ERASMUS UNIVERSITY, ROTTERDAM

Netherlands School of Economics

ECONOMETRIC INSTITUTE

Report 7311/E

A NEW CLASS OF LIMITED INFORMATION ESTIMATORS FOR SIMULTANEOUS EQUATION SYSTEMS

by W.J. Keller

July 4, 1973

Preliminary and Confidential

A NEW CLASS OF LIMITED INFORMATION ESTIMATORS

FOR SIMULTANEOUS EQUATION SYSTEMS

by W.J. Keller

CONTENTS

SUMMARY

Limited information maximum likelihood estimators of structural coefficients in a system of linear equations are derived for linear restrictions imposed on these coefficients and when the covariance matrix of the contemporaneous reduced form disturbances is known and possible singular. In the particular case of zero-nonzero restrictions this method results in the so-called 'Ω-class estimators'. If the index matrix of the Ω-class estimator coincides with the true covariance matrix of contemporaneous disturbances and if the latter are assumed to be normally distributed, the Ω-class estimator will correspond to the maximum likelihood estimator. Both assumptions are superfluous, however, for proving that the Ω-class estimator is consistent. The LIML and the TSLS estimators are shown to be members of the Ω-class, while two other Ω-class estimators are introduced, viz. the LIMLD and the OTSLS estimator, both with a diagonal index matrix. In small samples, the LIML estimator, then being based on a poor estimate of the covariance matrix, may be replaced by the LIMLD estimator. TSLS estimators are criticized on the grounds of being inconsistent with the stochastic nature of the endogenous variables and the arbitrariness of their choice. Still in large samples, the consistency and computational convenience could make the TSLS estimator attractive. Finally, it is shown that different TSLS estimators, resulting from different normalizations, can be considered as the limits of certain Ω-class estimators, under general conditions.

1. INTRODUCTION

The structural equation system of G equations is a mathematical representation of a theory which relates the G jointly endogenous variables y_{tg} (g = 1,..,G) for the T observations t = 1,..,T, to the K exogenous variables x_{tk} (t = 1,.., T; k = 1,..,K). The equations are assumed to be linear and stochastic. In matrix notation :

(1.1) $$Y\Gamma + XB = E$$

where Y is a TxG matrix of endogenous variables,
X a TxK matrix of exogenous variables,
E a TxG matrix of disturbances,
Γ a GxG matrix of unknown coefficients,
B a KxG matrix of unknown coefficients.

The theory includes restrictions on the coefficients in Γ and B as a means of distinguishing the equations from each other.

The purpose of this paper is to present methods of estimating the coefficients of any single equation of system (1.1), taking account of the restrictions imposed on the coefficients of that particular equation. This is called 'Limited Information estimation' (see KOOPMANS (1950)). Without loss of generality we may assume that the first equation is the one in which we are interested. This equation is written as follows :

(1.2) $$Y\gamma + X\beta = \varepsilon$$

where $\gamma = (\gamma_{11}, \gamma_{21}, .., \gamma_{G1})'$ a G vector
$\beta = (\beta_{11}, \beta_{21}, .., \beta_{K1})'$ a K vector
$\varepsilon = (\varepsilon_{11}, \varepsilon_{21}, .., \varepsilon_{T1})'$ a T vector

representing the first column of the matrices Γ, B and E respectively.

We make the following assumptions.

ASSUMPTION 1. The T values taken by the G jointly dependent variables $y_{t1}, .., y_{tG}$ are the realizations of a TxG stochastic matrix Y which satisfies (1.1), where X is a TxK matrix of exogenous variables, E a TxG matrix of disturbances, Γ and B unknown coefficient-matrices of order GxG and KxG, respectively, with Γ assumed to be nonsingular.

ASSUMPTION 2. The T rows of the TxG matrix E of disturbances are independent random drawings from a G-dimensional population with an expectation zero and unknown but finite covariance matrix Σ.

ASSUMPTION 3. The TxK matrix X of the values taken by the exogenous variables has rank K and consists of nonstochastic elements.

For the restrictions we make a distinction between the so-called 'normalization rule' or normalization restriction and the other restrictions.

ASSUMPTION 4. The normalization restriction for the equation (1.2) is

$$e_i'\gamma = 1 \tag{1.3}$$

where e_i is a Gx1 unit-vector consisting of a single 1 in the i-th position and zeroes everywhere else. This implies that γ_{i1} is taken equal to 1.

The other restrictions are assumed to be formulated in one of the following alternative modes, stated in assumption 5a and 5b, respectively.

ASSUMPTION 5a. The coefficient vectors γ and β satisfy the (homogeneous) linear restrictions

$$Q\gamma + R\beta = 0 \tag{1.4}$$

with known matrices Q and R of order sxG and sxK, respectively, where the rank of the partitioned matrix (Q,R) equals s, the restriction being linearly independent.

ASSUMPTION 5b. The coefficient vectors γ and β satisfy the zero-nonzero restrictions

$$\begin{aligned} \gamma &= (\gamma^{+\prime}, \gamma^{o\prime})' \quad \text{with} \quad \gamma^o = 0 \\ \beta &= (\beta^{+\prime}, \beta^{o\prime})' \quad \text{with} \quad \beta^o = 0 \end{aligned} \tag{1.5}$$

possibly after a rearrangement, where γ^+ and β^+ are G^+ and K^+ vectors, and γ^o and β^o are G^o and K^o vectors, respectively, so that

$$G = G^+ + G^o \quad \text{and} \quad K = K^+ + K^o$$

Note that zero-nonzero restrictions are a special case of linear restrictions.

Restrictions on the coefficients of the equation are required for the identification of the equation. We refer to MALINVAUD (1970), Chapter 18 for an extensive discussion of the identification problem. The following theorem is stated and proved in MALINVAUD (1970) p. 658.

THEOREM. *Under the assumptions* 1, 2, 3, 4 *and* 5a, *the equation* (1.2) *is identified if*

(1.6) $$\text{rank } (\Gamma'Q' + B'R') = G - 1$$

A necessary condition for identification of equation (1.2) *is*

(1.7) $$s \geq G - 1$$

Analogous results can be formulated for the case of zero-nonzero restrictions (assumption 5b).

ASSUMPTION 6. *The equation in question is identified by the restrictions on the coefficients, formulated in assumption* 4 *and* 5a, *or* 4 *and* 5b.

An identified equation is called *exactly identified* if

(1.8) $$s = G - 1$$

and *overidentified* if

(1.9) $$s > G - 1$$

for the case of linear restrictions (assumption 5a). For the case of zero-nonzero restrictions (assumption 5b), s in (1.8) and (1.9) has to be replaced by $K^o + G^o$, the number of restrictions in that case.

Since Γ is nonsingular, we can solve for the jointly dependent variables Y by postmultiplying both sides of (1.1) by Γ^{-1}.

(1.10) $$Y = -XB\Gamma^{-1} + E\Gamma^{-1}$$

or

(1.11) $$Y = X\Pi + U$$

with

(1.12) $$\Pi = -B\Gamma^{-1} \quad \text{a KxG matrix of coefficients}$$
$$U = E\Gamma^{-1} \quad \text{a TxG matrix of disturbances}$$

(1.11) is called the *reduced form* of system (1.1). From assumption 2 we find for the reduced form disturbances matrix U

(1.13) $$\mathcal{E}(U) = 0 \quad \text{(a zero matrix)}$$
$$C(U) = \Gamma^{-1}{}'\Sigma\Gamma^{-1} \otimes I = \Omega \otimes I \quad \text{(say)}$$

where $\mathcal{E}(A)$ is the matrix of expectations of the elements of matrix A, C(A) is the covariance matrix of a stochastic matrix A[1] and $\otimes$ the so-called Kronecker- or direct product of matrices[2]. Some properties of the vec-operator and the Kronecker product are mentioned in Appendix A.2. The GxG matrix Ω is the covariance matrix of contemporaneous reduced form disturbances. In the following we do not exclude singularity of Ω; Ω is assumed to be positive semidefinite, that is $x'\Omega x \geq 0$ for every x. The covariance matrix Ω will be singular if the structural system (1.1) contains one or more non-stochastic equations; e.g. the j-th equation :

(1.14) $$Y\gamma_j + X\beta_j = 0 \quad \text{where } \gamma_j = \Gamma e_j \text{ and } \beta_j = Be_j$$

The reduced form coefficient matrix Π is related to the structural coefficients γ and β, by (see (1.12))

(1.15) $$\Pi\gamma = -\beta$$

Without proof we now state the following result : in the exactly identified case, for each arbitrary KxG matrix A there are unique vectors γ and β, satisfying the restrictions imposed on them, for which [3]

(1.16) $$A\gamma = -\beta$$

In the overidentified case we have to take care of the restrictions on Π resulting from the restrictions on γ and β. Then, in general, it is not possible to find vectors γ and β which satisfy the restrictions, for which (1.16) holds good for each arbitrary KxG matrix A unless we restrict the choice of A.

