THE BALANCING AND RECONCILIATION OF INPUT-OUTPUT TABLES

by

W.I. Morrison

University College London

and

Dar Al Handasah Consultants

African Regional Conference (Session 3)

International Association for Research
in Income and Wealth

United Nations Economic Commission for Africa

Douala
November 1982

THE BALANCING AND RECONCILIATION OF INPUT-OUTPUT TABLES

1. Introduction

The construction of a series of input-output tables is increasingly being seen as an integral part of the production of national accounts data, and the United Nations' system of national accounts (SNA) incorporates an input-output approach as a central component (United Nations Statistical Office, 1968). This paper examines some of the problems involved in constructing input-output tables in the context of the SNA, and in particular focuses on the fundamental questions of balancing and reconciliation.

A wide range of methods has been proposed for use in balancing and reconciling input-output tables. Some of these are specific to the input-output approach, whereas others consist of more general algorithms borrowed or adapted for input-output purposes. In this paper it is suggested that in circumstances where some survey data are available the initial emphasis should be placed on techniques which involve the removal of major statistical discrepancies by the further investigation of the basic data. More mathematical approaches are thought to be most appropriate for making final adjustments to the tables, for estimating tables where very few data exist, or in circumstances where the up-dating of a matrix is the prime subject of concern. It is shown that the more relevant of the mathematical approaches rely to a greater or lesser extent on the judgement of the analysts involved with a particular exercise, and so it is argued that these techniques should be seen as extensions to the detailed examination of the basic data rather than as an alternative approach.

The basis of the input-output calculations in the UN SNA is the input-output data system, which distinguishes between accounts on commodities and accounts on industries. The system consists of a number of accounting blocks, and the relationship between these blocks, and the overall structure of the system, can be simplified and shown in diagrammatic form (see Figure 1).

Figure 1 The Input-Output Data in the SNA

	Commodities	Industries	Final buyers Net	Totals
Commodities		U	e	q
Industries	V			g
Value added		y'		η
Totals	a'	g'	η	

In this figure an upper case letter represents a matrix, a lower case letter a vector (with a prime superscript showing a row vector), and a Greek letter a scalar. Thus:

U = Absorption of commodities as intermediate inputs by industries (dimension j x k, where there are j commodities and k industries)

e = Net final use of commodities (dimension j x 1)

q = Domestic output of commodities (j x 1)

V = Production of commodities by industries (k x j)

y' = Primary inputs (value added) of industries (1 x k)

η = Sum of value added in each industry ≡ Sum of net final demand for each commodity

If we define the unit column vector i:

$$q = Ui + e \quad (1)$$

$$q = V'i \quad (2)$$

$$g = Vi \quad (3)$$

The matrix U is generally known as an absorption (or use) matrix, and the matrix V is generally called a make matrix. (The vectors e and y' may also be disaggregated and consist of matrices showing final demand or primary inputs.)

The arrangement of input-output data as recommended in the SNA is best seen as a long-term goal (McGilvray and Morrison, 1982). In many countries, and particularly in developing countries, the data available will enable only part of the framework to be completed. Gaps in the data will thus have to be filled, and in addition inconsistencies in the data will have to be removed if the three arithmetic identities (equations 1-3) are to hold. Stated simply, the balancing and reconciliation of input-output tables involves the successful completion of these equations. The starting point of this paper is an assumption that some attempt has been made to construct the absorption and make matrices, U and V.

2. Commodity balances

The UN SNA represents a compromise between the ultimate use of the framework as a base for economic model building, and the feasibility of assembling the component data. In order to minimise data problems and to simplify the survey work needed, the main focus is placed on obtaining data on the cost structure of industries (the U matrix) and the production of commodities by industries (the V matrix). It is felt that data on the intermediate and primary inputs which are used in the production of each commodity by each industry, and data on the intermediate and final uses to which the establishments of various industries send their products are not practical to collect (Aidenoff, 1970).

If the UN SNA approach is followed, in general there will thus be only one entry (if there is any entry) in each of the cells of the make and absorption matrices.[1] The entries in the make matrix will be expressed in terms of producers' prices, and those in the absorption matrix will be shown in most instances in terms of purchasers' prices. The difference between the two sets of prices reflects, at least in theory, the margins - trade and transport - which are added to the producers' prices and which are incorporated in the price paid by the purchaser.

Recall equation (1):

$$q = Ui + e \tag{1}$$

which indicates that the total output of any commodity contained in the vector q is equal to the total intermediate consumption of that commodity, together with final demand.

