Estimating Nonnegative Matrices from Marginal Data
Author(s): Michael Bacharach
Source: *International Economic Review*, Vol. 6, No. 3 (Sep., 1965), pp. 294-310
Published by: Wiley for the Economics Department of the University of Pennsylvania and Institute of Social and Economic Research, Osaka University
Stable URL: https://www.jstor.org/stable/2525582
Accessed: 08-11-2019 23:48 UTC

INTERNATIONAL ECONOMIC REVIEW
Vol. 6. No. 3, September, 1965

ESTIMATING NONNEGATIVE MATRICES FROM MARGINAL DATA*

BY MICHAEL BACHARACH

1. ECONOMIC BACKGROUND

IN EMPIRICAL WORK one is often in the position of knowing the values of the elements of a nonnegative matrix function of time at one date but only its row and column sums at another. The problem arises of using these marginal data to estimate the elements of the matrix at the second date. One simple hypothesis for estimation is $a_{ij}^B = r_i s_j a_{ij}$, where a_{ij} is the value of the (i, j)-th element of the matrix at the first date, which is observed, a_{ij}^B the corresponding value at the second date which is not observed; r_i, s_j are factors multiplying the i-th row and j-th column of A; and the estimation is subject to the restrictions $A^B \geqq 0$, $\sum_{j=1}^n a_{ij}^B = u_i$, $\sum_{i=1}^m a_{ij}^B = v_j$, $i = 1, \cdots, m$; $j = 1, \cdots, n$, where the u_i, v_j are the observed row and column sums at the second date. The matrix A^B will be called biproportional to the matrix A, the superscript B in the notation A^B serving as a reminder of this relationship. Any model in which one matrix is made biproportional to another will from now on be called a biproportional (matrix) model, and any problem of the above type a biproportional (matrix) problem.

The biproportional model has been used in the past as a basis for adjusting sample estimates of contingency matrices to known population values for their marginal totals [6, 8]. Recently it has been proposed by Stone [11] for the estimation of input-output coefficient matrices. Let $A(1)$ denote the value of such a matrix at one date, $A(2)$ its value at a second date, and $X(2)$ and $q(2)$ the transactions matrix and gross output vector at the second date. Letting $\hat{x}$ denote the diagonal matrix with a vector x on the main diagonal and adding a brim to the "hat" when the vector is represented by a combination of symbols, we may write $X(2) = A(2)\widehat{q(2)}$. Suppose we make the hypothesis $a_{ij}(2) = x_i y_j a_{ij}(1)$, i.e.

$$A(2) = \hat{x}A(1)\hat{y} \; ; \tag{1}$$

then $X(2) = \hat{x}A(1)(\hat{y}\hat{q})$. Let us now require our estimate of $X(2)$ likewise to be biproportional to $A(1)$. Observing $A(1)$ and the vectors $X(2)i$, $iX(2)$ of intermediate outputs and inputs at the second date, we have a biproportional estimation problem for $X(2)$. If we write a solu-

* Manuscript received July 25, 1964, revised March 1, 1965.

tion as $X^B(2)$, the natural estimate of $A(2)$, $X^B(2)\widehat{q(2)}^{-1}$, itself satisfies a relationship of the form postulated for $A(2)$ in (1).

The factors x_i, y_j yield immediate economic interpretations [11]: x_i represents the extent to which commodity i has, uniformly over using industries, been substituted for other commodities between the two dates; y_j the extent to which intermediate inputs have uniformly assumed increased weight in the input structure of the j-th commodity, i.e. the extent to which the j-th commodity has become uniformly less fabricated.

In this interpretation x_i and y_j must be regarded as partial effects. $y_j = 1$, for instance, need not follow from constancy of the j-th column sum of the coefficient matrix. The degree of fabrication may have remained constant because, say, a partial fabrication effect lower than unity has been cancelled by a tendency throughout the economy to substitute certain goods that bulked large in commodity j's input structure.

It is seen at once that if $\hat{x}A(1)\hat{y}$ is a biproportional estimate of $A(2)$ then so is $(\lambda\hat{x})A(1)\{(1/\lambda)\hat{y}\}$ for any nonzero scalar λ. This means that we may not interpret certain values of x (e.g. those greater than one) as indicating substitution-in and others substitution-out until we have applied some normalization which itself has an appropriate economic meaning.

Leontief, in [9], has also considered the hypothesis (1) for changes in input-output coefficients over time. His interpretation differs slightly from Stone's [11].

A solution of the above estimation problem fits all the observed data perfectly, so goodness of fit criteria cannot provide a way of assessing the performance of the model. In the Belgian input-output study of Paelinck and Waelbroeck, however, a direct observation of the matrix at the second date was available for comparison with an estimate obtained from the above model [4, 10]. This estimate fitted the directly observed matrix substantially better than the naive estimate of constant input-output coefficients, but several elements diverged seriously from their directly observed values. In some cases this was apparently the result of too high a degree of aggregation and was not attributable to the specific assumption (1). In others, the divergences clearly arose from falsity of the assumption of substitution effects uniform over using industries, e.g. in the case of coal, used as a material input in the coke industry but as a fuel input elsewhere. For the most part, however, the bad cells could, by their nature, have been identified in advance. Moreover, many of them involved nationalized fuel sectors, so that independent forecasts of their coefficients might, in further applications, be available more readily than most. Paelinck and Waelbroeck now followed up with a second test [5, 10] in which such *a priori* suspect cells were removed from the estimation problem. The method now yielded fewer errors elsewhere in the table, revealing that some of these had previously arisen to compensate those in the cells now removed. Inspection of the y_j's strongly suggested that often these too were of the nature of compensatory factors, provoked by errors originating either in over-aggregation or in the uniform substitution assumption contained in (1), rather than measures of true fabrication tendencies.