---

1) $C(A) = \mathcal{E}(\text{vec } A - \mathcal{E}\text{ vec } A)(\text{vec } A - \mathcal{E}\text{ vec } A)'$ where $\text{vec } A = (a_{11},.., a_{n1}, a_{12},.., a_{n2}, a_{13},..,...... a_{nm})$ if A is of order nxm.

2) The Kronecker product of an nxm matrix A and a matrix B is defined as

$$A \otimes B = \begin{pmatrix} a_{11}B & . & . & a_{1m}B \\ . & & & . \\ . & & & . \\ a_{n1}B & . & . & a_{nm}B \end{pmatrix}$$

3) MALINVAUD (1970) p.652, shows exceptional cases in which this result does not hold good.

2. MAXIMUM-LIKELIHOOD ESTIMATORS AS A FUNCTION OF Ω

We shall now develop maximum-likelihood (m.l.) estimators of the reduced form coefficient matrix Π, and the structural coefficients γ and β. We shall not derive m.l. estimators of the reduced form covariance matrix Ω; instead we formulate the estimators of Π, γ and β as functions of the unknown matrix Ω. Π, γ and β are estimated in two steps : first, Π is estimated as a function of γ and β, after which γ and β are estimated in a second step.

Our starting point is the reduced form (1.11)

$$Y = X\Pi + U$$
$$\mathcal{E}(U) = 0$$
$$C(U) = \Omega \otimes I$$

The GxG matrix Ω is symmetric and positive semidefinite. The rank of Ω is denoted as r ($\leq G$). Now we define

D an rxr diagonal matrix; $d_{11}, \ldots, d_{rr}$ are the r positive characteristic roots of Ω,

F a Gxr matrix whose columns are the characteristic vectors of Ω corresponding to the positive roots $d_{11}, \ldots, d_{rr}$,

Z a Gx(G-r) matrix whose columns are the characteristic vectors of Ω corresponding to the zero roots.

We assume orthonormality of (F,Z) :

(2.1) $$(F,Z)(F,Z)' = I, \text{ hence } FF' + ZZ' = I$$

(2.2) $$(F,Z)'(F,Z) = I, \text{ hence } F'F = I, Z'Z = I, F'Z = 0$$

The definitions of F,Z and D imply :

(2.3) $$\Omega F = FD$$

(2.4) $$\Omega Z = 0$$

so

(2.5) $$\Omega = FDF'$$

(2.6) $$D = F'\Omega F$$

The pseudo-inverse of Ω (also called the Moore-Penrose inverse)[4] is

(2.7) $$\Omega^P = FD^{-1}F'$$

hence

(2.8) $$\Omega^P Z = 0$$

---

4) See appendix A.1

We now separate the reduced form (1.11) into a stochastic and a nonstochastic part. Postmultiplying (1.11) by the nonsingular matrix (F,Z) :

$$Y(F,Z) = X\Pi(F,Z) + U(F,Z),$$

we obtain $YF = X\Pi F + UF$

and $YZ = X\Pi Z + UZ$

for which :

(2.9) $\mathcal{E}(UF) = \mathcal{E}(UZ) = 0$

(2.10) $C(UF) = \mathcal{E}(\text{vec } UF)(\text{vec } UF)' = (F'\otimes I)(\Omega \otimes I)(F \otimes I) = (D \otimes I)$

(2.11) $C(UZ) = \mathcal{E}(\text{vec } UZ)(\text{vec } UZ)' = (Z'\Omega Z \otimes I) = 0$

where (2.10) and (2.11) are based on properties of the 'vec' operator and the Kronecker product (see appendix A.2).Consequently, an equivalent reduced form representation is

(2.12) $YF = X\Pi F + UF$ with $\mathcal{E}(UF) = 0$ and $C(UF) = D \otimes I$

and

(2.13) $YZ = X\Pi Z$ with probability 1

Note that (2.13) serves as a restriction on Π.

Maximum-likelihood estimation requires specification of the disturbance distribution. Instead of assumptions on the distribution of U, we specify an r-dimensional Normal distribution for the rows of UF, the latter having a nonsingular covariance matrix.

ASSUMPTION 7. The T rows of the Txr matrix UF of disturbances are independent random drawings from an r-dimensional normally distributed population, with zero expectation and an unknown but nonsingular covariance matrix D.

Under assumption 7, the log-likelihood function of Π, given Ω, Y and X, can easily be found to be

$$L(\Pi \mid \Omega,Y,X) = \text{constant} - \frac{1}{2}\,\text{tr}((YF - X\Pi F)D^{-1}(YF - X\Pi F)')$$

or

(2.14) $$L(\Pi \mid \Omega,Y,X) = \text{constant} - \frac{1}{2}\,\text{tr}((Y - X\Pi)\Omega^{P}(Y - X\Pi)')$$

where 'tr' stands for the trace operator.

The restrictions on Π are

$$Q\gamma - R\Pi\gamma = 0 \tag{2.15}$$

$$YZ - X\Pi Z = 0 \tag{2.16}$$

where (2.15) is based on (1.4) and (1.15), and (2.16) is based on (2.13). Note that (2.15) will be not effective in the exactly identified case, i.e. in that case we do not restrict the choice of Π by (2.15). We will, however, treat the exactly identified case and the overidentified case simultaneously.

First we develop m.l. estimators of Π as a function of Ω, γ and β, by maximizing the log-likelihood subject to the restrictions. The Lagrangian function is

$$L_1 = tr((Y - X\Pi)\Omega^p(Y - X\Pi)') - 2\,\alpha'(Q\gamma - R\Pi\gamma) - 2\,tr(N(YZ - X\Pi Z)) \tag{2.17}$$

where α is an s vector and N a (G-r)xT matrix, both containing Lagrange multipliers. Differentiating L_1 with respect to Π, α and N and equating the results to zero yields :

$$-2X'Y\Omega^p + 2X'X\hat{\Pi}\Omega^p + 2R'\hat{\alpha}\gamma' + 2X'\hat{N}'Z' = 0 \tag{2.18}$$

$$Q\gamma - R\hat{\Pi}\gamma = 0 \tag{2.19}$$

$$YZ - X\hat{\Pi}Z = 0 \tag{2.20}$$

where we have replaced Π, α and N by $\hat{\Pi}$, $\hat{\alpha}$ and $\hat{N}$ respectively to indicate that we have solutions rather than variables.

In order to extricate $\hat{\Pi}$ from (2.18) we make use of the general solution representation of a matrix equation, in terms of pseudoinverses (see Appendix A.1):

$$\hat{\Pi} = (X'X)^{-1}(X'Y\Omega^p - R'\hat{\alpha}\gamma' - X'\hat{N}'Z')\Omega + H(I - \Omega^p\Omega)$$

or

$$\hat{\Pi} = (X'X)^{-1}X'YFF' - (X'X)^{-1}R'\hat{\alpha}\gamma'\Omega + HZZ' \tag{2.21}$$

provided that there is a solution to (2.18). (2.21) can be found by means of the results stated in (2.1) through (2.8). The matrix H of order KxG is arbitrary. Equation (2.18) can be solved if (consistency test)

$$(X'Y\Omega^p - R'\hat{\alpha}\gamma' - X'\hat{N}'Z')\Omega\Omega^p = (X'Y\Omega^p - R'\hat{\alpha}\gamma' - X'\hat{N}'Z')$$

or

$$R'\hat{\alpha}\gamma'ZZ' = - X'\hat{N}'Z' \tag{2.22}$$

Equation (2.22) represents a restriction on $\hat{N}$. In the case of (2.22) the consistency test shows that an $\hat{N}$ always exists, which satisfies (2.22).

Now we try to specify the matrix H by substituting (2.21) into (2.20)

$$YZ - X(X'X)^{-1}X'YFF'Z - X(X'X)^{-1}R'\hat{\alpha}\gamma'\Omega Z + XHZZ'Z = 0$$

or

(2.23) $$YZ = XHZ$$

hence

(2.24) $$H = (X'X)^{-1}X'YZZ' + J - JZZ' = (X'X)^{-1}X'YZZ' + JFF'$$

with J an arbitrary matrix of order KxG. The consistency test for (2.23) is

$$X(X'X)^{-1}X'YZZ'Z = YZ \text{ or } X(X'X)^{-1}X'YZ = YZ$$

which holds good with probability 1 in view of (2.13).

Substituting (2.24) into (2.21), we find

$$\hat{\Pi} = (X'X)^{-1}X'YFF' - (X'X)^{-1}R'\hat{\alpha}\gamma'\Omega + (X'X)^{-1}X'YZZ'ZZ' + JFF'ZZ'$$

or

(2.25) $$\hat{\Pi} = (X'X)^{-1}X'Y - (X'X)^{-1}R'\hat{\alpha}\gamma'\Omega$$

In order to eliminate $\hat{\alpha}$, we now introduce β. From (1.15) we derive

(2.26) $$\beta = -\hat{\Pi}\gamma = -(X'X)^{-1}X'Y\gamma + (X'X)^{-1}R'\hat{\alpha}\gamma'\Omega\gamma$$

hence

(2.27) $$R'\hat{\alpha} = (X'X)\beta(\gamma'\Omega\gamma)^{-1} + X'Y\gamma(\gamma'\Omega\gamma)^{-1}$$

hence

(2.28) $$\hat{\Pi} = (X'X)^{-1}X'Y - \beta(\gamma'\Omega\gamma)^{-1}\gamma'\Omega - (X'X)^{-1}X'Y\gamma(\gamma'\Omega\gamma)^{-1}\gamma'\Omega$$

after substitution of (2.27) into (2.25). We have now found m.l. estimator of Π as a function of the observations, and the unknown γ,β and Ω.

