Regional Science and Urban Economics 24 (1994) 3–25. North-Holland

On the methodology of input–output analysis

Thijs ten Raa*

Tilburg University, P.O. Box 90153, 5000 LE Tilburg, Netherlands

Final version received July 1993

The basic elements of input–output analysis, notably technical coefficients, quantity and value equations, and a total factor productivity growth measure, are derived as intermediate constructs when the problem of national income or product determination is directly related to input and output flow data. By embedding input–output concepts in a neoclassical framework, specification issues are resolved, notably the problems of construction of coefficients and of determination of value. Conversely, neoclassical concepts of marginal productivities can be related to a consistent input–output framework of data. Sources of substitution are identified.

JEL classification: C67; D57

1. Introduction

Input–output analysis and neoclassical economics seem to part as schools of thought, with little appreciation of each others' contributions. Neoclassical economists criticize the rigidity of the input–output model, particularly its assumption of fixed coefficients and the failure to explain factor rewards. Input–output analysis is perceived as a mechanical manipulation of data. On the other hand, the neoclassical concept of a smooth production function which maps factor inputs directly into some jelly output, and its associated marginal productivities, meets a cool reception in the world of input–output economists. Neoclassical economics is considered an elegant, but futile theory. I do not intend to contribute to the criticism, but will attempt to accommodate it. I take the neoclassical critique seriously and will rethink the methodology of input–output analysis. By relating input–output, including its statistical basis, to economic problems and applications, I hope to inject theoretical stucture. Open issues, such as the choice of model in the construction of coefficients, the relationship with fixed proportions, the consolidation of the quantity and value systems in a unifying framework, and

Correspondence to: Thijs ten Raa, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, Netherlands

*I wish to acknowledge intellectual debt to my teachers, William Baumol and Wassily Leontief. I thank Bert Steenge, an anonymous referee, and editor Konrad Stahl for stimulating remarks, and the Royal Netherlands Academy of Sciences for a senior fellowship.

This paper was refereed under the editorial control of Konrad Stahl and John Quigley.

SSDI 0166-0462(93)02029-3

the foundation of productivity measurement, can be enlightened by an open-minded reconsideration of the relationship between data and economic objectives. The basic assumptions and equations of input–output analysis may emerge in the process, but possibly modified.

The core sections of the paper are the next two. In section 2, the construction of an input–output matrix is related to the quantity and value equations in which they are put to use. The equations, in turn, are derived in section 3 from the formulation of an economic problem, such as the determination of the national product. Section 4 introduces substitution. Section 5 discusses patterns of specialization that trouble solutions to economic models at the interface of neoclassical economics and input–output analysis. Section 6 resolves the trouble in a framework of intercountry substitution. An intertemporal version of substitution is reflected in productivity growth. Section 7 embeds the concept in the same economic problem that was used to generate the equations of input–output analysis.

2. Quantity and value equations: Construction implications

The center-piece of input output analysis is a matrix,

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix},$$

of technical coefficients, a_{ij}, which describe the commodity inputs per units of commodity outputs. Here $(1,\dots,n)$ classifies the commodities. For example, if 1 is iron and 2 is automobiles, then a_{12} is the quantity of iron needed in the production of an automobile. The standard reference is Leontief (1966). The simplicity of the framework has attracted both economists and statisticians to the field of input–output analysis. The matrix of coefficients, A, is thus used as the point of departure both for economic analysis and for national or regional accounting. In economic analysis, two input–output equations are prominent, namely the following:

$$x = Ax + y, \tag{M}$$

$$p = pA + v. \tag{F}$$

Here x is a vector of gross outputs, y is a vector of net outputs, p is a row vector of prices, and v is a row vector of value-added coefficients. The first equation equilibrates supply and demand and the second equation balances revenues and costs. They are the so-called quantity and value equations of input–output analysis, also called the material and financial balances. The latter terminology is reflected in the notation indexing the equations (M) and (F). A well-known application of the input output equations is national planning, particularly the determination of output levels which are required to sustain a certain level of final demand. An example would be a study of the implications of an exports promotion program. The increase of exports would appear in the final demand vector, y, directly, and in the gross output vector, x, indirectly through equation (M). Another example would involve the tracing of price effects which result from an increase in the value-added coefficients associated with an energy shock. It is straightforward to assess the direct energy costs increase in the row vector of value-added coefficients, v, but the indirect price effects are determined through eq. (F). In either case, analysis amounts to the solution of the input–output equations. Mathematically, the inversion of the matrix $I-A$ is at stake, which defines the so-called Leontief inverse of the A-matrix.

Since input–output constitutes a more or less unified framework for economy wide analysis, statisticians use it for the organization of intersectoral data. If the above commodity classification, $(1,\ldots,n)$, can also be used for sectors, then it is natural to set up a so called transactions table,

$$T=\begin{bmatrix} t_{11}\ldots t_{1n} & t_{1n+1} \\ \vdots & \\ t_{n1}\ldots t_{nn} & t_{nn+1} \end{bmatrix},$$

The first row of the table displays the sales receipts of sector 1 from sectors $1,\ldots,n$ as well as the final demand compartments (household and government consumption, investment and net exports). Table T underlies matrix A, in the sense that by appropriate divisions of transactions elements of T, one may calculate the technical coefficients of A. Thus, the input–output coefficients matrix bridges the gap between economic analysis and national accounts. However, one must be careful not to consider the input–output coefficients matrix as the point of departure for analysis. If all interaction between input–output statisticians and economists were channeled through the single concept of an input–output coefficients matrix, the two departments of investigation would have their own dynamics, with little cross fertilization. Moreover, by taking an input–output coefficients matrix as the point of departure, one risks imposing a framework of analysis that simply does not fit reality.

