

ADAPTATION OF DETAILED INPUT-OUTPUT INFORMATION: RESTRUCTURING AND AGGREGATION

BY J. ASGER OLSEN

Danmarks Statistik

This article deals with problems of construction of "square" input-output tables from detailed commodity and industry data and operationalization of the tables for use in econometric modeling. The adaptation procedure suggested is quite new and involves perfect and imperfect aggregation, and suppression of insignificant cells of the tables.

Using slight modifications of well-known input-output methods (to make definitions conform to general concepts of network flow theory) it is shown that the construction and aggregation of tables, as well as the suppression of minor cells, can be viewed as still higher levels of the very same process: the search for a manageable model with roughly the same abstract properties as the original detailed, but overwhelmingly large, model.

Simultaneously a consistent input-output terminology is suggested with fewer symbols and more rules than usual.

The adaptation procedure has been applied successfully to the 1982 version of ADAM, the macroeconomic model operated by Danmarks Statistik (the Danish Central Bureau of Statistics).

INTRODUCTION

This article summarizes the theoretical and practical experience of two years effort to incorporate a complete input-output model into the econometric model ADAM. The "Annual Danish Aggregate Model" is operated by Danmarks Statistik, and it is extensively used by Danish government agencies for forecasting and planning purposes.¹

The starting point of our analysis is a detailed breakdown of the national production account, showing sources and uses of a large number of commodities and primary factors of production. Danmarks Statistik provides a yearly time series of such accounts from 1966 onwards.² However, even major computers of today cannot contain such an amount of data simultaneously, and thus we need methods of extracting summary information from the accounts, with a minimum loss of information.

The purpose of this article is to show how this method of extraction, in a simple and well-defined way, covers various well-known problems related to input-output analysis. In turn, these problems include theoretical questions of constructing "square" input-output tables from commodity accounts and perfect aggregation of industries, as well as practical questions of imperfect aggregation and suppression of small cells of the tables. The suppression of small cells is an

¹This institutional set-up is perhaps a little unusual. One of the advantages of it has proved to be that the ADAM group developed close relations to the national accountants of Danmarks Statistik. The combined view on construction, restructuring and aggregation of input-output tables presented here may be regarded as a product of this set-up.

²See Thage (1982) for a description in English of the Danish system of national accounts. The structure of the Danish system is quite similar to that of the Canadian system described in this journal by Lal (1982).

unorthodox simplification of input–output tables, which in the ADAM case turned out to be very competitive to further aggregation.

THE DATA BASE

The data base on which the input–output tables are to be constructed is assumed to be commodity accounts (“balances”) and information on the uses of primary factors of production. Commodity accounts are usually worked out for a large number of commodities. The debit side of a commodity account shows the total supply of the given “commodity” and its distribution according to the sources of supply (i.e. industries and imports, possibly by supplying country). The credit side of a commodity account shows the corresponding total use of the commodity and its distribution by user categories (i.e. industries and components of final demand). Of course, total supply must equal total use.

The commodity accounts and the primary factor input information are usually collected in a matrix diagram as in Table 1.

TABLE 1.
THE ACCOUNTING FRAMEWORK

From \ To	Commodities	Industries	Final Demand	Sum
Commodities		U	F	$q + m$
Industries	D			g
Imports	m'			$i'm$
Primary Factors		Y	Y_f	y
Sum	$(q + m)'$	g'	f'	

The dimensions of the Danish system are: approximately 2,500 commodities, 117 industries, 74 components of final demand, 1 import account and 6 types of primary factors.

The reader should be familiar with the diagram and the symbols used. The diagram bears a close relationship to the accounting framework proposed by the United Nations, and it is essentially identical to the table given by, e.g., Lal (1982).³ Table 1 is taken as the starting point of our construction of input–output tables.

THE FUNDAMENTAL ASSUMPTIONS

The matrix diagram of Table 1 offers a complete and detailed record of information concerning production and consumption of commodities. However, the amount of information contained in it is so large that it would be of little practical value without some kind of reduction by abstraction from details. Thus, we introduce simplifying assumptions for the purpose of reducing the vast body of data.

³Cf. United Nations (1968). See, e.g., Mr Lal's paper for a brief, but full, explanation of the diagram.