In the second step we introduce $\hat{\Pi}$ according to (2.28) into the log-likelihood function and maximize this function with respect to γ and β, subject to (1.4) which is equivalent to the remaining constraint (2.19) of the first step. We also make use of the normalization (1.3). Instead of maximizing the log-likelihood, we minimize (see (2.14))

(2.29) $$\text{tr}\ ((Y - X\hat{\Pi})\Omega^{p}(Y-X\hat{\Pi})').$$

Substituting (2.28) into (2.29) yields (after some manipulations)

(2.30) $$\text{tr}(\Omega^{p}Y'MY) + (\gamma'\Omega\gamma)^{-1}(\gamma'Y'X(X'X)^{-1}X'Y\gamma + 2\gamma'Y'X\beta + \beta'X'X\beta)$$

by making use of tr(AB) = tr(BA) if A and B are of order mxn and nxm, respectively. M is a TxT symmetric, idempotent matrix :

(2.31) $$M = I - X(X'X)^{-1}X'$$

From (2.30) we conclude that the m.l. estimators of γ and β can be obtained by minimizing

(2.32) $$\frac{(X(X'X)^{-1}X'Y\gamma + X\beta)'(X(X'X)^{-1}X'Y\gamma + X\beta)}{\gamma'\Omega\gamma}$$

subject to :

(2.33) $$Q\gamma + R\beta = 0 \quad \text{and} \quad e_i'\gamma = 1.$$

(2.32) is equal to the second part of (2.30). Note that the existence of (2.32) requires

(2.34) $$\gamma'\Omega\gamma \neq 0$$

The minimization problem stated in (2.32) and (2.33) can be solved by introducing a new Lagrangian with a vector α of s Lagrange multipliers :

(2.35) $$L_2 = \frac{(X(X'X)^{-1}X'Y\gamma + X\beta)'(X(X'X)^{-1}X'Y\gamma + X\beta)}{\gamma'\Omega\gamma} - 2\alpha'(Q\gamma + R\beta)$$

(2.35) is equal to

(2.36) $$L_2 = \frac{\rho'A_1\rho}{\rho'A_2\rho} \quad \text{where}$$

$$A_1 = \begin{pmatrix} Y'X(X'X)^{-1}X'Y & Y'X & Q' \\ X'Y & X'X & R' \\ Q & R & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} \Omega & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and}$$

$$\rho' = (\gamma', \beta', -\gamma'\Omega\gamma\alpha')$$

Differentiating L_2 with respect to ρ and equating the result to zero yields :

(2.37) $$(A_1 - \lambda A_2)\bar{\rho} = 0 \quad \text{where} \quad \lambda = \frac{\bar{\rho}'A_1\bar{\rho}}{\bar{\rho}'A_2\bar{\rho}} = L_2(\bar{\rho})$$

where we have replaced ρ by $\bar{\rho}$ to indicate that we have solutions rather than variables. The scalar λ can be found from the determinantal equation

(2.38) $\det (A_1 - \lambda A_2) = 0$

This determinantal equation has (G+K+s) roots. The definition of λ in (2.37) implies that :

(2.39) $$\lambda = \frac{(X(X'X)^{-1}X'Y\bar{\gamma} + X\bar{\beta})'(X(X'X)^{-1}X'Y\bar{\gamma} + X\bar{\beta})}{\bar{\gamma}'\Omega\bar{\gamma}}$$

where $\bar{\gamma}$ and $\bar{\beta}$ correspond to the parts of $\bar{\rho}$. Consequently we should take the smallest root in order to minimize (2.32). In order to distinguish this smallest root from the others, we denote it by $\hat{\lambda}$. Assumption 6 implies that the smallest root has a multiplicity of one : multiple roots $\hat{\lambda}$ would yield several solutions $\bar{\rho}$ all having the same value of the log-likelihood, precluding the identifiability of the equation.

Having obtained $\hat{\lambda}$ we arrive at a unique solution vector $\hat{\rho}$ from (2.37) and the normalization restriction.

Being interested in the consequences of the procedure outlined above for the exactly identified case and the overidentified case, we study the rank of matrix A_1. Using a result obtained by KHATRI (1968) :

$$\text{rank } A_1 = \text{rank} \begin{pmatrix} C_1'C_1 & C_2' \\ C_2 & 0 \end{pmatrix} = \text{rank } (C_1', C_2') + \text{rank } C_2$$

where $C_1 = (X(X'X)^{-1}X'Y, X)$ of order Tx(G+K)

$C_2 = (Q,R)$ of order sx(G+K)

hence

$$\text{rank } A_1 = \text{rank} \begin{pmatrix} Y'X(X'X)X' & Q' \\ X' & R' \end{pmatrix} + s = \text{rank} \begin{pmatrix} Y'X(X'X) & Q' \\ I & R' \end{pmatrix} \begin{pmatrix} X' & 0 \\ 0 & I \end{pmatrix} + s$$

from which we conclude

(2.40) $\text{rank } A_1 \leq K + 2s$

Since the matrix A_1 is of order G+K+s, it follows that A_1 is singular if

(2.41) $s < G$

Under assumption 6, requirement (2.41) is met only in the exactly identified case in view of (1.7) and (1.8). In that case $\lambda = 0$ is a solution of the determinantal equation (2.38). It is also the smallest root $\hat{\lambda}$, since (2.39) implies that λ is always positive or zero (and real). The smallest root $\hat{\lambda}$ being zero, we conclude that (2.32) has attained its absolute minimum rather than a constrained minimum, which implies $\hat{\alpha} = 0$. Writing out (2.37) for $\hat{\lambda} = \hat{\alpha} = 0$ yields :

(2.42) $$\hat{\Pi} = (X'X)^{-1}X'Y$$

(2.43) $$\hat{\beta} = -\hat{\Pi}\hat{\gamma}$$

while $\hat{\gamma}$ is determined by

(2.44) $$(Q - R\hat{\Pi})\hat{\gamma} = 0 \quad \text{and} \quad e_i'\hat{\gamma} = 1$$

If the equation is overidentified, we can obtain the $\hat{\gamma}$ and $\hat{\beta}$ by solving the determinantal equation (2.38). Unlike the exactly identified case, however, the estimators are then a function of Ω.

Finally we note that the function (2.32) that has to be minimized subject to constraints, contains in the numerator the residual sum of squares after replacing Y by $X(X'X)^{-1}X'Y$.

We now summarize our conclusions in a theorem.

THEOREM 1. _The Limited Information Maximum Likelihood estimator of_ γ _and_ β, _formulated as a function of the unknown positive semidefinite covariance matrix_ Ω, _given the assumptions_ 1, 2, 3, 4, 5a, 6 _and_ 7, _equals_ $\hat{\gamma}$ _and_ $\hat{\beta}$ _which minimize_

(2.45) $$\frac{(X(X'X)^{-1}X'Y\gamma + X\beta)'(X(X'X)^{-1}X'Y\gamma + X\beta)}{\gamma'\Omega\gamma}$$

subject to

$$Q\hat{\gamma} + R\hat{\beta} = 0$$
$$e_i'\hat{\gamma} = 1$$

and these $\hat{\gamma}$ _and_ $\hat{\beta}$ _are the solutions of the system_ :

(2.46) $$\begin{pmatrix} Y'X(X'X)^{-1}X'Y - \hat{\lambda}\,\Omega & Y'X & Q' \\ X'Y & X'X & R' \\ Q & R & 0 \end{pmatrix} \begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \\ \hat{\mu} \end{pmatrix} = 0$$

where $\hat{\lambda}$ _is the smallest root of the corresponding determinantal equation and_ μ _a vector of_ s _Lagrange multipliers_.

In order to indicate the estimation procedure we shall refer to the estimators found in this way as LIML(Ω) estimators, the estimators being functions of Ω.

3. ZERO-NONZERO RESTRICTIONS

The estimation procedure outlined above will now be worked out for the case of zero-nonzero restrictions (assumption 5b). In practice, zero-nonzero restrictions are assumed more frequently than linear restrictions. Furthermore, estimators are much easier to calculate in the former case than in the latter one.

We introduce the following partitions, corresponding to the partition of γ and β (see (1.5)):

$$(3.1)\qquad Y = (Y^+, Y^o) \quad ; \quad X = (X^+, X^o)$$

$$(3.2)\qquad \Omega = \begin{pmatrix} \Omega^{++} & \Omega^{+o} \\ \Omega^{o+} & \Omega^{oo} \end{pmatrix}$$

where Y^+, Y^o, X^+ and X^o are TxG^+, TxG^o, TxK^+ and TxK^o matrices, respectively and Ω^{++} is of order $G^+ \times G^+$. We assumed that the rows and columns in the matrices are rearranged in order to make the partition possible.

