

# COMPILATION OF INPUT-OUTPUT DATA FROM THE NATIONAL ACCOUNTS

(Received June 1994; revised October 1994)

**ABSTRACT** In this paper, a new method is presented to derive an input-output table from a system of make and use tables. The method, which we call 'activity technology', is mathematically equivalent to the well-known commodity technology, but chooses another unit, i.e. the activity. We will argue that, in the activity technology model, negatives can only arise from causes such as heterogeneity and errors in the data. To apply the activity technology, very detailed make and use matrices are required, as well as additional data on the input structures of certain activities. We will describe a method that can incorporate this additional data within the activity technology framework. Statistics Netherlands has adopted the method

**KEYWORDS:** Input-output tables, make and use tables, national accounts, secondary products

## 1. Introduction

Input-output tables serve two purposes. They are used as an integral part of the system of national accounts (SNA) being used by statisticians as a device to reconcile data from different sources and, thus, arrive at better estimates of, for instance, the national product. Input-output tables are also used by economists to perform analyses on the production structure of an economy. For both purposes, the situation has changed over the last few years.

In the 1968 revision of the SNA (UN, 1968), the traditional input-output table was replaced by two tables: the use matrix, describing the demand for commodities by industries, and the make matrix, describing the supply of commodities by industries. Commodity flows can be described more accurately in the new system and the integration of data from different sources can be performed on a more detailed level, i.e. the level of commodities (instead of industries). However, the economist who desires to calculate some Leontief multipliers to analyze direct and indirect effects of changes in final demand now faces the problem that a choice has to be made between several options to derive a traditional input-output table from the system of make and use matrices. While statisticians now have a much better framework than before, economists still disagree on the way to utilize the make and use matrices in economic analysis.

In this paper, we address the problem of the derivation of an input-output table from a system of make and use matrices. Our aim is to develop a method of construction of an input-output table that is suitable for analysis. Our leading principle is that the assumptions made in the construction of the input-output data should be consistent with the assumptions made in subsequent input-output analyses. We adopt the view expressed by Rainer and Richter (1992) that, to construct an analytical input-output table, we may need make and use tables that are different from the tables that are part of the national accounts. The requirements of input-output analysis may conflict with the concepts of the SNA. We will need to have an 'analytical' set of make and use matrices, on which the derivation of the input-output table is based. As will be shown, in our view, it is especially the classification of commodities that needs modification.

The new SNA of 1993 (UN et al., 1993) prescribes the construction of a commodity-by-commodity input-output table. However, it does not tell how this table should be constructed. Hence, the derivation of input-output tables from make and use tables is still a topical subject.

The two most important traditional methods of deriving an input-output table are called the 'commodity technology' and 'industry technology' models (for an overview of existing methods, see, for example, Kop Jansen & ten Raa (1990)). The commodity technology assumes that each commodity has its typical input structure, regardless of which industry is the producer. The industry technology assumes that each industry has its typical input structure, regardless of its product mix. Thus, we can say that the commodity technology attaches technology to the commodities, whereas the industry technology attaches technology to the industries.

Both methods--as will be explained in more detail in Section 2--have undesirable properties. The commodity technology results in unacceptable negative input-output coefficients, while the industry technology turns out to be inconsistent with the fundamentals of input-output analysis. To deal with the rigidities of these models, so-called 'mixed technology' models were introduced, with an appropriate mix of commodity technology and industry technology assumptions (Gigantes, 1970; Jhuang, 1990). These models, however, inherit the undesirable properties of both models, thus

making matters only worse.

It seems from these undesirable properties that it is not realistic to assume that technology is attached to either the commodities or the industries. It seems a much more realistic approach to connect technology to the production processes involved in the production of goods and services. However, production processes are not observable by means of economic statistics. Therefore, we have to derive the data about the processes from the data of the make and use tables.

In Section 3, we will describe how we can incorporate the concept of production processes into our input-output framework. Because firms or establishments are grouped into industries, and goods and services are grouped into commodities, we will group production processes into what we will call 'activities'. The activity concept will be an intermediate concept between industries and commodities. We have named our model the 'activity technology' model. The term 'activity' in our model relates more to the way that it is used in activity analysis or linear programming than to the meaning it is given in the SNA. The idea of using the activity unit for the input-output table originally came from de Boer (1987).

Our model will use the mathematical structure of the commodity technology model but, in our model, negative input-output coefficients can only emerge from causes such as heterogeneity in the defined activities and errors in the data. A practical application of the activity technology model requires very detailed (usually rectangular) make and use matrices, as well as the use of exogenous information on input structures of activities. In Section 4, we give a full mathematical description of the activity technology model, including a method that can be used to incorporate all this information to compile a very detailed input-output table.

