A Method of Estimating the Elements of an Interindustry Matrix Knowing the Row and Column Totals

R. C. GEARY

It is a pity, but a fact of life, that few (if any) countries produce IO tables annually. The more common situation is that many countries produce these tables at intervals of years and then only with long delays. However, economic analysts cannot do without IO tables of reasonable up-to-dateness. Their position often is that they have a full IO table relating to five or six years ago, but (with reasonable accuracy from production statistics and national income accounts) only the *marginal totals* for more recent annual tables. These margins consist of gross output of agriculture (possibly in several divisions) and of industries and services, elements of final demand and of primary inputs for each such branch of economic activity.

Here we are concerned with the interindustry part of the table. We assume that we know the row and column totals (all positive) of this part for a "current" year. Our problem is to estimate the non-zero elements of the current matrix. Provided that the number N of non-zero elements to be estimated exceeds $(m+n-1)$, m and n being row and column numbers, it is obvious that the number of solutions is infinite since the system has $(N-m-n+1)$ degrees of freedom.

However, we also assume as given in base year interindustry matrix and we propose to constrain our current estimated matrix to be as "like" the base matrix as possible. Current marginal totals are assumed expressed at base year prices.

The problem then is to estimate, in a reasonable way, at base year prices the missing interindustry matrix. One is allowed in the first instance to assume unchanged technology, which means that non-zero and zero entries to be estimated correspond to those in the basic table. In practice, it will be necessary, no doubt, to make many amendments, from special knowledge about new industries, increase in productivity etc. Here we deal only with the purely algebraic problem, as follows:—

Given the sums of the rows and columns of an m× n *matrix, to estimate the elements, having close regard to the values of these elements in a full base-year matrix, in particular that the non-zero and zero values correspond.*

It will be assumed further that all current row and column sums given are positive.

*The Stone-Brown Method**

This ingenious solution, sometimes called "the RAS method", is as follows. First, the elements in the basic matrix are adjusted proportionately, column by column, to sum to the given current column or totals. Second, the adjusted elements so found are adjusted proportionately, row by row, to sum to the given row totals. The iterative process is repeated until *both* row and column sums agree with the given current totals, to a given degree of accuracy.

It will be noted, that with this method, zeroes and non-zero elements will be in the same positions in the solution as in the basic table. Experience has shown that, when non-zero elements in the basic table are all positive, only a few iterations are usually required to obtain an accurate solution. The writer is unaware of any mathematical proof of this convergence, or of any direct way of reaching solution, i.e. one without iteration. If in the basic table all marginal totals are positive and non-zero elements positive the convergence proposition is intuitively appealing since the positive values of the elements in the solution have an upper limit. But is the possibility of oscillation to be ruled out? Does one reach the same limiting solution if one starts the process with rows instead of columns (as in description above)?

In order to deal with product-mix and/or the creation of dummy industries one must admit the possibility of negative elements in the interindustry part of the IO table.** Even though in practice the negative values are small, they seem to make non-convergence possible, using the Stone-Brown method, as the example given below shows.

A Least Squares Method

We present a completely different method here, the merits† of which are that:—

(i) a solution is obtainable without iteration, in fact from the solution of a system of linear simultaneous equations; and

*R. Stone and J. A. C. Brown: "A Long-term Growth Model for the British Economy", *Europe's Future in Figures* (North Holland Publishing Company, edited by R. C. Geary, pp. 287 *seq.*, 1962).

**In the C.S.O.—E. W. Henry IO Table for 1964 (Prl. 985) with interindustry dimensions 92 × 92 there are 20 negative entries, a few large.

†We refrain from mentioning the trivial merit that the method yields the obviously correct answer (i.e., proportionality) when m (or n) are unity. So does Stone-Brown.

(ii) it appears to be applicable even when some elements in the basic matrix are negative.