A prime illustration of an active interface between statistical and economic investigations is the construction of input–output matrices. If reality were to present itself through a simple input–output transactions table, T, the construction of a matrix of technical coefficients, A, would be a straightforward matter of divisions:

$$a_{ij}=t_{ij}\Big/\sum_{k=1}^{n+1} t_{jk}.$$

This situation is too simplistic. For one, the very existence of a transactions table presumes that commodities and sectors can be classified in the same way. Moreover, it suggests that sectors have a multitude of inputs, but only single outputs. To accommodate the obvious implications, Professor Stone has suggested accounting for inputs and outputs separately. Hence input flows are tabulated in a use table, U, and output flows in a make table, V. In the System of National Accounts (S.N.A.) proposed by the United Nations (1967), the convention is that the dimensions of matrices U and V are commodities × sectors and sectors × commodities, respectively. The inputs of sector 1 are listed in the first column of matrix U and the outputs in the first row of matrix V, and so on. The framework is general. In particular, the traditional transactions model is recovered if sectors can be identified with their primary commodity outputs or, more precisely, if make table V is diagonal. Then the transactions table, T, coincides with the use table, U, augmented by a column. The last column of T makes the row totals add to the gross outputs, as determined by the column totals of the make table, V. In this case, the matrix of technical coefficients is obtained by dividing the use part of the transactions matrix by the (diagonal) output matrix that is, $A = UV^{-1}$. In general, however, the make table, V, features non-zero off-diagonal elements, since sectors produce mixtures of outputs. The problem of constructing an input–output matrix, be it T or A, is therefore non-trivial. Alternative methods to deal with it exist and are described by Viet (this issue). These are the industry technology model, the commodity technology model, and many more. Alternative assumptions are made on the nature of the off-diagonal elements of make table V, also called secondary products, or on their input technologies. The choice of model is made on the basis of the reasonableness of the assumptions, as judged by the statisticians or the economists. Whatever model is employed, some matrix of technical coefficients, A, comes out ot it and is used in the equations of input–output analysis, particularly (M) and (F).

Implications of input–output applications for the statistical construction of an input–output matrix can be introduced by reporting some of my own experience at the Institute for Economic Analysis at New York University, where I had to construct input–output coefficients for non-fuel minerals. The coefficients were to be part of an enlargement of the United States Bureau of Economic Analysis (BEA) input–output matrix. The BEA constructs the input–output matrix, A, according to the so-called industry technology model, on the assumption that each industry has a specific input technology which is independent of the commodity composition of its output vector [Viet (this issue)]. This methodology is problematic. The resulting A-matrix is not invariant with respect to units of measurement. Invariance with respect to units of measurement simply means that, for example, the quantity of iron per automobile (a_{12} in section 2) ought to double when metric pounds are used instead of kilograms (one metric pound is 500 grams). The reason that the industry technology model fails to meet this requirement, is that it employs a concept of 'industry output', which intermingles different commodities. In other words, apples and oranges are added up. Wassily Leontief urged me to think of an alternative method for the construction of input–output coefficients, to comply with the requirements of invariance with respect to units of measurement. Now, if this were the only concern, one might propose setting all elements of the A-matrix equal to zero. This 'zero' method is well defined and invariant with respect to units of measurement. Intuitively, however, putting $A=0$ is nonsense. It is important to clarify this intuition. The 'zero' method is nonsensical in the context of the application because it invalidates the material balance equation, (M), or, for that matter, the financial balance equation, (F). The left-hand sides would no longer equal, but exceed the right-hand sides. Hence (M) and (F) impose restrictions on the construction of the matrix A.

The moral of my thought experiment is that the economic structure of input–output analyisis has implications for the statistical construction of the matrix. Kop Jansen and ten Raa (1990) list the elements of the structure, namely the material balance and the financial balance, as well as base year price invariance and a scale property. The material and financial balances are essentially eqs. (M) and (F) presented above. Only when they are observed do input–output matrices balance material requirements and financial accounts. The element of base year price invariance, (P), is essentially the invariance with respect to units of measurement, since a price system is basically a system of measures. The scale property, (S), is a counterpart of the latter in the real sphere. It requires that *if* an economy must have constant returns to scale and fixed proportions, *then* it must have constant coefficients. This logical requirement makes no assumptions on the observation of economies, but restricts the method of construction of the coefficients. More precisely, Kop Jansen and ten Raa (1990) have proved that the just described structure of input–output analysis, involving (M), (F), (P) and (S), not only imposes restrictions on the choice of model of construction, but determines it uniquely. By one theorem, in the real sphere, the combination of (M) and (S) is shown to imply that the A-matrix must be constructed by the so-called commodity technology model. By another theorem, in the nominal sphare, the combination of (F) and (P) is shown to imply the same result. The theorems do not necessarily favor the commodity technology model over alternative constructs. If, however, an alternative method of constructing input–output matrices is used, then one must be prepared to revise the basic structure of input–output analysis, since at least one of the equations or properties must be violated. For example, if one uses the U.S. input–output coefficients, which are constructed according to the industry technology model, but continues to decompose productivity growth rates by standard input–output analysis without revision, then a bias creeps in. This bias has been analyzed and estimated by ten Raa and Wolff (1991).

As mentioned before, another risk of considering an A-matrix as the point of departure for statistical and economic work, is that the frame need not fit reality. An input–output matrix, however constructed and however sensitive to the data, suggests economy-wide relations which need not hold. For the purpose of clarification, suppose we accept the basic structure of input–output analysis, say (M) and (S) and/or (F) and (P), and suppose we construct the input–output matrix accordingly. Then, following Kop Jansen and ten Raa (1990), the matrix A is constructed by the specifications of the commodity technology model. This model is defined by the assumption that each commodity has a unique input structure, irrespective of the sector of fabrication. Now it is well known that the commodity technology model has the problem of negatives. If applied mechanically, the formula yields some negative coefficients. The negatives are very small and are usually suppressed one way or another. It is natural to hypothesize that the negatives are due to errors of measurement. To his own surprise, ten Raa (1988) has rejected the hypothesis. In other words, the construction of coefficients which is consistent with the basic structure of input–output analysis yields negatives which cannot be ascribed to errors of measurement.