Using this partition, the function (2.45) which had to be minimized, becomes :

$$(3.3)\qquad \frac{(X(X'X)^{-1}X'Y^+\gamma^+ + X^+\beta^+)'(X(X'X)^{-1}X'Y^+\gamma^+ + X^+\beta^+)}{\gamma^{+\prime}\Omega^{++}\gamma^+}$$

Differentiating (3.3) with respect to β^+ and equating the result to zero yields

$$(3.4)\qquad \hat{\beta}^+ = -(X^{+\prime}X^+)^{-1}X^{+\prime}Y^+\gamma^+$$

Substitution of (3.4) into (3.3) leads to the following expression in γ^+

$$(3.5)\qquad \frac{\gamma^{+\prime}Y^{+\prime}(M^+ - M)Y^+\gamma^+}{\gamma^{+\prime}\Omega^{++}\gamma^+}, \qquad \text{where } M^+ = (I - X^+(X^{+\prime}X^+)^{-1}X^{+\prime}),$$

to be minimized with respect to γ^+ subject to the normalization restriction. Using (2.36) and (2.37), we obtain for $\hat{\gamma}^+$:

(3.6) $$(Y^{+\prime}(M^{+}-M)Y^{+} - \hat{\lambda}\,\Omega^{++})\hat{\gamma}^{+} = 0$$

where $\hat{\lambda}$ is the smallest root of the determinantal equation

(3.7) $$\det(Y^{+\prime}(M^{+}-M)Y^{+}-\lambda\,\Omega^{++}) = 0$$

It is also possible to find a system of equations for solving $\hat{\gamma}^{+}$ and $\hat{\beta}^{+}$ simultaneously, by combining (3.6) and (3.4) into :

(3.8) $$\begin{pmatrix} Y^{+\prime}X(X'X)^{-1}X'Y^{+} - \hat{\lambda}\Omega^{++} & Y^{+\prime}X^{+} \\ X^{+\prime}Y^{+} & X^{+\prime}X^{+} \end{pmatrix} \begin{pmatrix} \hat{\gamma}^{+} \\ \hat{\beta}^{+} \end{pmatrix} = 0$$

Making use of the normalization restriction, we can find a unique solution vector $(\hat{\gamma}^{+\prime},\hat{\beta}^{+\prime})'$ from equation (3.8). We shall assume that none of the elements of this solution vector is zero; excluding cases with probability zero, this assumption will generally hold good.

Using the normalization restriction, we can also solve the estimators explicitly. For that purpose we introduce a further partitioning (possibly after a rearrangement) :

(3.9) $Y^{+} = (y^{i}, Y^{i})$ where $y^{i} = Ye_{i}$ a T vector, Y^{i} a $Tx(G^{+}-1)$ matrix

$\gamma^{+} = (1,\gamma^{i\prime})'$ where γ^{i} a $(G^{+}-1)$ vector, and $e_{i}'\gamma = 1$

$\Omega^{++} = \begin{pmatrix} w_{ii} & w^{i\prime} \\ w^{i} & \Omega^{ii} \end{pmatrix}$ where w^{i} a $(G^{+}-1)$ vector, and Ω^{ii} a $(G^{+}-1)x(G^{+}-1)$ matrix

The equation (1.2) becomes

$$y^{i} + Y^{i}\gamma^{i} + X^{+}\beta^{+} = \varepsilon$$

Assuming that there exists a solution vector of the system (3.8) with non-zero elements, we conclude that the rank of the symmetric matrix appearing in (3.8) is $(G + K - 1)$ and that each column and row vector are linearly dependent on the other column and row vectors, respectively. This enables us to disregard a superfluous equation of system (3.8), and to write the column vector corresponding to the normalized position as a linear function of the other columns. Disregarding the first equation and writing out, yields

(3.10) $$\begin{pmatrix} Y^{i\prime}PY^{i}-\hat{\lambda}\Omega^{ii} & Y^{i\prime}X^{+} \\ X^{+\prime}Y^{i} & X^{+\prime}X^{+} \end{pmatrix} \begin{pmatrix} \hat{\gamma}^{i} \\ \hat{\beta}^{+} \end{pmatrix} = - \begin{pmatrix} Y^{i\prime}Py^{i} - \hat{\lambda}w^{i} \\ X^{+\prime}y^{i} \end{pmatrix}$$

where

(3.11) $$P = X(X'X)^{-1}X' ,$$

or

(3.12) $$\begin{pmatrix} \hat{\gamma}^i \\ \hat{\beta}^+ \end{pmatrix}_{\Omega} = - \begin{pmatrix} Y^{i\prime}PY^i - \hat{\lambda}\Omega^{ii} & Y^{i\prime}X^+ \\ X^{+\prime}Y^i & X^{+\prime}X^+ \end{pmatrix}^{-1} \begin{pmatrix} Y^{i\prime}Py^i - \lambda w^i \\ X^{+\prime}y^i \end{pmatrix}$$

We index the estimator vector with Ω in order to denote its dependence on the symmetric, positive semidefinite matrix Ω.

4. THE Ω-CLASS ESTIMATORS

4.1 Consistency of the Ω-class estimator

Thus far we developed maximum-likelihood estimators of the structural coefficients as a function of the unknown matrix Ω, the covariance matrix of the contemporaneous reduced form disturbances. In practice, the covariance matrix Ω is rarely known and the best thing we can do is to estimate the covariance matrix from the sample, or, if the sample is very limited, to assume a 'reasonable' covariance matrix which is predetermined. To distinguish the true, unknown covariance matrix Ω from the estimated or fixed matrix we write $\bar{\Omega}$ for the latter.

We now define __the__ __Ω-class__ __estimator__ for the structural coefficients as

(4.1) $$\begin{pmatrix} \hat{\gamma}^i \\ \hat{\beta}^+ \end{pmatrix}_{\bar{\Omega}} = - \begin{pmatrix} Y^{i\prime}PY^i - \hat{\lambda}\bar{\Omega}^{ii} & Y^{i\prime}X^+ \\ X^{+\prime}Y^i & X^{+\prime}X^+ \end{pmatrix}^{-1} \begin{pmatrix} Y^{i\prime}Py^i - \hat{\lambda}\bar{w}^i \\ X^{+\prime}y^i \end{pmatrix}$$

where $\hat{\lambda}$ is the smallest root of the determinantal equation

(4.2) $$\det(Y^{+\prime}(M^+ - M)Y^+ - \lambda\,\bar{\Omega}^{++}) = 0 ,$$

where $\bar{\Omega}$ is partitioned according to the partition of Ω (see (3.2) and (3.9)).

Note that the $\bar{\Omega}$-class estimator defined in (4.1) is a function of the symmetric, positive semidefinite matrix $\bar{\Omega}^{++}$ of order G^+xG^+, corresponding to the covariance matrix Ω^{++} of the rows of Y^+, where

(4.3) $$Y^+\gamma^+ + X^+\beta^+ = y^i + Y^i\gamma^i + X^+\beta^+ = \varepsilon$$

Before stating an important property of the Ω-class estimators, we make the following assumptions :

ASSUMPTION 8. The KxK matrix $(X'X)/T$ converges, as $T \to \infty$, to a positive definite matrix.

ASSUMPTION 9. The matrix $\bar{\Omega}^{++}$ is a symmetric, positive semidefinite matrix with at least one non-zero diagonal element. If the matrix $\bar{\Omega}^{++}$ consists of stochastic elements, it converges as $T \to \infty$, in probability to a symmetric, positive semidefinite matrix with at least one non-zero diagonal element.

Assumption 9 does not imply that the matrix $\bar{\Omega}^{++}$ converges in probability to the real matrix Ω^{++}. The assumption of one non-zero diagonal element is necessary in view of (2.34).

We now state the following theorem :

THEOREM 2. Under the assumptions 1, 2, 3, 4, 5b, 6, 8 and the assumption 9 for a $G^{+}xG^{+}$ matrix $\bar{\Omega}^{++}$, the Ω-class estimator defined in (4.1) and (4.2) is a consistent estimator of the structural coefficients :

$$(4.4) \qquad \operatorname*{plim}_{T\to\infty} \begin{pmatrix} \hat{\gamma}^{i} \\ \hat{\beta}^{+} \end{pmatrix}_{\bar{\Omega}} = \begin{pmatrix} \gamma^{i} \\ \beta^{+} \end{pmatrix}$$

and the smallest root of (4.2) divided by T, converges, as $T \to \infty$, in probability to zero :

$$(4.5) \qquad \operatorname*{plim}_{T \to \infty} \hat{\lambda}/T = 0$$

The proof is presented in Appendix A.3, where we first prove (4.5) and then use the result to demonstrate the asymptotic equivalence of the Ω-class estimator and the consistent 'k-class estimator'. Note the analogy of the Ω-class estimator to the so-called 'k-class estimator' (see, for example, THEIL (1971), P.504) :

$$\begin{pmatrix} \hat{\gamma}^i \\ \hat{\beta}^+ \end{pmatrix}_k = - \begin{pmatrix} Y^{i\prime}Y^i - kY^{i\prime}MY^i & Y^{i\prime}X^+ \\ X^{+\prime}Y^i & X^{+\prime}X^+ \end{pmatrix}^{-1} \begin{pmatrix} Y^{i\prime}y^i - kY^{i\prime}My^i \\ X^{+\prime}y^i \end{pmatrix}$$

Note also the difference between the Ω-class estimator and the k-class estimator. The latter class is defined on a scalar k, which must converge in probability to one in order to ensure consistency of the estimators. The Ω-class estimators are defined on a symmetric positive semidefinite matrix which can be chosen quite arbitrary; under very general conditions, all the members of the class are consistent estimators.