Let the basic interindustry $m\times n$ matrix have elements ξ_{ij}, some possibly zero, but non-zeroes numbering N. We assume that $N \geqslant m+n-1$. Let row and column totals be $\xi_{i.}$ and $\xi_{.j}$ respectively, grand total ξ For current interindustry $m\times n$ matrix let elements be x'_{ij} (reason for prime will appear presently) with value zero at every (i, j) position that ξ_{ij} has zero. Row and column totals are $x'_{i.}$ and $x'_{.j}$, grand total x', all given. The problem is to estimate the x'_{ij}.

Without loss of generality we can assume that $\xi = x'$. This merely involves changing all elements and marginal totals in the basic table in the ratio x'/ξ. No change is thereby made in the essential character of the basic matrix. Its revised ξ_{ij} are merely brought nearer in general order of magnitude to the x'_{ij}. In fact our method for finding the x'_{ij} will be to minimise the function z given by:—

$$2z = \sum_{i=1}^{m} \sum_{j=1}^{n} (x'_{ij} - \xi_{ij})^2 \tag{1}$$

It will be convenient to set:—

$$x'_{ij} - \xi_{ij} = x_{ij}, \tag{2}$$

with $x'_{i.} - \xi_{i.} = x_{i.}$ and $x'_{.j} - \xi_{.j} = x_{.j}$, so that the $x_{i.}$ and $x_{.j}$ are given with:—

$$\Sigma_i x_{i.} = 0 = \Sigma_j x_{.j} \tag{3}$$

Independent constraints are:

$$\left\{\begin{array}{ll} \text{(i)} & \sum_{j=1}^{n} x_{ij} = x_{i.},\ i = 1, 2, \ldots, m-1 \\ \text{(ii)} & \sum_{i=1}^{m} x_{ij} = x_{.j},\ j = 1, 2, \ldots, n-1 \\ \text{(iii)} & \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} = 0 \end{array}\right. \tag{4}$$

Total number of independent constraints is accordingly $(m+n-1)$ and number of degrees of freedom in the system is $(N-m-n+1)$. The problem then is to find the values of all variables which minimise z given by:

$$2z = \sum_{i=1}^{m} \sum_{j=1}^{n} x^2_{ij} - 2 \sum_{i=1}^{m-1} \lambda_i \sum_{j=1}^{n} x_{ij} - 2 \sum_{=1}^{n-1} \lambda_j \sum_{i=1}^{m} x_{ij} - 2\lambda \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}, \tag{5}$$

the $(m+n-1)\lambda$'s being Lagrange multipliers. The variables are the x_{ij}, numbering N and the λs, numbering $(m+n-1)$, $(N+m+n-1)$ in all, to be found from the same number of linear simultaneous equations, $(m+m-1)$ from (4) and N from $\frac{\delta z}{\delta x_{ij}}=0$.

No row or column is all zeros. We are at liberty to interchange rows and columns at will. We may therefore assume a non-zero in the position (m, n), i.e., x_{mn}, to be found.

$$0=\frac{\delta z}{\delta x_{ij}}=x_{ij}-\lambda_{i.}-\lambda_{.j}-\lambda, \quad \begin{cases} i=1, 2, \ldots, m-1 \\ j=1, 2, \ldots, n-1 \end{cases} \tag{6}$$

$$0=\frac{\delta z}{\delta x_{in}}=x_{in}-\lambda_{i.}-\lambda, \quad i=1, 2, \ldots, m-1 \tag{7}$$

$$0=\frac{\delta z}{\delta x_{mj}}=x_{mj}-\lambda_{.j}-\lambda, \quad j=1, 2, \ldots, n-1 \tag{8}$$

$$0=\frac{\delta z}{\delta x_{mn}}=x_{mn}-\lambda \tag{9}$$

Of course, only non-zero positions (i, j) will be involved. If all were non-zero the number of equations in (6)–(9) would be mn. There are *some* equations in all four series and the actual total is N.

To solve, substitute for x_{ij} giving by (6)–(9) in (4). There result $(m+n-1)$ linear equations in the $\lambda_{i.}$, $\lambda_{.j}$ and λ which are solved for the λ. The solutions in the x_{ij} are then found from (6)–(9).

The general solution is far simpler than might have appeared earlier when there appeared to be $(N+m+n-1)$ linear equations which might be beyond the capacity of a computer available. In fact only $(m+n-1)$ equations require computering.

