KURODA'S METHOD FOR
CONSTRUCTING CONSISTENT INPUT-OUTPUT DATA SETS

by

Peter J. Wilcoxen

Impact Research Centre, University of Melbourne

April 1989

This paper describes a method that can be used to resolve inconsistencies between input-output data from various sources. It belongs to the class of algorithms discussed by Bacharach (1970) and has been used extensively by Kuroda (1988) and Wilcoxen (1988).

In input-output analysis it is often necessary to use inconsistent data sets originating from different government agencies. For example, the table of interindustry transactions created by one agency may not be consistent with value-added and final demand vectors produced elsewhere. In this case, the investigator will be confronted with three pieces of data which do not agree: a table of interindustry transactions, a vector of commodity outputs, and a vector of gross outputs by industry. The task then becomes adjusting the transactions table to match the commodity and industry output vectors.

In the past, this problem has been solved by using the RAS method. RAS is an iterative algorithm which scales the rows and columns of the transactions table up and down repeatedly until the table's row and column sums agree with the target vectors. It has been shown that RAS will eventually converge, but the result will not necessarily be close in any economic sense to the original transactions table. The purpose of this paper is to define a measure of how far a new transactions table is from the original, and to derive an algorithm which will construct a table minimizing that distance.

Given an $n \times m$ matrix X^o of initial data, define r_{ij} and c_{ij} to be the shares of each element in the row and column sums of the original matrix:

$$r_{ij} = \frac{X_{ij}^o}{\sum_{j=1}^{m} X_{ij}^o}, \qquad c_{ij} = \frac{X_{ij}^o}{\sum_{i=1}^{n} X_{ij}^o} \tag{1}$$

Let R be a vector of target row totals, and C a vector of desired column totals. The following function can then be used to measure the distance between a revised matrix X and the original (embodied in r and c), where w and v are arbitrary sets of weights:

$$Q = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{X_{ij}}{R_i} - r_{ij}\right)^2 w_{ij} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{X_{ij}}{C_j} - c_{ij}\right)^2 v_{ij} \tag{2}$$

It is now possible to choose X to minimize this function subject to the following constraints:

$$R_i = \sum_{j=1}^{m} X_{ij} \tag{3}$$

$$C_j = \sum_{i=1}^{n} X_{ij} \tag{4}$$

The Lagrangian for this problem is:

$$L = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{X_{ij}}{R_i} - r_{ij}\right)^2 w_{ij} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{X_{ij}}{C_j} - c_{ij}\right)^2 v_{ij}$$

$$+ \sum_{i=1}^{n}\lambda_i\left(R_i - \sum_{j=1}^{m} X_{ij}\right) + \sum_{j=1}^{m}\mu_j\left(C_j - \sum_{i=1}^{n} X_{ij}\right) \tag{5}$$

Taking first order conditions gives:

$$\frac{\partial L}{\partial X_{ij}} = \left(\frac{X_{ij}}{R_i} - r_{ij} \right) \frac{w_{ij}}{R_i} + \left(\frac{X_{ij}}{C_i} - c_{ij} \right) \frac{v_{ij}}{C_i} - \lambda_i - \mu_j = 0 \tag{6}$$

Collect terms in X_{ij}, and for convenience make the following definitions:

$$S_{ij} = \left(\frac{w_{ij}}{R_i^2} + \frac{v_{ij}}{C_j^2} \right)^{-1} \tag{7}$$

$$G_{ij} = \frac{r_{ij} w_{ij}}{R_i} + \frac{c_{ij} v_{ij}}{C_j} \tag{8}$$

This allows the first order conditions to be rewritten as shown

$$X_{ij} = S_{ij} (\lambda_i + \mu_j + G_{ij}) \tag{9}$$

Both S and G depend only on initial data and the weights w and v, so it is only necessary to determine λ and μ, to find the optimal X_{ij}. These may be determined by applying the constraints:

$$R_i = \sum_{j=1}^{m} X_{ij} = \sum_{j=1}^{m} S_{ij} (\lambda_i + \mu_j + G_{ij}) \tag{10}$$

$$C_j = \sum_{i=1}^{n} X_{ij} = \sum_{i=1}^{n} S_{ij} (\lambda_i + \mu_j + G_{ij}) \tag{11}$$