Before doing so, we shall have to modify the matrix diagram in trivial ways. Firstly, we split the import account into a large number of accounts, namely one import account for each commodity. In addition we introduce one new account for each commodity, transmitting domestic production from industries to total commodity supply. This leaves three accounts for each commodity: Imports, domestic production and total supply. Secondly, we apply the practice, usual in the treatment of transactions tables, that the accounts are numbered in the same way row-by-row and column-by-column. The result is the transactions matrix shown in Table 2. (The symbol $\hat{}$ on a $(n \times 1)$ -vector represents diagonalization of the vector in a $(n \times n)$ diagonal matrix. The symbol $'$ denotes transposition of a vector or matrix)

TABLE 2.
THE FUNDAMENTAL SYSTEM MATRIX

From \ To	Primary Factors (1)	Import Commodities (2)	Domestic Commodities (3)	Commodities Total (4)	Industries (5)	Final Demand (6)	Sum
(1)					Y	Y_f	y
(2)				\hat{m}			m
(3)				\hat{q}			q
(4)					U	F	$q+m$
(5)			D				g
(6)							0
Sum	0	0	q'	$(q+m)'$	g'	f'	

At first glance these changes do not seem to be advantageous, since the new diagram is even larger than the original one. However, the augmentation is trivial, and soon we shall see how the matrix easily can be reduced. The matrix of Table 2 is called the compounded system matrix or briefly the system.⁴ It is denoted by the letter T . The system matrix is seen to be block-triangular, since the columns corresponding to external inputs y and m are identically zero, as are the rows corresponding to external outputs f . Now, define the vector of external inputs as

$$(1) \quad t^i = (y, m, 0, 0, 0, 0)$$

and the vector of external outputs as

$$(2) \quad t^0 = (0, 0, 0, 0, 0, f).$$

Due to the conservation law of each account, the column sums of T plus external inputs must equal the row sums of T plus external outputs:

$$(3) \quad T' t^i + t^i = T t + t^0 = t.$$

The vector t is called the throughput vector, since the j th element of t is the throughput or balance of account no. j . The vector t is

$$(4) \quad t = (y, m, q, m+q, g, f).$$

⁴The system matrix is identical to the flow matrix concept of general network flow theory. The matrix defines an open flow network with sources (y, m) and sinks f .

A common practice in an input-output context is to assume in general that all flows into any account of the system are proportional. The assumption requires specifically that the input mix of any account does not depend on the corresponding output mix. Formally, we assume that the coefficient matrix

$$(5) \quad \bar{T} = T\hat{t}^{-1}$$

is independent of t . We shall consistently use the bar ($\bar{\quad}$) to denote matrices and submatrices of coefficients (shares of system throughputs). Accordingly, $\bar{D} = D\hat{q}^{-1}$, $\bar{m} = \hat{m}(\widehat{m+q})^{-1}$, $\bar{U} = U\hat{g}^{-1}$ etc.⁵

The assumption of a constant input mix of any account can be given various interpretations. Concerning commodity accounts we assume that commodities are supplied with fixed market shares for imports (\bar{m}) as well as domestic industries (\bar{D}). Concerning industries, we assume that the input structure by commodities (\bar{U}) and primary factors (\bar{Y}) is constant, which amounts to the well-known "industry technology" assumption. Finally, we assume that the components of final demand are fixed-proportion commodity bundles (the proportions given by \bar{F} and \bar{Y}_f).

To solve the system equations, substitute the assumption (5) into the conservation conditions (3). This yields

$$(6) \quad \begin{aligned} t &= \bar{T}\hat{t} + t^0 \\ &= \bar{T}t + t^0, \end{aligned}$$

which may be solved to

$$(7) \quad t = (I - \bar{T})^{-1}t^0.$$

The matrix

$$(8) \quad Z = (I - \bar{T})^{-1}$$

is called the fundamental inverse, since it contains all relationships between all accounts of the fundamental system. Of course, in practice we would try to reduce the system before inverting the very large matrix $(I - \bar{T})$. However, the fundamental inverse has conceptual importance, and it is thus shown in Table 3 (proof omitted).^{6,7}

As long as the basic assumptions hold, the fundamental inverse defines the external inputs t^i as a unique linear mapping of the external outputs t^0 . This linear relationship is briefly called the transfer properties of the system. But since the only system inputs are factors y and imports m , and the only system outputs

⁵This is a minor departure from usual terminology, since the matrix \bar{U} is denoted B in the main literature. However, the author believes that the proposed symbols are better mnemonics: T (Total system), Y (Yields), U (Uses), F (Final uses), D (Domestic market shares), t (throughputs), m (imports), q (quantity produced) and g (gross production). The bar notation for coefficients leaves these symbols to be remembered only.

⁶It can be found by utilizing the fact that \bar{T} is block-triangular, in addition to the relationship $(I - \bar{T})Z = I$. The process of inversion is laborious, but not difficult.