From the form of (5) it is obvious that the limiting values of x_{ij} found yield a minimum value of z.

In input-output work the interindustry part is usually a square matrix, though latterly there has been a tendency towards commodity rows and industry columns, i.e., the matrix need not be square.

Some Elements of Current Matrix Known

It often happens that, in the course of estimating marginal totals, one has estimated some of the current interindustry elements. Indeed, these may be the larger elements and should, of course, be accepted as part of the solution. They can be dealt with easily using the present LS method.

The values of these known elements are first subtracted from current row and column totals. Exactly similar adjustments are made in the basic table. The positions are then treated as if they were zeros. The present LS method is then applied for the estimation of the remaining current non-zero elements. Finally the known elements are restored.

A Remark on Grand Total Equalisation

For the present LS method it will be recalled that in the basic table all elements and margins are altered proportionately to bring the grand total of the table also to x', without, of course, changing the row-column additiveness character of the table. The present method could be applied without such prior grand total equalisation. The only difference in the formal statement of the problem would be that the right side of (4) (iii) would be $(x' - \xi)$ instead of zero.

It was at first surmised intuitively that the results would be identical. This is not the case. The results of the two approaches were compared using the data of the first constructed example below. The drastic assumption was made that in the basic table the grand total was 36 (instead of 72), all elements and margins also being halved. The solution matrix x'_{ij} was as follows:

10·20	19·85	·	7·95	38
·	4·15	−0·15	·	4
6·80	·	20·15	3·05	30
17	24	20	11	72

The values of the eight elements will be seen to be rather similar to those of the equalisation case, but they are far from being identical. The mean absolute deviation is 1·3, not too serious with a grand total of 72. One cannot decide which result is the "better", only that the equalisation procedure of the text seems the more natural.

Applications

The first constructed example is as follows:

$i \backslash j$	1	2	3	4	$\xi_{i\cdot}$	$x'_{i\cdot}$	$x_{i\cdot}$
	ξ_{ij}						
1	8	7	·	5	20	38	18
2	·	9	−4	·	5	4	−1
3	17	·	19	11	47	30	−17
$\xi_{\cdot j}$	25	16	15	16	72		
$x'_{\cdot j}$	17	24	20	11		72	
$x_{\cdot j}$	−8	8	5	−5			0

It will be noted that the basic table values have been deemed to be grossed up to the marginal grand total 72 and that the basic table contains one negative entry, namely −4 in position (2, 3). We first try to work out the Stone-Brown solution, then we present our own. Here $m = 3$, $n = 4$.

Stone-Brown Method

The first two iterations are as follows:

5·44	10·5	·	3·4375	19·3775
·	13·5	−5·3333	·	8·1667
11·56	·	25·3333	7·5625	44·4558
17	24	20	11	72
10·6680	20·5909	·	6·7411	38
·	6·6122	−2·6122	·	4
7·8010	·	17·0956	5·1034	30
18·4690	27·2031	14·4834	11·8445	72

For those who work with a desk machine, a merit of the method is that it is self-checking at each iteration: note that the last column of the first iteration and the last row of the second add to 72.

The fifth iteration is:

10·9220	15·7900	·	6·9792	33·6912
·	8·2100	−9·6342	·	−1·4242
6·0780	·	29·6342	4·0208	39·7330
17	24	20	11	72

A glance at the final column (note the negative total!) is enough to indicate that continued iteration holds no prospect of convergence, obviously, a consequence of the negative entry in the basic table.