The above method has been extensively used in the Cambridge Growth Project for the estimation both of input-output matrices and of "make" matrices showing the principal and secondary products of industries. A make matrix allows the transformation of data on commodity outputs into information on the outputs of multi-product industries: one needs such a matrix when, for instance, final demand information together with a commodity input-commodity output matrix yields commodity gross output requirements, and one wishes to go on to derive factor requirements from production functions that relate to multi-product industries. These applications of the biproportional matrix model in the Cambridge Growth Project are described in [2].

Several applications of the same model have been made in Bénard's [3] and Waelbroeck's [13] studies of international trading matrices. Waelbroeck interprets a row multiplier r_i as an index of one country's "market shares" in the other countries whose trade is represented in the matrix; the constancy of a column multiplier s_j over the elements of its column is justified by appeal to the stability of other countries' market shares in a given country in response to changes in that country's total imports. Bénard tests biproportional estimates against observed trade matrices and concludes that estimates of this type may be useful over periods of up to 15 years. Waelbroeck obtains measures of the effect of the European Common Market in increasing intra-Market trade, by comparing with observed current volumes of intra-Market transactions the volumes appearing in estimates of current matrices that fit these matrices' observed marginals and are biproportional transformations of pre-Market matrices.

The model that we have called the biproportional matrix model is referred to in most of the literature [2, 3, 4, 5, 10, 11, 13] as the "RAS" model. This name comes from the notational form $\hat{r}A\hat{s}$—which we retain—originally employed by Stone [11]. We have used another name for two reasons. First, "biproportional" calls to mind more immediately the relationship between A and the final matrix. Secondly, we shall henceforth consider a slightly wider set of matrices to be solutions of a biproportional matrix problem that the set of nonnegative matrices having the prescribed row and column sums u, v and of the form $\hat{r}A\hat{s}$. We shall allow as solutions limits of sequences of matrices of the form $\hat{r}A\hat{s}$ when these limits are not themselves expressible in this form. In other words, our solution set will be the closure of the set

$$\{B \mid B = \hat{r}A\hat{s},\ B \geqq 0,\ Bi = u,\ iB = v\} .$$

A matrix B will from now on be called biproportional to A if it is the limit of a sequence $\{\widehat{R^t}A\widehat{S^t}\}$ for some sequences $\{R^t\}, \{S^t\}$ of vectors.

The author investigated numerous properties of the biproportional model and of its input-output and other applications (see [1]). The present paper sets out only to establish a few fundamental formal properties of solutions of the biproportional problem that have economic significance. It is shown in Theorems 1 and 2 that a solution is unique. Hence we do not need criteria to choose between solutions. Theorem 3 establishes conditions for the convergence to this nonnegative solution of a certain iterative procedure described in the next section which has often been successfully employed in practice. It is easily shown that a solution must preserve zero elements of the initial matrix, which is usually what one would wish. Various corollaries to Theorem 3 define the circumstances in which initially positive elements are driven to zero in a solution.

Most of the complications of the following analysis are due to possible complete or partial "disconnections" of the initial matrix: all the results can be proved much more rapidly in the strictly positive case.

2. THE PROBLEM (A, u, v) AND THE PROCESS (A, u, v)

In this section we formally define the biproportional matrix problem and the standard iterative solution method and draw attention to one or two simple but useful facts about them.

We begin by giving a precise statement of the biproportional matrix problem.

DEFINITION. The *problem* (A, u, v) is to find A^B such that

$$\begin{aligned} A^B &\geqq 0 , \\ A^B &= \lim_{t\to\infty} \widehat{R^t}A\widehat{S^t} , \\ A^B i = u , &\quad iA^B = v , \end{aligned}$$

for some sequences $\{R^t\}, \{S^t\}$ of vectors with $R_1^t = 1$ for all t; where

A is a given $m \times n$ matrix such that

$$a_i \geq 0 , \quad a^j \geq 0 , \qquad \text{for all } i, j , \tag{2}$$

and u, v are given m- and n-vectors such that

$$u > 0 , \quad v > 0 . \tag{3}$$

a_i, a^j denote the i-th row and j-th column of A. ≥ 0 denotes semipositivity. In subsequent Sections where we have occasion to consider A^B as a function of u, v, we shall sometimes write a solution of (A, u, v) as $A^B(u, v)$. A solution of the form $\hat{r}A\hat{s}$ will sometimes be called an "interior" solution; a solution not expressible in this form a "boundary" solution.

The normalization $R_1^t = 1$, which is convenient in proving a uniqueness result in Section 3, involves no real loss of generality. The definition covers any matrix of the form $\hat{r}A\hat{s}$ which is nonnegative and fits the constraints. For the conditions $u > 0$, $v > 0$ ensure that for all i, j, $r_i \neq 0$, $s_j \neq 0$ in such a solution, whence there is a number λ such that $\widehat{\lambda r}A\widehat{(1/\lambda)s}$ is nonnegative and fits the constraints and $\lambda r_1 = 1$. The definition generalizes this notion of a solution by also admitting nonnegative, constraint-fitting limits of sequences of matrices of the form $\hat{r}A\hat{s}$, $r_1 = 1$.

The specifications (2) and (3) differ only trivially from the general case of nonnegative A, u, v. If some a_i vanishes, a solution exists only if u_i vanishes too. If u_i vanishes, then all questions concerning a solution reduce to questions about a solution of $(\bar{A}, \bar{u}, v)$, where bars denote the suppression of the i-th row of A and the i-th element of u. If there is no solution to the latter problem, there is clearly none to the first. Conversely, if, in an obvious notation, a solution of $(\bar{A}, \bar{u}, v)$ is given by the sequences $\{\bar{R}^t\}$, $\{S^t\}$, then the original problem is solved by the sequences $\{R^t\}, \{S^t\}$ with $R_i^t = 0$ for all t. Similar remarks apply to the a^j's and v_j's.

