4-1-1974

Methods of Estimating the Input Structure for Separate Groups of Hetergeneous Industries

G. J. A. Mensink

C58r

SIXTH INTERNATIONAL CONFERENCE ON input-output techniques

(Vienna, 22 - 26 April 1974)

Session 3 : Studies of national applications

REGIONAL RESEARCH INSTITUTE

METHODS OF ESTIMATING THE INPUT STRUCTURE FOR SEPARATE PRODUCT GROUPS OF HETEROGENEOUS INDUSTRIES

by

G.J.A. Mensink
Netherlands Central Bureau of Statistics
Voorburg
Netherlands

Received on 19.3.1974

id.74-1855

CONTENTS: PAGE:

The construction of detailed input-output tables is often hampered for lack of data on cost structures of separate product groups of heterogeneous industries. In this article methods will be described which may be helpful in estimating those cost structures, using data on inputs and outputs from individual census forms as the basic material. Limitations of the methods will be discussed and illustrated from some experience, gained by the Netherlands Central Bureau of Statistics in the last few years.

1. Introduction

In the construction of input-output tables basic statistics on the cost structure of separate industries and products play an indispensable role. When detailed information of this kind is lacking, recourse must be had to assumptions. In "A System of National Accounts" (S.N.A.) the United Nations propose some mechanical methods, all relating to the handling of 'non-characteristic' products of industries, and to be combined as far as possible with detailed information about cost structures of this subsidiary production. Limiting assumptions behind these methods are the assumption of a commodity technology on the one hand, the assumption of an industry technology on the other. These are extreme assumptions which "will only give the same results if there is no subsidiary production of any kind, that is to say if the problem they ar designed to resolve does not exist." 1)

Besides these problems of the transfer of non-characteristic products to other categories, there is the difficulty that an industry may have more than one characteristic product, the cost structure of each being different. Apparently, the constructor of an input-output table faces no problem, when such a heterogeneous industry constitutes one row and column in the table to be compiled. Really, however, it must be said that in those cases the existing problem of the unknown cost structure per product is passed on to the user of the table. As a rule, the latter will solve the problem by the assumption of an industry technology, which means "that an industry is assumed to have the same input structure whatever its product mix." 2)

It will be clear that, as far as this assumption is not fulfilled, the usefulness of the table for planning purposes is diminished. Therefore, the constructor of the table still faces the problem to sub-divide, as far as possible, the rows and columns regarding homogeneous groups, so that the user's assumption of a given technology for each group is close to reality.

The most troublesome examples of such heterogeneous industries surely are agriculture and construction, for which the International Standard Industrial Classification gives a very broad division indeed, even at the group level: 3)

group 1110 Agricultural and livestock production
" 1120 Agricultural services
" 1130 Hunting, trapping and game propagation
" 5000 Construction.

The reason why a further subdivision of these groups is not given is clear enough: no known methods of collection of basic statistics will provide information on outputs and related inputs for groups of agricultural products or parts of construction.

In this article regression methods will be described which may be useful in efforts to subdivide, in addition to the rows, also the columns regarding such heterogeneous industries in the input-output table. In essence, the problem here is not different from that which is met when the output and related inputs of non-characteristic products must be transfered. The possibility of obtaining useful information to this end from the individual census records has been indicated already in the S.N.A. where it is stated that "census returns may be regrouped to show the input structures in various industries of groups of establishments which are not much engaged in uncharacteristic production." 4)

1) United Nations, [9], page 39.
2) United Nations, [9], page 39.
3) United Nations, [10], pp. 27-40
4) United Nations, [9], p. 37.

of heterogeneous industries into more rows and columns

For heterogeneous industries a regrouping of firms into categories producing different commodities will generally be impossible, as the output of most firms will be of a mixed character. When, however, the product mix is different between firms, regression analysis may be applied to the census data of individual firms to estimate the input structure for separate products or activities.

The Netherlands Central Bureau of Statistics applied the method formerly in an international comparison of labour productivity, where it was neccessary to have estimates of value added per unit of separate activities of heterogeneous industries. 5) In chapter 2 the description of the method will be extended to comprise the whole input structure of separate activities. Furthermore, the consequences for the estimating procedure of introducing restrictions that some cost percentages should be zero for technological reasons, and no negative percentages should occur, will be indicated.

In chapter 3 a number of limitations of the methods will be discussed and illustrated by a description and some numerical results of the experience, gained by the Netherlands Central Bureau of Statistics in the last few years. The experience relates to applications of the methods to census data of firms belonging to the groups 'agricultural and livestock production' and 'manufature of furniture and fixtures, except primarily of metal' 6). For agriculture the records of the census of farming results 1965/'66 have been used, for wooden furniture the data have been taken from the census of production 1968 of the manufacture of wood products. As an example, the results obtained for the wooden furniture industry will be presented in full in chapter 4.

From the discussions in chapters 3 and 4 it will be seen that not all problems of estimating cost structures seem to be capable of tackling by the methods described. In general, however, the results obtained may be regarded as promising for this kind of attack.

2. Exposition of the methods

Suppose that the firms within a heterogeneous industry produce, in separated production processes, a number of goods in proportions which are different between firms. The cost structure of each kind of output is assumed to be the same for each firm, except for random disturbances.

The assumption made here with regard to the cost structures is less rigid than the assumption of 'a commodity technology', as mentioned in the S.N.A. as one method for the transfer of non-characteristic products of industries. The latter means "that a commodity is assumed to have the same input structure in whichever industry it is produced."7) The assumption in this article only relates to the production of commodities within the (part of the) industry under consideration, leaving open the possibility that different input structures may occur in the production of these commodities by other industries or by the same industry in e.g. other regions or other size classes of firms.

5) G.J.A. Mensink, [5], especially p. 39.
6) United Nations, [10], pp. 27-40.
7) United Nations, [9], p. 39.

The main difference, however, lies in the way in which the assumption is exercised. For the transfer of non-characteristic products it is handled as a mechanical means of effecting transfers, which is applied to the commodidies defined in the input and output tables from which the input-output table has to be constructed. In our case the product groups to be distinguished can, within certain limits set by the assumption of separated production processes, be detailed in such a way as to make them as homogeneous as possible with regard to cost structure. The assumption can therefore be more realistic and opens perspectives for just an increase of the number of rows and columns in the input-output table.

2.1 Equation-by-equation ordinary least-squares

The assumptions made give rise to the following system of linear equations for firm k:

$$(1)\quad \begin{cases} Y_{1k} = \alpha_{11}X_{1k}+\alpha_{12}X_{2k}+\ldots\ldots\ldots\ldots+\alpha_{1N}X_{Nk}+u_{1k} \\ Y_{2k} = \alpha_{21}X_{1k}+\alpha_{22}X_{2k}+\ldots\ldots\ldots\ldots+\alpha_{2N}X_{Nk}+u_{2k} \\ \vdots \qquad\qquad\qquad \vdots \\ Y_{Mk} = \alpha_{M1}X_{1k}+\alpha_{M2}X_{2k}+\ldots\ldots\ldots\ldots+\alpha_{MN}X_{Nk}+u_{Mk} \end{cases}$$

in which

Y_{mk} = value of input of item m (m=1,2,...,M) by firm k;

X_{nk} = value of output of item n (n=1,2,..,N) by firm k;

α_{mn} = value of input of item m (m=1,2,...,M) needed for an output of X_n = 1 (n=1,2,...,N);

u_{mk} = residual input of item m by firm k.