One reason that might account for the rejection is that ten Raa (1988) assumes that the variances of the errors are known. If the variances have to be estimated from a sample, it is more difficult to reject and the requirement of non-negative coefficients may be salvaged. I shall detail this approach after the introduction of multiple observations in section 4. The nature of the problem of negatives is easy to understand. Imagine that sector 1 is pure, producing a single output, commodity 1, but that sector 2 produces not only commodity 2, but also commodity 1, as a secondary product. The input coefficients for commodity 1 are revealed by sector 1. The input coefficients for commodity 2 are obtained by purification of sector 2. More precisely, evaluation of the commodity technology model formula of the matrix of input–output coefficients involves inversion of the output flow table, which in this example amounts to subtraction of the secondary product and the associated input requirements from sector 2. The input requirements associated with the secondary product (commodity 1) are given by the already computed input coefficients for this commodity. The problem of negatives emerges when sector 1 uses an input which is not used by sector 2. This input must be subtracted from sector 2 in the process of purification. The subtraction creates a negative input coefficient for this input in sector 2. Statistically, the problem can be ascribed to an error of measurement only by hypothesizing that either the entire input cell in sector 1 creating the negative in sector 2 or the zero entry in sector 2 is fake. I find this difficult to accept. A more reasonable approach of the problem seems to me to accept the possibility of coexisting technologies for the production of commodity 1 in sectors 1 and 2.

When this point of view is adopted, there is no point is constructing a single matrix of technical coefficients. The question of what remains to be done cannot be answered in general, but depends on the economic issue that is addressed, as well as data availability. A good example is the issue of profit maximization, which can be modeled in an input–output like framework, without technical coefficients. The relationship with input–output coefficients matrices will be discussed, but is not essential to the model and certainly imposes no non-negativity requirements.

3. Quantity and value equations: Economic origins

Optimization naturally is at the core of neoclassical economics. However, formulating an input–output problem in terms of a linear programming model does not necessarily introduce neoclassical elements in input–output models. Since the 1960s, linear programming formulations of Leontief-type systems, in which real and price systems can be viewed as a set of constraints, are a normal part of standard texts. The Dorfman et al. (1958) model considers the maximization of the value of a given bill of final demands by choice of prices and the minimization of labor input by choice of gross outputs. As ten Raa and Mohnen (this issue) argue, these combinations of objectives and instruments are not neoclassical. The problem is that the Leontief-type systems have been taken for granted and that the linear programs have been chosen to make the systems dual to each other.

Instead of taking an input–output coefficients matrix as a point of departure, I propose to relate data directly to economic problems, without imposing a preconceived input–output structure. To illuminate, consider the problem of the maximization of the national product, y, valued at the world price vector, p, subject to a constraint on resources, say a labor force, N. Here y is a commodity vector, which lumps together the familiar categories of the national product, namely household and government demand, investment, and net exports. In view of the last item, y may have negative components. Other items in final demand may have negative elements as well. Investment, for example, is comprised of gross fixed capital formation and net inventory change. If inventory depletions exceed inventory additions and if the negative net inventory change exceeds new capital formation, as may well occur in severe recessions, elements of this component will also be negative. Individual elements of private consumption may also be negative if consumers sell used materials (cars, clothing, etc.). p is an exogenous row vector of given world prices. N is the exogenous number of workers. The input–output data comprise a use table, U_0, a make table, V_0, and a sectoral employment row vector, L_0. If we drop the subscripts, the data turn variables, namely U, V, and L. We now relate the subjects of the economic problem, the national product, y, and the labor force, N, to the variables, (U, V, L). Recalling the dimensions of U (commodity × sector) and V (sector × commodity), we see that $V^{\mathrm{T}}-U$ is the net output matrix (commodity × sector) and that aggregation over sectors yields the net output vector. Hence $y=(V^{\mathrm{T}}-U)e$, where e is the vector with all entries equal to unity and T denotes transposition. Similarly, the resource constraint reads $Le \leqq N$. It remains to restrict the variables, (U, V, L), in agreement with the production possibility set. Certainly feasible are the observed values, (U_0, V_0, L_0). If we assume constant returns to scale, then $(U, V, L)=(U_0\hat{s}, \hat{s}V_0, L_0\hat{s})$ is also feasible for any non-negative vector of scales, s. (Here $\hat{s}$ is the associated diagonal matrix.) Thus, the problem is written as

$$\max_{s \geqq 0} py$$

subject to

$$y=(V^{\mathrm{T}}-U)e, \quad Le \leqq N, \quad (U, V, L)=(U_0\hat{s}, \hat{s}V_0, L_0\hat{s}).$$

As far as I am concerned, we are done. We have related the economic problem directly to the data and economic analysis can be performed; see ten Raa and Mohnen (this issue). Input–output analysis is implicit. To reveal it, consider the following change of variables:

$$x=V^{\mathrm{T}}e=(\hat{s}V_0)^{\mathrm{T}}e=V_0^{\mathrm{T}}\hat{s}e=V_0^{\mathrm{T}}s \quad \text{or} \quad s=V_0^{-\mathrm{T}}x=(V_0^{\mathrm{T}})^{-1}x.$$

The change of variables is one-to-one only if the make table, V_0, is invertable. In other words, the recognition of standard input output analysis is possible only if there are equally many commodities and sectors. This condition is not needed for a head-on analysis of the above economic problem. I now invite the reader to go through a number of steps, which are mathematically trivial, but not so methodologically. The change of variables turns the problem into

$$\max_{x \in X} py$$

subject to

$$y=(V^{\mathrm{T}}-U)e, \quad Le \leqq N, \quad (U, V, L)=(U_0\widehat{V_0{}^{\mathrm{T}}x}, \widehat{V_0{}^{\mathrm{T}}x}V_0, L_0\widehat{V_0{}^{\mathrm{T}}x},$$

where X is the cone spanned by the rows of V_0, because of $x=V_0^{\mathrm{T}}s$, $s \geqq 0$.