In Section 5, we will briefly describe an application of the activity technology model to the Dutch make and use matrices of 1987. This empirical work has been described in much more detail by Konijn (1994), who also gave a much more elaborate discussion about commodity technology, industry technology and activity technology. Recently, an input-output table for 1990 was also compiled based on the activity technology model (Central Bureau voor de Statistiek, 1994). Statistics Netherlands has adopted the methodology described in this paper for the regular compilation of commodity-by-commodity input-output tables in the future.

## 2. Existing Methods

Usually, there are two types of input-output tables distinguished. On the one hand, we have commodity-by-commodity--also often called functional--tables, which describe technological relationships between commodities. On the other hand, we have industry-by-industry--or institutional--tables, which describe interindustry relationships. There are also two methods for the derivation of input-output tables, i.e. the methods based on the so-called commodity technology and industry technology assumptions. Traditionally, it is said that, from both assumptions, both types of input-output table can be derived, thus yielding commodity-by-commodity and industry-by-industry variants of commodity technology and industry technology assumptions (see UN, 1968).

The commodity technology and industry technology methods can be seen as transformation methods: secondary products and their (assumed) inputs are transferred from the industry that actually produces the products to the industries where the respective commodities are produced as the primary product. The commodity technology assumes that, in this transformation process, each commodity has its own typical input structure, irrespective of where it is produced. The industry technology assumes that each industry has its own typical input structure, irrespective of its product mix. Both assumptions are assumptions about 'how' commodities are produced--that is why they are called 'technology assumptions'.

It can relatively easily be understood that these technology assumptions are not used to construct industry-by-industry tables. In constructing industry-by-industry tables, assumptions are made on the origins and destinations of products and not on the technology of production. Hence, we find the traditional presentation of methods as referred to above to be incorrect (see Konijn (1994) for more details).

In the Introduction, we stated that we take as our leading principle that the assumptions made in the construction of the input-output data should be consistent with the assumptions made in subsequent input-output analyses. One of the most important assumptions of input-output analysis is the assumption of homogeneous production. This assumption implies effectively that it is assumed that the set of commodities contained in a row or column of the input-output table is as homogeneous as possible. The homogeneity of commodities should be understood especially in

terms of the way that the commodities are produced. To satisfy this assumption, we should build an input-output table that is as homogeneous as possible. An industry-by-industry table does not fulfil this requirement. Therefore, an input-output table that will be used in (traditional) input-output analysis should be of the commodity-by-commodity type.

Which assumption should then be used to construct this commodity-by-commodity input-output table? Again, we have to make sure that the assumptions used in the construction of the table are consistent with the assumptions used in input-output analysis. The consequences of this principle can best be explained with some mathematics.

Let  $U$  be the commodity-by-industry use matrix.  $U$  describes the intermediate inputs of commodities used by the industries. Let  $V'$  be the commodity-by-industry make matrix.  $V'$  describes the production of commodities by industries. We will also have matrices of final demand, primary inputs and imports, but we leave them for the moment for what they are. The desired input-output coefficient matrix, of dimensions commodity-by-commodity, is denoted with  $A$ .

Let us suppose that  $x$  is some vector of commodity outputs. Then,  $Ax$  is the vector of commodity inputs required to produce  $x$ , according to elementary input-output analysis. Each column of the make matrix  $V'$ , say  $V'.j$ , is such a commodity output vector, where the corresponding commodity input vector is the corresponding column of the use matrix  $U$ , i.e.  $U.j$ . For  $A$  to be consistent with the make and use matrices, we therefore need to have

$$U.j = AV'.j \text{ for all } j$$

or

$$U = AV'$$

The input-output matrix  $A$  has to fulfil this equation to achieve full consistency between the data in the make, use and input-output matrices, and the way that the input-output matrix will be used in input-output analysis.

The attentive reader will have noticed that this fundamental equation is equal to the equation of the commodity technology model. Thus, we have shown that the commodity technology model is the only model that is consistent with the fundamentals of input-output. Kop Jansen and ten Raa (1990) came to the same conclusion, based on an axiomatic approach, starting from four desirable properties. We have shown above the same result in an easier and clearer way. The commodity technology model is basically nothing more than the traditional Leontief model with a commodity-by-commodity matrix. (The above derivation was first given by van Rijckeghem (1967).)

This result clearly shows the invalidity of the industry technology or, in fact, any other model for deriving a commodity-by-commodity input-output table. If the industry technology assumption were used, we would obtain a situation where the input-output model gives results that do not correspond to the actual situation if actual demand vectors are given as input. This does not mean that we reject the industry technology on empirical grounds. It may well be that the industry technology assumption is actually more plausible than the commodity technology assumption.