Having found an interesting class of estimators, we now look for interesting species belonging to this class. Below we give some examples of such estimators; among them, we shall come across well-known estimators (Limited Information Maximum Likelihood and Two Stage Least Squares) as members of the Ω-class estimators.

4.2 The Limited Information Maximum Likelihood (LIML) estimator

Since the Ω-class is deduced from a Maximum Likelihood procedure based on a known covariance matrix Ω, it seems reasonable to choose for $\bar{\Omega}$ an estimate of the true covariance matrix. A well-known consistent estimator of the covariance matrix of the contemporaneous reduced form disturbances is[5]:

(4.6) $$\hat{\Omega} = \frac{1}{T} Y'MY \quad \text{so} \quad \hat{\Omega}^{++} = \frac{1}{T} Y^{+\prime}MY^+$$

where $$M = I - X(X'X)^{-1}X' = I - P$$

Substituting

(4.7) $$\bar{\Omega}^{++} = \frac{1}{T} Y^{+\prime}MY^+, \quad \text{hence} \quad \bar{\Omega}^{ii} = \frac{1}{T} Y^{i\prime}MY^i \quad \text{and} \quad \bar{w}^i = \frac{1}{T} Y^{i\prime}My^i,$$

into (4.1) and (4.2) yields,

(4.8) $$\begin{pmatrix} \hat{\gamma}^i \\ \hat{\beta}^+ \end{pmatrix} = - \begin{pmatrix} Y^{i\prime}Y^i - (1+\frac{\hat{\lambda}}{T})Y^{i\prime}MY^i & Y^{i\prime}X^+ \\ X^{+\prime}Y^i & X^{+\prime}X^+ \end{pmatrix}^{-1} \begin{pmatrix} Y^{i\prime}y^i - (1+\frac{\hat{\lambda}}{T})Y^{i\prime}My^i \\ X^{+\prime}y^i \end{pmatrix}$$

---

5) Sometimes the divisor 1/(T-K) is used instead of 1/T but this makes no difference to the results mentioned here.

and

$$(4.9) \qquad \det(Y^{+\prime}(M^{+}-M)Y^{+} - \frac{\lambda}{T} Y^{+\prime}MY^{+}) = 0,$$

respectively, where $\hat{\lambda}$ is the smallest root of the determinantal equation (4.9). In formulae (4.8) and (4.9) we recognize the formulae for the Limited Information Maximum Likelihood estimator (see, for example, JOHNSTON (1963) p.260).

Thus the LIML estimator is a member of the Ω-class estimator. Note that (4.8) and (4.9) result from minimizing (see (3.5)) :

$$(4.10) \qquad \frac{\gamma^{+\prime}Y^{+\prime}(M^{+} - M)Y^{+}\gamma^{+}}{\gamma^{+\prime}Y^{+\prime}MY^{+}\gamma^{+}}$$

subject to the normalization restriction.

A warning against the use of the LIML estimator in small samples seems to be in order. If the number of observations is so small that

$$(4.11) \qquad T < K + G^{+},$$

the estimator $\hat{\Omega}^{++}$ will be singular for lack of degrees of freedom :

$$\text{rank } Y^{+\prime}MY^{+} \leq \text{rank } (I-X(X'X)^{-1}X') = tr(I - X(X'X)^{-1}X') = T - K$$

Consequently if condition (4.11) is complied with, the $G^{+}xG^{+}$ matrix has rank smaller than G^{+}.

4.3 An alternative to the LIML estimator

We now assume that the contemporaneous reduced form disturbances are uncorrelated, so that Ω^{++} is a diagonal matrix. We use the diagonal elements of the estimator mentioned in (4.6), as estimator of the variances :

$$(4.12) \qquad \hat{\Omega}^{++} = \text{diag}(\frac{1}{T} Y^{+\prime}MY^{+}) = \bar{\Omega}^{++}$$

Substituting the partitions of $\bar{\Omega}^{++}$ in (4.1) and (4.2) gives

$$(4.13) \qquad \begin{pmatrix} \hat{\gamma}^{i} \\ \hat{\beta}^{+} \end{pmatrix} = - \begin{pmatrix} Y^{i\prime}PY^{i} - \frac{\hat{\lambda}}{T}\,\text{diag}(Y^{i\prime}MY^{i}) & Y^{i\prime}X^{+} \\ X^{+\prime}Y^{i} & X^{+\prime}X^{+} \end{pmatrix}^{-1} \begin{pmatrix} Y^{i\prime}Py^{i} \\ X^{+\prime}y^{i} \end{pmatrix}$$

and

(4.14) $$\det(Y^{+\prime}(M^{+} - M)Y^{+} - \frac{\lambda}{T}\,\mathrm{diag}(Y^{+\prime}MY^{+})) = 0 ,$$

respectively, where $\hat{\lambda}$ is the smallest root of the determinantal equation (4.14).

Note that (4.13) and (4.14) result from minimizing (see (3.5)):

(4.15) $$\frac{\gamma^{+\prime}Y^{+\prime}(M^{+} - M)Y^{+}\gamma^{+}}{\gamma^{+\prime}(\mathrm{diag}\; Y^{+\prime}MY^{+})\gamma^{+}}$$

subject to the normalization restriction.

We shall call this estimator Limited Information Maximum Likelihood estimator based on a Diagonal covariance matrix, in short LIMLD. If the number of observations is small in particular if condition (4.11) is complied with, the LIMLD estimator could be preferred to the LIML estimator, since the latter is based on an unreliable estimate of the covariance matrix.

4.4 The Two-Stage Least Squares (TSLS) estimator

In order to find the TSLS estimator as a member of the Ω-class estimators, we must start with a somewhat unusual assumption for the covariance matrix Ω^{++} :

(4.16) $$\Omega^{++} = \sigma^2(e_j e_j') \quad \text{with} \quad \sigma^2 \neq 0$$

where e_j is a G^{+}x1 unit vector consisting of a single one in the j-th position and zeroes everywhere else. In this case the covariance matrix Ω^{++} is singular; the endogenous variables represented in the matrix Y^{+} are assumed to be nonstochastic except for the j-th variable y^j which is assumed to be stochastic with variance σ^2.

The Ω-class estimator with $\bar{\Omega}^{++}$ equal to (4.16) can be found by substituting $\bar{\Omega}^{++}$ into (4.1) and (4.2) but this procedure does not lead to the most appropriate expression of the estimator. Instead of starting with (4.1) and (4.2) we go back to the derivation of those formulae from (3.8). We introduce a further partition if $i \neq j$; after a rearrangement :

(4.17) $$Y^{+} = (Y^j,\ y^j) = (y^i,\ Y^i) \quad \text{where } y^j = Y^{+}e_j,\ Y^j \text{ a } Tx(G^{+}-1) \text{ matrix}$$

so the last column of $\bar{Y}^{+}$ consists of the stochastic variable (y^j) and the first column represents the 'normalized' variable (y^i). Following the arguments after (3.8) and (3.9) with the exception that we now ignore the last equation instead of the first one, we find

$$(4.18) \qquad \begin{pmatrix} \hat{\gamma}^i \\ \hat{\beta}^+ \end{pmatrix} = - \begin{pmatrix} Y^{j\prime} P Y^i & Y^{j\prime} X^+ \\ X^{+\prime} Y^i & X^{+\prime} X^+ \end{pmatrix}^{-1} \begin{pmatrix} Y^{j\prime} P y^i \\ X^{+\prime} y^i \end{pmatrix}$$

It is now not necessary to solve a determinantal equation. Equation (4.18) results from minimizing

$$(4.19) \qquad \frac{\gamma^{+\prime} Y^{+\prime} (M^+ - M) Y^+ \gamma^+}{(\gamma_{j1})^2}$$

subject to the normalization restriction; γ_{j1} is the coefficient belonging to the stochastic variable y^j.

If i = j, it will not be necessary to partition (4.17); the normalized variable coincides with the stochastic variable. In this case the estimator equals (4.18) with i = j and the expression to be minimized is the same as in (4.19), with the exception that $\gamma_{j1} = \gamma_{i1} = 1$ by definition.