Elimination of (U, V, L) by substitution of the last constraint yields

$$\max_{x \in X} py$$

subject to

$$y=[(\widehat{V_0^{-T}x}\, V_0)^T - U_0 \widehat{V_0^{-T}x}]e, \quad L_0 \widehat{V_0^{-T}x}\, e \leqq N.$$

This can be simplified to

$$\max_{x \in X} py$$

subject to

$$y = V_0^T V_0^{-T} x - U_0 V_0^{-T} x, \quad L_0 V_0^{-T} x \leqq N$$

or

$$\max_{x \in X} py$$

subject to

$$x = Ax + y, \quad lx \leqq N,$$

where we introduced the shorthand $A=U_0V_0^{-T}$ and $l=L_0V_0^{-T}$. This notation may be recognized as the formula for technical coefficients according to the commodity technology model. We see that standard input–output analysis *may* provide the correct formulation of an economic problem, but that it is by no means trivial. In particular, the method of construction of the input–output matrix is implicit in the formulation of the economic problem. The analysis involved a listing of the economic variables of the problem, (y, N), plus a relationship with the data, (U_0, V_0, L_0), and the associated variables, (U, V, L). In proceeding this way, technical coefficients merely fall out as values of mappings defined on the data. The mappings are determined by the formulation of the economic problem.

The material balance equation, (M), has emerged in the course of rewriting the economic problem. I shall now discuss the emergence of the financial balance equation, (F). Let us assume, in addition to constant returns to scale, free disposability and that output vectors span non-negative space. The problem becomes

$$\max_{x \geqq 0} py$$

subject to

$$x \geqq Ax + y, \quad lx \leqq N.$$

Although the material balance constraint set is widened by the replacement of the equality sign, this will not affect the solution, for it is easy to show that the constraint is binding. The latter problem, also called the primal program, lends itself to a more convenient formulation of the so-called dual program. Applying Schrijver (1986, p. 90), the dual program becomes

$$\min_{w \geqq 0} wN$$

subject to

$$p \leqq pA + wl.$$

Here w is the Lagrange multiplier associated with the labor force constraint, or marginal productivity of labor. An increase in N by one unit in the primal problem would increase the solution value of py. The latter increase is w, which, therefore, is also called the shadow wage rate. In contrast to standard input–output analysis, the wage rate is related to the quantity system through the concept of marginal productivity. This is the central place where a neoclassical framework provides structure to input–output analysis. The extension to other factor inputs, such as capital, is obvious. Although there is non-substitutability at the sectoral levels, changes in the composition of final demand allow for factor intensity variation and, therefore, full employment. The rewards are the marginal productivities with respect to the national product.

By the main theorem of linear programming, the solutions to the primal and dual programs yield equal values:

$$py = wN.$$

This is the equality between the national product and income, in our single-factor economy with zero operating surplus. Now, by the two inequalities of the primal program,

$$(p - pA - wl)x \geqq py - wN = 0.$$

A last step to establish the value equations involves a new concept. Define active and sleeping sectors as follows. Sector $i \in I$ (the active sectors) if $x_i > 0$, and $i \in II$ (the sleeping sectors) if $x_i = 0$. Then I and II partition the sectors and

$$x=\begin{pmatrix} x_I \\ x_{II} \end{pmatrix},$$

with x_I strictly positive and x_{II} zero. The last inequality becomes

$$(p-pA-wl)_I x_I+0\geqq 0,$$

and by the constraint of the dual program,

$$(p-pA-wl)_I=0$$

or

$$p_I=(pA)_I+wl_I.$$

We have now arrived at the value equations of input–output analysis, previously indicated by (F), the financial balance. They are not standing by themselves, but follow from the same economic problem that was also used to establish the quantity equations or the material balance, (M). Once more, it is important to note the unification brought about by the neoclassical framework of profit maximization. The equations not only emerge as primal and dual constraints to a common problem, but the quantities and the value are determined jointly. In particular, the wage rate is the marginal productivity of labor and the prices are consistent with it through competitive cost equations. Only active sectors are relevant in the determination of value. This is a methodological point that would have been overlooked if the basic structure of input–output analyis had been taken for granted rather than derived.

The various assumptions that I made in the course of the derivation were introduced to reveal the implicit role of standard input–output analysis, but are not really necessary to the investigation of the economic problem. For example, the shadow prices of factor inputs are presented directly as the Lagrange multipliers to the program and are thus available without the necessity to set up value equations.

4. Substitution

Input–output economics was invented by Wassily Leontief and the so-called Leontief production function is defined by the absence of substitution. Many economists therefore think that in input–output analysis there is no substitution. Substitution will be analyzed as an issue of changing coefficients, A. Let us investigate the thought that input–output analysis excludes substitution. As before, we have input and output matrices, U and V, and observations U_0 and V_0. We have seen that economic problems may be expressed in terms of U and V and that in the course of analysis an input–output matrix based on U_0 and V_0 may emerge. The latter dependence is denoted by writing $A(U_0, V_0)$. Taken as a mapping, A is a model of construction, a device that tells you how to manipulate the arguments. Whatever the model of construction, if we change the data, (U_0, V_0), then the coefficients, $A(U_0, V_0)$, change as well, except when the new observations, say U_1 and V_1, are collinear, with the collinearity given by the old coefficients: $U_1 = A(U_0, V_0)V_1$. This change involves *substitution* of inputs if and only if components in a column of A move in opposite directions. In other words, the very fact that coefficients change with data changes indicates the presence of substitution in input–output analysis. This is not what is meant by the neoclassical critics. Instead, they refer to the use of coefficients based on one observation. In scenario analysis the constructed coefficients matrix, $A(U_0, V_0)$, is applied to some hypothetical (U, V), and feasibility of the latter reads

$$Ue = A(U_0, V_0)\,V^{\mathrm{T}}e.$$

This equation can be shown to be equivalent to the material balance, (M), using the change of variables, $x = V^{\mathrm{T}}e$. The equation holds trivially for $(U, V) = (U_0, V_0)$, at least when model A is an established one. For example, if A is the commodity technology model, then $A(U_0, V_0) = U_0 V_0^{-\mathrm{T}}$ and the above feasibility equation reads $Ue = U_0 V_0^{-\mathrm{T}} V^{\mathrm{T}} e$, which is true for $(U, V) = (U_0, V_0)$. If this equation is required for all feasible (U, V), including hypothetical ones, then inputs and outputs must be proportional (with coefficients $U_0 V_0^{-\mathrm{T}}$) and, therefore, substitution is assumed away indeed. This methodology may make sense if there is only one observation, (U_0, V_0), and even then merely reflects an extreme restriction of data availability. Otherwise the neoclassical critique becomes most relevant and substitution becomes unavoidable.