Equation (2):

$$q = V'i \tag{2}$$

indicated that the total output of any commodity was equal to the sum over all industries of the commodity in question. Clearly, if the elements of U are expressed in purchasers' prices and the elements of V are in producers' prices, the right hand side of equation (1) will not equal the right hand side of equation (2). For this equality to hold, a vector of sectoral margins (g) has either to be subtracted from the right hand side of equation (1) or added to the right hand side of equation (2), depending whether a balance is to be made in terms of producers' or purchasers' prices. This gives:

---

[1] This means that much of the debate which has taken place on the balancing of regional input-output tables is only of limited relevance, since the literature is based on an assumption that in general each cell in the table will contain two entries, one relating to the reported sales and one related to the reported purchases of establishments in a sector. Reconciliation of these conflicting elements is thus the basic problem. For further details on this point see Jensen and McGaurr (1976; 1977), Gerking (1976a; 1979a).

$$q^* = Ui + e - g = V'i \quad (4)$$

where the * superscript indicates producers' prices, and

$$q = V'i + g = Ui + e \quad (5)$$

which is the balance in purchasers' prices.

In most input-output exercises, especially in circumstances where only partial data are available, commodity balancing will initially involve an approach based on equation (4), since the absorption matrix, U, can only be expressed in terms of producers' prices if the relative proportion of the purchases of each commodity passing through the trade sector, the differential margins on these commodities and the transport margins on all commodity inputs are known. Such data rarely exist, even in countries with well-developed statistical services, and proportionality assumptions are inevitably made.

A further problem in moving towards the initial commodity balances concerns the treatment of imports. In the same way that it is difficult to estimate the proportion of commodity inputs into each industry which have passed through the trade sector, it is difficult to estimate the proportion which consists of imports. This is the case even when detailed survey returns are available, since many imported commodities are obtained from wholesalers or agents. The problem of distinguishing between domestic and imported commodities applies both to the elements of the absorption matrix and to final demand consumption, so that a further modification to the initial balance equation is needed, in the form of a vector of imports (m). Thus equation (4) must be further modified to provide a base for the development of commodity balances:

$$q^* = Ui + e - g - m \quad (6)$$

Lal (1982) discusses the question of commodity balancing in the Canadian context, although his conclusions have general validity. He identifies three separate phases in the preparation of commodity balances. The first, and obvious one, is a system of classification for commodities. The second is to fill the gaps in the available data bases, and the third is to ensure that supply matches demand for each commodity. The classification system is extremely important, but even when international conventions have been adopted problems will still occur in practice. One set of problems concerns the need to link production data with trade statistics and the difficulties of mapping a data set based on the BTN or the SITC onto ISIC data were referred to in another paper prepared for this conference (Morrison, 1982). But other problems concern the difficulties of allocating products to an appropriate sector, since this process often requires a very detailed knowledge of industrial technologies. The more disaggregation, the greater the problems. Experience would suggest that the commodities which are most difficult to allocate precisely are those in the chemical sectors, but metal products and food products (where there is overlap with agriculture) also pose problems for the analyst.

The second phase in commodity balancing - the completion of the data sets - requires the integration of a wide range of data of differing quality, and may involve ad hoc studies and additional surveys in order to complete the basic framework. Thus at this stage a preliminary estimate of margins might be made by allocating many of the initial discrepancies to the margins column, although clearly this can only be a first step and will always require subsequent verification. (It could even lead to margins which appear to be negative, so the procedure should not be adopted in a purely mechanical way.)

The major part of the commodity balancing process involves the detailed examination of each of the rows identified in the analysis, in order to remove inconsistencies between the total supply and the total demand for each commodity. The words of Lal carry a message which will be only too familiar to anyone who has undertaken this exercise. "There are no ready-made statistical approaches to solving such imbalances. The only approach is laborious investigation; one has to go back to the basic records to locate the sources of such imbalances" (Lal, 1982, p. 13).

But first, the distribution of demand will need to be looked at, to see if any obvious consuming sectors have been inadvertently omitted. All large values should be examined, to see if there has been over-estimation. The unique features of certain sectors will be noted, so that, for example, account can be taken of those rows with large numbers of small producers (such as bakeries or clothing), or those rows which export the majority of their output (for example, mining) or which sell to a small or a large number of other sectors.

When commodity balancing is undertaken a useful exercise is to make a distinction between sectors which are especially easy to identify or distinguish and all others. The former category will consist of sectors which contain only one or two key establishments, or which are especially important in terms of their contribution to export earnings. If these sectors contain few establishments balancing and checking is often easier, and so there are advantages to be gained in starting the balancing process with these "high profile" sectors, which then gives a base for subsequent analysis and balancing work.