Present LS Method

The 6 ($=m+n-1$) equations, found by substituting for the x_{ij} from (6)–(9) in (4) are:

	$\lambda_{2.}$	$\lambda_{1.}$	$\lambda_{.1}$	$\lambda_{.2}$	$\lambda_{.3}$	λ	Const.
	·	3	1	1	·	3	18
	2	·	·	1	1	2	−1
	·	1	2	·	·	2	−8
	1	1	·	2	·	2	8
	1	·	·	·	2	2	5
	2	3	2	2	2	8	0
Solution	−0·8	13·4	−1·5	6·9	12·1	−9·2	

e.g., the first row of figures means the equation:

$$3\lambda_{1.}+\lambda_{.1}+\lambda_{.2}+3\lambda = 18$$

As a general remark: the coefficients of the λ depend on the numbers of zeroes and non-zeroes in the basic table and not on what the non-zero values are. The solution (very easily worked on a desk machine because of the smallness of coefficients) is given as the last line of the table. By a happy arithmetical chance, the figures are exact and not merely "correct to one decimal place". The values of x_{ij} are found from (6)–(9). Finally the required x'_{ij} from $(x_{ij}+\xi_{ij})$:

Solution matrix x'_{ij}

10·7	18·1	·	9·2	38
·	5·9	−1·9	·	4
6·3	·	21·9	1·8	30
17	24	20	11	72

This LS solution bears little resemblance to any of the Stone-Brown iterations. It satisfactorily reproduces a negative value at position (2, 3); this may not always be the case.

In the second constructed example (worked by F. S. Ó Muircheartaigh) the object is to compare the present LS method with the RAS method in conditions in which the latter applies, i.e., all basic non-zero elements positive. Data are:

$i \backslash j$	1	2	3	$\xi_{i\cdot}$	$x'_{i\cdot}$	$x_{i\cdot}$
		ξ_{ij}				
1	10	15	20	45	48	3
2	21	·	15	36	41	5
3	30	37	41	108	100	−8
$\xi_{\cdot j}$	61	52	76	189		
$x'_{\cdot j}$	56	50	83		189	
$x_{\cdot j}$	−5	−2	7			0

As in the first example we start the process after the basic table elements have been raised proportionately to grand current total x' (=189), thus preserving the row/column additive character of the basic table. This example is more realistic than the first in that the row and column totals, basic (adjusted) and current, are closer to equality.

Using the RAS method the fifth iteration yielded the solution (to one decimal place):

9·4	15·8	22·8	48
21·9	·	19·1	41
24·7	34·2	41·1	100
56	50	83	189

With this sample we took the opportunity of comparing the results of RAS when one starts with rows as well as columns. After five iterations of each the results are identical (as above), though not absolutely.

As regards the present LS method, the 5 equations found by for the x_{ij} from (6)–(9) in (4) are:

	$\lambda_{2\cdot}$	$\lambda_{1\cdot}$	$\lambda_{\cdot 1}$	$\lambda_{\cdot 2}$	λ	Const.
	·	3	1	1	3	3
	2	·	1	·	2	5
	1	1	3	·	3	−5
	·	1	·	2	2	−2
	2	3	3	2	8	0
Solution	5·6833	3·6667	−4	−2·25	−0·5833	

The values of x_{ij} are found from (6)–(9). Finally the required x'_{ij} from $(x_{ij}+\xi_{ij})$; to one decimal place:

9·1	15·8	23·1	48
21·5	·	19·5	41
25·4	34·2	40·4	100
56	50	83	189

The RS estimates are reassuringly similar to those of RAS, given earlier. Perhaps the main point of the exercise is that they are not identical.

Concluding Remark

The present method, or any other of similar intent (i.e., for estimating the interindustry matrix from known marginal totals), is not a reliable substitute for actual enumeration. In fact, experience has shown that such estimations can be wide of the mark. If one has to use the matrix in one's model its status is that of a set of behaviouralistic equations (which would be identities if the matrix were known), adding to the imprecision of one's results or inferences. This statement is true even when, as one knows, all complete IO tables contain a considerable amount of estimation.

The author is aware that there is a considerable literature on coefficient estimation in general, and on RAS in particular, with which, he confesses, he is not fully familiar. It was his lack of success in proving the convergence of RAS iteration that led him to seek another solution. It is possible that this LS solution, which is simple and obvious, is not new. It is also possible that the convergence of RAS has been established. The author would welcome enlightenment on both points.

The author is much indebted to E. W. Henry, F. S. Ó Muircheartaigh and R. O'Connor for helpful comments.

Economic and Social Research Institute,
Dublin.