Consider a second observation, (U_1, V_1). It would be a coincidence if $A(U_1, V_1) = A(U_0, V_0)$. If, in scenario analysis, some weighted average of $A(U_0, V_0)$ and $A(U_1, V_1)$, say $\bar{A}$, were applied to (U, V), then feasibility would read $Ue = \bar{A}V^{\mathrm{T}}e$ and substitution would still be absent. However common, this approach seems inconsistent to me, at least in a non-stochastic world. For example, the observations $(U, V) = (U_0, V_0)$ and $(U, V) = (U_1, V_1)$ need not be feasible under $\bar{A}$, not even under standard technology assumptions like constant returns to scale and free disposability. A simple illustration is given by the following pair of observations,

$$(U_0, V_0) = \left(\begin{bmatrix} 1/3 & 0 \\ 0 & 2/3 \end{bmatrix}, I\right) \quad \text{and} \quad (U_1, V_1) = \left(\begin{bmatrix} 2/3 & 0 \\ 0 & 1/3 \end{bmatrix}, I\right).$$

Then

$$A(U_0, V_0)=\begin{bmatrix}1/3 & 0\\ 0 & 2/3\end{bmatrix} \quad \text{and} \quad A(U_1, V_1)=\begin{bmatrix}2/3 & 0\\ 0 & 1/3\end{bmatrix}.$$

Hence

$$\bar{A}=\begin{bmatrix}a & 0\\ 0 & b\end{bmatrix}$$

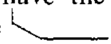
with $1/3<a,\ b<2/3$. (U_0, V_0) is feasible under $\bar{A}$ if and only if $U_0 e \geqq \bar{A} V_0^{\mathrm{T}} e$ (using free disposability of inputs) or

$$\begin{bmatrix}1/3\\ 2/3\end{bmatrix} \geqq \begin{bmatrix}a\\ b\end{bmatrix}.$$

(U_1, V_1) is feasible under $\bar{A}$ if and only if

$$\begin{bmatrix}2/3\\ 1/3\end{bmatrix} \geqq \begin{bmatrix}a\\ b\end{bmatrix}.$$

These two conditions cannot be met by $a, b>1/3$. This completes the demonstration that a weighted average of the coefficients is inconsistent with feasibility of the observed flows.

It seems more appropriate to declare (U, V) feasible if it can be decomposed in two terms, say $(U, V)=(U^0, V^0)+(U^1, V^1)$, with (U^0, V^0) feasible with respect to $A(U_0, V_0)$ and (U^1, V^1) feasible with respect $A(U_1, V_1)$. In input space isoquants no longer have the familiar L-shape of a Leontief production function, but look like  . Such an isoquant features an interval of perfect substitution.

A prime setting for this elementary type of substitution is a model of international trade between countries with different technologies, that is $A(U_0, V_0) \neq A(U_1, V_1)$, where 0 now represents the home country and 1 the foreign country. Although each country may be incapable of substituting inputs, reallocations of activity brings it about at a global level. Trade mitigates substitution and, when modeled properly, input–output thus loses its problematic features of excess supplies and zero prices for some inputs, at least when net output proportions are fixed, as we shall see in section 6. Neoclassical features are thus introduced without having to go all the way to the concept of a smooth production function. In my view, a Cobb–Douglas production function, or any other functon with smooth isoquants, is generated only in a world with infinitely many observations. The above shape of an isoquant is modified further by more kinks, and eventually becomes smooth.

Bert Steenge has pointed out that my discussion of two (or more) observations is unambiguous only in a non-stochastic setting. He also suggested that if (U_0, V_0) and (U_1, V_1) are viewed as realizations of a random variable, $\bar{A}$ can be interpreted as an approximation of the 'true' input–output matrix. In this case, infeasibilities of observations under $\bar{A}$ could be ascribed to errors. A discussion of the implications of allowing for a stochastic approach may be as follows. Input observations would be related to output observations through the 'true' matrix, say $\mathscr{A}$, and discrepancies are collected in error terms, say ε:

$$(U_0, U_1) = \mathscr{A}(V_0^{\mathrm{T}}, V_1^{\mathrm{T}}) + (\varepsilon_0, \varepsilon_1).$$

(The extension to more than two observations is obvious.) A would be an estimator of $\mathscr{A}$, and hence a function of (U_0, U_1) and (V_0, V_1). It could be ordinary least squares, or restricted ordinary least squares, if the true matrix is non-negative. The framework is consistent with the axioms of Kop Jansen and ten Raa (1990), as well as non-negativity requirements. Non-negativity is less likely to be rejected as in ten Raa (1988), since the variance–covariance matrix is no longer known, but must be estimated from (U_0, U_1) and (V_0, V_1). If non-negativity continues to be rejected in the presence of many observations, my preferences would be to drop the notion of a common 'true' matrix by admitting different coefficients, that is substitution. Moreover, I would not enforce non-negativity on each realization. In fact, explicit evaluation of the coefficients is not necessary in economic analysis, not even when proportions are assumed to be constant. I refer to the analysis of ten Raa and Mohnen (this issue) for an illustration.

A multitude of observations and underlying techniques is one source of substitution between factor inputs. Another source is the commodity composition of final demand or, more precisely, its variability. Neoclassical economists exploit this source of substitution as well, but in a rather implicit manner, through the concept of an aggregated commodity. It is illuminating to establish the relationship a bit more clearly. It suffices to consider one observation, (U_0, V_0, L_0, K_0), including sectoral labor and capital employment row vectors. Turning to variables by dropping the zero subscripts and introducing, as before, technical coefficients $A = U_0 V_0^{-\mathrm{T}}$, $l = L_0 V_0^{-\mathrm{T}}$ and $k = K_0 V_0^{-\mathrm{T}}$, the factor requirements of a bill of final goods, y, becomes $l(I-A)^{-1}y$ and $k(I-A)^{-1}y$, which clearly vary with the composition of vector y. This simple source of substitution is sufficient to obtain full employment of resources (ten Raa and Mohnen, this issue).