As the items m=1...M comprise all the inputs inclusive of value added, the sum total of the input coefficients α_{mn} for output n equals unity, so

$$(2)\qquad \sum_{m=1}^{M} \alpha_{mn} = 1 \qquad (n=1....N).$$

For a sample of K firms the equations regarding the input category m can be expressed in matrix notation as

(3) $$Y_m = X\alpha_m + u_m \qquad (m = 1....M)$$

where

$$X = \begin{bmatrix} X_{11} & \text{-----------} & X_{N1} \\ X_{12} & \text{-----------} & X_{N2} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ X_{1K} & \text{-----------} & X_{NK} \end{bmatrix}; \qquad Y_m = \begin{bmatrix} Y_{m1} \\ Y_{m2} \\ \cdot \\ \cdot \\ \cdot \\ Y_{mK} \end{bmatrix}$$

$$\alpha_m = \begin{bmatrix} \alpha_{m1} \\ \alpha_{m2} \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{mN} \end{bmatrix}; \qquad u_m = \begin{bmatrix} u_{m1} \\ u_{m2} \\ \cdot \\ \cdot \\ \cdot \\ u_{mK} \end{bmatrix};$$

The set of M regression equations (3) can be written as

$$\begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ Y_M \end{bmatrix} = \begin{bmatrix} X & 0 & \text{--------} & 0 \\ 0 & X & & \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & X & 0 \\ 0 & \text{--------} & 0 & X \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \alpha_M \end{bmatrix} + \begin{bmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ u_M \end{bmatrix}$$

or, compactly, in the familiar form

(4) $$Y = Z\alpha + u$$

where the number of observations is MK and the number of explanatory variables is MN.

Though, for $K > N$ and X having rank N, the application of ordinary least-squares to (4) does result in an unbiased estimate of the vector α, the estimate is not efficient, because the covariance matrix of the disturbance u does not take the simple form $\sigma^2 I$.

In general, when it is assumed besides $E(u) = 0$ that the covariance matrix of u is given by

(5) $$E(uu') = V$$

where V is a positive definite matrix the order of which equals the number of obeervations, best linear unbiased estimates will be obtained by the generalized least-squares estimator:[8)]

(6) $$\hat{\alpha} = (Z'V^{-1}Z)^{-1}Z'V^{-1}Y$$

with covariance matrix

(7) $$\text{var}\ \hat{\alpha} = (Z'V^{-1}Z)^{-1}$$

For our specific problem we assume that

(8) $$E(u_m u'_{m'}) = \sigma_{mm'} I \qquad (m,m'=1....M)$$

For m=m' this assumption implies that the disturbances of the single equation regarding input m have homogeneous variance σ_{mm} and are uncorrelated between firms. For $m \neq m'$ (8) states that the disturbances of the equations regarding the inputs m and m' have the same covariance $\sigma_{mm'}$ for all firms, whereas for disturbances concerning different firms these covariances are zero.

From (8) it follows that in our case

(9) $$V = \begin{bmatrix} E(u_1u'_1) & \cdots & E\ u_1u'_M \\ \vdots & & \vdots \\ E(u_Mu'_1) & \cdots & E\ u_Mu'_M \end{bmatrix} = \begin{bmatrix} \sigma_{11}I & & \sigma_{1M}I \\ \vdots & & \\ \sigma_{M1}I & & \sigma_{MM}I \end{bmatrix} = \Sigma \otimes I$$

where

(10) $$\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1M} \\ \vdots & & \vdots \\ \sigma_{M1} & \cdots & \sigma_{MM} \end{bmatrix}$$

and $\otimes$ denotes Kronecker multiplication.[9)]

A special feature in our case is that the disturbances are linearly dependent and therefore Σ and V are singular. This is caused by the fact that by definition the sum total of a firm's inputs equals the sum total of its outputs, so for firm k

(11) $$\sum_{m=1}^{M} Y_{mk} = \sum_{n=1}^{N} X_{nk}$$

8) See e.g. H. Theil,[8], section 6.1 or J. Johnston,[2], section 7.1. Generalized least-squares amounts to ordinary least-squares, applied after a transformation of the data in such a way that the covariance matrix of the transformed disturbance indeed takes the form $\sigma^2 I$. The transformation is accomplished by pre-multiplying both sides of (4) by a nonsingular matrix P defined in such a way that $P'P=V^{-1}$. The estimator (6) results now from minimizing the sum of transformed squares $(PY-PZ\hat{\alpha})'(PY-PZ\hat{\alpha})$.

9) The Kronecker product of two matrices A and B, A being of order m x n and B of order p x q, is given by

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$$

The resulting matrix has order mp x nq.

As from (1) and (2) it follows that

$$\sum_{m=1}^{M} Y_{mk} = X_{1k}\sum_{m=1}^{M} \alpha_{m1} + \ldots\ldots + X_{Nk}\sum_{m=1}^{M} \alpha_{mN} + \sum_{m=1}^{M} u_{mk} = \sum_{n=1}^{N} X_{nk} + \sum_{m=1}^{M} u_{mk} \ ,$$

it is easily seen that, because of (11),

$$\sum_{m=1}^{M} u_{mk} = 0$$

This implies that also

$$u_{m'k}(u_{1k}+u_{2k}+\ldots\ldots+u_{Mk}) = 0$$

and so, taking expectations of both sides:

$$\text{(12)} \qquad \sigma_{m'1}+\sigma_{m'2}+\ldots\ldots+\sigma_{m'M} = 0 \qquad (m' = 1\ldots\ldots M).$$

When, however, (12) is the only linear dependence between the disturbances of the system (4), the generalized least-squares technique comes down to deleting from the system the equation regarding one of the inputs, as indicated in general by Theil 10). The estimates of the parameters of the deleted equation, for which we take the equation regarding input number 1, can be computed afterwards by

$$\hat{\alpha}_{1n} = 1 - \sum_{m\neq 1} \hat{\alpha}_{mn} \qquad (n = 1\ldots\ldots N),$$

its standard errors can be obtained by applying the formula for the variance of a sum.

For this reason, from now onwards the equations (4) through (7) will be regarded upon as having reference to the input categories 2...M, so exclusive of the category 1.

The covariance matrix of the disturbance now reads

$$V = E(uu') = \Sigma \otimes I$$

where

$$\text{(13)} \quad \Sigma = \begin{bmatrix} \sigma_{22} & \text{------------} & \sigma_{2M} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \sigma_{M2} & \text{------------} & \sigma_{MM} \end{bmatrix}$$

is assumed to be positive definite and the unit matrix is of order K.

The inverse of V is given by

$$\text{(14)} \qquad V^{-1} = \Sigma^{-1} \otimes I$$

using the property that $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, for matrices A and B being both nonsingular.

---

10) H. Theil, [8], section 6.7.

According to (6) and (7) the generalized least-squares estimator and its covariance matrix now come to read

$$(15) \qquad \hat{\alpha} = \left[Z'(\Sigma^{-1} \otimes I)Z\right]^{-1} Z'(\Sigma^{-1} \otimes I)Y$$

and

$$(16) \qquad \operatorname{var} \hat{\alpha} = \left[Z'(\Sigma^{-1} \otimes I)Z\right]^{-1}$$

respectively.