Balancing is an iterative process, since any amendments introduced in one part of the data set have implications elsewhere. For example, increasing the demand for any commodity may mean that the supply of an industry's output may also have to be increased. This in turn may have further implications for other sectors in the economy, or for the margins in the sector. If output figures are changed, it may be necessary to change valued added, which in turn will influence GDP. But above all, any changes introduced must be consistent with the establishment level data available, and once these data have been checked they should be regarded as fixed. Obviously, there will be more scope to introduce changes if the available data only relate to a small proportion of the number of known establishments in a sector - but even here there are usually certain structural parameters which restrict the extent of the changes which can be made.

The process of commodity balancing forces the analyst to reconcile data at a series of levels, from the micro to the macro. Establishment data will have to be reconciled with aggregate trade statistics, data for different sectors will have to be matched, and the sum of all final demand, equal to the sum of all value added, will have to be related to estimates of gross domestic product at market prices and factor cost. This can pose further difficulties, particularly if survey-based work results in major revisions of GDP, or if data permitting the allocation of indirect taxes to individual sectors are unavailable.

It is not the intention of this paper to examine all the problems involved in constructing input-output tables, but in this discussion on balancing some mention must be made of secondary production. Whether an input-output table is balanced before or after secondary production is transferred - assuming that secondary production is to be transferred - is to some extent a matter for the analyst to decide. However, the methodology adopted, and the extent of secondary production, will have implications for the balancing process, and if an establishment level data base is to be maintained it will generally be preferable to undertake the commodity balancing before the transfer of secondary production. This will enable the data for any defined sector to be related directly to a set of establishment records, and will thereby reduce the scope for introducing arbitrary modifications.

The balancing and reconciliation of input-output tables is thus a painstaking process involving the detailed investigation of micro data. It should not involve the use of a sector or a group of sectors as a sump to achieve an artificial balance. Equally, mechanical adjustment procedures will certainly produce a balanced table - but the table will be a display of assumptions and will be of no statistical value (Lal, 1982). Yet there do exist certain adjustment procedures which can be of value when input-output tables are being constructed, and it is to these that we now turn.

3.0 General approaches to balancing

There inevitably comes a time in an input-output study when the limits of the available data are reached, even though a global balance has not been achieved. Clearly, the fewer the data available, the sooner this point is reached, and even if a series of micro data sets is accessible the data may relate to different time periods, so that at the very least some deflation or inflation will be needed. In this way discrepancies can be introduced - discrepancies which cannot be removed through simple data examination. Alternative methods have thus to be adopted to obtain a balanced input-output table.

Three general approaches to the production of input-output tables can be identified. The first involves the judgement of the analyst, and can best be described as a pragmatic approach. The second, which may also involve the judgement of the analyst, requires the implementation of one of the methods available for solving what is generally known as the constrained matrix problem. The third, best described as a stochastic approach, relies rather more on statistical estimation procedures.

3.1 The pragmatic approach

It is easy to criticise a pragmatic approach, on the grounds that it is subjective, that it is unscientific, that it requires too many arbitrary assumptions, and so on. Yet pragmatism has much to offer, especially if a team of people has been involved in the various stages of an input-output exercise over a period of time. In this way the analysts will have required a feel for the data, and an understanding of the economic structure of the spatial unit for which the input-output table is being constructed. Discussion and compromise can then make a major contribution.

Pragmatism has a major part to play in input-output exercises, and should not therefore be ruled out of order. What must be remembered, however, is that the construction of input-output tables should be seen as part of the wider and long-term planning process. Thus new data sets and revisions of old data will be available at various times, so enabling input-output tables to be up-dated. In this context, it is important to note that detailed records should be kept of all decisions made, so that earlier assumptions can be modified and improved upon as later data become available.

3.2 The constrained matrix approach

Methods available for solving the constrained matrix problem range from the purely mechanical to the highly subjective. The basic problem involves the production of an interaction matrix, given a certain amount of prior information which can be used to constrain the estimation procedure. The problem may be approached in two ways (Bacharach, 1970). One involves the specification of a simple form, which gives a solution matrix as a function of some given matrix. The other requires that the solution matrix be obtained by minimising a criterion (e.g. distance), subject to the specified constraints. Sometimes the two approaches can be regarded as identical.