We next define the process that provides the standard method of solution used in practice [2, 3, 10, 13]. Starting with the given matrix A, one first multiplies each row by a scalar that will make the row sum equal the row constraint, next multiplies each column of the resulting matrix A^1 by a scalar that will make its sum equal its constraint. This gives a matrix A^2 that serves as starting point for the next iteration. The formal definition is as follows:

DEFINITION. The *process* (A, u, v) is

$$A^{2t+1} = \widehat{r^{t+1}}A^{2t} , \tag{4}$$

$$(5)\qquad A^{2t+2} = A^{2t+1}\widehat{s^{t+1}} = \widehat{r^{t+1}}A^{2t}\widehat{s^{t+1}}\,,$$

where

$$(6)\qquad r_i^{t+1} = \frac{u_i}{\sum_{j=1}^{n} a_{ij}^{2t}}\,,$$

$$(7)\qquad s_j^{t+1} = \frac{v_j}{\sum_{i=1}^{m} a_{ij}^{2t+1}}\,.$$

t takes the values $0, 1, 2, \cdots$. We start the process by setting $A^1 = A$. The sequence of matrices A^t is so computed that, for all t,

$$(8)\qquad A^{2t+1}\dot{\imath} = u\,,$$

$$(9)\qquad \dot{\imath}A^{2t+2} = v\,.$$

We shall find the notation $r^t(A, u, v)$ convenient or, where there is no ambiguity, $r^t(A)$. Note that

$$(10)\qquad r^{t+1}(A) = r^1(A^{2t})\,.$$

A few obvious properties of the sequences $\{A^{2t}\}$, $\{A^{2t+1}\}$, $\{r^t\}$, $\{s^t\}$ are worth pointing out. (2) and (3) ensure that r^1 is defined and is positive, whence by (4) A^1 has finite semipositive rows and columns, whence by (7) s^1 is defined and is positive. So $A^2 = \widehat{r^1}A\widehat{s^1}$ has finite semipositive rows; by an easy induction

$$(11)\qquad a_i^{2t},\ a^{2t+1,j} \geq 0, \qquad t = 0, 1, 2, \cdots$$

$$(12)\qquad r^t,\ s^t > 0, \qquad t = 0, 1, 2, \cdots$$

where $a^{2t+1,j}$ denotes the jth column of A^{2t+1}; and (12) means that

$$(13)\qquad a_{ij}^{2t} = 0 \quad \text{or} \quad a_{ij}^{2t+1} = 0 \quad \text{implies} \quad a_{ij} = 0, \qquad t = 0, 1, 2, \cdots$$

that is, a positive element of A can never vanish in the course of the process (A, u, v).

3. UNIQUENESS

A matrix A will be called *disconnected* if there are nonvacuous index sets I, I', J, J', such that $a_{ij} = 0$ for $i \in I$, $j \in J'$ and for $i \in I'$, $j \in J$, where prime denotes complementarity. We shall also write this condition, in an obvious notation for partitions of A, as $A_{IJ'} = 0$, $A_{I'J} = 0$. The concept generalizes decomposability, as I may differ from J, and indeed the matrix need not be square. A matrix is called *connected* if it is not disconnected.

We first prove the following theorem on the uniqueness of an interior solution.

THEOREM 1. *If A is connected, then a solution A^B of (A, u, v) expressible as $A^B = \hat{r}A\hat{s}$ is unique.*

PROOF. We first note that if A is connected then the provision that a solution be nonnegative means that the r, s of a solution of the form $\hat{r}A\hat{s}$ are either both positive or both negative. Clearly neither can contain zeros; suppose then that some r_i are positive and the rest negative, say $r_I > 0$, $r_{I'} < 0$. Then certainly s cannot be positive or negative, so say $s_J > 0$, $s_{J'} < 0$. Then $A_{IJ'}$, $A_{I'J}$ are null, for otherwise the solution contains negative elements; thus A is disconnected.

Suppose that there are two solutions of the stated form. Then there exist ρ and σ such that

$$\hat{\rho}A^B\sigma = A^B i = u \,, \tag{14}$$

$$\rho A^B\hat{\sigma} = iA^B = v \,, \tag{15}$$

where A^B is the first solution of the original problem, and ρ, σ are either both positive or both negative as A^B is connected; first assume

$$\rho > 0 \,, \quad \sigma > 0 \,. \tag{16}$$

Reorder the elements of ρ, σ so that

$$\rho_1 \geqq \rho_2 \geqq \cdots \geqq \rho_m \,, \tag{17}$$

$$\sigma_1 \leqq \sigma_2 \leqq \cdots \leqq \sigma_n \,. \tag{18}$$

To prove that the two solutions are the same, it suffices to show that if $a^B_{ij} > 0$ then $\rho_i\sigma_j = 1$. Suppose the contrary, for instance that for some (i_0, j_0)

$$a^B_{i_0j_0} > 0 \,, \quad \rho_{i_0}\sigma_{j_0} > 1 \,.$$

We first show that this implies either $\rho_1\sigma_1 < 1$ or $\rho_1\sigma_1 > 1$. If $j_0 = 1$ then $\rho_1\sigma_1 > 1$ by (17). If $j_0 > 1$ then there must be a $j_1 < j_0$ such that $a^B_{i_0j_1} > 0$, $\rho_{i_0}\sigma_{j_1} < 1$, for otherwise $\sum_{j=1}^n \rho_{i_0}\sigma_j a^B_{i_0j} > \sum_{j=1}^n a^B_{i_0j}$ by (18), contrary to (14). Suppose that $j_1 > 1$. Then, similarly, there is an $i_1 < i_0$ such that $a^B_{i_1j_1} > 0$, $r_{i_1j_1} > 1$. In the sequence $j_0, i_0, j_1, i_1, \cdots$ there is a first member equal to 1. If this is an i_t we have $a^B_{1j_t} > 0$, $\rho_1\sigma_{j_t} > 1$, whence $\rho_1\sigma_1 < 1$ by (18) and (14). If it is a j_t, we have $a^B_{i_{t-1}1} > 0$, $\rho_{i_{t-1}}\sigma_1 < 1$, whence $\rho_1\sigma_1 > 1$ by (17) and (15).