In formula (4.18) with i = j, we recognize the so-called Two Stage Least Squares (TSLS) estimator. In order to distinguish the other estimators defined by (4.18) with $i \neq j$ from it, we shall use the abbreviation <u>TSLS(j)</u> where the index j corresponds to the non-zero position on the diagonal of the index matrix $\bar{\Omega}^{++}$. As can easily be seen, the TSLS(j) estimator corresponds to the estimator defined by the following procedure :

1. estimate the coefficients by means of TSLS subject to $\gamma_{j1} = 1$,
2. divide these estimates by the estimate of γ_{i1}

We have now reached the following conclusion. The TSLS estimator is a member of the Ω-class estimators, with an index matrix

$$(4.20) \qquad \bar{\Omega}^{++} = \sigma^2 (e_i e_i') \quad \text{where } \sigma^2 \neq 0 \text{ and } e_i \text{ is a } (G^+ x1) \text{ unit vector,}$$

i.e. a matrix with one non-zero element on the diagonal, which position corresponds to the normalized variable. The TSLS estimator is a maximum-likelihood estimator under the assumptions that the true covariance matrix Ω^{++} equals $\bar{\Omega}^{++}$ as defined in (4.20), together with the Normality assumption 7.

Note that the assumption for Ω^{++} can hardly be called 'reasonable'. First, it is arbitrary since the normalization restriction can be chosen arbitrarily. Second, it is at variance with the simultaneous nature of the system of equation in which stochastic variables are related to other stochastic variables. We will return to these problems in section 5.

4.5 The Orthogonal Two-Stage Least Squares (OTSLS) estimator

The TSLS estimator is an example of an Ω-class estimator based on a fixed rather than an estimated index matrix as in the case of the LIML and LIMLD estimator. We now shall present another member of the Ω-class estimators based on a fixed index matrix. In particular, we assume that the contemporaneous reduced form disturbances are mutually uncorrelated and have equal variances :

(4.21) $\Omega^{++} = \sigma^2 I$ where $\sigma^2 \neq 0$, I is a $G^+ x G^+$ identity matrix.

The Ω-class estimator with an index matrix $\bar{\Omega}^{++}$ corresponding to (4.21) can be obtained by substituting this matrix into (4.1) and (4.2):

(4.22) $$\begin{pmatrix} \hat{\gamma}^i \\ \hat{\beta}^+ \end{pmatrix} = - \begin{pmatrix} Y^{i\prime}PY^i - \hat{\lambda} I & Y^{i\prime}X^+ \\ X^{+\prime}Y^i & X^{+\prime}X^+ \end{pmatrix}^{-1} \begin{pmatrix} Y^{i\prime}Py^i \\ X^{+\prime}y^i \end{pmatrix}$$

(4.23) $$\det(Y^{+\prime}(M^+ - M)Y^+ - \lambda\, I) = 0 \quad \text{with smallest root } \hat{\lambda}$$

The factor σ^2 is incorporated in the smallest root $\hat{\lambda}$ so that it has not to be known a priori and separately.

The corresponding minimization problem is

(4.24) $$\text{Minimize} \quad \frac{\gamma^{+\prime}Y^{+\prime}(M^+ - M)Y^+\gamma^+}{\gamma^{+\prime}\gamma^+}$$

subject to the normalization restriction.

Note that the estimator (4.22) is not independent of units of measurements of the variables, in contradiction to the other estimators mentioned here. In general, changing the units of scale for one variable does not result in a corresponding change in the estimated coefficient of that variable.

For reasons given in the next section, we shall call the estimator (4.22) the Orthogonal Two-Stage Least Squares (OTSLS) estimator.

5. AN 'ERRORS IN VARIABLES MODEL' APPROACH OF THE Ω-CLASS ESTIMATORS

In this section we come across the Ω-class estimator as an 'Errors in Variables Model' estimator in a transformed equation. This view enables us to interpret the results of the preceding section more clearly.

Let the equation under consideration be

(5.1) $$Y^+\gamma^+ + X^+\beta^+ = \varepsilon$$

where

(5.2) $$Y^+ = X\Pi^+ + U^+$$

(5.3) $$C(U^+) = (\Omega^{++} \otimes I)$$

Π^+ and U^+ are matrices of order KxG^+ resp. TxG^+, corresponding to the partitions of Y as defined in (3.1). For the symmetrix idempotent matrix

(5.4) $$P = X(X'X)^{-1}X'$$

a KxT matrix A_1 exists such that *)

(5.5) $$A_1'A_1 = P \quad \text{and } A_1A_1' = I$$

Define

(5.6) $$A = T^{-\frac{1}{2}}.A_1$$

and consider the transformed equation (5.1)

(5.7) $$AY^+\gamma^+ + AX^+\beta^+ = A\varepsilon$$

then, using $C(\varepsilon) = \sigma^2_{11}\, I$, where σ^2_{11} is an element of Σ , we obtain :

(5.8) $$C(A\varepsilon) = T^{-1}\, \sigma^2_{11}\, (A_1IA_1') = T^{-1}.\, \sigma^2_{11}\, I$$

(5.9) $$C(AY^+) = C(AU^+) = (\Omega^{++} \otimes T^{-1}A_1A_1') = T^{-1}(\Omega^{++} \otimes I)$$

The implications of the transformation can be shown by writing out (5.7)

(5.10) $$Ay^i = - AY^i\gamma^i - AX^+\beta^+ + A\varepsilon$$

and estimating the coefficients by means of ordinary least squares regression :

(5.11) $$\begin{pmatrix} \hat{\gamma}^i \\ \hat{\beta}^+ \end{pmatrix} = -\begin{pmatrix} Y^{i\prime}A'AY^i & Y^{i\prime}A'AX^+ \\ X^{+\prime}A'AY^i & X^{+\prime}A'AX^+ \end{pmatrix}^{-1} \begin{pmatrix} Y^{i\prime}A'Ay^i \\ X^{+\prime}A'Ay^i \end{pmatrix} = -\begin{pmatrix} Y^{i\prime}PY^i & Y^{i\prime}X^+ \\ X^{+\prime}Y^i & X^{+\prime}X^+ \end{pmatrix}^{-1} \begin{pmatrix} Y^{i\prime}Py^i \\ X^{+\prime}y^i \end{pmatrix}$$

In this expression we recognize the TSLS estimator. Note the difference between the TSLS procedure and the one followed here; in the first stage the transformation

*) Consider the K orthonormal characteristic vectors of P corresponding to the root 1.

matrix A is used here instead of the matrix P used in the TSLS procedure. The reason for using the transformation matrix A rather than the matrix P is the serial correlation caused by P :

$$(5.12) \qquad C(P\varepsilon) = \sigma_{11}^{2}\ (PIP') = \sigma_{11}^{2}\ P$$

$$(5.13) \qquad C(PY^{+}) = (\Omega^{++} \otimes PP') = (\Omega^{++} \otimes P)$$

We have now formulated the TSLS estimation procedure as ordinary least squares regression in equation (5.10) which is based on equation (5.7). However, is the use of ordinary least squares regression in equation (5.7) justified ? When the sample size increases, the transformed variables in (5.7) become more and more deterministic (see (5.8) and (5.9)) and the TSLS estimates converge to the true values. But what if the sample size is small ? Ordinary least squares regression is justified if we assume that the right hand variables in (5.10) are uncorrelated with disturbances $A\varepsilon$. From

$$(5.14) \qquad \mathcal{E}\{(\text{vec } AY - \mathcal{E}\text{vec } AY)(\text{vec } AE - \mathcal{E}\text{vec } AE)'\} =$$
$$\mathcal{E}\,(\text{vec } AY.\ \text{vec}'AE) = \frac{1}{T}\,(\Omega\Gamma \otimes I)$$

we conclude that, in general, this is not the case.

Inspection of equation (5.7) and (5.9) shows us that the relation between the transformed variables corresponds to the so-called 'Errors in Variables Model' (see, e.g., MALINVAUD, chapter 10). The functional relation between the deterministic parts of the transformed variables is

$$(5.15) \qquad AX\Pi^{+}\gamma^{+} + AX^{+}\beta^{+} = 0$$

The observed values are related to the deterministic parts, according to

$$(5.16) \qquad AY^{+} = AX\Pi^{+} + AU^{+} \quad \text{and} \quad AX^{+} = AX^{+} + 0$$

where the covariance matrix of AU^{+} is expressed by (5.9).
Given the covariance matrix Ω^{++}, we are able to estimate γ^{+} and β^{+} in equation (5.7) corresponding to the principle of the so-called 'weighted regression' (see, for example, MALINVAUD (1970), chapter 10 and KOOPMANS (1937)).
According to this principle, the estimators $\hat{\gamma}^{+}$ and $\hat{\beta}^{+}$ are found as those values γ^{+} and β^{+}, satisfying the normalization restriction, for which

$$(5.17)\qquad \frac{\begin{pmatrix}\gamma^+\\ \beta^+\end{pmatrix}' \Big((AY^+,AX^+)'(AY^+,AX^+)\Big)\begin{pmatrix}\gamma^+\\ \beta^+\end{pmatrix}}{\begin{pmatrix}\gamma^+\\ \beta^+\end{pmatrix}' \begin{pmatrix}\Omega^{++} & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\gamma^+\\ \beta^+\end{pmatrix}} = \frac{\begin{pmatrix}\gamma^+\\ \beta^+\end{pmatrix}' \begin{pmatrix}Y^{+\prime}PY^+ & Y^{+\prime}X^+\\ X^{+\prime}Y^+ & X^{+\prime}X^+\end{pmatrix}\begin{pmatrix}\gamma^+\\ \beta^+\end{pmatrix}}{\gamma^{+\prime}\Omega^{++}\gamma^+}$$

is minimized. Comparing (5.17) with (3.3) shows us that the weighted regression estimator for equation (5.7), for a given covariance matrix Ω^{++}, equals the Ω-class estimator with index matrix Ω^{++}. Consistency of the weighted regression estimator is garanteed by the transformation A.