However, the commodity technology also is not free from flaws. To be able to calculate  $A$ , we need to assume that there are as many commodities as industries. Then, we can calculate  $A$  using  $A = U(V')^{-1}$ . However, this way of calculating cannot guarantee a non-negative input-output matrix. Kop Jansen and ten Raa (1990, p. 220) concluded that the non-negativity property is inconsistent with their four axioms. Therefore, let us have a closer look at this problem of negative coefficients.

Applying the commodity technology assumption (that each commodity has its typical input structure, wherever it is produced) for calculating the input-output table can be described as a transformation process of secondary products, as we have seen. If the commodity technology assumption is correct, then the transformation procedure must result in a non-negative matrix  $A$ . If the assumption is false, i.e. some commodity is produced in more than one way, then  $A$  may contain some negative coefficients. As an example from the real world (see Konijn, 1994), let us consider oil products produced by a refinery (as primary products) and by a petrochemical firm (as secondary products). The refinery produces the oil products directly from crude oil, whereas the petrochemical firm uses naphtha substances (probably delivered by the refinery). Using the input structure of the refinery to determine the inputs used by the petrochemical firm will yield a negative input of crude oil in the input column of the petrochemical firm. It is clear that an

unrealistic assumption has been made here: the assumption that the oil products have the same input structure, regardless of where they are produced. It should be noted that the problem would not arise if two commodities for the oil products were distinguished.

Theoretically, the existence of more technologies for one product is the only reason for the emergence of negatives. Practically speaking, however, there is another problem that causes negative input-output coefficients: the problem that products are to be aggregated into groups of products. In the make and use matrices, we have to classify all the products in a limited number of commodities. Then, it may happen that products found in the same commodity can actually be different (or differently produced) products. Applying the commodity technology assumption on this aggregated level implies assuming that these different products are produced in the same way. This faulty assumption may introduce negative coefficients.

Thus, we can conclude that the problem of negatives is actually a classification problem. By choosing another commodity classification the amount of negatives can be reduced. The commodity classification (of the descriptive tables) is based on several, sometimes conflicting criteria. The homogeneity of the input structures is only one of them. Another criterion may be the destination of a product, for example. However, such a criterion is not relevant for the derivation of an analytical input-output table. For our purpose, only the homogeneity of the input structures of products is relevant. To return to the distinction between descriptive and analytical database systems, we can conclude that, in the analytical system (set up especially for the derivation of the input-output table on the basis of the commodity technology), we need another classification scheme of commodities, other than that in the descriptive system.

### **3. Activity Technology Model**

We will now turn to the introduction of the concept of activities. As noted in the Introduction, it seems much more realistic to let technology depend on the production processes rather than on commodities or industries: a commodity may be produced by different industries in different ways, while an industry may have several distinguishable technologies.

A production process can be defined as a process that transforms goods, services and primary inputs into other goods and services. For our purpose, i.e. for input-output analysis, we can characterize a production process by an input structure (a column of input coefficients) and a set of goods or services produced according to that input structure. In our view, the production process is a micro-economic concept. For practical use, we aggregate production processes in 'activities'. Thus, an activity is a set of production processes with input structures as homogeneous as possible.

The new SNA of 1993 defines the concept of 'units of homogeneous production' as 'a producer unit in which only a single [non-ancillary] productive activity is carried out' (UN et al., 1993, p. 118). This concept comes close to our concept of production processes, except that we have defined our concept in terms of input-output analysis. The SNA recommends that the unit of homogeneous production is used as the basis for input-output analysis but does not elaborate on how this should be done. Aggregates of units of homogeneous production are also sometimes called 'branches'. Hence, activities and branches are closely related concepts.

We will build our input-output model around the concept of activities. Since an input-output table should represent the technology of an economy, the input structures of activities should be contained in the input-output table. The input structure of an activity can be formulated in terms of commodities (then it describes which products are used by that activity) or in terms of activities (then it describes by which activities the products used by the activity are produced). Clearly, a square input-output matrix, on which the calculation of Leontief multipliers can be based, now should be of dimensions activity-by-activity. In our model, we assume the following.

- Industries may employ several activities. To be precise, an industry is described as a linear combination of activities. Therefore, an industry may have secondary activities (in which secondary products are produced) besides its primary activity.
- An activity may be employed by several industries.
- An activity may produce several commodities, thus incorporating by-production and joint production.
- A commodity can only be produced by one activity. This is because, if a commodity is produced in two different ways, we should distinguish two commodities which are assigned to two different activities.