However, $\rho_1\sigma_1 < 1$ implies $\rho_i\sigma_1 < 1$ for all i by (17), whence $\sum_{i=1}^m \rho_i\sigma_i a^B_{i1} < \sum_{i=1}^m a^B_{i1}$, contradicting (15); and $\rho_1\sigma_1 > 1$ is ruled out for a similar reason. An argument dual to the above applies in the case of the opposite inequality $\rho_{i_0}\sigma_{j_0} < 1$. If in place of (16) we had assumed $\rho < 0$, $\sigma < 0$ and $\hat{\rho}A^B\hat{\sigma}$ differed from A^m, so would $\widehat{-\rho}A^B\widehat{-\sigma}$, contrary to what we have just shown.

We next generalize Theorem 1 to cover boundary solutions as well as interior ones. A is still restricted to be connected.

THEOREM 2. *If A is connected, then a solution of (A, u, v) is unique.*

PROOF. Let the sequence $\{R^t, S^t\}$ yield a solution A^B of (A, u, v). By definition,

$$\lim_{t\to\infty} \widehat{R^t}A\widehat{S^t} = A^B , \tag{19}$$

$$R_1^t = 1 , \qquad \text{for all } t . \tag{20}$$

Let (r, s) be a limit point of $\{R^t, S^t\}$. Then

$$r_i s_j a_{ij} = a_{ij}^B \tag{21}$$

for all i, j for which the left hand side is defined. Now suppose the set of infinite elements of r,

$$I = \{i \mid r_i = \pm\infty\}$$

is not empty. For any i, $a_{ij} > 0$ for some j, by (2), hence the set $\{j \mid s_j = 0\} = J'$ say, is also nonempty. But I' is nonempty by the normalization (20), hence so finally is J.

Then $A_{IJ} = 0$ (otherwise some $a_{ij}^B = \infty$), so that $A_{IJ}^B = 0$ by (19). But $A_{I'J'}^B = 0$ from (21) and the definitions of I', J'; thus the solution A^B is disconnected. Very similar arguments show that A^B is disconnected if we begin by supposing $\{j \mid s_j = \pm\infty\}$ to be nonempty, instead of I. Note that if both these sets are empty then the solution A^B is of the kind already dealt with in Theorem 1.

Theorem 1 can now be applied to show that any connected solution A^B is unique. But suppose that A^B is disconnected. Then we have, for some nonvacuous sets I, I', J, J'.

$$A_{IJ'} = 0 , \quad A_{I'J} \neq 0 ,$$

$$A_{IJ'}^B = 0 , \quad A_{I'J}^B = 0 , \quad A_{IJ}^B \text{ connected} . \tag{22}$$

In this case, (22) must hold for any solution A^* say. Clearly $A_{IJ'}^* = 0$. Next, $A_{IJ'}^B = 0$, $A_{I'J}^B = 0$ imply that

$$U_I = V_J , \quad U_{I'} = V_{J'} ,$$

where $U_I = \sum_{i\in I} u_i$, etc., whence $A_{I'J}^* = 0$. Suppose then that A_{IJ}^* were disconnected. Then we would have, similarly, $U_{I_1} = V_{J_2}$, $U_{I_2} = V_{J_1}$, where I_1, I_2 partition I and J_1, J_2 partition J, and A^B_{IJ} would also be disconnected, contrary to (22). But Theorem 1 now shows that A^B_{IJ} is the unique value for this partition of any nonnegative solution, being a connected solution of (A_{IJ}, u_I, v_J), in an obvious notation. We know that $A^B_{IJ'}, A^B_{I'J} = 0$ in any solution, so applying the above argument concerning A^B_{IJ} to each such connected partition of A^B in turn clinches the corollary.

Finally, we relax the requirement that A be connected.

COROLLARY 1. *A solution of* (A, u, v) *is unique.*

PROOF. If A is disconnected, then by suitable partitioning of A_{IJ}, $A_{I'J'}$ in the case that either of these is disconnected, and continuing in the same way, after a finite number of partitionings and by suitable permutations of rows and of columns, we shall have A in block-diagonal form with each diagonal partition connected. Then Theorem 2 can easily be applied to each diagonal partition considered as a matrix.

4. CONVERGENCE OF THE PROCESS

In this section we establish, in Theorem 3, necessary and sufficient conditions for the convergence of the process (A, u, v) to a solution of the problem (A, u, v). These enable us easily to obtain necessary and sufficient conditions for the existence of such a solution. The strategy is to show first, in the Lemma, that if A is connected $\min_i r_i^{t+m-1}(A)$ is strictly greater than $\min_i r_i^t(A)$ if the latter is less than $\max_i r_i^t(A)$. That is, if the row multipliers are not all equal at the t-th iteration, the least of them will have increased by the $(t+m-1)$-th. But the increase $\min_i r_i^{t+m-1}(A) - \min_i r_i^t(A)$ depends continuously on A. We consider (case (i)) a connected limit point A^B of $\{A^{2t}\}$ and suppose that $\max r_i^1(A^B) - \min r_i^1(A^B) > 0$, whence we know from the Lemma that $\min_i r_i^m(A^B) - \min_i r_i^1(A^B) = h > 0$. Then on the one hand the increase in $\min_i r_i^t(A)$ must be close to h for A^{2t} near A^B. On the other hand it must be close to zero for A^{2t} near A^B. This shows that $r^1(A^B) = \lambda i$, and it is only left to show that the conditions asserted in the theorem to be sufficient imply that $\lambda = 1$.