Most neoclassical economists think of a combination of the two sources of substitution when modeling a national production function. The effects of choice of techniques, and hence alternative technical coefficients, on the relationship between net outputs and factor inputs are envisaged. It should now be clear that methodologically this is a problem of the effects of coefficients variation on $(I-A)^{-1}$, the Leontief inverse of A. Thus, the input–output counterpart of neoclassical substitution is variation of the Leontief inverse. This problem is analyzed in a stochastic setting by Kop Jansen and Steel (this issue).

In a neoclassical framework, substitutability of inputs is determined by their marginal productivities. For factor inputs the commodity composition effect of final demand is, as mentioned, one source of substitution. If we denote the solution to the primal program, py, by Q, then, by the main theorem of linear programming, $Q=wL_0+rK_0$, where w and r are the Lagrange multipliers of the labor and capital constraints which fulfill $w=\partial Q/\partial N$ and $r=\partial Q/\partial K$. So even though there may be no substitution of inputs within sectors, the possibility of varying components of the net output vector in solving the economic program yields substitutability of factor inputs.

Conceptually, substitution is modeled by constructing the hybrid economy comprising all the observed techniques. The first application is in Carter (1970). In the solution, only one technique of the observed ones will be active and the others are worse as valued by the shadow prices of the material balance and factor input constraints, by the phenomenon of complimentary slackness. So all you know is such types of inequalities. When the techniques are not finite but constitute a continuum and can thus be parameterized, the superiority of the active techniques in terms of value can be assigned first-order conditions yielding the equality between relative prices and marginal rates of substitution. However, I dislike this idealization and, therefore, refrain from relating coefficients changes to the dual prices.

5. Specialization

When an economic problem is formulated mathematically as a linear program, bang-bang behavior is to be expected. Typically, the number of active variables is no more than the number of constraints. When the value of the national product is maximized subject to the material balance constraints and the primary factor constraints, the former are binding and can be used to express national product components in sectoral activity levels. This elimination procedure leaves only the primary factor constraints to bind the sectoral activity levels and, therefore, the number of active sectors will match the number of primary factors. For example, in section 3, where only labor was considered, an extreme form of specialization in only one sector results.

The extreme behavior of input–output type models is believed to be caused by the assumption of non-substitutability of inputs of a technique. This belief is false. In fact, substitution makes things worse. My elaboration comes in three fold. First, I shall reproduce the argument of non-substitutability as a source of extreme behavior. Second, I shall discuss the consequences of the introduction of substitution in the maximization problem of this paper. Third, I shall comment on a neoclassical approach to the problem of specialization.

A macro-economic production function relates net output of a national economy directly to its factor inputs. Gross outputs and intermediate inputs are implicitly eliminated by Leontief inversion. Although no one would argue that this procedure yields a Leontief production function for an aggregate measure of output, many applied input–output studies exhibit this behavior. Whenever the proportions of the final bill of goods are fixed by consumption and trade coefficients, the fixed primary input proportions can be associated with components of the vector of final goods and be weighted. In this case the national economy is implicitly modeled by a Leontief macro-economic production function and bang-bang behavior emerges in the form of some zero shadow prices of factor inputs. These observations pertain to standard linear programming formulations of the input–output model, such as Dorfman et al. (1958), but not to our approach. The proportions of final goods many vary freely and, therefore, the aggregate factor intensities also. Since the objective function is a linear valuation of net outputs, the latter are perfect substitutes and, since factor intensities vary across net outputs, there is some degree of substitution between factor inputs as well. So the extreme behavior of our model is not the usual phenomenon of Leontief-type models. As a matter of fact, the introduction of substitution makes things worse. The best way to understand this is to go back to the very first formulation of the economic problem in section 3. As we have seen in section 4, the introduction of substitution merely increases the dimension of the activity space (or the vector of scales, s). The linear programming result that the active number of activities matches the number of primary inputs remains. The disaggregation of a sectoral activity into a number of activity vectors that comes with the introduction of substitution cannot increase the number of active sectors, but has the possibility of concentrating activities in fewer sectors. By going to the limit of neoclassical production functions, one cannot escape this logic. A fine example is Diewert and Morrison (1986). To avoid specialization, they impose a translog structure not on the production function or inputs of the economy, but on the national product function, or outputs of the economy. This procedure eliminates specialization, but it is brute force for four reasons. First, a peculiar jointness of net output is implicitly assumed. Second, since the signs of the components cannot flip when a translog function is imposed, the pattern of trade must be considered given. Third, estimation at the net output side of the economy requires the assumption that the observed flows are consistent with perfect competition. Fourth, even when the previous point is taken for granted, the required concavity assumptions are inconsistent at the output level, as admitted in a footnote by Diewert (1982, p. 576).

Specialization is a serous 'problem' that plagues input–output as well as neoclassical models of national product determination. In applications additional constraints are considered, such as the non-tradability of certain commodities, and the number of active sectors is increased accordingly. An intermediate device is to model imports as imperfect substitutes. Although such practices remedy the extreme nature of corner solutions, the economics of specialization must be accepted. The best objection against our linear programming approach is that the pattern of specialization is dependent on the coefficients of the objective functions, such as the world terms of trade, and that variations in the latter cannot be anticipated, so that diversification is a safe policy. However, without imposing a peculiar jointness on net outputs, Gilchrist and St. Louis (this issue) are able to address diversification by taking into account the fluctuations in the terms of trade. Their study is regional economic. Patterns of specialization, as predicted by international trade theories, are best tested in regional economics since impediments to trade are less prevalent between regions than between nations.

