCE method, changes in the ordering are possible.

Proof. For the RAS, for all columns j,

$$\frac{v_{i_2 j}}{v_{i_1 j}} = \frac{r_{i_2} u_{i_2 j} s_j}{r_{i_1} u_{i_1 j} s_j}$$
$$= \frac{r_{i_2}}{r_{i_1}} \frac{u_{i_2 j}}{u_{i_1 j}}.$$

So for given i_1, i_2, the vector of elements $v_{i_2 j}/v_{i_1 j}$, $j = 1, \ldots, J$, is a scalar multiple of the vector of elements $u_{i_2 j}/u_{i_1 j}$, by a positive scaling factor r_{i_2}/r_{i_1}. So the two vectors have the same ordering.

This establishes the positive part of the proposition, that concerning the RAS. The negative part, concerning the MSCE method, is readily established by counterexample. Consider initial estimates

$$\begin{bmatrix} 0.4142 & 0.6667 \\ 0.5858 & 1.3333 \end{bmatrix}$$

together with target row totals $\begin{bmatrix} 1.9191 & 1.0809 \end{bmatrix}'$ and target column totals $\begin{bmatrix} 1.0000 & 2.0000 \end{bmatrix}$. The initial column coefficients are

$$\begin{bmatrix} 0.4142 & 0.3333 \\ 0.5858 & 0.6667 \end{bmatrix},$$

so initially, the input 2 to input 1 usage ratio is $0.5858/0.4142 = 1.414$ for industry 1 and $0.6667/0.3333 = 2.000$ for industry 2; so the industry 2 to industry 1 relative intensity ratio is $2.000/1.414 = 1.414$; so industry 2 is relatively more input-2-intensive than industry 1. After rebalancing, the RAS yields revised estimates of

$$\begin{bmatrix} 0.6919 & 1.2272 \\ 0.3081 & 0.7728 \end{bmatrix},$$

so the revised input-output coefficients are

$$\begin{bmatrix} 0.6919 & 0.6136 \\ 0.3081 & 0.3864 \end{bmatrix},$$

the input 2 to input 1 usage ratios are $0.3081/0.6919 = 0.4453$ for industry 1, and $0.3864/ 0.6136 = 0.6297$ for industry 2, and the industry 2 to industry 1 relative intensity ratio is $0.6297/0.4453 = 0.4142$, so again industry 2 is relatively more input-2-intensive than industry

1; indeed, the relative intensity ratio is unchanged (this holds not just for this example but generally). In contrast, the MSCE method yields revised estimates of

$$\begin{bmatrix} 0.5858 & 1.3333 \\ 0.4142 & 0.6667 \end{bmatrix},$$

so the revised input-output coefficients are

$$\begin{bmatrix} 0.5858 & 0.6667 \\ 0.4142 & 0.3333 \end{bmatrix},$$

the input 2 to input 1 usage ratios are $0.4142/0.5858 = 0.7071$ for industry 1, and $0.3333/0.6667 = 0.5000$ for industry 2, and the industry 2 to industry 1 relative intensity ratio is $0.5000/0.7071 = 0.7071$, so now industry 2 is relatively less input-2-intensive than industry 1; indeed, the relative intensity ratio takes the reciprocal of its initial value (this is artificially neat—we specified the problem so as to ensure that exact reciprocity—but reversals in ordering are possible in realistic as well as in artificial examples). ∎

Proposition 6 *The RAS is a cross entropy minimization model.*

First proof. Instead of combining the column cross-entropies by simple addition, as in equation (12), combine them as a weighted sum,

$$\sum_j \frac{v_{\bullet j}}{v_{\bullet \bullet}} \mathrm{CE}_j$$

$$= \sum_i \sum_j \frac{v_{\bullet j}}{v_{\bullet \bullet}} b_{ij} \log \frac{b_{ij}}{a_{ij}}.$$

Minimizing this objective function is equivalent to minimizing

$$\sum_i \sum_j d_{ij} \log \frac{d_{ij}}{c_{ij}},$$

where c_{ij} represents the initial estimate $u_{ij}/u_{\bullet \bullet}$ of the share of the (i,j)th element in the matrix total, and d_{ij} the final estimate $v_{ij}/v_{\bullet \bullet}$. And that, as discussed immediately below, yields the RAS model.