⁷Interpreting the table, the following formulae may be helpful:

$$(I - \bar{U}\bar{D}\bar{q})^{-1} = \bar{U}(I - \bar{D}\bar{q}\bar{U})^{-1}\bar{D}\bar{q} + I = \bar{U}\bar{D}(I - \bar{q}\bar{U}\bar{D})^{-1} + I.$$

The formulae are easily shown, using the well-known expansion $(I - X)^{-1} = I + X + X^2 + X^3 + \dots$

TABLE 3.
THE FUNDAMENTAL INVERSE

From \ To	Primary Factors (1)	Import Commodities (2)	Domestic Commodities (3)
(1)	I	0	$\bar{Y}\bar{D}(I - \bar{q}\bar{U}\bar{D})^{-1}$
(2)	0	I	$\bar{m}\bar{U}\bar{D}(I - \bar{q}\bar{U}\bar{D})^{-1}$
(3)	0	0	$(I - \bar{q}\bar{U}\bar{D})^{-1}$
(4)	0	0	$\bar{U}\bar{D}(I - \bar{q}\bar{U}\bar{D})^{-1}$
(5)	0	0	$\bar{D}(I - \bar{q}\bar{U}\bar{D})^{-1}$
(6)	0	0	0

From \ To	Commodities Total (4)	Industries (5)	Final Demand (6)
(1)	$\bar{Y}\bar{D}\bar{q}(I - \bar{U}\bar{D}\bar{q})^{-1}$	$\bar{Y}(I - \bar{D}\bar{q}\bar{U})^{-1}$	$\bar{Y}\bar{D}\bar{q}(I - \bar{U}\bar{D}\bar{q})^{-1}\bar{F} + \bar{Y}_f$
(2)	$\bar{m}(I - \bar{U}\bar{D}\bar{q})^{-1}$	$\bar{m}\bar{U}(I - \bar{D}\bar{q}\bar{U})^{-1}$	$\bar{m}(I - \bar{U}\bar{D}\bar{q})^{-1}\bar{F}$
(3)	$\bar{q}(I - \bar{U}\bar{D}\bar{q})^{-1}$	$\bar{q}\bar{U}(I - \bar{D}\bar{q}\bar{U})^{-1}$	$\bar{q}(I - \bar{U}\bar{D}\bar{q})^{-1}\bar{F}$
(4)	$(I - \bar{U}\bar{D}\bar{q})^{-1}$	$\bar{U}(I - \bar{D}\bar{q}\bar{U})^{-1}$	$(I - \bar{U}\bar{D}\bar{q})^{-1}\bar{F}$
(5)	$\bar{D}\bar{q}(I - \bar{U}\bar{D}\bar{q})^{-1}$	$(I - \bar{D}\bar{q}\bar{U})^{-1}$	$\bar{D}\bar{q}(I - \bar{U}\bar{D}\bar{q})^{-1}\bar{F}$
(6)	0	0	I

are the final demands f , the transfer properties are determined by the two upper right submatrices $\bar{Y}\bar{D}\bar{q}(I - \bar{U}\bar{D}\bar{q})^{-1}\bar{F} + \bar{Y}_f$ and $\bar{m}(I - \bar{U}\bar{D}\bar{q})^{-1}\bar{F}$ alone. Together, these submatrices constitute the transfer matrix of the system. We denote the accounts of system inputs and outputs by the common term open accounts, since they are the "ports" through which the system interacts with the environment. The remaining accounts with throughputs q , $q + m$ and g are denoted intermediate or closed accounts, since they represent internal system transactions only.

This is the key to our reduction of the system. Any system matrix having the same transfer properties as the fundamental system is said to be equivalent to this system. There will be several kinds of equivalent systems, each having advantages and drawbacks, and this is indeed reflected in the literature.⁸ The author believes that the method outlined here is the only way to treat these equivalent systems in an organized manner.

EQUIVALENT SYSTEM FORMULATIONS

The simplest equivalent system that we can always achieve is the reduced system. That is, the system with no intermediate accounts. It is given by the two upper right transfer matrices (re-scaled to levels), and it is shown in Table 4. However, this system is not widely used. This must be due to two factors. Firstly, it is very vulnerable to modifications of the basic assumptions. For example, if the import share of steel would change, it would not be clear which cells to modify, since practically all cells of the matrices have some indirect content of

⁸Cf. e.g. United Nations (1968) or (1973).

TABLE 4
THE REDUCED SYSTEM⁹

From \ To	Primary Factors	Import Commodities	Final Demand	Sum
Primary Factors			$\bar{Y}\bar{D}\bar{q}(I - \bar{U}\bar{D}\bar{q})^{-1}F + Y_f$	y
Import Commodities			$\bar{m}(I - \bar{U}\bar{D}\bar{q})^{-1}F$	m
Final Demand				0
Sum	0	0	f	

steel. Secondly, some of the throughputs of intermediate accounts may have an interest of their own, e.g. the gross production of industries. Thus, some kind of compromise between the fundamental and the reduced systems has to be found. We shall examine some of the possibilities.