We now consider the case (case (ii)) of a disconnected limit point A^B. The reasoning for the connected case is applied to show that the partition of $r^1(A^B)$ corresponding to A^B's k-th connected partition, $r_k^1(A^B)$ say, has the form $\lambda_k i$ for a positive scalar λ_1. It is easily shown that if $\min_k \lambda_k = 1$ then the condition of the theorem implies that A^B is the solution. We suppose then that $\min_k \lambda_k < 1$. Letting I denote the set of rows i for which $r_i^1(A^B) = \min_k \lambda_k$ and J the set of columns j

connected with I, it is easy to show that $s_j^1(A^B) = 1/\min_k \lambda_k$ for j in J and $A^B_{I'J} = 0$. We now establish that, if $A_{I'J} \neq 0$, then $iA^{2t}_{I'J}i$ is increasing whenever it is near enough to zero. This is used to show that there can be no limit point in which the partition (I', J) is null; but we know that $A^B_{I'J} = 0$. The contradiction shows that $A_{I'J} = 0$. But $\min_k \lambda_k < 1$ implies $\sum_{i \in I} u_i / \sum_{j \in J} v_j < 1$, and together with $A_{I'J} = 0$ this constitutes a violation of the sufficient condition for convergence stated in Theorem 3.

LEMMA. *If A is connected and* $\min_i r_i^{t'} < \max_i r_i^{t'}$, *then* $\min_i r_i^t > \min_i r_i^{t'}$ *and* $\max_i r_i^t < \max_i r_i^{t'}$ *for* $t \geqq t' + m - 1$, *for all* t'.

PROOF. The difficulty lies entirely in the possible nonuniqueness of $\min_i r_i^{t'}$. Consider for the time being a problem in which the initial matrix A satisfies one set of constraints, namely

$$iA = v .$$

Reorder if necessary so that

$$r_1^1 = \min_i r_i^1 ,$$

and suppose that

$$r_1^1 < \max_i r_1^1 .$$

Define $\underline{I} = \{i \mid r_i^1 = r_1^1\}$.

Now (7), (4) give

$$s_j^{t+1} = \frac{v_j}{\sum_{i=1}^{m} r_i^{t+1} a_{ij}^{2t}} = \frac{1}{\sum_{i=1}^{m} r_i^{t+1} \left(\frac{a_{ij}^{2t}}{v_j} \right)} ,$$

which is the reciprocal of a convex combination of the r_i^{t+1} by (9) and (11). So

$$s_j^{t+1} \leqq \frac{1}{\min_i r_i^{t+1}} \tag{23}$$

and in particular $s_j^1 \leqq 1/\min_i r_i^1 = 1/r_1^1$.

Define $\bar{J} = \{j \mid s_j^1 = 1/r_1^1\}$. Then

$$a_{ij} = 0 \qquad \text{for } i \in \underline{I}', j \in \bar{J} . \tag{24}$$

But similarly (6), (5) give

$$r_i^{t+2} = \frac{1}{\sum_{j=1}^{n} s_j^{t+1} \left(\frac{a_{ij}^{2t+1}}{u_i} \right)} ; \tag{25}$$

hence as before, by (8), (11),

$$r_i^{t+2} \geqq \frac{1}{\max_j s_j^{t+1}} \tag{26}$$

and in particular $r_i^2 \geqq 1/\max_j s_j^1 \geqq r_1^1$. Define $\underline{\underline{I}} = \{i \mid r_i^2 = r_1^1\}$, then $a_{ij}^1 = 0$ —and therefore $a_{ij} = 0$ by (13)—for $i \in \underline{\underline{I}}$, $j \in \bar{J}'$. But A is connected, so this together with (24) implies that $\underline{\underline{I}}$ is a proper subset of $\underline{I}$. Continuing, it is clear that the minimum of r_i^t must be unique and equal to r_1^1 for some $t \leqq m - 1$. Say

$$r_1^t = r_1^1$$

Now $s_j^t \leqq 1/r_1^1$, and equality occurs only for $j \in \bar{J}_t = \{j \mid a_{ij} = 0 \text{ for all } i > 1\}$. But $a_{1j} > 0$ for some $j \in \bar{J}_t'$ by connection, hence $r_1^{t+1} > r_1^t$ by an application of (25). Again, for $i > 1$, r_i^{t+1} is the reciprocal of a weighted average of the s_j^t in which positive weights occur only for $j \in \bar{J}_t'$, i.e. for j such that $s_j^t < 1/r_1^1$; so $r_i^{t+1} > r_1^t$ for $i > 1$ too, and we have $\min_i r_i^{t+1} > \min_i r_i^t$.

Thus for some $t \leqq m$, $\min_i r_i^t > \min_i r_i^1$. But (23), (26) easily give $\min_i r_i^t$ nondecreasing, hence $\min_i r_i^t > \min_i r_i^1$ for all $t \geqq m$. We have assumed that the columns of A already sum to v. But if in a general problem (A, u, v) the matrix A is connected, then so is A^{2t} by (13). Applying the above result to A^{2t} establishes the Lemma.

We are now in a position to prove the main theorem, on convergence of the iterative process.

THEOREM 3. *The process (A, u, v) converges to the unique solution of the problem (A, u, v) if and only if $A_{I'J} = 0$ implies*

$$U_{I'} \leqq V_{J'}\,, \tag{27}$$

$$U_I \geqq V_J\,, \tag{28}$$

where $U_I = \sum_{i \in I} u_i$, $V_J = \sum_{j \in J} v_j$, etc.

Note that if we take $I' = \{1, \cdots, m\}$ then $A_{I'J} = 0$ for $J = \{0\}$ by (2), and (27) gives $iu \leqq iv$, and if we take $J = \{1, \cdots, n\}$ then $A_{I'J} = 0$ for $I' = \{0\}$ and (28) gives $iu \geqq iv$. Thus one condition covered by the theorem is the important one

$$iu = iv\,. \tag{29}$$

PROOF. The a_{ij}^{2t} are nonnegative and bounded above by $\max_j v_j$, from (9) and (11). Thus $\{A^{2t}\}$ has a finite limit point A^B say. We shall show first that if the conditions of the theorem are met, A^B is the solution.