We now try to interpret the various Ω-class estimators mentioned in the preceding section.
The LIML estimation procedure corresponds to a weighted regression for equation (5.7), with an estimated covariance matrix. The LIMLD estimation procedure corresponds to a weighted regression for equation (5.7), with an estimated covariance matrix which is assumed to be diagonal. The OTSLS estimation procedure corresponds to the so-called orthogonal regression for equation (5.7) (see, for example, MALINVAUD (1970) p.35). Finally, we consider the TSLS(j) ($j=1,..,G^+$) estimation procedures. They correspond to the G^+ so-called elementary regressions for equation (5.7), disregarding the fact that the variables AY^j (see (4.17)) are stochastic and correlated with $A\varepsilon$. In particular in the TSLS procedure (where i=j) we erroneously project the Ay^i vector on the plane formed by the vectors of the 'independent' variables AY^i and AX^+, whereas the variables AY^i are, in general, not independent of $A\varepsilon$!

There is a relationship between the G^+ elementary regression estimators of equation (5.7) which are equivalent to the different TSLS(j) estimators, and the weighted regression estimator of equation (5.7) which is equivalent to the Ω-class estimator. The TSLS(j) ($j=1,..,G^+$) estimators can be considered as the limits of a certain class of weighted regression estimators, under general conditions. First, we introduce the following assumptions :

ASSUMPTION 10. All G^+ different TSLS(j) estimators are non-negative, possibly after changing the sign of some of the variables Y^+ and X^+, i.e. [6)]

$$\begin{pmatrix}\hat{\gamma}^+\\ \hat{\beta}^+\end{pmatrix}_j \geq 0 \quad \text{for } j = 1,..,G^+$$

where $()_j$ denotes the TSLS(j) estimator.

---

6) We denote $A \geq 0$ if all elements of A are non-negative.

Assumption 10 corresponds with the case where the signs of the estimates $\hat{\gamma}^+$ and $\hat{\beta}^+$ in the different TSLS(j) (j=1,..,G^+) procedures are compatible.

ASSUMPTION 11. The index matrix $\bar{\Omega}^{++}$ of order $G^+ \times G^+$ is a symmetric, positive semidefinite matrix with at least one non-zero diagonal element and for which

$$\bar{\Omega}^{++} \geq 0 \tag{5.19}$$

We now state

THEOREM 3. The Ω-class estimator of γ^+ and β^+, defined for the index matrix $\bar{\Omega}^{++}$, given the assumptions 1, 2, 3, 4, 5b, 6 and 10 (compatibility of the TSLS(j) estimates) and assumption 11 (non-negative index matrix), is confined to that space angle constructed on the different TSLS(j) (j=1,..,G^+) estimate vectors as edges, viz.

$$\begin{pmatrix} \hat{\gamma}^+ \\ \hat{\beta}^+ \end{pmatrix}_{\bar{\Omega}^{++}} = \sum_{j=1}^{G^+} \delta_j \begin{pmatrix} \hat{\gamma}^+ \\ \hat{\beta}^+ \end{pmatrix}_j , \qquad \delta_j \geq 0 \quad \text{and} \quad \sum_{j=1}^{G^+} \delta_j = 1 \tag{5.20}$$

where

$\begin{pmatrix} \hat{\gamma}^+ \\ \hat{\beta}^+ \end{pmatrix}_{\bar{\Omega}^{++}}$ is the Ω-class estimator defined on the index matrix $\bar{\Omega}^{++}$

$\begin{pmatrix} \hat{\gamma}^+ \\ \hat{\beta}^+ \end{pmatrix}_j$ is the TSLS(j) estimator with non-negative elements in view of assumption 10

This theorem is based on a related theorem concerning the position of a weighted regression vector with respect to the different elementary regression vectors (see KOOPMANS (1937), p.101). In appendix C we present the proof of theorem 3, which is a somewhat more general and modernised version of the one offered by KOOPMANS, concerning the weighted regression. As a corrollary we find that under assumption 10 the LIMLD- and the OTSLS estimators are confined to the space angle constructed on the different TSLS(j) estimate vectors as edges.

6. SUMMARY AND CONCLUSIONS

We shall now try to summarize the results and formulate some conclusions. We derived maximum-likelihood estimators for the structural coefficients in the case of linear restrictions imposed on these coefficients and a known positive semidefinite covariance matrix of the contemporaneous reduced form disturbances. In the case of zero-nonzero restrictions we introduced the Ω-class estimators based on the maximum-likelihood estimators mentioned above. This class of estimators is defined on a symmetric, positive semidefinite matrix. If this index matrix equals the true covariance matrix of the contemporaneous reduced form disturbances and if the disturbances are assumed to be normally distributed, the Ω-class estimator will correspond to the maximum-likelihood estimator. Both assumptions, however, are superfluous for proving that the Ω-class estimator is consistent. The well-known LIML and TSLS estimator are shown to be members of the Ω-class. Besides, two other Ω-class estimators are introduced, viz. the OTSLS and the LIMLD estimator, both with a diagonal index matrix. Finally, a relationship was derived between the different TSLS(j), $j=1,..,G^+$, estimators, resulting from different normalizations, and other members of the Ω-class.

In practice, a choice has to be made from the different methods of estimation. The theoretical considerations presented in this paper lead us to suggest the following.
In general, the LIML estimator would have to be preferred to the LIMLD, OTSLS and the different TSLS(j) estimators, the latter estimators being based on less general assumptions about the covariance matrix. If, however, the sample is small, with the number of observations T about as large as $K + G^+$, the LIML estimator could be unreliable since it is based on a poor estimate of the covariance matrix. In this case we suggest that a diagonal covariance matrix be used as approximation to the unknown covariance matrix of the contemporaneous reduced form disturbances. Since the OTSLS estimator is not independent of the units of measurements of the variables, we recommend the use of the LIMLD estimator in small samples.

Finally, we mention the G^+ different TSLS(j) estimators. We have found that these estimators are theoretically justified only if almost all the endogenous variables are nonstochastic, which is at variance with the simultaneous nature of the system. We have also shown that the commonly used TSLS estimator is rather arbitrary, given the existence of G^+-1 other TSLS(j) estimators, none of which could be preferred to the others on reasonable grounds. Consequently, we did not recommend the TSLS(j) estimators, although practical considerations could lead to their being used. In large samples, in particular, the consistency and computational convenience could make the TSLS(j) estimators attractive. Further research on small sample properties of the Ω-class estimators should provide more detailed answers to the important question of which estimator is to be preferred.

REFERENCES

Dhrymes, P.J. (1970), Econometrics : Statistical Foundations and Applications, Harper and Row, New York.

Gantmacher, F.R. (1959), Matrix Theory, Vol. II, Chelsea Publ. Comp., New York.

Goldberger, A.S. (1964), Econometric Theory, John Wiley and Sons, New York.

Graybill, F.A. (1969), Introduction to Matrices with Applications in Statistics, Wadsworth Publ. Comp., Belmont, California.

Johnston, J. (1963), Econometric Methods, McGraw-Hill Book Comp., New York.

Khatri, C.G. (1968), 'Some Results for the Singular Normal Multivariate Regression Models', Sankhya, ser.A, V.30, pp. 267-280.

Koopmans, T. (1937), Linear Regression Analysis of Economic Time Series, De Erven F. Bohn N.V., Haarlem.

Koopmans, T.C. (ed.)(1950), Statistical Inference in Dynamic Economic Models, John Wiley and Sons, New York.

Malinvaud, E. (1970), Statistical Methods of Econometrics, second revised edition, North-Holland Publ. Comp., Amsterdam.

Neudecker, H. (1969), 'Some Theorems on Matrix Differentiation with Special Reference to Kronecker Matrix Products', American Statistical Association Journal, V.64, pp. 953-963.

Theil, H. (1971), Principles of Econometrics, John Wiley and Sons, New York.

APPENDIX

A.1. The pseudo inverse of a matrix

Where A is a nxm matrix, there exists a unique matrix A^p of order mxn, called the pseudo inverse of A (or Moore-Penrose inverse) which obeys the following four conditions :

1) $AA^pA = A$
2) $A^pAA^p = A^p$
3) A^pA is symmetric
4) AA^p is symmetric

We mention some properties of the pseudo inverse (see GRAYBILL (1969) chapter 6)

1) $(A')^p = (A^p)'$
2) $(A^p)^p = A$
3) $(A'A)^p = A^pA'^p$
4) $(AA^p)^p = AA^p$
5) rank A^p = rank A = rank of AA^p, A^pA, AA^pA and A^pAA^p

A necessary and sufficient condition for the matrix equation

$$AXB = C$$

to be consistent, is

$$AA^pCB^pB = C$$

and the solution can be written as

$$X = A^pCB^p + H - A^pAHBB^p$$

where the matrix H, with the dimensions of X, can be chosen arbitrarily.