Now, let us try the following procedure: We simply delete rows and columns of the fundamental inverse corresponding to intermediate accounts, except the accounts for domestic commodities, which we leave unaffected. This yields a new system inverse Z^d , still having the same two upper right submatrices and thereby the same transfer properties. Reversing equation (8) using the new Z^d yields the system coefficient matrix

$$(9) \quad \bar{T}^d = I - (Z^d)^{-1}.$$

The resulting equivalent domestic commodity system is shown in Table 5.¹⁰

TABLE 5
THE EQUIVALENT DOMESTIC COMMODITY SYSTEM

From \ To	Primary Factors	Import Commodities	Domestic Commodities	Final Demand	Sum
Primary Factors			$\bar{Y}D$	Y_f	y
Import Commodities			$\bar{m}\bar{U}D$	$\bar{m}F$	m
Domestic Commodities			$\bar{q}\bar{U}D$	$\bar{q}F$	q
Final Demand					0
Sum	0	0	q'	f'	

Alternatively, we might have deleted the intermediate rows and columns of the fundamental inverse, except those for total commodities. This would yield another inverse Z^c , from which the equivalent commodity system T^c could be found analogously. This system is given in Table 6.

The commodity system has the same dimensions as the domestic commodity system. Both systems rely heavily on the assumption of constant \bar{D} coefficients, but input structures and import shares may be modified. However, the commodity

⁹By the formulae of note 7,

$$\bar{Y}\bar{D}\bar{q}(I - \bar{U}\bar{D}\bar{q})^{-1}F = \bar{Y}\bar{D}(I - \bar{q}\bar{U}\bar{D})^{-1}\bar{q}F = \bar{Y}(I - \bar{D}\bar{q}\bar{U})^{-1}\bar{D}\bar{q}F.$$

Thus, finding the reduced system does not require inversion of the huge matrix $(I - \bar{U}\bar{D}\bar{q})$.

¹⁰The easiest way to demonstrate equivalence is to solve the domestic commodity system from table 5 and use the formulae from note 9.

TABLE 6
THE EQUIVALENT COMMODITY SYSTEM

From \ To	Primary Factors	Import Commodities	Commodities Total	Final Demand	Sum
Primary Factors			$\bar{Y}D$	Y_f	y
Import Commodities			\bar{m}		m
Commodities Total			$\bar{U}D$	F	$q+m$
Final Demand					0
Sum	0	0	$(q+m)'$	f'	

system is preferable by virtue of the larger number of zeros. Still, it may be too large for practical purposes.

To achieve a substantial reduction, we must return to the fundamental inverse and delete both types of intermediate commodity accounts, maintaining the industry accounts. By the now usual procedure, we get the equivalent industry system T'' shown in Table 7.

TABLE 7
THE EQUIVALENT INDUSTRY SYSTEM

From \ To	Primary Factors	Import Commodities	Industries	Final Demand	Sum
Primary Factors			Y	Y_f	y
Import Commodities			$\bar{m}U$	$\bar{m}F$	m
Industries			$\bar{D}\bar{q}U$	$\bar{D}\bar{q}F$	g
Final Demand					0
Sum	0	0	g'	f'	

Like the former reduced systems, the industry system relies heavily on the assumption of constant \bar{D} coefficients, but in addition we may have difficulty in varying import shares, since domestic production and imports are no longer classified in the same way. This has led some input-output theorists to reclassify imports by "competing industry" using the \bar{D} coefficients. Other methods are available and possibly preferable, but they are outside the scope of this paper.¹¹ Despite the difficulties involved in modeling import shares, the industry system has become the preferred compromise between applicability and detail in many countries. This choice is probably forced by the vast dimensions of both commodity systems—or, alternatively, by problems concerning the aggregation they necessitate.

Perhaps the future will call for mixed commodity-industry systems, in which selected commodities are open for variations in import shares, while less important commodities are treated on the less accurate industry basis. The method of construction of such systems should now be obvious.¹²

¹¹The method used in the Danish econometric model ADAM is summarized in Dam (1984).

¹²In fact, such a mixed system has been integrated in the Danish input-output system, c.f. Thage (1982).

ALTERNATIVE "TECHNOLOGY ASSUMPTIONS"

All models considered so far are of the "industry technology" type. Indeed, we demonstrated that the industry technology assumption fits naturally into the general input-output assumption of constant input coefficients of any system account. However, this fact implies nothing about the realism of such an assumption.

The alternative "commodity technology" assumption is widely advocated as a more meaningful one.¹³ In terms of the fundamental system matrix T , this assertion implies that the determination of submatrices D , U and Y by the general assumption (5) is abandoned. Such an extension of analysis is perfectly possible, and the reader will find the proposed framework useful for this purpose.

However, we do not elaborate the extension here. In the author's view, the choice of "technology assumption" is an interesting theoretical question, but the practical importance of it is easily exaggerated. The two "technology assumptions" differ, but only to the extent that a number of commodities are produced in more than one industry as "by-products". As industries are usually classified according to type of product, the "by-products" problem is likely to be of minor importance, quantitatively.