Second proof. Rather than combine multiple column-specific cross entropies, we may use a single cross-entropy measure,

$$\mathrm{CE}_U = \sum_i \sum_j \frac{v_{ij}}{v_{\bullet \bullet}} \log \frac{v_{ij}/v_{\bullet \bullet}}{u_{ij}/u_{\bullet \bullet}}$$

$$= \sum_i \sum_j d_{ij} \log \frac{d_{ij}}{c_{ij}}.$$

With this objective function, the optimization problem is to minimize CE_U subject to the conditions

$$\sum_i v_{ij} = v_{\bullet j},$$

$$\sum_j v_{ij} = v_{i\bullet}.$$

Defining a Lagrangian

$$\mathcal{L} = \sum_i \sum_j d_{ij} \log \frac{d_{ij}}{c_{ij}} + \sum_i \lambda_i \left(\sum_j d_{ij} - \frac{v_{i\bullet}}{v_{\bullet\bullet}} \right) + \sum_j \mu_j \left(\sum_i d_{ij} - \frac{v_{\bullet j}}{v_{\bullet\bullet}} \right),$$

we obtain first order conditions

$$\log \frac{d_{ij}}{c_{ij}} = -1 - \lambda_i - \mu_j,$$

which, by convexity, ensure a global minimum. By inspection, the solution is of the biproportional functional form

$$d_{ij} = r_i c_{ij} s_j,$$

so it is the RAS model solution. ∎

As noted in section 2 above, that the RAS is a minimum cross-entropy solution has been well known for a long time. The "second proof" given above is the traditional one; the "first proof" is motivated by the IO coefficients interpretation of the problem.

5   Generalized entropy optimization and the RAS

Going beyond maximum entropy and minimum cross entropy as presented so far, GJR [11] and Golan, Judge and Miller ([10]; henceforward *GJM*) present a model they call *generalised maximum entropy*, *GME*, or *generalised cross entropy*, *GCE* (GJR [11] uses only the former name, but GJM[10] uses both). The details of the presentation differ slightly, GJR [11] taking a simpler, and GJM [10] a more general, approach. For the present purpose it is sufficient to present the simpler (GJR [11]) approach.

In the GJR [11] presentation, each coefficient b_{ij} is considered to be the expected value $\sum_m p_{mij} z_m$ of a random variable with support (range) $\mathbf{z} = \begin{bmatrix} z_1 & z_2 & \dots & z_M \end{bmatrix}$ and corresponding probabilities $\mathbf{p}_{ij} = \begin{bmatrix} p_{1ij} & p_{2ij} & \dots & p_{Mij} \end{bmatrix}$. Given prior estimates q_{mij} of the probability distributions, the generalised cross entropy model entails minimizing the objective function

$$H = \sum_m \sum_i \sum_j p_{mij} \log \frac{p_{mij}}{q_{mij}}$$

13

subject to constraints

$$\sum_m p_{mij} = 1, \tag{13}$$

$$\sum_i \sum_m p_{mij} z_m = 1, \tag{14}$$

$$\sum_j \sum_m p_{mij} z_m v_{\bullet j} = v_{i\bullet} \tag{15}$$

The coefficients b_{ij} are then recovered:

$$b_{ij} = \sum_m p_{mij} z_m.$$

It is not entirely clear what these probability distributions represent, or how the derivation of the IO coefficients as expected values of random variables is to be interpreted. Indeed, their proponents are not clear on how we ought to take them. On the one hand, GJR [11] suggest that we should take them seriously as probability distributions:

> *Note, however, we have estimates of the p_{mij} and this provides information concerning a probability distribution and an uncertainty measure for each of the $[b_{ij}]$. This is important since RAS techniques do not permit this and questions of this type are paramount when working with real data*

(their emphasis). On the other hand (in a more general context), GJM [10] describe the random variables in the GME and GCE models as "merely conceptual devices used to express the prior and sample knowledge in a mutually compatible format."

The GME/GCE approach is a large topic for investigation; but we confine ourselves here to a few preliminary points.

- If we choose to take the probability distributions seriously as probability distributions and not merely as conceptual devices (for example, if we are fortunate enough to possess some information about the distribution of the IO coefficients), then we need to consider the merits of the GME/GCE approach relative to traditional econometric approaches (see further Preckel [17]).