A.2 The Kronecker product and the 'vec' operator

The Kronecker-, or direct product is defined as

$$(A \otimes B) = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{21}B & \cdots & a_{2m}B \\ \vdots & & & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nm}B \end{pmatrix}$$

where A is of order nxm. The 'vec' operator is defined as

$$\text{vec } A = (a_{11}, a_{21}, \ldots, a_{n1}, a_{12}, a_{22}, \ldots \quad , a_{nm})'$$

We mention some properties of the Kronecker product and 'vec' operator :

1) $(A \otimes B)(C \otimes D) = (AC \otimes BD)$
2) $(A \otimes B)^{-1} = (A^{-1} \otimes B^{-1})$
3) $(A \otimes B)' = (A' \otimes B')$
4) $(A \otimes (B \otimes C)) = ((A \otimes B) \otimes C)$
5) $\text{vec } ABC = (C' \otimes A) \text{ vec } B$

The interested reader is referred to NEUDECKER (1969).

A.3 Proof of theorem 2

First we prove that

$$\underset{T \to \infty}{\text{plim}} \frac{\hat{\lambda}}{T} = 0$$

where $\hat{\lambda}$ is the smallest root of

$$\det(Y^{+\prime}(M^{+} - M)Y^{+} - \lambda\, \bar{\Omega}^{++}) = 0 \quad \text{or,}$$

$$\det\left(\frac{Y^{+\prime}(M^{+} - M)Y^{+}}{T} - \frac{\lambda}{T} \bar{\Omega}^{++}\right) = 0$$

Under the assumptions of theorem 2 it can be proved (see, e.g., GOLDBERGER (1964) p. 344 or DHRYMES (1970) p.354) [7] that

---

7) Both proofs contain minor errors : DHRYMES p. 354 defines X_2^{*} erroneously; X_2^{*} must be $X_2 - X_1(X_1'X_1)^{-1}X_1'X_2$. GOLDBERGER p.344 states that $\Sigma_{22.1}$ (his notation) is positive semidefinite, which must be positive definite in view of Q being positive definite, where Q = plim (X'X)/T

$$\underset{T \to \infty}{\text{plim}} \frac{Y^{+\prime}(M^{+} - M)Y^{+}}{T}$$

exists and has rank $G^{+} - 1$.

Since we assume that $\hat{\lambda}$ is a single root of the determinantal equation, it is a continuous function of the elements of the matrices in the determinantal equation. This implies that

$$\hat{\nu} = \underset{T \to \infty}{\text{plim}} \hat{\lambda}/T$$

is the smallest root of

$$\det\left(\underset{T \to \infty}{\text{plim}} \frac{Y^{+\prime}(M^{+} - M)Y^{+}}{T} - \nu \underset{T \to \infty}{\text{plim}} \bar{\Omega}^{++}\right) = 0$$

Inspection of this determinantal equation shows that, under the assumptions of theorem 2,

$$\underset{T \to \infty}{\text{plim}} \hat{\lambda}/T = 0$$

Now consider the k-class estimator (see, e.g. THEIL (1971), p.504), where

$$\underset{T \to \infty}{\text{plim}} k = 1$$

Then (see, for example, THEIL (1971), p. 505) :

$$\underset{T \to \infty}{\text{plim}} \begin{pmatrix} \hat{\gamma}^{i} \\ \hat{\beta}^{+} \end{pmatrix}_{k} = \underset{T \to \infty}{\text{plim}} \left[\begin{pmatrix} \frac{Y^{i\prime}Y^{i}}{T} - k\frac{Y^{i\prime}MY^{i}}{T} & \frac{Y^{i\prime}X^{+}}{T} \\ \frac{X^{+\prime}Y^{i}}{T} & \frac{X^{+\prime}X^{+}}{T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{Y^{i\prime}y^{i}}{T} - k\frac{Y^{i\prime}My^{i}}{T} \\ \frac{X^{+\prime}y^{i}}{T} \end{pmatrix} \right] =$$

$$\left(\underset{T \to \infty}{\text{plim}} \begin{bmatrix} \frac{Y^{i\prime}Y^{i}}{T} - k\frac{Y^{i\prime}MY^{i}}{T} & \frac{Y^{i\prime}X^{+}}{T} \\ \frac{X^{+\prime}Y^{i}}{T} & \frac{X^{+\prime}X^{+}}{T} \end{bmatrix} \right)^{-1} \left(\underset{T \to \infty}{\text{plim}} \begin{bmatrix} \frac{Y^{i\prime}y^{i}}{T} - k\frac{Y^{i\prime}My^{i}}{T} \\ \frac{X^{+\prime}y^{i}}{T} \end{bmatrix} \right) =$$

$$= - \left(\underset{T \to \infty}{\text{plim}} \begin{bmatrix} \frac{Y^{i\prime}PY^{i}}{T} & \frac{Y^{i\prime}X^{+}}{T} \\ \frac{X^{+\prime}Y^{i}}{T} & \frac{X^{+\prime}X^{+}}{T} \end{bmatrix} \right)^{-1} \left(\underset{T \to \infty}{\text{plim}} \begin{bmatrix} \frac{Y^{i\prime}Py^{i}}{T} \\ \frac{X^{+\prime}y^{i}}{T} \end{bmatrix} \right) = \underset{T \to \infty}{\text{plim}} \begin{pmatrix} \hat{\gamma}^{i} \\ \hat{\beta}^{+} \end{pmatrix}_{\bar{\Omega}} = \begin{pmatrix} \gamma^{i} \\ \beta^{+} \end{pmatrix}$$

which shows the consistency of the Ω-class estimator.

A.4 Proof of theorem 3

We first mention the following theorem (see GANTMACHER (1959) p.66) :

THEOREM. A non-negative square matrix $A : A \geq 0$, always has a non-negative characteristic root λ^{*} such that the moduli of all the characteristic roots of A do not exceed λ^{*}. To this 'maximum' characteristic root λ^{*} there corresponds a non-negative characteristic vector x :

$$Ax = \lambda^{*}x \quad \text{where} \quad \lambda^{*} \geq 0 \quad \text{and} \quad x \geq 0$$

The Ω-class estimator, defined for an index matrix $\bar{\Omega}^{++}$, can be derived from the following set of equations (see (3.8)) :

$$(A_1 - \hat{\lambda} A_2)x = 0 \quad \text{where}$$

$$A_1 = \begin{pmatrix} Y^{i\prime}PY^{i} & Y^{i\prime}X^{+} \\ X^{+\prime}Y^{i} & X^{+\prime}X^{+} \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} \bar{\Omega}^{++} & 0 \\ 0 & 0 \end{pmatrix}, \quad x = \begin{pmatrix} \hat{\gamma}^{+} \\ \hat{\beta}^{+} \end{pmatrix}_{\bar{\Omega}}$$

and $\hat{\lambda}$ is the smallest root of the corresponding determinantal equation. All roots λ are real and either positive or zero, as we found in section 2.

First we consider the case where A_1^{-1} exists, corresponding to $\hat{\lambda} > 0$. Then

$$(A_2 A_1^{-1} - \frac{1}{\hat{\lambda}} I)\delta = 0 \qquad \text{where} \quad \delta = A_1 x, \ a(G^{+}+K^{+}) - \text{vector}$$

so that $1/\hat{\lambda}$ corresponds to the maximum characteristic root of the matrix $A_2 A_1^{-1}$. The first G^{+} column vectors of the matrix A_1^{-1} contain the different TSLS(j) ($j=1,..,G^{+}$) estimators, apart from a normalization factor; this can be easily seen by substituting $\bar{\Omega}^{++}$ from (4.16) into A_2 and solving the characteristic equation.

The matrix $A_2A_1^{-1}$ can be written as

$$A_2A_1^{-1} = \begin{pmatrix} \bar{\Omega}^{++}A^{11} & \bar{\Omega}^{++}A^{12} \\ 0 & 0 \end{pmatrix} \quad \text{where} \quad A_1^{-1} = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}$$

Because of the symmetry of the matrix A_1^{-1}, the partitions A^{11} and A^{12} contain the elements of the TSLS(j) estimates. Under the assumptions 10 and 11 we find that

$$A_2A_1^{-1} \geq 0$$

From the theorem mentioned above, we conclude

$$\delta \geq 0$$

and from the partition of $A_2A_1^{-1}$, it follows that

$$(\delta_{G^+}, \delta_{G^++1}, \ldots, \delta_{G^++K^+}) = 0$$

Therefore

$$x = A_1^{-1}\delta \quad \text{where} \quad \delta = (\delta_1, \ldots, \delta_{G^+}, 0, \ldots, 0)' \geq 0$$

The condition that the elements of the δ vector add up to 1, is derived from the normalization restriction.

For the case where A_1 is singular, all the Ω-class estimators coincide, including the different TSLS estimators. This is the exactly identified case.