To the author, the basic problem is rather a different one, namely that about 2,500 commodity input structures are wanted, whereas only 117 independent input structures can be estimated from industry data (using Danish system dimensions). Thus, in the case of "commodity technology" assumptions, we are forced to aggregate commodities into 117 "characteristic commodities," while in the case of "industry technology" assumptions we need not aggregate commodities, since we relate the 117 input structures available directly to industries. This means that the two "technology assumptions" are properly to be regarded as alternative aggregation procedures having roughly the same properties, but each having some drawbacks as to the other.

However, none of the subsequent results depend critically on the "technology assumption." The reader may readily substitute the words "characteristic commodity" and "commodity group" for "industry" and "branch of production" below (and, of course, modify definitions (10) through (13) appropriately).

AGGREGATION OF INDUSTRIES

In the following, we take the industry system as the basis for our analysis. For brevity, we introduce the symbols

$$(10) \quad A = \bar{D}\bar{q}U$$

$$(11) \quad E = \bar{D}\bar{q}F,$$

$$(12) \quad M = \bar{m}U, \quad \text{and}$$

$$(13) \quad M_f = \bar{m}F.$$

The system can now be written as in Table 8.

¹³See e.g. U.N. (1973).

TABLE 8
THE INDUSTRY SYSTEM (REPEATED)

From \ To	Primary Factors	Import Commodities	Industries	Final Demand	Sum
Primary Factors			Y	Y_f	y
Import Commodities			M	M_f	m
Industries			A	E	g
Final Demand					0
Sum	0	0	g'	f'	

We still use the bar to denote coefficient matrices. The transfer properties of the industry system are now determined by the two matrices $\bar{Y}(I - \bar{A})^{-1}\bar{E} + \bar{Y}_f$ and $\bar{M}(I - \bar{A})^{-1}\bar{E} + \bar{M}_f$.

While the industry system of Table 8 will probably satisfy the input-output theorist, it might still be too large for the econometrician, since he needs time-series of economic data. Thus, he has to store, say, ten observations of the tables, and this may require further reductions in the table size.

In this case, the first thing to do is to aggregate imports into components, e.g. commodity groups, factors into factor classes, and final demand into broad categories. Since the number of categories of final demand determines the degrees of freedom of the quantity model, and the number of system inputs determines the degrees of freedom of the dual price model, this aggregation of the open accounts is certainly critical. However, we shall not treat this problem here, since it is of a philosophical or professional, rather than a technical nature.¹⁴ Instead, we focus on the more technical question of aggregating industries into main branches of production—given the preferred choice of aggregation of system inputs and outputs.

PERFECT AGGREGATION

We now ask the following question: is it possible to aggregate industries into main branches of production in such a way that the transfer properties of the main branch system are identical to the transfer properties of the industry system? A sufficient condition for such an equivalence obviously is, that the transfer matrix of the system is unchanged, i.e. that

$$(14) \quad \bar{Y}(I - \bar{A})^{-1}\bar{E} = \bar{Y}^*(I - \bar{A}^*)^{-1}\bar{E}^*,$$

and

$$(15) \quad \bar{M}(I - \bar{A})^{-1}\bar{E} = \bar{M}^*(I - \bar{A}^*)^{-1}\bar{E}^*,$$

where the asterisk (*) indicates, that the industry dimension(s) of the matrix are aggregated to a main branch level.¹⁵

¹⁴In certain quite special cases, however, perfect aggregation of demands or inputs is possible, c.f. J. A. Olsen (1982).

¹⁵If the system outputs f vary freely, the conditions (14) and (15) are also necessary. If the variation of the f elements is subject to linear restrictions, the conditions might be weakened.

It is well-known from the theory of aggregation of input-output models¹⁶ that conditions (14) and (15) hold if and only if any two industries to be aggregated into a main branch satisfy either

(16) The two industries have identical input structure, i.e. identical columns in matrices \bar{A} , \bar{Y} and \bar{M} ,

or

(17) The two industries produce proportionally, i.e. their relative shares of the main branch production are constant.

While the first condition is well explored in the literature, the significance of the second has been quite overlooked. This has probably been due to the lack of a total system view of the problem.

The key to the importance of condition (17) is the fact that gross productions of industries, according to the model itself, are determined as a linear map of final demands f , and since the number of industries is usually larger than the number of demand categories, this imposes linear restrictions on the variation of gross productions g_j . However, this need not imply the desired proportionality, since linear restrictions in general may be more subtle. The author believes that the following special cases are the only two that guarantee proportionality:

If, for a given main branch, either

(18) only one of the industries in the main branch supplies outside that branch,

or

(19) all industries in the main branch have an identical output structure, i.e. proportional rows in the matrix (\bar{A}, \bar{E}) , except for the intra-branch supplies,

then all industries in the main branch produce proportionally.¹⁷

While the condition of an identical input structure relates to so-called horizontal aggregation, the condition (18) relates to vertical aggregation, and the output structure condition (19) may relate to both types. The simplest example of two industries fulfilling the vertical aggregation condition is the case of a "chain": If the whole output of some industry is used as input in a single other industry, the two industries produce proportionally, due to the assumption of constant technical coefficients.