- If we choose to take the probability distributions as mere conceptual devices, we need to consider the merits of the GME/GCE approach relative to other methods of differentially weighting information, such as Byron's [6] method incorporating weights into the objective function.

14

- Unless we possess actual information about the coefficient distributions p_{mij}, we need to synthesize them. If we have prior coefficient estimates a_{ij}, then we can use them to impose expected value restrictions

$$\sum_m p_{mij} z_m = a_{ij}, \tag{16}$$

but in general the distributions remain underdetermined. Derivation of such distributions subject to the expected value restrictions (16) is a natural task for maximum entropy methods. For the GCE model as formulated by GJR [11], for example, a maximum entropy approach leads naturally to the Maxwell-Boltzmann distribution,

$$p_{mij} = \Omega_{ij}^{-1} e^{-\mu z_m},$$

where

$$\Omega_{ij} = \sum_m e^{\mu_{ij} z_m},$$

and μ_{ij} is a parameter determined implicitly by the restriction (16) (see, e.g., Kapur and Kesavan [12, subsection 3.2.1]).

- Taking as given the decision to adopt a GCE-like approach, deriving the IO coefficients as expected values of random variables, one may still question that feature of the GCE model that treats the variables underlying the different coefficients as independent. But by definition, the coefficients a_{ij} are subject to the restrictions

$$\sum_i a_{ij} = 1,$$

so arguably the corresponding random variables should be subject to like restrictions. More precisely, a single vector-valued random variable $\mathbf{A}_j = \begin{bmatrix} A_{1j} & \dots & A_{Ij} \end{bmatrix}'$ should underlie the coefficients $\begin{bmatrix} a_{1j} & \dots & a_{Ij} \end{bmatrix}'$, with the property

$$\sum_i A_{ij} = 1.$$

- In light of the discussion in previous sections of the relative merits of the RAS and MSCE models, we may find it more convenient and fruitful to apply a GCE-like approach to the entire use matrix, rather than on a column-by-column basis.

Consistent with these observations, we may define a "generalised generalised cross entropy" (*generalised GCE*) approach to the matrix balancing problem. With this approach, we treat the initial and final complete coefficient matrices c_{ij}, d_{ij} as expectations of random variables C_{ij}, D_{ij} with probability distributions p_m, q_m over a support $\mathbf{z} = \begin{bmatrix} z_1 & \dots & z_M \end{bmatrix}$, where

$$z_m = \begin{bmatrix} z_{m11} & \cdots & z_{m1J} \\ \vdots & & \vdots \\ z_{mI1} & \cdots & z_{mIJ} \end{bmatrix},$$

15

subject to the restrictions

$$\sum_i \sum_j C_{ij} \equiv 1,$$

$$\sum_i \sum_j D_{ij} \equiv 1,$$

that is, for all m,

$$\sum_i \sum_j z_{mij} = 1,$$

$$\sum_i \sum_j z_{mij} = 1.$$

We associate with the random variables y, z probability distributions p, q where p_m denotes the probability that C assumes the value z_m, and q_m denotes the probability that D assumes the value z_m. We require that the expected value of C be equal to c, that is, for all i, j,

$$\sum_m p_m z_{mij} = c_{ij}. \tag{17}$$

The model then entails constructing coefficients d_{ij} as expectations of random variables D_{ij} minimizing the cross-entropy

$$CE = \sum_m q_m \log \frac{q_m}{p_m}$$

subject to the restrictions

$$\sum_m \sum_i p_m z_{mij} = \frac{v_{\bullet j}}{v_{\bullet \bullet}} = d_{\bullet j},$$

$$\sum_m \sum_j p_m z_{mij} = \frac{v_{i\bullet}}{v_{\bullet \bullet}} = d_{i\bullet}.$$

Now constructing suitable prior probability distributions for this generalized generalized model might seem at first a daunting task. But in fact nothing could be easier; and as a bonus, doing it in the easiest way leads us directly to:

Proposition 7 *The RAS is a generalised GCE model.*

Proof. Let z_{gh} denote the matrix with (g, h)th element equal to 1 and all other elements equal to 0, and let $z_{11}, \ldots, z_{1J}, \ldots, z_{I1}, \ldots, z_{IJ}$ be the support for C and D. Note that, putting z_{ghij} for the (i, j)th element of z_{gh}, we have

$$z_{ghij} = \begin{cases} 1, & g = i \text{ and } h = j, \\ 0, & \text{otherwise.} \end{cases}$$

Then the condition (17) on the expectation of the prior estimates is satisfied uniquely by

$$p_{gh} = c_{gh}, \quad g = 1, \ldots, I, \quad h = 1, \ldots, J,$$

and the generalised GCE model becomes: minimize

$$CE = \sum_g \sum_h q_{gh} \log \frac{q_{gh}}{p_{gh}}$$

subject to the constraints

$$\sum_g \sum_h \sum_i p_{gh} z_{ghij} = \sum_i p_{ij} = d_{\bullet j},$$

$$\sum_g \sum_h \sum_j p_{gh} z_{ghij} = \sum_j p_{ij} = d_{i\bullet}.$$

But this is just the cross-entropy formulation of the RAS model. ∎

If one must apply the GCE in its original formulation, then it appears impossible to derive the RAS model as a special case. Comparison of GCE models with the RAS remains a subject for future research.

6 Useful uses of entropy optimization principles

As noted above, entropy optimization methods do not supersede the RAS; on the contrary, the RAS is itself an entropy optimization method. For the most commonly encountered cases in matrix balancing, it is probably the method of choice.

This does not mean however that the entropy-theoretic approach to matrix balancing is barren, or that the entropy-theoretic foundations of the RAS are an intellectual curiosity of no practical importance. Knowledge of these foundations is useful, not in superseding the RAS, but in extending it. Knowledge of the entropy-theoretic foundation enables us to taylor methods to problems, rather than forcing into the RAS framework problems that do not well fit into it.

For example, one candidate for this treatment is an extension of the use matrix balancing problem. Suppose we have two variants of the matrix, one at tax-exclusive and the other at tax-inclusive prices. We have target totals for total costs (tax-inclusive) and total sales (tax-exclusive) for each sector. Accordingly, we wish to rebalance the matrices to impose target column totals in the tax-inclusive matrix, and target row totals in the tax-exclusive matrix, while maintaining the tax rates implicit in the original pair of matrices.

One entropy-theoretic approach is to minimize a cross-entropy type objective function

$$I = \sum_i \sum_j v_{ij}^X \log \frac{v_{ij}^X}{u_{ij}^X} + \sum_i \sum_j v_{ij}^I \log \frac{v_{ij}^I}{u_{ij}^I},$$

17

where v_{ij}^X denote final tax-exclusive values, v_{ij}^I, final tax-inclusive values, and u_{ij}^X and u_{ij}^I the corresponding initial estimates. This leads to a solution of the form

$$v_{ij}^X = \frac{1}{\Omega_{ij}^X} e^{-\lambda_i - t_{ij}\nu_{ij}} u_{ij}^X,$$

$$v_{ij}^I = \frac{1}{\Omega_{ij}^I} e^{-\mu_j + \nu_{ij}} u_{ij}^I,$$

where t_{ij} are powers of taxes (ratios of tax-inclusive to tax-exclusive values), and the parameters λ_i, μ_j, and ν_{ij} are given by the conditions

$$\sum_j v_{ij}^X = v_{i\bullet}^X,$$

$$\sum_i v_{ij}^I = v_{\bullet j}^I,$$

$$t_{ij}v_{ij}^X = v_{ij}^I.$$

This is easier, simpler, and a better fit to the problem than the alternative RAS-based approach, applying the RAS to the tax-exclusive and tax-inclusive matrices separately, since that requires construction of tax-exclusive column totals and tax-inclusive row totals, values that lie outside the original problem and that cannot well be estimated in advance of the matrix balancing.

As an other example, consider the problem of disaggregating an IO use matrix, given data for disaggregate row and column totals, and initial estimates for the disaggregate matrix (these might be data from another period, or even from another country). We call the original matrix the *control matrix*, and the initial disaggregate estimates the *reference matrix*. We require that the final estimates meet the *reaggregation condition*, that on reaggregating sectors, we recover the control matrix. Here the entropy-theoretic approach leads to a tri-proportional model, in which each element of the final matrix is related to the corresponding element of the reference matrix by three factors: a row scaling factor, a column scaling factor, and a block scaling factor, where the block scaling factor is shared by all elements in the block of elements corresponding to a single element of the control matrix (see further the forthcoming PhD thesis by Jing Liu, Purdue University).