A^B has semipositive rows, for clearly $a_{ij}^B \geqq 0$, and if $a_i^B = 0$, some r_i^t

would be forced arbitrarily large. This would contradict the fact that $\max_i r_i^t$ is nonincreasing, easily established from (23), (26). Similarly, A^B has semipositive columns: for if $a^{B,j} = 0$ then for some t $a^{2t,j}$ and hence $a^{2t+1,j}$ are arbitrarily small, and so s_j^{t+1} is arbitrarily large.

Case (i). Suppose A^B is connected. We shall show first that $r^1(A^B, u, v) = r^1(A^B) = \lambda i$ for some $\lambda > 0$, then that $\lambda = 1$. Suppose that on the contrary the elements of $r^1(A^B)$ are not all equal, then by the Lemma

$$\min_i r_i^m(A^B) = \min_i r_i^1(A^B) + h, \qquad \text{for some } h > 0 .$$

$r^1(B)$ is clearly a continuous function of B for a matrix B with semipositive and rows and columns. It follows that $\min_i r_i^1(B)$ is also continuous at B. (This is obvious if the minimum is unique. Let $\min_i r_i^1(B) = r_1^1(B) = \cdots = r_k^1(B)$. Then each of $r_1^1(C), \cdots, r_k^1(C)$ is within any e of its value at B for C near enough B, and the minimum at C is found among them if C is near to B.) $r^t(B), t > 1$, is also a continuous function of B, being obtained by a finite sequence of permissible algebraic operations on B, u, v. Thus $\min_i r_i^t(B)$ and finally $[\min_i r_i^t(B) - \min_i r_i^1(B)]$ are continuous functions of B.

Consider two points $A^{2t'}, A^{2t}$ of a subsequence of $\{A^{2t}\}$ having A^B as a limit. By the above continuity property there exists $\varepsilon > 0$ such that

$$(30) \qquad |a_{ij}^{2t'} - a_{ij}^B| \leqq \varepsilon, \qquad \text{for all } i, j ,$$

implies that $[\min_i r_i^m(A^{2t'}) - \min_i r_i^1(A^{2t'})]$ is arbitrarily close to h, e.g.

$$(31) \qquad \min_i r_i^m(A^{2t'}) - \min_i r_i^1(A^{2t'}) > \frac{1}{2}h .$$

Hence by the nondecreasing property of $\min_i r_i^t$ and (10)

$$(32) \qquad \min_i r_i^1(A^{2t}) - \min_i r_i^1(A^{2t'}) > \frac{1}{2}h, \qquad t \geqq t' + m - 1 .$$

On the other hand, $\min_i r_i^1(A^{2t'})$ and hence $\min_i r_i^1(A^{2t})$ are both brought arbitrarily near to $\min_i r_i^1(A^B)$ and thus to each other for $A^{2t'}$ near enough A^B, and in particular there exists $\eta > 0$ such that

$$(33) \qquad \min_i r_i^1(A^{2t}) - \min_i r_i^1(A^{2t'}) < \frac{1}{2}h$$

provided that

$$(34) \qquad |a_{ij}^{2t'} - a_{ij}^B| \leqq \eta, \qquad \text{for all } i, j .$$

But we can certainly find t' such that both (30) and (34) are satisfied.

Then the contradiction (32), (33) shows that

$$r^1(A^B) = \lambda i \; , \tag{35}$$

and $\lambda > 0$ since A^B has semipositive rows.

Now (29) implies $\lambda = 1$ easily. Thus A^B satisfies both sets of constraints, but it is certainly nonnegative and of the form $\lim_{t\to\infty}(\widehat{R^t}A\widehat{S^t})$ with $R_1^t = 1$ for all t; hence it is unique by Theorem 2. Thus $\{A^{2t}\}$ has as its sole limit point, hence as its limit, the unique solution of (A, u, v).

Case (ii). Assume A^B is disconnected, and let $k = 1, \cdots, K$ denote the subsets of rows or columns appearing in its connected partitions; then in an obvious notation $A_{kk'}$ has semipositive rows and columns or is null according as $k = k'$ or $k \neq k'$. The case (i) argument up to (35) goes through for each A_{kk}^{2t}, and we get $r_k^1(A^B) = \lambda_k i$, where $r_k^1(A^B)$ denotes the k-th partition of $r^1(A^B)$ and λ_k a positive number. Order so that $\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_K$. Then $iu = iv$ implies $\lambda_1 \leqq 1$ and $\lambda_K \geqq 1$, and

$$s_k^1(A^B) = \frac{1}{\lambda_k} i \; , \tag{36}$$

since each element of $s_k^1(A^B)$ is the reciprocal of a convex combination of the elements of $r_k^1(A^B)$ by the disconnection assumption.

If $\lambda_1 = 1$, A^B is the solution; if $\lambda_1 < 1$, we shall show that the condition of the theorem is violated. First let $\lambda_1 = 1$; then $iu = iv$ implies $\lambda_K = 1$ and we have $r^t \to i$ as in Case (i). (This is what happens if, for instance, the initial matrix A is disconnected, and each connected partition satisfies the condition of the theorem. Note the theorem does not restrict A to be connected.)

Suppose on the other hand that $\lambda_1 < 1$. Define $I = \{i \mid r_i^1(A^B) = \lambda_1\}$, $J = \{j \mid s_j^1(A^B) = 1/\lambda_1\}$. Then clearly

$$\frac{1}{\lambda_1} = \frac{V_J}{U_I} \; . \tag{37}$$

Let a limit point whose partition (I', J), vanishes be denoted by A^*. Then for any A^*

$$\sum_{j\in J} s_j^1(A^*) i a_{Ij}^{*1} = V_J \; ,$$

where a_{Ij}^{*1} denotes the j-th column of A_{IJ}^{*1}, and A^{*1} denotes the first matrix of the sequence generated by (A^*, u, v). Hence

$$\sum_{j\in J}\left[s_j^1(A^*)\frac{i a_{Ij}^{*1}}{iA_{IJ}^{*1} i}\right] = \frac{V_J}{iA_{IJ}^{*1} i} = \frac{V_J}{U_I - iA_{IJ'}^{*1} i} \; .$$