¹⁶See e.g. Theil (1957) or (1971).

¹⁷Proof: We prove the case in which the main branch does not supply other industries, i.e. the only extra-branch supplies are final demands by branch industries, denoted e_b , which is a subvector of $e = \bar{E}f$. We partition the A -matrix into our branch b and the remaining industries r and rearrange, so

$$A = \begin{bmatrix} A_{bb} & 0 \\ A_{rb} & A_{rr} \end{bmatrix}.$$

The zero submatrix will remain in the inverse $(I - \bar{A})^{-1}$. Thus, $g_b = (I - \bar{A}_{bb})^{-1}e_b$. If the e_b elements are bound to move proportionally, the elements of g_b must as well. The simplest case of this is that only one of the elements is non-zero. This proves (18). Another case, guaranteeing that the elements of e_b move proportionally, is proportionality of all rows in E_b . This proves (19). Obviously, the necessary proportionality of extra-branch supplies still holds if we substitute for the assumption $A_{br} = 0$ an assumption of proportional rows of A_{br} , but this proof is omitted.

If, for a given aggregation, the assumptions of an identical input structure, identical output structure or the vertical aggregation condition hold inside branches, the aggregated main branch system is equivalent to the fundamental system. In empirical work, however, such assumptions are not likely to hold exactly, but rather to some degree. Thus, we abandon the requirement of strict equivalence to the fundamental system, focusing our attention on measuring the “deviation” of a proposed main branch system from the fundamental system. This means, in the words of Theil (1954), that we leave the field of perfect aggregation and enter into the field of imperfect aggregation.

IMPERFECT AGGREGATION

The reductions of the fundamental system discussed so far imply no real sacrifice of information, since the transfer properties of all the proposed systems are identical. Nevertheless, in practice we may wish to push reductions even further, necessitating such a sacrifice of information. This will be the case if we regard the benefit of the system reduction as greater than the cost of minor distortions of the fundamental transfer properties. This question is the traditional problem of finding the relevant level of abstraction in a given scientific context. The problem has no theoretical solution, but practice has shown that many empirical economists have been willing to accept substantial “aggregation error” to achieve the desired simplification.¹⁸

To provide the best basis for our choice of abstraction level, we need a measure of the aggregation “errors.” An obvious choice of such a measure is the vectors of “aggregation bias” on the system inputs:

$$(20) \quad b_y = (\bar{Y}(I - \bar{A})^{-1} \bar{E} - \bar{Y}^*(I - \bar{A}^*)^{-1} \bar{E}^*)f$$

$$(21) \quad b_m = (\bar{M}(I - \bar{A})^{-1} \bar{E} - \bar{M}^*(I - \bar{A}^*)^{-1} \bar{E}^*)f$$

These vectors of aggregation bias are null-vectors in the base year, but they should be calculated for relevant alternative values of the f vector, e.g. using historical *ex post* predictions.¹⁹

In empirical works by the author concerning the aggregation level of the Danish econometric model ADAM, the aggregation biases (20) and (21) proved to be applicable measures of the quality of alternative aggregations of industries.²⁰ However, their strength lies in the testing phase of proposed aggregations, rather than in the constructive phase. In the latter phase, proposing reasonable aggregation keys, more intuition has to be involved. In this context, the author found it very convenient to take the pattern of zeros in the industry system matrix as the

¹⁸This acceptance must be viewed in the light of a number of empirical investigations showing that the “technical coefficients” are not particularly stable. Danish evidence is summarized in F. Lauritzen (1982).

¹⁹In this case, the aggregation biases (20) and (21) can be viewed as generalizations of Theil’s measure of aggregation bias, c.f. Theil (1957) or (1971). Please note that Theil’s concept of “first order bias” loses significance from our point of view. A detailed treatment of the aggregation bias measure, including proposed decompositions, is given in J. A. Olsen (1982).