Remembering that the matrix Z has the form

$$(17) \qquad Z = \begin{bmatrix} X & & & 0 \\ & X & & \\ & & \ddots & \\ 0 & & & X \end{bmatrix} = I \otimes X$$

where the unit matrix has the order M-1, (15) and (16) can be reduced[11] by substituting the expressions

$$(18) \qquad \left[Z'(\Sigma^{-1} \otimes I)Z\right]^{-1} =$$

$$= \left[(I \otimes X')(\Sigma^{-1} \otimes I)(I \otimes X)\right]^{-1} =$$

$$= \left[\Sigma^{-1} \otimes X'X\right]^{-1} = \Sigma \otimes (X'X)^{-1}$$

and

$$(19) \qquad Z'(\Sigma^{-1} \otimes I)Y = (I \otimes X')(\Sigma^{-1} \otimes I)Y = (\Sigma^{-1} \otimes X')Y$$

Using these expressions (15) comes to read

$$(20) \qquad \hat{\alpha} = \left[\Sigma \otimes (X'X)^{-1}\right](\Sigma^{-1} \otimes X')Y = \left[I \otimes (X'X)^{-1}X'\right]Y =$$

$$= \begin{bmatrix} (X'X)^{-1}X' & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & (X'X)^{-1}X' \end{bmatrix} \cdot \begin{bmatrix} Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_M \end{bmatrix}$$

and (16) becomes

11) The condensed way of deriving the reduction has been taken from H.Theil, 8 , pp. 309-310, where use has been made of the properties that $(A_1 \otimes B_1)(A_2 \otimes B_2)....(A_N \otimes B_N) = (A_1.A_2....A_N) \otimes (B_1.B_2....B_N)$, provided that the products $(A_1.A_2....A_N)$ and $(B_1.B_2....B_N)$ both exist, and that $(A \otimes B)' = A' \otimes B'$.

$$(21)\quad \operatorname{var}\hat{\alpha} = \Sigma \otimes (X'X)^{-1} = \begin{bmatrix} \sigma_{22}(X'X)^{-1} & \cdots & \sigma_{2M}(X'X)^{-1} \\ \vdots & & \vdots \\ \sigma_{M2}(X'X)^{-1} & \cdots & \sigma_{MM}(X'X)^{-1} \end{bmatrix}$$

From (20) and (21) it can be seen that in this case the generalized least-squares estimates and their covariances coincide exactly with the results being obtained when the ordinary least-squares method is applied to each equation separately. This is caused by the fact that the observation matrix X is the same in all equations, involving that the matrix Z in partitioned form has the same matrix X as its main diagonal elements.

This result is independent from the circumstance that just the equation regarding the input category 1 has been deleted from the calculations. The same result would have been obtained when another input instead of the input 1 were excluded. This fact, together with the result that the parameter estimates for the deleted equation regarding input i are obtained as $1 - \sum_{m \neq i} \hat{\alpha}_{mn}$ (n=1...N), implies that, when in this case ordinary least-squares is applied separately to all equations, the sum total $\sum_{m=1}^{M} \hat{\alpha}_{mn}$ must be equal to unity for all n. That this is true indeed can be shown as follows.

Summation of the equation-by-equation least-squares estimates

$$\hat{\alpha}_m = (X'X)^{-1}X'Y_m$$

over all m = 1....M yields

$$(22)\qquad \sum_{m=1}^{M} \hat{\alpha}_m = (X'X)^{-1}X'\sum_{m=1}^{M} Y_m$$

From (11) it follows that

$$(23)\quad \sum_m Y_m = \begin{bmatrix} \sum_m Y_{m1} \\ \sum_m Y_{m2} \\ \vdots \\ \sum_m Y_{mK} \end{bmatrix} = \begin{bmatrix} \sum_n X_{n1} \\ \sum_n X_{n2} \\ \vdots \\ \sum_n X_{nK} \end{bmatrix} = X \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

where the unit column vector has order N x 1.

Substituting (23) in (22) gives

$$\sum_m \hat{\alpha}_m = (X'X)^{-1}X'X \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

and so

$$(24) \qquad \sum_m \hat{\alpha}_m = \begin{bmatrix} \sum_m \hat{\alpha}_{m1} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \sum_m \hat{\alpha}_{mN} \end{bmatrix} = \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}$$

The conclusion has been reached in this section that, when in the model discussed the observation matrix X is the same in all equations, best linear unbiased estimates of the α_{mn} are obtained by applying ordinary least-squares equation-by-equation, and these estimates fulfil the criterion (24).

Unfortunately, the application of this simple method is often hampered by two practical problems.

At first it may happen that a certain type of input cannot be, for technical reasons, an input factor for one or more of the distinguished types of output. In agriculture and livestock production for instance we know that fertilizers are not used up in poultry farming and feeding-stuffs do not constitute a direct input for the output of cereals. Those input coefficients have to be put zero beforehand, which problem necessitates to follow the joint estimating procedure, to be described further in section 2.2.

Secondly, neither the equation-by-equation least-squares estimator, nor the joint estimator to be discussed in section 2.2, will ensure the estimates of all α_{mn} to be non-negative, as they should be. 12) Introduction of these inequalities as restrictions renders the problem into a case of quadratic programming, the solution of which will be indicated in section 2.3.

2.2 Joint generalized least-squares

When one or more of the α_{mn} have to be put equal to zero for technical reasons, the estimate of the vector α can be required to satisfy these restrictions. The generalized least-squares (g.l.s.) estimator satisfying the constraints will be denoted by a, as opposed to the unconstrained g.l.s. estimator $\hat{\alpha}$ in section 2.1.

In our case the constraints may be of two kinds, dependent from whether the restriction regards a parameter of the equation which has been deleted from the direct application of g.l.s. (input number 1), or a parameter of one of the other equations. In the latter case a constraint can be expressed directly as e.g. $a_{23} = 0$ or, in matrix notation:

$$(25) \qquad [0010.....0]\, a = 0$$

12) Only for profits, when taken as a separate input category, negative percentages might occur.

where the row vector of (M-1)N elements has unity in the position corresponding to the position of a_{23} in a and zero's elsewhere.
A constraint, however, as e.g. $a_{12}=0$ has to be expressed in terms of the coefficients occurring in the vector a (which is exclusive of the coefficients of the equation 1) and therefore comes to read

$$\sum_{m\neq 1}^{M} a_{m2} = 1$$

or, in matrix notation:

$$(26) \qquad \left[010....010....010....0\right] a = 1$$

where the row vector of (M-1)N elements now has unity in the positions corresponding to the positions of a_{22} , a_{32} , ... a_{M2} in a , and zero's elsewhere.

The difference is important because in general a restriction that a parameter estimate should be zero, can be imposed as well by introducing the restriction explicitly in the estimating procedure, as by deleting the variable associated with that parameter from the set of explanatory variables 13). In our case the latter procedure can be followed only as far as no regression coefficients of the first (deleted) equation have to be put equal to zero, because a constraint of the type (26) has to be introduced always explicitly. Both procedures will be described here, starting with the explicit introduction of all the constraints.