But the left hand side is a convex combination of $s_j^1(A^*)$, none of which can exceed $1/\lambda_1$ by the easily established nondecreasing property of $\max_j s_j^t$; and the right hand side is not less than $1/\lambda_1$ by the nonnegativity of $A^{*1}_{IJ'}$ and (37). Therefore $s_j^1(A^*) = 1/\lambda_1$ for all $j \in J$, $A^{*1}_{IJ'} = 0$, and by an application of (36) $r_i^1(A^*) = \lambda_1$ for all $i \in I$. That is, if $\bar{I}$ denotes the set of i for which $r_i^1(A^*) = \lambda_1$, $\bar{I} \supseteqq I$. But then $A^B_{\bar{I}'J} = 0$, and applying to A^B what we have just shown for A^* we get $r_i^1(A^B) = \lambda_1$ for all $i \in \bar{I}$. The definition of I then implies that $\bar{I} \subseteqq I$. We have shown that, for any A^*, $r_i^1(A^*) = \lambda_1$ if and only if $i \in I$.

It follows from the continuity property of $r^1(B)$ that if A^{2t} is near enough to some A^* then for any $(i, j) \in (I', J)$, $a_{ij}^{2t+2}/a_{ij}^{2t}$ is arbitrarily near λ_k/λ_1 for some $\lambda_k > \lambda_1$. Suppose now that

$$A_{I'J} \neq 0 . \tag{38}$$

Then $iA^{2t}_{I'J}i > 0$ for all t and there is a positive number ε^*, depending on A^*, such that

$$| a_{ij}^{2t} - a_{ij}^* | \leqq \varepsilon^* \quad \text{for all } i, j \text{ implies} \quad iA^{2t+2}_{I'J}i > iA^{2t}_{I'J}i .$$

The set of limit points of $\{A^{2t}\}$ is finite. For let A^B, A^{BB} be any two that are disconnected in the same way. Then $r_i^1(A^{BB}) = \lambda_k$ for $i \in k$ and A^B_{kk}, A^{BB}_{kk} are both solutions of $(A_{kk}, (1/\lambda_k)u_k, v_k)$ and so are one and the same. But the number of distinct disconnections of a matrix is finite.

$\min_{A^*} \varepsilon^*$ exists and is positive as the set of A^*'s is finite. But then there is a positive number $\varepsilon \leqq \min \varepsilon^*$ such that, if $a_{ij}^{2t} \leqq \varepsilon$ for all $(i, j) \in (I', J)$ then, for some A^*, $| a_{ij}^{2t} - a_{ij}^* | < \min \varepsilon^*$ for all (i, j). For otherwise there is a subsequence of $\{A^{2t}\}$ whose partition $(I' J)$ tends to zero but none of whose limit points is an A^*, contrary to the definition of the set of A^*'s. It follows that

$$a_{ij}^{2t} \leqq \varepsilon \text{ for all } (i, j) \in (I', J) \text{ implies } i\, A^{2t+2}_{I'J}i > i\, A^{2t}_{I'J}i . \tag{39}$$

By (38), $iA^{2t}_{I'J}i > \min_i r_i^1 \min_j s_j^1 \eta$ for $t = 0$ and for some positive number $\eta \leqq \varepsilon$. Assume that $iA^{2t}_{I'J}i > \min_i r_i^1 \min_j s_j^1 \eta$. If, on the one hand, $iA^{2t}_{I'J}i > \eta$, then $iA^{2t+2}_{I'J}i > \min_i r_i^1 \min_j s_j^1 \eta$ by the fact that $a_{ij}^{2t+2} = r_i^t s_j^t a_{ij}^{2t}$ and the monotone properties of $\min_i r_i^t$, $\min_j s^t$. If, on the other hand, $iA^{2t}_{I'J}i \leqq \eta$, then $iA^{2t+2}_{I'J}i > iA^{2t}_{I'J}i > \min_i r_i^1 \min_j s_j^1 \eta$ by (39) and the fact that $\eta \leqq \varepsilon$. We have shown that $iA^{2t}_{I'J}i > \min_i r_i^1 \min_j s_j^1 \eta$ for all t so that no limit point A^* exists, contrary to the assumption that A^B is such a limit point. The assumption (38) that $A_{I'J} \neq 0$ is therefore false. But $\lambda_1 < 1$ implies $U_1 < V_j$ by (37), which in view of $A_{I'J} = 0$ contradicts the condition of the theorem.

We have shown the sufficiency of (27)–(28). Necessity is easily shown. If $A_{I'J} = 0$ then in any solution A^B

$$V_{J'} = \sum_{i=1}^{m} \sum_{j\in J'} a_{ij}^B \geqq \sum_{i\in I'} \sum_{j\in J'} a_{ij}^B ,$$

and the last summation $= U_{I'}$ by hypothesis. This is (27), and (28) follows from a similar argument.

COROLLARY 2. *A solution of* (A, u, v) *exists if and only if* (27)–(28) *holds.*

PROOF. Necessity was shown in proving the necessity part of Theorem 3. Sufficiency follows from the sufficiency part of Theorem 3.

5. SOME COROLLARIES

In this section we present various corollaries of the fundamental uniqueness and convergence-existence theorems of the last two sections. The first of these connects our conditions for the existence of a nonnegative matrix A^B obeying the row and column constraints and biproportional to a given matrix A, with those for the existence of a nonnegative matrix B obeying the same constraints but otherwise constrained only to preserve zero elements of a given matrix, A. The latter problem has been studied in a more general form by Fréchet [7] and by others referred to in [12]. The two sets of conditions are one and the same.

COROLLARY 3. *There exists a matrix B satisfying*

$$B \geqq 0 ,$$
$$Bi = u, iB = v ,$$
$$a_{ij} = 0 \textit{ implies } b_{ij} = 0$$

if and only if conditions (27)–(28) *are fulfilled.*

PROOF. Sufficiency is obvious. Necessity is proved by an immediate generalization of the proof of the necessity part of Theorem 3.