The final point implies that we have to adapt the given commodity classification in cases where a commodity is produced in more than one way. The task of the compiler of an input-output table is to decide which commodities are produced by which activities and which industries employ which activities. The number of activities is not a priori determined, as we will see. By-products and joint products can easily be incorporated into our model, by assigning a by-product to the same activity as the main product to which it belongs, for example. Thus, the by-product is given the same input structure as the main product. It should be noted, however, that no assumptions are made on the connection between the by-product and the main product. For example, there is no assumption of fixed output proportions per activity, which is an assumption frequently made in joint production models (such as the von Neumann model). In fact, we cannot say anything about the output proportions of the commodities within an activity. If an activity output vector is calculated in an analysis, it is impossible, without any further assumptions, to relate that back to commodity outputs.

The input structures of the activities are determined in a transformation process of inputs similar to that of the commodity technology model. The input structure of a secondary activity is assumed to be equal to the input structure of that activity when employed as a primary activity. According to this assumption, the secondary activities and the assumed inputs are transferred to the industries that have that activity as a primary activity. Thus, the input structures of the industries are the basis of the calculation of the input structures of the activities.

The activity classification is derived from the commodity classification. To arrive at the activity classification, two steps are involved. First, we need to disaggregate the current (say descriptive) commodity classification. Ideally, we should split up each commodity into as many commodities as there are ways to produce it. Secondly, we assign the commodities to activities, with regard to their input structure. Thus, for example, chairs and tables would first be divided into wooden chairs, metal chairs, wooden tables and metal tables. Then, if it were impossible to distinguish between the input structures of wooden tables and wooden chairs, we would assign them both into the activity of wooden furniture. Similarly, the metal chairs and tables would be assigned to the activity metal furniture. (This is not a real-world example.)

To return to the example mentioned in the previous section, we should now distinguish two commodity groups, i.e. oil products produced by the refinery and oil products produced by the petrochemical firm. Although the products may really be the same thing, the distinction is necessary to achieve a correct description of the inputs used in the production of the oil products. The commodity 'oil products produced by the refinery' becomes the primary activity of the refineries (such that the input structure of the refineries, after the transformation of the inputs corresponding to secondary activities, becomes the input structure of that primary activity). The commodity 'oil products produced by the petrochemical firm', if considered to be a by-product of the petrochemical firm, may be classified under the main activity of the firm, such that it obtains the same input structure as that main activity. If it is more likely to be a subsidiary product which bears no relation to the main activity of the firm, then it becomes necessary to introduce a separate activity of the oil products produced by the petrochemical firm, for which an input structure should be exogeneously estimated. We will come back to this later.

The first step of the above procedure, i.e. the disaggregation of commodities into as many commodities as there are ways to produce it, is rather difficult to formalize in a set of rules. For example, we do not know in advance how many ways there are to produce a certain commodity. If a commodity is produced in more than one way, then this will become clear from negative elements arising in the transformation process. Thus, we can use the negatives to determine how many commodities we need to distinguish to describe correctly the inputs used to produce that commodity.

Let us assume now that this disaggregation of commodities has already taken place, so that, from now on, when we speak of commodities we refer to the new commodity classification. We will have more commodities than activities. We will also have more activities than industries, since a primary activity can be defined for every industry, while there will be some activities without a primary producer (such as services on a contract basis; see Konijn, 1994).

To derive a symmetric activity-by-activity table, the (intermediate and final) demand for commodities as described in the use matrix should be reclassified into activities. This is a more difficult problem, because information is needed or an assumption should be made on which activities produced the commodities demanded. This is a serious drawback but is a necessary consequence of our desire to describe the technological relationships of an economy.

It is important that, in our model, each activity has, by definition, its own input structure, where we had before the rather unrealistic assumption that each commodity has its own input structure. In the commodity technology model, negative coefficients could appear, which becomes now, in theory, impossible. This is because, if a negative coefficient were to appear, it would be a sign that a particular product was produced in more than one way, implying that we should have distinguished two commodities.

In practice, however, negatives can appear, since heterogeneity may be present: two different products (with different input structures) may be classified in one commodity group. Furthermore, if there appears a negative input-output coefficient and there is no reason to assume that the respective commodities have different input structures, then the negative is a clear indication of an error in the make or use matrix.

It seems that Kop Jansen and ten Raa (1990), in stating that non-negativity is inconsistent with their four axioms (and, thus, with the commodity technology model), have neglected the classification problem. If we assume that the commodities are strictly classified according to their input structures (which is impossible in practice, of course), a non-negative matrix should result.