Define diagonal matricies S^R and S^C as indicated:

$$S_{ii}^R = \sum_{j=1}^{m} S_{ij}, \qquad S_{jj}^C = \sum_{i=1}^{n} S_{ij} \tag{12}$$

In matrix notation, the constraints can now be expressed as

$$S^R \cdot \lambda + S \cdot \mu = R - A \tag{13}$$

$$S' \cdot \lambda + S^C \cdot \mu = C - B \tag{14}$$

where A and B are vectors defined as follows:

$$A_i = \sum_{j=1}^{m} S_{ij} G_{ij}, \qquad B_i = \sum_{i=1}^{n} S_{ij} G_{ij} \tag{15}$$

With more manipulation, it is possible to derive explicit formulae for λ and μ. For computational purposes, however, it is better to arrange the equations into the following system, which can be solved easily by any competent numerical package:

$$\begin{bmatrix} S^R & S \\ S' & S^C \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} R-A \\ C-B \end{bmatrix} \tag{16}$$

Armed with the values of λ and μ, the optimal choice of X_{ij} can be computed directly.

It is worthwhile to examine a few of the possible weighting schemes that can be used. The most obvious approach is to weight all errors equally, which means that $w_{ij}=v_{ij}=1$ for all i and j. On the other hand, the following choice of weights results in a drastic simplification of the revision formula:

$$w_{ij} = \frac{1}{2}R_i^2, \qquad v_{ij} = \frac{1}{2}C_j^2 \tag{17}$$

This means that $S_{ij}=1$ for all i and j. Moreover, the following is true of G_{ij}:

$$G_{ij} = \frac{1}{2}(r_{ij}R_i + c_{ij}C_j) \tag{18}$$

which is simply the average of the values obtained by applying the original shares to the target row and column sums. Furthermore, it can be shown that all of the following are true:

$$\sum_{i=1}^{n}\lambda_i = 0, \qquad \sum_{j=1}^{m}\mu_j = 0, \tag{19}$$

$$S_{ij}^R = m, \qquad S_{ij}^C = n. \tag{20}$$

This means that the revision formula has a particularly simple form:

$$X_{ij} = G_{ij} + \frac{1}{m}(R_i - \sum_{j=1}^{m}G_{ij}) + \frac{1}{n}(C_j - \sum_{i=1}^{n}G_{ij}) \tag{21}$$

Thus, the revised X_{ij} is just G_{ij} (which has the interpretation above) adjusted to correct the row and column sums. It is important to note that this method does not guarantee that all elements of X will be nonnegative.

Kuroda proposes a different weighting scheme, with w and v determined by the following equations:

$$w_{ij} = \frac{1}{r_{ij}^2}, \qquad v_{ij} = \frac{1}{c_{ij}^2}. \tag{22}$$

This choice of weights causes Q to be a function of the percentage changes in the coefficients:

$$Q = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{X_{ij}/R_i}{r_{ij}} - 1\right)^2 + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{X_{ij}/C_j}{c_{ij}} - 1\right)^2 \tag{23}$$

In most cases, this will ensure that all elements of X are positive, since making one negative would require a change of more than one hundred percent, resulting in a large value of Q. It is possible, however, for negative numbers to arise if the row and column targets differ substantially from the corresponding totals of the initial array.

REFERENCES

Bachrach, Michael (1970), *Biproportional Matricies and Input-Output Change*, Cambridge: Cambridge University Press.

Kuroda, M. (1988), "A Method of Estimation for the Updating Transaction Matrix in the Input-Output Relationships," in *Statistical Data Bank Systems*, K. Uno and S. Shishido (eds.), Amsterdam: North Holland.

Wilcoxen, P.J. (1988), "The Effects of Environmental Regulation and Energy Prices on U.S. Economic Performance," Ph.D. Thesis, Harvard University.