In matrix notation all constraints of the type (25) and (26) can be combined to

$$R\, a = r$$

where matrix R has the order g x (M-1)N and r is a column vector of g elements, g being the number of constraints.
The constrained g.l.s. estimator is usually expressed as 14)

$$(27) \qquad a = \hat{\alpha} + CR'(RCR')^{-1}(r-R\hat{\alpha})$$

and has the covariance matrix

$$(28) \qquad \text{var } a = C - CR'(RCR')^{-1}RC$$

where

$$(29) \qquad C = (Z'V^{-1}Z)^{-1}$$

and $\hat{\alpha}$ is the unconstrained estimator (6).
Though this article is concerned with the application, not with the derivation of these formulae it is necessary, in view of the problem to be discussed in section 2.3, to indicate the way in which formula (27) is obtained.

---

13) See e.g. J. Johnston, [2], pp.158-159. For generalized least-squares this holds true as well as for ordinary least-squares, since the former is equivalent to ordinary least-squares applied to transformed data.

14) See e.g. H. Theil, [8], sections 6.8 and 7.3.

The aim of the procedure is to minimize, instead of

$$(PY-PZ\hat{\alpha})'(PY-PZ\hat{\alpha})$$

as in the unconstrained case (see footnote 8), the expression

$$\tfrac{1}{2}(PY-PZa)'(PY-PZa) - \lambda'(Ra-r) \tag{30}$$

where λ is a column vector of g Lagrange multipliers and the factor $\frac{1}{2}$ before the left-hand term has been introduced for convenience. The system of normal equations obtained from differentiating (30) with respect to a and λ and equating to zero, now comes to read

$$\begin{bmatrix} C^{-1} & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} a \\ -\lambda \end{bmatrix} = \begin{bmatrix} Z'V^{-1}Y \\ r \end{bmatrix}$$

from which it follows that

$$\begin{bmatrix} a \\ -\lambda \end{bmatrix} = \begin{bmatrix} C^{-1} & R' \\ R & 0 \end{bmatrix}^{-1} \begin{bmatrix} Z'V^{-1}Y \\ r \end{bmatrix} \tag{31}$$

with C as defined in (29).

The inverse matrix in (31) is given by

$$\begin{bmatrix} C^{-1} & R' \\ R & 0 \end{bmatrix}^{-1} = \begin{bmatrix} C-CR'(RCR')^{-1}RC & CR'(RCR')^{-1} \\ (RCR')^{-1}RC & -(RCR')^{-1} \end{bmatrix} \tag{32}$$

After substitution of (32) in (31), the subestimator a according to (31) can be reduced to(27), using $\hat{\alpha} = CZ'V^{-1}Y$ from (6).

In our case, with V^{-1} and Z as given by (14) and (17) respectively, C reduces to

$$C = \Sigma \otimes (X'X)^{-1}$$

from (18). As the covariance matrix Σ is unknown, its elements are replaced by the estimates

$$s_{mm'} = \frac{e'_m e_{m'}}{K-N} \qquad (m,m'=2....M), \tag{33}$$

the $e'_m e_{m'}$ indicating the sums of squares and cross-products of residuals from the equation-by-equation least-squares application described in section 2.1.

The alternative procedure of deleting an explanatory variable from an equation whenever its regression coefficient has to be put equal to zero, has the advantage that the constraints have not to be introduced explicitly. However, this can only happen at the cost of the matrix Z taking no longer the form (17), since not all output categories will appear now as explaining variables in all equations.

This implies that the reduction of (15) and (16) to (20) and (21) respectively, does not hold any longer. The former have to be applied with

$$(34) \qquad Z = \begin{bmatrix} X_2 & & 0 \\ & \ddots & \\ 0 & & X_M \end{bmatrix},$$

the submatrices X_m (m=2...M) now being different from X in that the columns regarding outputs for which the commodity m cannot be an input factor, have been deleted. Wth Z as given in (34), the equations (15) and (16) result in the constrained estimator

$$(35) \qquad a = \begin{bmatrix} \sigma^{22}X_2'X_2 & \cdots & \sigma^{2M}X_2'X_M \\ \vdots & & \vdots \\ \sigma^{M2}X_M'X_2 & \cdots & \sigma^{MM}X_M'X_M \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{m=2}^{M} \sigma^{2m}X_2'Y_m \\ \vdots \\ \sum_{m=2}^{M} \sigma^{Mm}X_M'Y_m \end{bmatrix}$$

with covariance matrix

$$(36) \qquad \text{var } a = \begin{bmatrix} \sigma^{22}X_2'X_2 & \cdots & \sigma^{2M}X_2'X_M \\ \vdots & & \vdots \\ \sigma^{M2}X_M'X_2 & \cdots & \sigma^{MM}X_M'X_M \end{bmatrix}^{-1}$$

where the $\sigma^{mm'}$ (m,m'=2...M) indicate the elements of the inverse of (13)(so not of (10)!).

The results of (35) and (36) coincide exactly with those of (27) and (28) respectively, provided that the same estimate of the covariance matrix Σ will be used in both cases. Minor differences may occur when, in the last alternative, the elements of Σ have been estimated from equation-by-equation least-squares after deletion of the explanatory variables associated with zero coefficients.

It may be repeated in the end, that the procedure of (35) and (36) can be followed only when no constraints of the type (26) have to be introduced.

2.3 The quadratic programming case

When non-negativity restrictions are imposed, the solution of the estimating procedure has to be obtained by quadratic programming. For a full discussion of the methods used in these problems of maximizing (or minimizing) a quadratic function subject to linear inequality constraints, reference must be made to the specific literature. 15)

15) See e.g. J.C.G. Boot,[1].

In this section only the way is indicated in which the optimal solution is obtained according to the Theil-Van de Panne procedure 16), which has been followed in the example of the wooden furniture industry in chapter 4.

In our case the inequality restrictions take the form

$$(37) \qquad a^{*}_{mn} \geq 0 \qquad (m=1....M;\ n=1....N),$$

denoting by a^{*} the g.l.s. estimator satisfying not only the equality restrictions discussed in section 2.2 (if any), but also the non-negativity restrictions. The sign * will be dropped for intermediate results, satisfying only part of the inequality constraints or satisfying them without being optimal. (Of course the fulfilment of the equality restrictions of section 2.2 implies that the corresponding constraints of (37) are satisfied with the strict equality sign.)

The principles underlying the Theil-Van de Panne procedure are the following. When estimates have been obtained according to the methods described before, i.e. without regard to any inequality constraint, some of these constraints may be violated by the solution obtained. In that case, as Theil and Van de Panne show, at least one of these violated constraints will be binding (i.e. fulfilled in equality form) in the optimal solution. Theil and Van de Panne concentrate on the way in which the set of constraints binding in the optimal solution can be found, thereby reducing the programming problem to a maximization problem with only equality constraints such as discussed in section 2.2.

The procedure they proved to be leading to the optimal solution, if it exists, runs as follows. As a starting point take the solution obtained with due regard to equality constraints, if any, but without observance of any inequality constraint. So in our case the starting point is either the solution (20), viz when no regression coefficients have been put equal to zero beforehand, or the solution (27) (or (35)), when one or more of such equality constraints had to be imposed.

If this solution is feasible, i.e. satisfies all the inequality constraints too, it is the optimal solution. If, however, one or more of the inequality constraints are violated by this solution, which means in our case that one or more coefficients a_{mn} are found to be negative, further steps have to be made.

First, each of the violated constraints is successively taken as binding. This means in our case that the estimating procedure according to (27) or (35) is repeatedly applied, each time with one of the violated constraints taken in equality form. 17) So when the solution of the starting round includes a number of, say, 3 negative coefficients,

16) H. Theil and C. van de Panne, [7]. We used the description of their solution technique by J.C.G. Boot, [1], chapter 5.