As said, in a practical application, it will not be possible to achieve a perfect activity classification, implying that (small) negatives will still remain. All large negatives should have been deleted by refining the activity classification. The remaining small negatives are then the results of remaining heterogeneities or errors in the make and use matrices. Since it is difficult at this point in the compilation process to distinguish between the two sources of negatives, it might be considered to adjust the values of use and make matrices. Since the negatives are only small, the required adjustments may be expected to be small as well. The compiler of the make and use tables should then decide whether or not the adjustments are within the confidence margins of the respective values (for a discussion on locating errors in make and use matrices, see Steenge (1990)).

Above, we stated that, in theory, no negatives can arise in our activity technology model but that, in practice, small negatives will remain. However, it can easily be demonstrated that it is always possible to make a non-negative input-output matrix, simply by making activities equal to industries. Then, all the production of an industry is assumed to be produced by one (main) activity, so no secondary production exists! Since no secondary production exists, no transformation of inputs is necessary, so no negatives can occur. In fact, the input-output table will equal the use matrix.

Of course, if negatives do not arise, we cannot simply conclude that we have made a correct technological assumption. We then have no objective grounds to reject the assumption made. In principle, positive elements in the input-output table can be just as wrong as are negative elements. However, we have no a priori criterion to detect them.

We have also noted above that, to derive an activity-by-activity input-output matrix, we need to reclassify the demand for commodities, as recorded in the use matrix, into activities. If we use the simple market shares assumption for this reclassification and if activities are defined equivalent to industries, it is not hard to show that the resulting input-output matrix is equal to the matrix that would be obtained when the so-called industry-by-industry variant of the industry technology model would have been applied. (We argued above that this method, in fact, has little to do with the industry technology assumption. For convenience's sake, we adopt the traditional terminology here.) This is a remarkable result in view of the way that we have set up our framework. This result shows that, in cases of insufficient resources to apply the activity technology in full, the industry-by-industry variant of the industry technology model is the second-best solution. Also, if the industry classification is already very detailed and homogeneous, then the assumption that activities and industries are equivalent is not too unrealistic.

We have seen that, to apply the commodity technology, we needed an equality between the numbers of industries and commodities. Now, we would need an equality between the numbers of industries and activities, since the input structures of the industries are used to determine the input structures of the activities. If the number of activities needed to achieve a classification that satisfies the requirement of homogeneity of input structures is larger than the number of industries available, then it seems hardly possible to apply the activity technology. Consequently, we will end up again with a number of negative coefficients, since we have to aggregate activities with different input structures. It is necessary to apply the transformation procedure at a level that is as detailed as possible. The next section, which will also give a formal presentation of the activity technology model, discusses this problem and gives a solution: we have to add exogenous information on the input structures of certain activities. The availability of such information becomes a determinant for

the classification of activities and, thus, for the amount of negatives.

#### 4. Mathematical Formulation of the Activity Technology Model

We stated in the Introduction that using input-output techniques may require a different data set-up from that employed in the (descriptive) make and use matrices. We have encountered the commodity classification as the most important source of changes for the analytical data set. However, there may be other adaptations necessary. For instance, we need another valuation in the use matrix: basic prices instead of producer's prices. This means that we have to deduct from every element of the use matrix the indirect taxes and subsidies. Konijn (1994) describes a number of preliminary adjustments. From now on, we suppose that all adjustments have already taken place; we now consider only the classification changes.

These can be performed in two steps, as mentioned in Section 3. First, we modify the commodity classification. Secondly, we assign the commodities to activities. In the first step, existing commodities are ideally split up into as many commodities as there are ways to produce that commodity. The new commodities may then become industry dependent, i.e. the industry where it is produced is the criterion to distinguish commodities. When this is completed, in the second step, the (new) commodities can be assigned uniquely to an activity, according to the input structure.

As before, it is assumed that the first step of this classification procedure has already taken place. Thus, the commodity classification used below is the 'new' classification. All we have to do is assign the commodities to activities.

We have already encountered some notation. We repeat them here for convenience and add other important variables:

U the commodity-by-industry table of intermediate demand;  
 U the activity-by-industry table of intermediate demand;  
 V' the commodity-by-industry table of production;  
 V' the activity-by-industry table of production;  
 y' industry vector of primary inputs;  
 e commodity vector of final demand;  
 e activity vector of final demand;  
 q<sub>d</sub> commodity vector of total domestic production;  
 q<sub>d</sub> activity vector of total domestic production;  
 q<sub>m</sub> commodity vector of imports;  
 q<sub>m</sub> activity vector of imports;  
 q commodity vector of total supply;  
 q activity vector of total supply;  
 g industry vector of total domestic production;  
 A commodity-by-activity input-output coefficient matrix;  
 A activity-by-activity input-output coefficient matrix;  
 Z commodity-by-activity input-output flow matrix;  
 Z activity-by-activity input-output flow matrix;  
 w' activity vector of primary input coefficients;  
 z' activity vector of primary inputs.