A variety of mathematical methods have been used to solved the problem, but by far the best known technique is the RAS approach, which produces a balanced input-output matrix, given a base matrix and constraints on the row and column sums of the new matrix (for a full exposition, see Bacharach, 1970). Dissatisfaction with the mechanical nature of the row and column scaling process in the basic RAS model led Lecomber (1975) to argue for the inclusion of more information ("even that which is not fully reliable or appropriate"), and to suggest a generalised RAS approach which enabled individual elements to be constrained in addition to the row and column totals.

Other approaches to the solution of the constrained matrix problem have involved entropy maximisation techniques (Wilson, 1974), an additive model (Friedlander, 1961), linear programming (Matuszewski, Pitts and Sawyer, 1964) and the concept of information inaccuracy (Theil, 1967). But as pointed out by Morrison and Thumann (1980), a feature common to all these techniques is that they cannot incorporate additional linear constraints relating to subsets of elements in the new matrix, or qualitative information.

Morrison and Thumann (1980) demonstrated that an appropriate objective function could be developed, which permitted the estimation of a new input-output matrix as the solution of a constrained optimisation problem, and it was shown that linear solutions could be obtained and computed. In their formulation the new matrix elements are expressed as linear combinations of the base matrix elements, and the coefficients are determined by the constraints. Treatments of this kind have also been outlined by Stone (1961; 1970), Byron (1978) and van der Ploeg (1982), and a common feature is that all permit the inclusion of subjective data. Moreover, as shown by Hyman and Morrison (1980), the choice of the objective function can also determine the relative weight which is placed on the base matrix or the new matrix.

These methods should be regarded as further weapons to be included in the pragmatic armoury, so that the subjective component should be seen as a strength rather than a weakness. If there is a weakness of these methods it is that negative elements can appear in the new matrix, although experience to date would suggest that these can usually be removed with little difficulty by imposing non-negativity constraints and re-estimating.

Even this problem can be overcome if a quadratic programming formulation is adopted, and recent work by Harrigan and Buchanan (1982) has shown that such a formulation can also permit the introduction of upper and lower bounds to the constraints. This represents a further move towards the incorporation of a wide variety of data sources and subjective estimates in the estimation procedure, and would appear to be an indicator of future developments in the field, since Harrigan and Buchanan, building on earlier work by Hildreth (1957) and Bachem and Korte (1978), demonstrate that computational capacity is no longer a constraint on the estimation procedure.

3.3 Stochastic approaches

The stochastic approach to the estimation of input-output data, as suggested by Gerking (1976b; 1976c), is, strictly speaking, not relevant to the question of balancing, since much of the debate has concerned the most appropriate methodology for estimating coefficient data in individual columns of a table. The basic approach involves estimating the technical coefficients by cross-sectional analysis, based on the assumption that all firms in each sector have the same production function. The methodology has come in for some criticism, and modifications have been proposed by Brown and Giarratani (1979) and Hanseman and Gustafson (1981). It would seem that the approach has little to offer in sectors where the number of establishments is small, and even where the sample available is large, the basic assumptions appear to be somewhat daunting.

In an attempt to overcome some of the earlier criticisms, Gerking (1979a) extended the approach to include row, column and non-negativity constraints to the minimum variance reconciliation method adopted, although the methodology could well be difficult to implement for computational reasons if the number of sectors in the model is large. Interestingly, Gerking (1979a) also discusses ways in which a priori information of a qualitative nature can be included in the estimation procedure, but here too further empirical testing is needed.

4.0 Conclusions

The importance of the balancing stage in input-output studies cannot be over-emphasised, and the value of a reliable base matrix is all the greater when subsequent applications work is undertaken. Professional judgement has a major role to play in any balancing exercise, and of the mathematical approaches to balancing, those which permit the inclusion of additional information in the form of constraints on the estimation procedure have most to offer. Mechanical techniques, involving the proportionate allocation of differences over all sectors, have few attractions, and more recent stochastic approaches also appear to have little to contribute to the balancing stage (which is not to reject the use of this approach in different contexts - for example, in forecasting).

Finally, this paper has discussed approaches to balancing and reconciliation of input-output tables in those situations where some form of make and absorption matrices is available. There are often circumstances in which such matrices do not exist, even in rudimentary form, and here it is the constrained matrix approaches which also have most to offer, since these methods permit the inclusion of the maximum amount of available information in the estimation process. Equally, constrained matrix approaches have much relevance when additonal input-output tables (for example, an imports matrix or a projected matrix) are being estimated. The construction of input-output tables is very much a long-term process. Bench-mark tables will be produced, but there should also be regular revision of the data sets. In this process different methodologies and techniques will contribute at different stages, but practical solutions and professional judgement will always be required even in the most sophisticated model-building exercises.