17) It must be emphasized that the application of (35) can be repeated only as far as no coefficient of the first (deleted) equation has to be put equal to zero; see section 2.2.

for instance

$$a_{22} < 0$$
$$a_{34} < 0$$
$$a_{35} < 0 \quad ,$$

the procedure according to (27) or (35) has to be repeated three times in this step, each time taking one of the equalities $a_{22} = 0$ $a_{34} = 0$ and $a_{35} = 0$ as a constraint (in addition to the equality constraints which eventually have been imposed in the starting round for technological reasons already).

If still no feasible solution is obtained in this step, a next round of calculations according to (27) or (35) is necessary, every time with two inequality constraints taken in equality form, viz one found violated in the starting round, combined with one of those, violated in the former round when the first one was taken as binding. Continuing our example, say that with $a_{22}=0$ we find two negative coefficients, for instance

$$a_{24} < 0$$
$$a_{35} < 0 \quad .$$

Supposing that the other trials in this round, viz $a_{34}=0$ and $a_{35}=0$ do not result in a feasible solution either, the following combinations of two binding constraints have to be tried:

$$a_{22} = 0 \text{ and } a_{24} = 0 \qquad (1)$$
$$a_{22} = 0 \text{ and } a_{35} = 0 \qquad (2)$$

and accordingly with a_{34} respectively a_{35} taken as the first one in a number of combinations of two equality constraints.

If none of these combinations of two binding constraints gives a feasible solution either, a next round of calculations is necessary, each time with three inequality constraints taken as binding, and so on[18].

When, in a round, a feasible solution is obtained, this is the optimal solution provided that it is the only feasible solution in this round. When more than one feasible solution is obtained in a round, one of them is optimal and Theil and Van de Panne give the following rule to decide which one it is. A feasible solution is optimal if and only if for each inequality constraint h taken as binding in this solution holds that h is violated when it is deleted from the set of binding constraints.

Boot, considering as the criterion for decisions the signs of the Lagrange multipliers associated with the binding constraints, proves that a feasible solution is optimal if and only if the Lagrange multipliers associated with the inequality constraints taken as binding, are all

---

18) The fact that an inequality constraint taken as binding may be violated itself when one of the other constraints is taken as binding in its stead, reduces the number of calculations. In the example above e.g. a_{35} was found to be negative with a_{22} taken as zero. When it is also found that a_{22} is negative with a_{35} taken zero, which is one of the necessary trials in the example, this does not give rise to two, but to only one combination of two binding constraints in the next round.

positive[19].

In both our formulae (27) and (35) the Lagrange multipliers are suppressed and therefore the Theil- Van de Panne criterion will have to be applied when one of these formulae is used. Boot's criterion can be applied only when calculations are made according to the extended form (31).

The quadratic programming procedure of introducing an ever-increasing number of equality constraints until the optimal solution has been obtained, has different implications for the equations (27) (or (31)) and (35). When (27) or (31) is applied, the order of the square matrix to be inverted is rising by one in each following step. An inverse needed in one step can be obtained from the corresponding inverse of the preceding step by applying the general rule[20] that the inverse of a positive definite symmetric matrix E, bordered by a column vector e, a row vector e' and an additional element ε equals

$$(38)\quad \begin{bmatrix} E & e \\ e' & \varepsilon \end{bmatrix}^{-1} = \begin{bmatrix} E^{-1} + \dfrac{E^{-1}ee'E^{-1}}{\varepsilon - e'E^{-1}e} & -\dfrac{E^{-1}e}{\varepsilon - e'E^{-1}e} \\ -\dfrac{e'E^{-1}}{\varepsilon - e'E^{-1}e} & \dfrac{1}{\varepsilon - e'E^{-1}e} \end{bmatrix}$$

When applying (35), where the vector a consists of less elements according as the number of coefficients put equal to zero rises in each following step, the right-hand matrix (and vector) have correspondingly diminishing order. Here the inverse of the matrix of lower order can be obtained from the corresponding inverse of the preceding step by applying the following rule[20]. When the matrix E_{-d} is obtained by deleting row and column d from E , the elements of $(E_{-d})^{-1}$ can be expressed in the corresponding elements of E^{-1} as

$$(39)\qquad e^{ij}_{-d} = e^{ij} - \frac{e^{id}e^{dj}}{e^{dd}} \qquad (i,j \neq d)$$

Instead of writing (35) with matrix and vectors of diminishing order in successive rounds, one may as well maintain both vectors as in the starting round, giving on the other hand a suitable extension to the inverse matrix. This is accomplished by inserting in the inverse, computed for each trial, rows and columns of zero's in the positions corresponding to the a_{mn} which have been put equal to zero in addition to the situation of the starting round. The multiplication of these rows with the column vector results in zero as it should be, whereas the product of each other row and the column vector is not influenced by the extension. It can be verified easily that inverse matrices of

19) J.C.G. Boot, [1], section 5.4. Of course, the 'positivity' criterion holds for the particular way in which Boot introduced the restrictions and Lagrange multipliers. When our inequality restrictions are written in the form $Ra^{*} \geqslant r$ (this means that e.g. the constraint $\sum_{m \neq 1}^{M} a^{*}_{m2} \leqslant 1$ is expressed as $[0\text{-}10\ldots0\text{-}10\ldots]a^{*} \geqslant -1$), the minus sign before λ' in (30) ascertains that in our formulation too the positivity criterion holds. The criterion has no reference to the Lagrange multipliers associated with the equality restrictions of section 2.2, which may be positive or negative in the optimal solution.

20) See e.g. J.C.G. Boot, [1], pp. 19-20.

this extended form result automatically when, in successive rounds, the principle of rule (39) is applied unrestrictedly, i.e. inclusive of i=d and j=d.

Which of the formulae (27), (31) and (35) will be preferred is merely a question of computer programming. 21) Formula (27) has the advantage that the matrix (RCR') to be inverted has the lower order g x g, but the inversion is only part of the computations. The application of the formula (31) seems to be the most elegant way and offers the opportunity of using Boot's criterion for decisions on the optimal solution. The matrix to be inverted has much higher order here, but the inversion in the starting round can be accomplished by applying equation (32). In formula (35) the order of the matrix to be inverted is lower than in (31) and the inversion in the starting round may be simplified by a suitable partition. Furthermore the application of rule (39) in subsequent steps seems to be easier in operation than rule (38), which has to be applied when (27) or (31) are used. A disadvantage of applying (35) is the eventuality that in one or more of the trials it will be necessary to impose constraints of the type (26), which have to be introduced explicitly along the same lines as in (27) or (31).

In our example of the wooden furniture industry we applied formula (35). To reduce the probability of being necessitated to introduce constraints of the type (26), we chose the input category 'value added' as the input number 1 to be excluded from the application of (35) and were successful in finding ultimately none of the a^{*}_{1n} (n=1...N) to be zero.

3. Limitations and special features

The linear form of the equations (1) is implied by the assumption of a cost structure for each kind of output which is the same for each firm. In view of the practical use of the results the advantages of this assumption are clear: the division by output category of total inputs of the regarded firms (whether these constitute a whole industry or a sample of it) can be estimated simply by applying the estimated cost percentages to the total value of output by product. In case of a sample, expansion of the detailed input estimates to the industry as a whole does not find more difficulties than does the raising of total output itself.