We have the following relationships for the unaggregated (defined in commodities) system:

$$U_i + e = V'_i + q_m = q_d + q_m = q$$

$$i'U + y' = i'V' = g$$

where  $i$  is the vector of ones. From this system, the aggregated data set, consequently denoted with a  $\sim$  symbol, is derived. We define the aggregation matrix  $P$  as the activity-by-commodity matrix with

$$P_{ij} = 1 \text{ if commodity } j \text{ is produced by activity } i$$

$$P_{ij} = 0 \text{ otherwise}$$

We will have  $\sum_i P_{ik} = 1$ , for all  $k$ , since a commodity can be produced in only one activity. An activity, however, can produce more than one commodity. Using  $P$ , we can transform the above equations as follows:

$$U_i + e = V'_i + q_m = q_d + q_m = q$$

$$i'U + y' = i'V' = g$$

with  $U = PU$ ,  $e = Pe$ , and similarly for  $V'$ ,  $q_m$ ,  $q_d$  and  $q$ .

From this system, we will derive the input-output matrices. We can extend the fundamental formula derived in Section 2 to the activity system. The following relationships should hold:

$$U = AV'$$

$$U = AV'$$

$$y' = w'V'$$

$$Z = A(q_d)$$

$$Z = A(q_d)$$

$$Z' = W'(q_d)$$

where  $x$  denotes the diagonal matrix with the elements of  $x$  on the diagonal. We will concentrate from now on on deriving  $A$  and  $w'$ . The other variables follow from these two, using the above formulas and from  $A = PA$ .

The analytical make matrix  $V'$  has dimensions of activity-by-industry, i.e. it is rectangular. Let us suppose that we have  $a$  activities and  $i$  industries. We have seen that  $a > i$ . Thus, we can allocate  $i$  activities as primary activity to an industry but, for  $(a-i)$  activities, we cannot find a primary producer. However, if we have the required information, we can fill this gap with exogenous information on the input structures of the activities for which no primary producer exists.

We can partition the make matrix as

$$V_1 \quad V' = \begin{pmatrix} & \\ & \end{pmatrix} \quad V_2$$

with  $V_1$  a heavily diagonal square  $(i \times i)$  matrix, and  $V_2$  an  $(a-i) \times i$  matrix. The rows of  $V_1$  correspond to those activities that have a primary producer, where activities and their primary producers have corresponding row and column numbers, and the rows of  $V_2$  correspond to those that could not be assigned a primary producer. (In practice, a permutation of the rows and columns may be necessary to achieve this form, but this does not affect the following.) The activity vector of domestic production ( $q_d$ ) can be partitioned similarly

$$q_{d1} \\ q_d = \begin{pmatrix} & \\ & \end{pmatrix} \quad q_{d2}$$

Now, we introduce new columns that represent what may be called 'pseudo-sectors' producing solely the total domestic production of one of the activities of  $V_2$ , i.e.

$$V_1 \quad \theta \\ V^* = \begin{pmatrix} & \\ & \end{pmatrix} \quad V_2 \quad q_{d2}$$

Similarly, we augment the original use matrix  $U$  and the vector of primary inputs ( $y'$ ) with estimations of the input structures of the activities of  $V_2$ , denoted by  $S$  and  $s'$  respectively, multiplied by the total domestic production of those activities, i.e.

$$U^* = (U \quad Sq_2) \text{ and}$$

$$y^* = (y' \quad s'q_{d2})$$

On these augmented matrices  $U^*$  and  $V^*$ , we simply apply the usual formula to obtain  $A$  as

[multiple line equation(s) cannot be represented in ASCII text]

The estimated input structures ( $S$ ) return in the derived input-output matrix  $A$ . It is easy to show that the matrix  $A$  derived in this way fulfils  $U = AV'$ :

[multiple line equation(s) cannot be represented in ASCII text]

Similarly, we obtain

[multiple line equation(s) cannot be represented in ASCII text]

for which, of course, it also will follow that  $y' = w'V'$  (we leave that to the reader).

Summarizing, we augment the use and make matrices with columns that represent pseudo-sectors producing those activities that do not have a primary producer. The information contained in the extra columns of the use matrix returns unchanged in the input-output table, as a result of the nature of the transformation process applied. Thus, we fix a priori the inputs that are involved with those activities, throughout the economy. This information can come from exogenous sources. Sometimes, we will have to make assumptions on the input structures of activities. However, it can be argued that even adding weak information in some cases is better than not adding information at all (which would immediately imply aggregation). For example, let us consider the oil products example again: if the oil products produced by the petrochemical firm bear no relationship to the main activity of the firm, or to the oil products produced by the refinery, then it is necessary to add an input structure for that particular secondary product. We can make a hypothesis about its input structure and check whether or not it is plausible, by performing the transformation procedure. If the transformation of inputs yields no negative coefficients, then we cannot reject our hypothesis of that input structure.