Using the entropy-theoretic approach and suitable, models suited to such problems can be formulated and implemented remarkably readily. Not only are they easily developed, but they are also remarkably transparent; using a general formula such as GJM (eq. 3.3.3), a scaling factor characterization of the solution can often be written down merely by inspection (this is true for the examples above). In bringing to the attention of economists further relevant works in the non-economic literature, GJR and GJM have helped put a powerful toolkit in the hands of economic data base developers.

7 Conclusions

To summarize the foregoing:

- The RAS is an entropy optimization method, and has long been known to be so.

- For the matrix filling problem, in general, the entropy optimization method of choice is proportional allocation.

- For the matrix balancing problem, in general, the entropy optimization method of choice is the RAS.

- If, following the GCE approach, we treat matrix elements as expected values of discrete random variables, the method of choice (in the absence of distributional data) is equivalent to the RAS.

- Entropy theory may fruitfully be used, not in attempting to supplant the RAS, but in extending and adapting it to problems that do not well fit the traditional matrix balancing framework.

References

[1] M. Bacharach. *Biproportional Matrices & Input-Output Change*. Number 16 in University of Cambridge Department of Applied Economics Monographs. Cambridge University Press, 1970.

[2] J. Bénard. Réseau des échanges internationaux et planification ouverte. *Economie Appliquée*, 16:249–76, 1963.

[3] L.M. Bregman. Proof of the convergence of Sheleikhovskii's method for a problem with transportation constraints. *USSR Computational Mathematics and Mathematical Physics*, 1(1):191–204, 1967.

[4] A. Brooke, D. Kendrick, A. Meeraus, and R. Raman. *GAMS Release 2.25 Version 92 Language Guide*. GAMS Development Corporation, Washington, D.C., 1997.

[5] J.P. Burg. *Maximum Entropy Spectral Analysis*. PhD thesis, Department of Geophysics, Stanford University, 1975.

[6] R.P. Byron. The estimation of large social account matrices. *Journal of the Royal Statistical Society, Series A*, 141:359–67, 1978.

[7] W.E. Deming and F.F. Stephan. On a least-squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11:427–44, 1940.

[8] University of Cambridge Department of Applied Economics. *A Computable Model of Economic Growth*. Number 1 in A Programme for Growth. Chapman and Hall, London, 1962.

[9] University of Cambridge Department of Applied Economics. *Input-Output Relationships 1954–66*. Number 3 in A Programme for Growth. Chapman and Hall, London, 1963.

[10] A. Golan, G. Judge, and D. Miller. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Wiley, Chichester, 1996.

[11] A. Golan, G. Judge, and S. Robinson. Recovering information in the case of underdetermined problems and incomplete data. *Review of Economics and Statistics*, 76:541–9, 1994.

[12] J.N. Kapur and H.K. Kesavan. *Entropy Optimization Principles with Applications*. Academic Press, New York, 1992.

[13] J. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35:917–26, 1956.

[14] A.N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Transactions in Information Theory*, IT-14:662–4, 1968.

[15] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.

[16] W.W. Leontief. *Structure of American economy, 1919-1929; an empirical application of equilibrium analysis*. Harvard, 1941.

[17] P.V. Preckel. Least squares and entropy as penalty functions. Staff Paper 98–16, Department of Agricultural Economics, Purdue University, 1998.

[18] M.H. Schneider and S.A. Zenios. A comparative study of algorithms for matrix balancing. *Operations Research*, 38:439–55, 1990.

[19] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–659, 1948.

[20] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Annals of Mathematical Statistics*, 35:876–9, 1964.

[21] R. Stone. Multiple classifications in social accounting. *Bulletin de l'Institut International de Statistique*, 39(3):215–33, 1962.

[22] H. Theil. *Economics and Information Theory*. North-Holland, Amsterdam, 1967.

[23] P. Uribe, C.G. de Leeuw, and H. Theil. The information approach to the prediction of interregional trade flows. Report 6507, Econometrics Institute of the Netherlands School of Economics, Rotterdam, 1965.