Of course, however, these advantages can be enjoyed only when the underlying assumption is close to reality. In this chapter some problems associated with the assumption of equal cost structures will be discussed and illustrated from the experience gained up till now by the Netherlands Central Bureau of Statistics.

---

21) As indicated in the beginning of section 2.3, the Theil-Van de Panne procedure is only one of the techniques available for solving quadratic programming problems. Other algorithms, e.g. those based on the Simplex technique, have been programmed for computers too. Reference must be made to Boot [1].

3.1 Differences in cost structure between sub-populations

It may happen that cost structures differ systematically according to one or more characteristics of the firms in an industry. To allow for such differences without giving up the advantages of the linear system, stratification by these aspects might be applied in order to arrive at sub-populations with homogeneous cost structures for each output category. In all those cases the methods of chapter 2 should be applied separately to each of the sub-populations distinguished. Aggregation to the national level can, for each output category, be achieved by adding up the estimated total values of each input category in all sub-populations.

It should be stressed that estimates of the national average cost structure for each output category cannot be obtained in such a case by applying regression techniques directly to the data of (a proportionally stratified sample of) all the firms in the industry. The regression coefficients would be 'averages' of the coefficients in the particular strata indeed, but essentially different from the concept which is needed here[22].

In our calculations for agriculture[23] we had a distinction by type and size of farms. Four types were distinguished, viz. arable farms, pasture farms, mixed holdings on sandy soils and mixed holdings outside sandy soils. (The sample for 1965/'66 consisted of 6,142 farms having a size of at least 4 ha.) For arable farms the estimated cost percentages for one of the major output categories, viz. cereals, were as follows (standard errors in brackets):

Arable farms Size class (ha)	4-< 15		15-<30		30-<50		50-<100	
	cost percentages for cereals:							
Seeds	8.7	(1.3)	11.5	(1.1)	11.6	(1.0)	10.1	(1.4)
Fertilizers	21.1	(2.2)	19.6	(1.5)	18.7	(1.2)	17.6	(2.0)
Other material inputs	15.4	(7.4)	11.3	(3.3)	17.9	(3.5)	18.1	(3.1)
Contract work	27.7	(6.6)	19.5	(3.9)	12.6	(2.6)	12.0	(4.4)
Value added	24.7	(15.9)	39.5	(4.6)	41.8	(3.6)	42.3	(6.8)
T o t a l	97.6		101.4		102.6		100.2	

These figures suggest that the relative importance of contract work and fertilizers diminishes and the value added percentage rises with an increase of the agricultural area.

An example of differences by type of farm is formed by the cost structures of the output category 'dairy farming'. For mixed holdings on sandy soils and pasture farms the estimates were as follows:

Mixed holdings on sandy soils Size class (ha)	4-<7		7-<15		15-<30		30-<100	
	cost percentages for 'dairy farming'							
Feeding-stuffs	32.8	(0.8)	32.6	(0.3)	32.2	(0.3)	31.4	(0.8)
Fertilizers	6.1	(0.4)	7.3	(0.4)	8.3	(0.5)	11.0	(0.7)
Seeds	1.1	(0.2)	0.4	(0.2)	0.1	(0.3)	-0.1	(0.4)
Other material inputs	3.2	(2.2)	7.6	(1.6)	11.8	(1.8)	9.6	(1.7)
Contract work	2.0	(1.0)	3.2	(0.5)	2.8	(0.5)	5.4	(0.8)
Value added	55.9	(7.8)	47.3	(3.0)	44.4	(2.0)	39.9	(2.2)
T o t a l	101.0		98.2		99.5		97.3	

22) See for discussions on this point e.g. H. Theil, [6].

23) Our experience on agriculture relates to the application of ordinary equation-by-equation least squares only, though some coefficients have been put equal to zero beforehand for technological reasons. This implies that the estimates of the cost percentages need not total up to 100 exactly here.

Size class (ha)	Pasture farms 4-<7		7-<15		15-<30		30-<100	
	cost percentages for 'dairy farming'							
Feeding-stuffs	34.6	(1.1)	30.9	(0.6)	27.4	(0.5)	23.5	(0.8)
Fertilizers	6.2	(0.3)	6.5	(0.2)	6.8	(0.2)	6.6	(0.3)
Seeds	0.30	(0.06)	0.05	(0.05)	0.10	(0.02)	0.07	(0.03)
Other material inputs	11.3	(1.7)	13.0	(0.3)	12.4	(0.2)	12.5	(0.3)
Contract work	1.6	(0.3)	2.4	(0.1)	2.2	(0.1)	2.5	(0.2)
Value added	48.3	(2.5)	50.8	(0.7)	53.8	(0.5)	52.2	(0.9)
T o t a l	102.2		103.6		102.7		100.5	

In the middle-size class 15-<30 ha the input percentages for feeding-stuffs and fertilizers are higher on mixed holdings than on pasture farms, at the cost of a lower value added percentage.
Moreover, the percentage of feeding-stuffs is falling on mixed holdings more slowly and the input of fertilizers rising more sharply with an increase of size than on pasture farms. Complementary the value added percentage is falling on mixed holdings as opposed to a rise on pasture farms, when size increases.

It is noteworthy that in our applications of regression analysis to agricultural data the firms were arranged in a regional order. In this way the Durbin-Watson test for positively autocorrelated disturbances might throw light on regional differences in cost structure. A striking example is the regression of feeding-stuffs input on 'dairy farming' output for pasture farms[24]. By size class of farms the resulting cost percentages were, as mentioned before:

4-< 7 ha	34.6	(1.1)
7-<15 ha	30.9	(0.6)
15-< 30 ha	27.4	(0.5)
30-<100 ha	23.5	(0.8)

The low values of the "d"-statistic, found in some of these cases, gave rise to a distinction between two regional strata, brought about with the help of agricultural experts. (Of course a distinction between regions should find a base in real differences in cost structure, caused by quality of soil, market conditions or other factors.) Estimates of the feeding-stuffs percentages for each of these regions separately were as follows:

	region 1		region 2	
4-< 7 ha	38.0	(1.0)	30.0	(1.6)
7-< 15 ha	34.7	(0.4)	24.6	(0.5)
15-< 30 ha	33.5	(0.4)	23.4	(0.3)
30-<100 ha	33.9	(2.3)	22.3	(0.4)

As our calculations had a preliminary character, we did not come to separate regressions by region for the other input categories in this case.

The assumption of a cost structure for each kind of output which is the same for all firms may be less realistic as regards the division of value added between (roughly) 'compensation to employees' and 'operating surplus'. Possibilities to substitute capital for labour, and differences in profit-earning capacity between firms raise problems for which it remains to be seen whether they can be overcome by a suitable stratification. As yet our experience is restricted to the handling of value added as one input category.

24) The input of feeding-stuffs used for 'dairy farming' was separately known for each pasture farm, and consequently the output of 'dairy farming' was the only explanatory variable in this case.

3.2 The problem of vertical integration

When the degree of vertical integration differs between firms, the cost structure for a final output category will also be different. In several ways the application of the methods of chapter 2 can be secured in those cases.

A simple device is to make separate estimates for sub-populations by degree of vertical integration, if such a classification can be brought about from the available data. So in our analysis of the manufacture of wooden furniture we might have estimated cost structures separately for firms using (besides veneers, plywood and the like) roundwood as their main input of timber and firms using mainly sawn wood.