The availability of data on input structures of activities, or the possibility to make hypotheses about them, determines the level of disaggregation (i.e. the number of activities) that we can obtain. Within the limits of the data availability, we are free to choose the number of activities in the input-output table--it is not necessarily equal to the number of industries.

In practice, the determination of the number of activities will be an iterative process. Starting from some initial classification scheme, we calculate the input-output matrix and analyze the negative elements. The negatives give a clear indication of the heterogeneities remaining in the commodity classification (as long as we assume that the data are correct). That gives us a clue as to where to introduce new activities.

We have seen that we calculate  $A$  through  $A = U*(V^*)^{-1}$ . This matrix will generally not be non-negative. Even if we have refined the activity classification such that it becomes very homogeneous, (small) negatives will still arise. They result either from remaining heterogeneities or from errors in the data. We can calculate the changes in the use matrix that would be necessary to obtain a non-negative input-output table, and evaluate the proposed changes in terms of the reliability of the use matrix. For this, we should first make the derived input-output table non-negative. We can do this by putting all the negatives to zero, and adjusting the other elements of the table to keep the row and column totals unaffected. How this can be done in practice is explained by Konijn (1994).

Let us suppose that we can make a matrix  $A_n > / 0$ , which is close to  $A$ . We could then write  $U = A_n V' + \text{epsilon}$ , with epsilon a matrix containing all the changes we should make in  $U$  to derive  $A_n$ . We can interpret epsilon as the 'error term' in the activity technology model. It takes care of all the remaining heterogeneities and errors in the use matrix. We can easily calculate the use matrix that would yield  $A_n$  in the transformation process, by calculating  $U_n = A_n V$ . Then,  $\text{epsilon} = U - U_n$ . Now, if a particular element of epsilon is too large, i.e. it proposes a change in  $U$  that cannot be accepted, then we have to conclude that the heterogeneity in the activity classification still plays too important a role, and that we should look for ways to change the activity classification or the allocation of commodities to activities.

## **5. Application to the Dutch Economy of 1987**

Within the scope of this paper, it is impossible to give a detailed account of the application of the activity technology model to the Dutch economy of 1987. The interested reader can find it in Konijn (1994). Here, we will suffice with some general remarks.

The make and use matrices compiled by Statistics Netherlands, as part of the national accounts, comprise at their most detailed (and confidential) level about 800 commodities and 250 industries. From this system, an activity-by-activity input-output table is compiled that comprises 313 activities, which is the most detailed table ever derived for The Netherlands. For 65 activities, additional input information was collected or constructed.

Of these 65 activities, 13 are activities within the agricultural and horticultural industry. Traditionally, agriculture and horticulture constitute one row and column in the input-output table. The breakdown of this industry is a major improvement for input-output analysis, since the homogeneity of the described flows is drastically increased.

Three activities are distinguished for the production of oil products, corresponding to the three main producers, each using different inputs. Also, three activities are distinguished for the production of plastics. There are a few industries, such as the sheltered workshops, that do not have one primary product. For such industries, new commodities are created that contain all or part of the production of these industries. These commodities then become the primary products of these industries.

Trade margins and services are divided into two commodities: trade margins and services produced by trade firms, and trade margins and services produced by non-trade firms (usually as a secondary activity). The same goes, more or less, for transport margins. Finally, services on a contract basis, i.e. services put out by one firm to another, are treated as a separate activity, for which an input structure was estimated with a high labour input.

The application of the activity technology model is an iterative process, as noted in the previous section. In our application, we iterated 11 times before we accepted the remaining negatives as being the 'error term'. In the first version of the input-output table, 300 activities were defined. The total negative value in this first version was 12.1 billion guilders, which is 1.51% of the total production value of the economy. In the last version of the input-output table, the number of activities had grown to 313, so that an input structure had to be collected or constructed for another 13 activities. The total negative value in the table was reduced to 5.9 billion guilders, or 0.73% of the total production value. Thus, the enlargement of the table has indeed significantly reduced the problem of the negatives.

## **6. Final Remarks**

We have described a method for derivating an analytical input-output table from a system of make and use matrices. It has the mathematical structure of the commodity technology model but, instead of applying it to commodities, we apply it to our new concept of activities. Therefore, we have called it the activity technology model. In theory, this model cannot yield any negative input-output coefficients but, in practice, small negatives will remain, as a result of heterogeneity of the activity classification and errors in the data.