The next corollary specifies the circumstances in which a positive element of A may be driven to zero in a solution A^B.

COROLLARY 4. *Let* $a_{ij} > 0$. *Then* $a_{ij}^B = 0$ *if and only if* (27)–(28) *hold as equations for some nonempty* I', J *and* $i \in I$, $j \in J'$.

PROOF. Sufficiency is obvious. Necessity may be shown by going through the proof of Theorem 2 using the normalization $R_i^t = 1$.

We have the following further result relating the vanishing of a positive element of A to the status of A^B as a boundary solution.

COROLLARY 5. *$a_{ij} > 0$ and $a^B_{ij} = 0$ for some (i, j) if and only if A^B is a boundary solution.*

PROOF. Necessity follows easily from the definition of an interior solution. Conversely, suppose A^B is a boundary solution. Then for some connected partition A_{kk} of A the corresponding partition A^B_{kk} of A^B cannot be expressed in the form $\widehat{r_k} A_{kk} \widehat{s_k}$. Applying the method of proof of Theorem 2 to A_{kk} shows that A^B_{kk} is disconnected, proving the sufficiency statement.

We next show that a solution A^B of a problem (A, u, v) may always be expressed in the form $\hat{r}\mathring{A}\hat{s}$ for a matrix $\mathring{A}$ whose nonzero partition is the corresponding partition of A. By permutations of rows and columns we may always express a solution A^B of (A, u, v) in the form

$$A^B = \begin{bmatrix} A^B_{11} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & A^B_{kk} \end{bmatrix},$$

where each diagonal partition A_{kk} is connected, $k = 1, \cdots, K$. Suppose this to have been done. Then we have

COROLLARY 6. *The solution A^B of (A, u, v) has the form*

$$A^B = \begin{bmatrix} \widehat{r_1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \widehat{r_K} \end{bmatrix} \begin{bmatrix} A_{11} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & A_{KK} \end{bmatrix} \begin{bmatrix} \widehat{s_1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \widehat{s_K} \end{bmatrix},$$

where A_{kk} is the partition of A corresponding to the partition A^B_{kk} of A^B, and the vector-pair (r_k, s_k) is unique up to the transformation $(\lambda_k r_k, (1/\lambda_k) s_k)$, where λ_k is a nonzero real number, $k = 1, \cdots, K$.

PROOF. The partition A^B_{kk} is a connected nonnegative solution of (A_{kk}, u_k, v_k), in an obvious notation. Hence by the proof of Theorem 2, $A^B_{kk} = \widehat{r_k} A_{kk} \widehat{s_k}$ for some vectors r_k, s_k unique up to multiplication by reciprocal real numbers.

Let us write

$$\begin{bmatrix} A_{11} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & A_{KK} \end{bmatrix} = \mathring{A}. \tag{40}$$

We note the obvious fact that $\mathring{A} = A$ is equivalent to A^B's being an interior solution.

Finally, Corollary 7 relates the connectedness of A^B to the equality of $\mathring{A}$ with A; hence to A^B's status as an interior solution.

COROLLARY 7. *Let A be connected. A^B is connected if and only if $\mathring{A} = A$.*

PROOF. Sufficiency is obvious. If A^B is connected then $K = 1$ in the expression (40) for $\mathring{A}$ and $\mathring{A} = A_{11}$ is an $(m \times n)$-order partition of A and therefore equal to A.

We may also prove Corollary 7 by a different route. If A is connected, then we know from the proof of Theorem 2 that A^B is connected only if it is an interior solution, and the sufficiency of this condition is clear. The result then follows from the observation above that $\mathring{A} = A$ if and only if A^B is interior.

University of Cambridge, England

REFERENCES

[1] BACHARACH, M., *Biproportional Matrices and Input-Output Change*, Department of Applied Economics, Cambridge University, duplicated (1965).

[2] BATES J. AND M. BACHARACH, *Input-Output Relationships, 1954-1966*, (*A Programme for Growth* 3), (London: Chapman and Hall, 1963).

[3] BÉNARD, J., "Réseau des échanges internationaux et planification ouverte", *Economie Appliquée*, XVI (April-June, 1963), 249-76.

[4] BROWN, J. A. C., "The Belgian RAS Experiment", Growth Project Paper No. 108, Department of Applied Economics, Cambridge University, duplicated (1962).

[5] ———, "The Second Belgian RAS Experiment", Growth Project Paper No. 113, Department of Applied Economics, Cambridge University, duplicated (1962).

[6] DEMING, W. E. 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*, XI (1940), 427-44.

[7] FRÉCHET, M., "Sur les tableaux dont les marges et des bornes sont données", *Review of the International Statistical Institute*, XXVIII (1960), 10-32.

[8] FRIEDLANDER, D., "A Technique for Estimating a Contingency Table, Given the Marginal Totals and Some Supplementary Data", *Journal of the Royal Statistical Society*, Series A, CXXIV, Part 3 (1961), 412-20.

[9] LEONTIEF, W. W., *The Structure of the American Economy, 1919-1939*, (New York: Oxford University Press, 1941).

[10] PAELINCK, J. AND J. WAELBROECK, "Étude empirique sur l'évolution de coefficients 'input-output'", *Economie Appliquée*, XVI (January-March, 1963), 81-111.

[11] STONE R. AND J. A. C. BROWN, *A Computable Model of Economic Growth*, (*A Programme for Growth* 1), (London: Chapman and Hall, 1962).

[12] THIONET, P., "Sur certaines variantes des projections du tableau d'échanges interindustrielles", *Bulletin of the International Statistical Institute*, Proceedings of the 34th Session (Ottawa, 1963), XL-Book 1, 119-31.

[13] WAELBROECK, J., "Les échanges des pays du Marché Commun avec les pays du monde extérieur", *Intégration Européenne et Réalité Economique*, (Collège de l'Europe, Tempelhof, Bruges, 1964).