According to another device the whole production process is subdivided in a number of successive activities, the output of each constituting (part of) the input of one or more of the following stages, or being sold. In order that the methods of chapter 2 will be applicable in this case, output data of each of the distinguished stages must be known for each firm, which will not be the case very often. These data are necessary because intermediate outputs have to be comprised in the inputs as well as in het outputs. An example is the agricultural production on mixed holdings, where feeding-stuffs for cattle can be produced on the own farm or purchased from outside. In our calculations the value of own-produced feeding-stuffs was included in the outputs as well as, together with purchased feeding-stuffs, in the inputs. For the manufacture of wooden furniture this method would imply that sawing timber be distinguished as a separate activity, which could not be performed as separate data on the output of this activity were not known.

This difficulty might have been overcome by still another method which assumes a 'commodity-technology'. This assumption implies that the cost structure for working up round wood into sawn wood by the wooden furniture industry equals the cost structure for these activities by proper sawing-mills. When the latter is known, the inputs for these sawing-mill activities by a furniture firm can be estimated from its input of round wood. After deducting, for each firm, these inputs from total inputs by category and putting the output of sawing-mill activities (as far as used as an input for furniture making) in its stead, the resulting data represent the inputs needed for the proper furniture making. So these data can be combined with the data of firms having only sawn wood as their input of timber, to estimate the cost structures of the activities of working up sawn wood into the distinguished kinds of furniture.

In our experimental calculations for wooden furniture we did not apply this last method either, but followed a way which can be seen as a partial application of the first method. From our calculations we excluded eight firms employing mainly roundwood and also three firms whose input of timber consisted of further processed wooden parts of furniture. So the results, reported on in chapter 4, only refer to the cost structures of firms working up sawn wood into furniture.

3.3 Joint products

The production of two or more of the output items of a heterogeneous industry may be related to each other. In those cases the methods described in this article cannot be of help in estimating the cost structure for each of these related products separately. Several situations can be distinguished.

In case of a very close relationship as in the often cited example of coke and gas, the output-ratio is (almost) the same for each firm involving that, because of multicollinearity, regression coefficients for each of these outputs separately cannot be obtained at all or will be very unreliable.

Therefore, coke and gas should not be handled as separate outputs, but the commodity group of coke and gas should be seen as the output of the coke making process. When this process is only one of the activities of the regarded industry, the cost structure of this process can be estimated by taking the output of the commodity group coke and gas as one of the explanatory variables in the methods of chapter 2. When, in the input-output table to be constructed, coke and gas should yet have to be separated, the necessary transfer of inputs and outputs has to be effected by making complementary assumptions.

In cases as in the livestock industry, where the relationship between the output items milk and meat is less closely, the output-ratio of these products may be different between firms indeed, but the assumptions underlying the methods of chapter 2 cannot be fulfilled then.
For differences in the ratio of the outputs can be brought about only by differences in the ratios of the inputs, implying different cost structures contrary to the assumptions of chapter 2. When, however, it is expected that each firm aims at the same ratio of the outputs (e.g. because the prevailing market conditions are the same for each firm), the situation turns into the case of coke and gas, enabling the estimation of the cost structure of the process as a whole[25]. In other cases sub-divisions by size, region or other criteria may perhaps be helpful in obtaining strata, in each of which the output-ratio of the process is approximately the same for all firms. In our calculations for dairy farming the additional regional stratification as mentioned in section 3.2 has partly been suggested by this aspect.

3.4 Distribution of imported inputs over output categories.

In "A System of National Accounts" a distinction has been made between competitive and complementary commodities. Competitive commodities are those for which there is a domestic industry and which may, therefore, be either home produced or imported. Complementary commodities can be obtained only from imports[26].

The methods proposed in this study, may be helpful for estimating the distribution of complementary inputs of a heterogeneous industry over separate output categories. For the imported part of competitive inputs separately this opportunity is lacking, because in the application of these methods the imported part and the home produced part of a competitive input cannot be taken as separate input categories. A practical reason is that, as a rule, the available data on the inputs of individual firms are distinguished by kind but not by origin (imported or home produced). However, the theoretical assumptions underlying these methods prevent from doing so even when the proportion of imports in a competitive input would be known for each firm. For the assumption of equal cost-structures for all firms would include in this case that for each output separately the share which imports constitute from a competitive input category, should be the same for all firms. It is very difficult to find arguments in favour of such an assumption. Only for the total of imports and domestic production of a competitive input category the assumption may be realistic.

25) In this case the circumstance that the ratio of the outputs of the process may change in the future, entails an additional problem in prediction by input-output techniques.

26) United Nations, [9], paragraphs 3.5 and 3.32.

4. A complete example: results obtained for the wooden furniture industry.

As mentioned in the Introduction, our calculations for the wooden furniture industry have been based on data taken from the census of production 1968. Though the census included all firms employing at least 10 persons and producing mainly wooden furniture, only the firms employing 50 or more persons furnished detailed data on inputs and outputs and could therefore be taken in account. Within the population of these larger firms no further sub-divisions have been introduced.

In defining the input- and output-categories to be distinguished in the analysis we were bound by the limits set by the questionnaire of the census. Seven output categories were distinguished, viz.

cupboards, sideboards, book-cases, etc.;
chairs and other seats;
bedroom furniture;
fancy tables, hat-racks, umbrella-stands and other small furniture;
other furniture (including component parts of wooden furniture);
other products;
work done on commission.

After exclusion of the eight firms employing mainly roundwood (see section 3.2), six input categories were left, viz.

sawn wood;
veneers, panelwood, chipboard, flaxboard, plywood, etc.;
component parts of wooden furniture;
other materials;
fuel;
value added[27] and payments for work done on materials given out.

Value added and payments for work done were combined in order to place more reliance on the assumption of equal cost structures for all firms. After further exclusion of three firms whose input of timber only consisted of component parts of wooden furniture, and one firm mainly producing a special kind of furniture viz. laboratory furniture, the assumption was considered fulfilled for the specification of inputs and outputs as mentioned above.

Table 1 presents the results obtained by the ordinary least-squares method, applied to the data of the remaining 73 firms. Each row shows the estimates of the regression coefficients obtained by taking a separate kind of input as the dependent variable. All estimates have been multiplied by 100 in order that the figures can be interpreted as cost percentages (instead of perunages).

As four of the input categories do not constitute cost factors for the output category 'work done on commission', this output had to be excluded from the set of explanatory variables in these four equations. As a consequence the property that the estimates of the cost percentages per product group should exactly total up to 100 (see section 2.1), is not necessarily fulfilled in this case. Indeed there are deviations from 100, especially for the output category 'work done on commission'.

---

27) Value added comprises all factors included in what is usually described as 'net output according to the census of production'.

T a b l e 1

Cost percentages per product group of wooden furniture industry in the Netherlands, 1968, resulting from equation-by-equation least squares (standard errors in brackets).