We claim that an input-output table based on our method is more suitable for input-output analysis than is an input-output table based on any other method. However, we must admit that this claim cannot be supported at this moment with numerical proof. We base our claim on theoretical arguments: the table is much more homogeneous and detailed, so it gives a better imputation of inputs to outputs.

The method has two main drawbacks. First, a given vector of final demand for commodities, for which the total requirements are to be calculated, needs to be transformed into a vector of final demand for activities. For this transformation, assumptions normally will be required. This drawback is a necessary consequence of our wish to describe the technology of an economy: the total requirements can only be calculated if the consequences for the production processes are known.

The second drawback of our method is its laboriousness. It has to be applied to very detailed make and use tables; we have to add extra information on input structures of activities; we need to evaluate closely the cause of a particular negative to decide whether or not it is caused by an erroneous assumption or by a data error; we have to remove all the remaining small negatives; etc. The compilation of an input-output table cannot be achieved from published aggregated make and use tables, since we need reference to the underlying data. However, the adoption of the activity technology model by Statistics Netherlands proves that this is not prohibitive.

Also, one might argue that our method seems rather arbitrary: it is not clear which criteria are used to distinguish activities, assign commodities to industries, etc. This line of argument, however, seems to suggest that there should be a new mathematical formula, by which the input-output table can easily be calculated. Instead, the crux of our method is the fact that the process of constructing input-output tables cannot be described in a unique formula. It is a process of looking carefully at the data, at the results of the transformations applied, at the plausibility of the assumptions used and at the classifications, using additional data as much as practically possible but also using a good amount of plain common sense. The compilation of input-output tables is a

mix of statistics and model building: the model building part is easy to describe, while the statistical part of the process is much more difficult.

In our opinion, one of the main points of this study is that the activity technology model provides us with a consistent method for treating the problem of the negatives. All the solutions to this problem which have been presented thus far have lacked a theoretical foundation. We hope that we have provided the theory necessary to understand the problem. By explicitly looking at the production processes rather than just the commodities, and by regarding the commodity classification of the make and use matrices not as a datum but as an instrument, we may have found a novel way to look at problems of multiple production and multiple technologies.

## **References**

de Boer, S. (1987) De transformatie van make en use tabellen tot homogene input-output tabellen, Working paper 499-87-PS.E8 (Voorburg, Statistics Netherlands) (in Dutch).

Centraal Bureau voor de Statistiek (1994) De produktiestructuur van de Nederlandse Volkshuishouding, deel XXI: input-outputtabellen en aanbod-en gebruiktabellen 1991 (Voorburg, Statistics Netherlands) (in Dutch).

Gigantes, T. (1970) The representation of technology in input-output systems, in: A. P. Carter & A. Brody (eds.) Contributions to Input-Output Analysis (Amsterdam, North-Holland), pp. 270-290.

Juhuang, H. (1990) A method of deriving input-output coefficients under the assumption of mixed technology, Economic Systems Research, 2, pp. 43-46.

Konijn, P. J. A. (1994) The make and use of commodities by industries; on the compilation of input-output data from the national accounts (Enschede, Faculty of Public Administration and Public Policy, University of Twente).

Kop Jansen, P. & ten Raa, T. (1990) The choice of model in the construction of input-output matrices, International Economic Review, 31, pp. 213 - 227.

Rainer, N. & Richter, J. (1992) Some aspects of the analytical use of descriptive make and absorption tables, Economic Systems Research, 4, pp. 159-172.

van Rijckeghem, W. (1967) An exact method for determining the technology matrix in a situation with secondary products, Review of Economics and Statistics, 49, pp. 607-608.

Steenge, A. E. (1990) The commodity technology revisited: theoretical basis and an application to error location in the make-use framework, Economic Modelling, 7, pp. 376-387.

United Nations (1968) A System of National Accounts, Series F, no. 2, Rev. 3, (New York, United Nations).

United Nations, CEC, IMF, OECD & Worldbank (1993) System of National Accounts.

~ ~ ~ ~ ~ ~ ~ ~

By P. J. A. KONIJN & A. E. STEENGE

P. J. A. Konijn, Statistics Netherlands, Department of National Accounts, PO Box 959, 2270 AZ Voorburg, The Netherlands. A. E. Steenge, University of Twente, Faculty of Public Administration and Public Policy. This paper is a revision of a paper presented at the 10th International Conference on Input-Output Techniques, Seville, Spain, 1993. The views expressed in this paper reflect those of the authors and do not necessarily reflect the views of their employers. The authors wish to thank the Department of National Accounts of Statistics Netherlands for allowing them to carry out this research at the Statistics Netherlands office with their data. We would also like to thank two anonymous referees for their useful comments.