²⁰A complete investigation of aggregation bias should also involve the dual “price bias,” defined as some row vector of input prices left-multiplied on the bracketed coefficient matrix of (20) and (21).

starting point. This pattern of zeros can be regarded as determining the qualitative structure of inter-industry supplies.²¹

The first step constructing an aggregation key, then, is to re-arrange the industries in such a way that the inter-industry matrix A becomes block-triangular. This amounts to an organized way of examining the possibilities of vertical aggregation. The re-arranging procedure is well known from the theory of macroeconomic models as “causal ordering,” and thus the necessary algorithm is a standard facility of packages of econometric software.²²

The second step of the procedure is to examine to what degree the blocks of industries fulfill the conditions of perfect aggregation. If this degree is readily acceptable, we are through. If not, we should still prefer to aggregate inside blocks rather than across blocks, since in the first case we preserve the main characteristics of inter-industry structure, while in the second case we do not.

We may extend the procedure by a third step, aggregating those of the blocks which supply the same component of final demand.²³ This amounts to an effort to preserve the zeros of the original transfer matrix in the aggregated transfer matrix.²⁴

Aggregation keys, constructed in the proposed way, in the ADAM case turned out to be superior to other keys. The 117 industries of the fundamental system of the Danish national accounts were aggregated into the 19 main branches of the ADAM model, causing less than 1 percent aggregation bias and no loss of predictive power (using 5-year *ex post* predictions).²⁵

RESETTING OF MINOR CELLS

Aggregation of the tables is not the only way to reduce the body of data. An alternative way is to ignore the minor cells of the tables, thereby increasing the number of zeros. As the zeros need not be stored, this is a simplification as well as the reduction of table dimensions.

In the case of the ADAM model, a procedure of resetting the small cells to zero turned out to be indeed competitive to further aggregation. The method is a natural extension of the above-mentioned structural aggregation procedure, which tends to concentrate inter-industry supplies in relatively few, but large cells of the main branch system. In the ADAM system matrix, e.g., 20 percent of the cells cover 90 percent of the amounts of commodity flow. Thus, in this case, a number of the small cells can be reset to zero, causing only inferior changes in the transfer properties. This conclusion is further supported by the

²¹The pattern of zeros determines the directed graph (“digraph”) of the system, see e.g. Harary *et al.* (1965). The brief introduction by Defourny and Thorbecke (1984) to social accounting matrix applications of graph theory is strongly recommended.

²²To make the main inter-industry structure appear clearly, it may be necessary to ignore small cells of the tables.

²³This should not be done if the blocks supply several categories of final demand.

²⁴In the language of graph theory, we wish to preserve the reachability properties of the original system digraph.

²⁵C.f. J. A. Olsen (1982).

empirical findings that the small coefficients of the tables tend to be relatively unstable.²⁶

The suppression of small cells should be done without affecting throughputs of the system accounts. This means that a cell cannot be reset to zero without a number of derived resettings. If, for example, we wish to set the cell (i, j) to zero, we have to increase some other cell in row no. i as well as in column no. j by the same amount as the reduction in cell (i, j) , say cells (i, l) and (k, j) . Finally we have to decrease cell (k, l) correspondingly to preserve throughputs, as shown in Figure 1.

	To	j	...	l
From		j	...	l
i		reset	...	+
.		.	.	.
k		+	...	-

Figure 1. The principle of resetting

The derived resettings are arbitrary, but they can be chosen with more or less flair. As in the case of aggregation, they are part of an abstraction process, requiring both caution and skill. However, two observations can be made. Firstly, the intra-industry (diagonal) element (i, i) is a harmless place to put derived resettings, if possible. Secondly, the derived resettings should be kept inside the submatrices A , E , M and M_f , leaving submarginals unaffected.

The process of determining derived resettings is quite laborious, when done by hand. Methods for automatic resetting have not yet been fully developed, and perhaps such methods are not even desirable, since they replace consciousness by electronics.²⁷ Anyway, once the resetting procedure has been implemented, it can be repeated automatically in updatings, provided that the inter-industry structure does not change dramatically.

The resetting procedure was a major success in the ADAM case, since the original number of 550 positive cells of the system matrix was reduced to a third, causing negligible changes in transfer properties. These changes can be measured analogously to (20) and (21), replacing the aggregated matrices by reset matrices.

CONCLUSION AND RESULTS

The implementation of integrated systems of national accounts and input-output tables based on detailed commodity flows, as recommended by the S.N.A., dramatically increases the amount of data available for econometric analysis. This increase creates, in turn, a growing demand for principles of abstraction,

²⁶For Danish evidence, see F. Lauritzen (1982).

²⁷The well-known RAS method will probably apply, since it preserves zeros as well as marginals, see e.g. United Nations (1973).

i.e. rules telling us how to extract the main informational content of the data, relevant for various special applications.