5. References

Aidenoff, A. (1970) 'Input-output data in the United Nations System of National Accounts'. Chapter 17, pp.349-368 of Carter, A.P. and Brody, A., Applications of Input-Output Analysis, North-Holland, Amsterdam.

Bacharach, M. (1970) Biproportional Matrices and Input-Output Change, Cambridge University Press, Cambridge.

Bachem, A. and Korte, B. (1978) 'An algorithm for quadratic optimisation over transportation polytopes'. Z Agnew Mathematik und Mechanik, 58, 459-461.

Brown, D.M. and Giarratani, F. (1979) 'Input-output as a simple econometric model: a comment'. The Review of Economics and Statistics, 61, 621-623.

Byron, R.P. (1978) 'The estimation of large social account matrices'. Journal of the Royal Statistical Society, Series A, 141, 359-367.

Friedlander, D. (1961) 'A technique for estimating a contingency table, given the marginal totals and some supplementary data', Journal of the Royal Statistical Society, Series A, 124, 412-420.

Gerking, S.D. (1976a) 'Reconciling 'rows only' and 'columns only' coefficients in an input-output model', International Regional Science Review, 1, 30-46.

Gerking, S.D. (1976b) 'Input-output as a simple econometric model', The Review of Economics and Statistics, 58, 274-282.

Gerking, S.D. (1976c) Estimation of Stochastic Input-Output Models. Martinus Nijhoff, Leiden.

Gerking, S.D. (1979a) 'Reconciling reconciliation procedures in regional input-output analysis', International Regional Science Review, 4, 23-36.

Gerking, S.D. (1979b) 'Input-output as a simple econometric model: Reply', The Review of Economics and Statistics, 61, 623-626.

Hanseman, D.J. and Gustafson, E.F. (1981) 'Stochastic input-output analysis: a comment', The Review of Economics and Statistics, 63, 468-470.

Harrigan, F.J. and Buchanan, J.T. (1982) 'A quadratic programming formulation of a constrained matrix problem', Fraser of Allander Institute, University of Strathclyde, Glasgow.

Hildreth, C. (1957) 'A quadratic programming procedure', Naval Research Logistic Quarterly, 4, 79-85.

Hyman, G. and Morrison, W.I. (1980) 'Linear approaches to up-dating interaction matrices', Sistemi Urbani, 2-3, 243-256.

Jensen, R.C. and McGaurr, D. (1976) 'Reconciliation of purchases and sales estimates in an input-output table', Urban Studies, 13, 59-65.

Jensen, R.C. and McGaurr, D. (1977) 'Reconciliation techniques in input-output analysis: some comparisons and implications', Urban Studies, 14, 327-338.

Lal, K. (1982) 'Compilation of input-output tables: Canada', pp.8-36 of Skolka, J.V. (1982)

Lecomber, J.R.C. (1975) 'A critique of methods of adjusting, updating and projecting matrices, together with some new proposals'. Chapter 1 of Allen, R.I.G. and Gossling, W. (Eds.) Estimating and Projecting Input-Output Coefficients, Input-Output Publishing Company, London.

Matuszewski, T.I., Pitts, P.R. and Sawyer, J.A. (1964) 'Linear programming estimates of changes in input-output coefficients', Canadian Journal of Economics and Political Science, 3, 203-210.

McGilvray, J.W. and Morrison, W.I. (1982) 'The compilation of input-output tables in developing countries', pp.235-251 of Skolka, J.V. (1982).

Morrison, W.I. (1982) 'Micro-data bases and the preparation of national accounts'. Paper prepared for African Regional Conference, IARIW/UNECA, Douala.

Morrison, W.I. and Thumann, R.G. (1980) 'A Lagrangian multiplier approach to the solution of a special constrained matrix problem', Journal of Regional Science, 20, 279-292.

Ploeg, F. van der (1982) 'Reliability and the adjustment of sequences of large economic account matrices', Journal of the Royal Statistical Society, Series A, 145, 169-194.

Skolka, J.V. (Ed.) (1982) Compilation of Input-Output Tables. Lecture Notes in Economics and Mathematical Systems 203, Springer-Verlag, Berlin.

Stone, R. (1961) Input-Output and National Accounts, OECD, Paris.

Stone, R. (1970) Mathematical Models of the Economy and Other Essays. Chapman and Hall, London.

Theil, H. (1967) Economics and Information Theory. North-Holland, Amsterdam.

United Nations Statistical Office (1968) A System of National Accounts. Studies in Methods, Series F, No.2, Rev.3, United Nations, New York.

Wilson, A.G. (1974) Urban and Regional Models in Geography and Planning. Wiley, London.