Input category	Product group: Cupboards, side-boards, book-cases, etc.	Chairs and other seats	Bedroom furniture	Fancy tables hat-racks and other small furniture	Other furniture incl. component parts of wooden furniture	Other products	Work done on commission
Sawn wood	4.39 (0.98)	4.77 (0.76)	5.49 (2.39)	8.51 (4.12)	4.69 (1.26)	-4.50 (9.75)	
Veneers, panelwood, chipboard, flaxboard, plywood, etc.	15.35 (1.31)	-0.09 (1.01)	14.72 (3.18)	9.44 (5.52)	14.22 (1.67)	-28.88 (12.96)	
Component parts of wooden furniture	5.86 (1.40)	3.06 (1.09)	-0.17 (3.41)	11.96 (5.92)	1.00 (1.79)	5.12 (13.90)	
Other materials	7.89 (2.70)	35.86 (2.09)	16.12 (6.54)	-4.23 (11.37)	20.02 (3.44)	29.80 (26.68)	
Fuel	1.68 (0.12)	0.92 (0.10)	1.75 (0.30)	1.28 (0.52)	1.08 (0.17)	-0.08 (1.24)	2.74 (2.33)
Value added and payments for work done on materials given out	64.95 (2.12)	55.42 (1.64)	62.02 (5.13)	73.38 (8.93)	57.68 (2.97)	100.72 (21.04)	139.35 (39.68)
T o t a l	100.12	99.94	99.93	100.34	98.69	102.18	142.09

T a b l e 2

Idem, estimated by joint generalized least squares, without non-negativity restrictions (standard errors in brackets)

Input category	Product group						
	Cupboards, etc.	Chairs and other seats	Bedroom furniture	Fancy tables, etc.	Other furniture	Other products	Work done on commission
Sawn wood	4.39 (0,99)	4.77 (0.76)	5.49 (2.39)	8.51 (4.16)	4.69 (1.26)	-4.50 (9.75)	
Veneers, etc.	15.35 (1.30)	-0.09 (1.01)	14.72 (3.15)	9.44 (5.48)	14.22 (1.66)	-28.88 (12.87)	
Component parts	5.86 (1.40)	3.06 (1.09)	-0.17 (3.41)	11.96 (5.92)	1.00 (1.79)	5.12 (13.90)	
Other materials	7.89 (2.67)	35.86 (2.07)	16.12 (6.48)	-4.23 (11.27)	20.02 (3.40)	29.80 (26.44)	
Fuel	1.68 (0.12)	0.92 (0.10)	1.75 (0.30)	1.28 (0.52)	1.08 (0.17)	-0.08 (1.23)	2.75 (2.19)
Value added, etc.	64.83 (2.11)	55.48 (1.64)	62.09 (5.12)	73.04 (8.91)	58.99 (2.69)	98.54 (20.91)	97.25 (2.19)
T o t a l	100	100	100	100	100	100	100

Table 3

Idem, estimated by joint generalized least squares, subject to non-negativity restrictions

Input category	Product group						
	Cupboards, etc.	Chairs and other seats	Bedroom furniture	Fancy tables, etc.	Other furniture	Other products	Work done on commission
Sawn wood	4.44	4.79	5.52	7.48	4.51	0	
Veneers, etc.	15.52	0	14.76	8.64	12.92	0	
Component parts	5.87	3.08	0	11.15	1.09	3.46	
Other materials	7.67	35.72	16.00	0	20.70	12.92	
Fuel	1.69	0.92	1.74	1.28	1.07	0.18	2.75
Value added, etc.	64.81	55.49	61.98	71.45	59.71	83.44	97.25
Total	100	100	100	100	100	100	100

Another feature is that no less than six of the estimates have got a negative value, though the standard errors of those estimates are relatively very high.

The residuals of the least squares applications of Table 1 have been used to estimate the covariances of the disturbances according to formula (33). (The denominator K-N has been put invariably equal to 73-6 = 67.) The resulting covariance matrix has been employed in the applications of joint generalized least squares.

In Table 2 the results are presented of the joint g.l.s. method (formulae (35) and (36)) without imposing non-negativity-restrictions. Besides the fact that this procedure ascertains the sum total of the estimates to be equal to 100, the only important difference as compared with Table 1 is the lower value added percentage for 'work done on commission'. The number of negative coefficients remains 6. As a matter of fact the percentage estimates on the first four rows are exactly the same in both tables. This is a consequence of the circumstance that in this particular case one and the same explanatory variable has been excluded from these four equations.

The results of Table 2 formed the starting point for the application of quadratic programming as described in section 2.3. The ultimate results, in which all negative coefficients have vanished, are presented in Table 3. Standard errors could not be given in this case because the unequality constrained estimator is biased and it is very difficult to obtain its sampling distribution[28].

Comparison with the former results shows that 5 coefficients having a negative sign in Table 2, have turned into zero values; one has got a positive sign.

The results of Table 3 don't look unrealistic. E.g. the high percentage of 'veneers, etc.' in the production of 'cupboards, etc.' as opposed to the low percentage for the product group 'fancy tables, etc.' and a zero percentage for 'chairs and other seats' don't disagree with what might be expected. The importance of 'other materials' for the production of chairs is to be explained from the fact that this input category includes fabrics and the like for upholstering. The low share of wooden materials in the inputs for 'other products' might indicate that this product group has little to do with proper furniture making. In Tables 1 and 2, however, the estimates for this product group all have high standard errors, which might raise the question whether this product group indeed has a uniform cost structure for all firms.

In conclusion it may be stated that the results obtained hitherto demonstrate the usefulness of the methods described. Ar the same time, however, the examples underline that the methods should not be applied automatically in each case. The fact that in the example of the wooden furniture industry a number of firms had to be excluded from the calculations, demonstrates the necessity of looking after the fulfilment of the assumptions underlying the methods. In each case the classification of inputs and outputs and(eventual) stratification of firms by size, region, etc. s should be brought about in such a way as to ascertain the assumptions to be approximately fulfilled. Attempts in this field might reveal necessary adaptations in the classifications as they have been used up till then in the collection of data.

---

28) See e.g. G.G. Judge and T. Takayama, [3], and M.C. Lovell and E. Prescott,[4].

References

[1.] J.C.G. Boot, Quadratic Programming; Algorithms-Anomalies-Applications, North-Holland Publishing Company, Amsterdam 1964.

[2.] J. Johnston, Econometric Methods, Mc Graw-Hill, New York, second edition, 1972.

[3.] G.G. Judge and T. Takayama, Inequality restrictions in regression analysis, Journal of the American Statistical Association, volume 61, 1966, pp. 166-181.

[4.] M.C. Lovell and E. Prescott, Multiple regression with inequality constraints: pretesting bias, hypothesis testing and efficiency, Journal of the American Statistical Association, volume 65, 1970, pp. 913-925

[5.] G.J.A. Mensink, Comparisons of labour productivity in the United Kingdom and the Netherlands 1958, Statistical Studies no. 18, Netherlands Central Bureau of Statistics, The Hague, 1966.

[6.] H. Theil, Linear Aggregation of Economic Relations, North-Holland Publishing Company, Amsterdam, 1954.

[7.] H. Theil and C. van de Panne, Quadratic programming as an extension of conventional quadratic maximization, Management Science, volume 7, 1960, pp. 1-20.

[8.] H. Theil, Principles of Econometrics, North-Holland Publishing Company, Amsterdam, London, 1971.

[9.] United Nations, A System of National Accounts, Studies in Methods, Series F, No. 2, Rev. 3, New York, 1968.

[10.] United Nations, International Standard Industrial Classification of all Economic Activities, Statistical Papers, Series M, No.4, Rev. 2, New York, 1968.