In the case of the Danish econometric model ADAM, such an extractive procedure was developed successfully. This is clearly demonstrated in Table 9,

TABLE 9
PREDICTIONS OF ADAM SYSTEM INPUTS
MILLIONS OF DKR, 1980 PRICES

	Observed 1975 (1)	Prediction Errors		Bias (2)-(3), (4)	Bias pct. (5)
		ADAM (2)	Detailed (3)		
Imports, by SITC					
0 food	7,160	315	309	6	0.1
1 beverages	1,084	133	105	28	2.6
2+4 crude materials	5,673	-1,604	-1,748	144	2.5
32 coal	1,310	-147	-179	32	2.4
333 crude oil	10,504	-3,521	-3,502	-19	-0.2
3, nei oil products	15,499	-477	-441	-36	-0.2
5 chemicals	7,863	-130	-197	67	0.9
67-69 metals	7,835	-1,100	-1,059	-41	-0.5
6, nei misc. products	9,974	-820	-756	-64	-0.6
78 motor vehicles	3,805	-108	34	-142	-3.7
79 ships, aircraft	4,273	3,187	3,187	0	0.0
7, nei machinery	15,484	193	204	-11	-0.1
8+9 misc. goods	8,232	104	-39	143	1.7
services	6,604	1,386	1,445	-59	-0.9
Duties	57,904	-444	-358	-86	-0.1
Factor income	282,540	3,034	2,996	38	0.0
	Observed 1980 (1)	Prediction Errors		Bias (2)-(3) (4)	Bias, pct. (5)
		ADAM (2)	Detailed (3)		
Imports, by SITC					
0 food	10,233	1,766	1,750	16	0.2
1 beverages	1,019	55	40	15	1.5
2+4 crude materials	7,248	312	236	76	1.0
32 coal	2,911	1,564	1,544	20	0.7
333 crude oil	7,584	-2,504	-2,513	9	0.1
3, nei oil products	13,958	-2,573	-2,587	14	0.1
5 chemicals	10,756	385	513	-128	-1.2
67-69 metals	9,964	1,105	1,077	28	0.3
6, nei misc. products	11,278	-521	-624	103	0.9
78 motor vehicles	2,915	-73	-113	40	1.4
79 ships, aircraft	1,005	-4,045	-4,022	-23	-2.3
7, nei machinery	18,552	2,270	2,141	129	0.7
8+9 misc. goods	11,696	2,255	2,269	-14	-0.1
services	8,415	875	898	-23	-0.3
Duties	57,797	-978	-803	-175	-0.3
Factor income	325,818	105	196	-91	-0.0

Prediction errors are observed minus predicted values, using 5 year old coefficient matrices (i.e. 1975 inputs were predicted using 1970 coefficients and 1980 using 1975 coefficients). Please note that prediction errors and biases must sum to zero.

showing *ex post* predictions of the ADAM system inputs. The ADAM inputs were predicted from 27 categories of final demand, using matrices of 117 industries and 19 ADAM main branches, alternatively. In both cases, prediction errors are quite large, as we should expect from a simple assumption of constant coefficients. However, there is little doubt that the aggregated and reset ADAM matrices reflect the structural changes in the economy in a satisfactory way, for macroeconomic purposes.

REFERENCES

- Dam, Poul Uffe (1984), The Danish Macroeconomic Model ADAM—A survey. *Economic Modelling* (forthcoming). Manuscript available from Danmarks Statistik.
- Defourny, J. and Thorbecke, E. (1984), Structural path analysis and multiplier decomposition within a social accounting matrix framework. *Economic Journal* 94 (March 1984), pp. 111-136.
- Harary, F., Norman and Cartwright (1965), *Structural Models—an Introduction to the Theory of Directed Graphs*, Wiley.
- Lal, Kishori (1982), Compilation of Input-output Tables: Canada. *Review of Income and Wealth* 1982:4, pp. 411-430. Also printed in Skolka (ed., 1982).
- Lauritzen, Finn (1982), An Investigation of Danish Input-output Tables, 1966-75. *Scand. J. Econ.*, 84 (3), pp. 405-20.
- Olsen, J. Asger (1982), Om aggregeringsniveauer i empiriske makromodeller (On aggregation levels in empirical macro models). Danmarks Statistik (mimeographed, in Danish. Available from Danmarks Statistik).
- Skolka, Jiri (ed., 1982), Compilation of Input-output Tables, Proceedings of the 17th General Conference of the IARIW. Lecture notes in Economics and Mathematical Systems, vol. 203. Springer Verlag, Berlin.
- Thage, Bent (1982), Techniques in the Compilation of Danish Input-output Tables: A new approach to the treatment of imports. in Skolka (ed., 1982).
- Theil, Henri (1954), *Linear Aggregation of Economic Relations*, North Holland.
- Theil, Henri (1957), Linear Aggregation in Input-output Analysis, *Econometrica* 25, pp. 111-122.
- Theil, Henri (1971), *Economics and Information Theory*, North Holland.
- United Nations (1968), *A System of National Accounts*, Series F, no. 2, rev. 3, New York.
- United Nations (1973), *Input-output Tables and Analysis*, series F, no. 14, New York.