6. Closing the model

The possibility of importing commodities admits negative components to the net output vector of a national economy and, therefore, the specialization in a number of sectors. Under conditions of national product maximization and free trade, the structure of an economy thus degenerates into a number of columns and the usual input–output multiplier effects evaporate from the national economy. It is only at the level of the world economy that the circularity of production and traditional input–output results re-emerge. The international division of labor is a vehicle for substitution. Commodities can be produced in different national economies with varying input proportions. As noted in section 4, substitution is modeled by constructing the hybrid economy comprising all the observed techniques. The commodity × sector use tables are stacked next to each other and the sector × commodity make tables are stacked under each other. In other words, sectors in different countries are treated as separate sectors. Since the system of National Accounts makes a distinction between commodities and sectors anyway, identifying the latter with pairs of input columns and output rows, there is no need to classify sectors across countries. Their numbers and order may vary; we only have to put them next to each other.

Thus let U_0, V_0, and L_0 be the use, make, and employment tables of the world obtained by stacking the national ones. Let s be the column vector of sectoral activity levels. The number of components is the sum of the numbers of sectors in the various countries. The sign pattern of the economic variable s will reveal the pattern of specialization between countries in the different commodity markets. The net output of the world will be $(V_0^{\mathrm{T}} - U_0)s$, assuming constant returns to scale. The total labor requirements are $L_0 s$.

For ease of exposition, I assume that labor is mobile, so that activities are constrained by $L_0 s \leqq N$, the world labor force. The alternative assumption of immobile labor could be accommodated by treating workers of different countries as different factor inputs.

Unlike the national economic analysis expounded in section 3, it does not make sense to maximize world net output at given world prices. Components of net output would be negative. They might be forced to be positive by adding constraints, but those constraints would be binding and their specification would drive the allocation of activity in a direct, mechanical manner. The world economy is closed and all trades cancel out in its net output vector. The net output must be related to preferences rather than some exogenous price system. Thus, let the desired net output proportions be given by a vector a with a non-negative shares for the commodities, adding to unity. There is no objection against declaring the status quo net output proportions desirable. The level of the desired net output vector is variable in the economic analysis and, in fact, constitutes the objective function:

$$\max_{s,\, c \geqq 0} c$$

subject to

$$(V_0^{\mathrm{T}} - U_0)s \geqq ac, \quad L_0 s \leqq N.$$

This program determines the pattern of specialization of countries. Net output is non-negative at the world level, by the constraint that imposes the desired proportions, but may have negative components for individual countries. Commodity prices are endogenous. In fact, they are the shadow prices associated with the net output constraints. The dual program reads

$$\min_{p,\, w \geqq 0} wN$$

subject to

$$pV_0^{\mathrm{T}} \leqq pU_0 + wL_0, \quad pa = 1.$$

Note that this dual program is basically the same as the dual program to the national product maximization program of section 3. The only essential difference is that the commodity price vector is now variable. The derivation of the value equations of input–output analysis is unaffected. As before, the solutions to the primal and dual programs yield equal values:

$$c = wN.$$

This is the equality between the world product and income. By the constraints of the primal program and the normalization of prices,

$$(pV_0^{\mathrm{T}} - pU_0 - wL_0)s \geqq pac - wN = 0.$$

As before, partition sectors in active ones ($s_i > 0$) and sleeping ones ($s_i = 0$), constituting index sets I and II, respectively. Then

$$(pV_0^{\mathrm{T}} - pU_0 - wL_0)_I s_I + 0 \geqq 0,$$

and by the constraint of the dual program,

$$(pV_0^{\mathrm{T}} - pU_0 - wL_0)_I = 0$$

or

$$pV_{0I}^{\mathrm{T}} = pU_{0I} + wL_{0I},$$

where I selects the active columns. It can be shown that the number of active comumns need not be greater than the number of commodities. Moreover, if the desired net output proportions (a) are not so extreme that some component can be supplied as a by-product, the number of active sectors must be at least equal to the number of commodities. Thus, V_{0I}^{T} will be a square matrix. Moreover, if primary output is dominant, it is invertible and we obtain

$$p = pA_I + wl_I,$$

where $A_I = U_{0I}V_{0I}^{-\mathrm{T}}$ and $l_I = L_{0I}V_{0I}^{-\mathrm{T}}$, the input coefficients according to the commodity technology model as applied to the active sectors. Consequently,

$$p = wl_I(I - A_I)^{-1},$$

the Marxian labor values as determined by the coefficients of the active sectors. Note that this general equilibrium price does not depend on the assumed desired net output proportions, a. As a matter of fact, not even the selection of active sectors, I, depends on a. In the dual program, a only normalizes prices. It can be shown that the price vector that solves the dual program is independent of the normalization constants listed in the vector a. Consequently, a component of the dual constraint is binding or not binding, whatever a. Thus, the classification of break-even and unprofitable sectors is independent of a. By the above analysis the break-even sectors are precisely the active sectors, and the profitable sectors are the sleeping sectors. This shows that the classification of sectors in active and sleeping ones is independent of the preferences. The application of the theory of linear programming thus provides a simple proof of the substitution theorem of

Samuelson (1951), by which an economy with a single factor input will not substitute techniques in response to changing demand conditions.

Note that the dual variables (prices) are positive. In the same way that the dual constraints were shown to be binding when the primal variables (activity levels) are positive, we can now conclude that the primal constraints are binding,

$$(V_0^{\mathrm{T}} - U_0)s = ac,$$

and, therefore,

$$(V_0^{\mathrm{T}} - U_0)_I s_I = ac.$$

By change of variable, $V_0^{\mathrm{T}} s_I = x$, we now have

$$x = A_I x + ac,$$

the traditional input–output equation, featuring circularity of production and the consequent multiplier effects:

$$x = (I - A_I)^{-1} ac.$$

The technical coefficients are not determined by aggregation of sectors across countries and submission of the world use and make tables to the standard formula, but by the best practice techniques selected by the linear program. While the selection is robust in our simple world model with one factor input, the situation becomes more complicated when more factor inputs are introduced. The input–output relations are maintained, but the set of active sectors may vary to accommodate factor scarcities.

Although the national product maximization program expounded in section 3 and the desired consumption level program for the world economy of this section would make no sense in each other's contexts (the national program entails negative net outputs and the world program excludes them), they are consistent. The general equilibrium model of this section subsumes the partial equilibrium model of section 3 if the prices which were considered there are the solution to the world model. Otherwise the extreme patterns of national net exports would yield excess supplies or demands in the world markets.

7. Productivity growth

One might argue that neoclassical economics provides a reduced form of the input–output model. Leontief inversion is presumed implicitly in neoclassical models and the value of net output attainable for given levels of resources is written by a simple function in lieu of the solution to a linear program as in section 3. Since the neoclassical production function is essentially the value function of a linear program, the marginal productivities of the resources are the Largrange multipliers to the linear program. To introduce productivity more precisely, recall the data of an economy: use and make tables, labor and capital employment row vectors, as well as total endowment stocks. As before, net output is $y=(V^{T}-U)e$, while factor inputs are stocks N and K. Roughly speaking, productivity is net output divided by factor input, hence profitability growth is the change of net output minus the change of factor input. The traditional measure of total factor productivity growth is

$$(p\,\mathrm{d}y - w\,\mathrm{d}N - r\,\mathrm{d}K)/(py),$$

where w is the wage rate and r the rental rate of capital. Input–output economists [Wolff (this issue)] consider w and r exogenous and commodity prices p endogenous, using the value equations of section 2,

$$p = pA + wl + rk. \qquad \text{(F')}$$

The traditional measure of productivity growth is rather mechanical, but can provide a theoretical foundation by embedding the input–output relationships in the neoclassical framework of profit maximization. Productivity is properly defined only if there is a criterion, or objective function, to measure the contributions of factor endowments. In my view, factor productivity *is* w or r, the shadow prices or Lagrange multipliers to a maximization problem. After all, shadow prices measure the contributions of factor inputs. Consequently, since factor productivity growth ought to be the growth of factor productivity, it must be $\mathrm{d}w$ or $\mathrm{d}r$. Hence factor productivity growth rates are changes in shadow prices resulting from changes in the data $(U_0, V_0, L_0, K_0, N, K)$. Since $\mathrm{d}w$ and $\mathrm{d}r$ are per unit of factor input, total factor productivity growth is $N\,\mathrm{d}w + K\,\mathrm{d}r$, or, relative to national product or income,

$$(N\,\mathrm{d}w + K\,\mathrm{d}r)/(py).$$

This shadow-price-based measure of total factor productivity growth can be used as a foundation for the traditional measure through the main theorem of linear programmity of section 3:

$$py = wN + rK.$$

Here w and r are endogenous marginal productivities associated with unit increases in N and K, but p is the exogenous row vector specifying the criterion (national product of world prices, for example). Hence differentiation yields

$$p\,dy = w\,dN + N\,dw + r\,dK + K\,dr.$$

Substitution into my direct definition of total factor productivity growth, $(N\,dw + K\,dr)/(py)$, yields the traditional measure outlined at the beginning of this section.

Note that in defining and deriving total factor productivity growth, I made no appeal to the traditional value equations, (F′). p is exogenous and w and r are the shadow prices associated with a maximization problem. It is an open question which value equations they fulfill. If there are no lower bounds to net output, they they fulfill the value equations restricted to the active sectors in the maximization problem, as defined in section 3. Typically, their number is the number of constraints, which is only two! This case is relevant to the measurement of productivity growth of an open economy under free trade. Under alternative regimes, commodity prices pick up tariffs and the consequent full prices observe a more compete system of traditional value equations. These tariffs are endogenous, see ten Raa and Mohnen (this issue). Input–output economists, by using the full price vector in evaluating total factor productivity growth, implicitly take trade restrictions for granted. The precise relationship between trade regimes and the measurement of total factor productivity growth is an open issue and presently under investigation.

8. Conclusion

In this paper we have used the neoclassical concept of profit maximization as a framework for the input and output data of an economy. The basic elements of input–output analysis, notably technical coefficients, and the quantity and value equations, emerged as intermediate constructs in the course of analysis. The technical coefficients construction methodology is forced by the quantity and value equations, and the latter are derived from the primal and dual constraints to the problem of profit maximization. Value-added coefficients are no longer exogenous but related to the quantity system as shadow prices. Their rates of change can be used to define factor productivity growth rates and the traditional input–output measure of total factor productivity growth has thus been provided a neoclassical foundation.

References

Carter, A.P., 1970, Structural change in the American economy (Harvard University Press, Cambridge, MA).

Diewert, W.E., 1982, Duality approaches to microeconomic theory, in: K.J. Arrow and M.D. Intriligator, eds., Handbook of mathematical economics, vol. II (North-Holland, Amsterdam).

Diewert, W.E. and C.J. Morrison, 1986, Adjusting output and productivity indexes for changes in the terms of trade, Economic Journal 96, 659–679.
Dorfman, R., P.A. Samuelson and R.M. Solow, 1958, Linear programming and economic analysis (McGraw-Hill, New York).
Kop Jansen, P. and Th. ten Raa, 1990, The choice of model in the construction of input–output coefficients matrices, International Economic Review 31, no. 1, 213–227.
Leontief, W., 1966, Input–output economics (Oxford University Press, New York).
ten Raa, Th., 1988, An alternative treatment of secondary products in input–output analysis: Frustration, Review of Economics and Statistics 70, no. 3, 535–538.
ten Raa, Th. and E. Wolff, 1991, Secondary products and the measurement of productivity growth, Regional Science and Urban Economics 21, 581–615.
Samuelson, P.A., 1951, Abstract of a theorem concerning substitution in open Leontief models, in: T.C. Koopmans, ed., Activity analysis of production and allocation, Cowles Commission Monograph no. 13 (Wiley, New York).
Schrijver, A., 1986, Theory of linear and interger programming (Wiley, Chichester).
United Nations, 1967, Proposals for the revision of SNA, 1952, Document E/CN.3/356.