

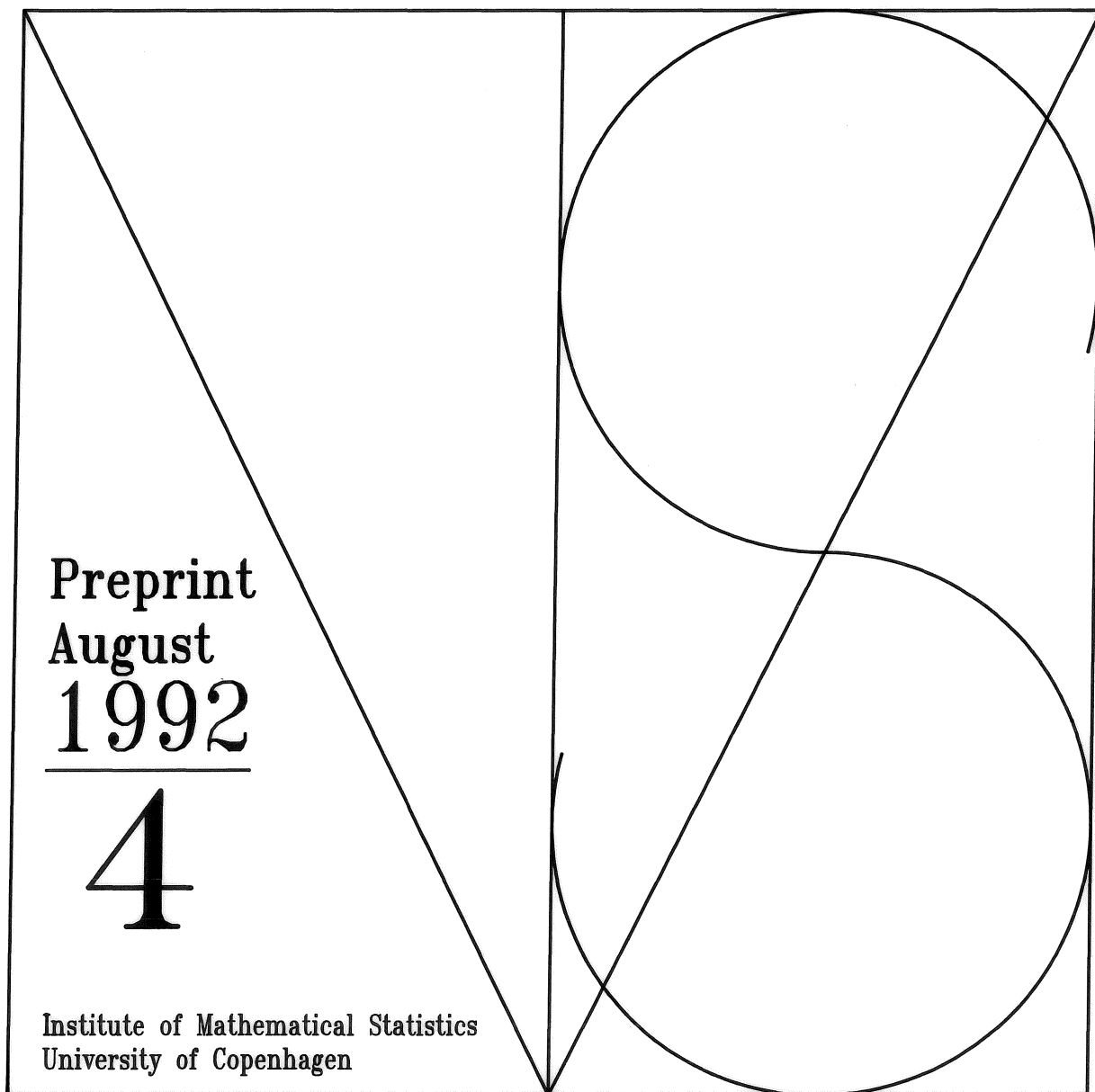
Søren Johansen

Identifying Restrictions
of Linear Equations

Preprint
August
1992

4

Institute of Mathematical Statistics
University of Copenhagen



IDENTIFYING RESTRICTIONS OF LINEAR EQUATIONS

S. Johansen

University of Copenhagen
Institute of
Mathematical Statistics

August 1992

Abstract

The main result of this paper is a simple algebraic condition on a set of linear restrictions in a simultaneous equations model that guarantees that most linear structures defined by the restrictions are identified.

The connection to the identification of cointegrating relations, common trends and short-run dynamics in a vector autoregressive model is pointed out. A general switching algorithm involving eigenvalue problems is suggested for estimation of the identified equations, and the asymptotic distribution is given.

1. Introduction and basic definitions

The problem of identification is met in econometrics in connection with the construction of systems of simultaneous linear equations that allow the coefficients to be estimated. Consider as an example the simultaneous linear equations

$$(1) \quad \beta' X_t = \Gamma y_t + B z_t = \epsilon_t$$

where the p -dimensional process $X_t = (y_t', z_t')$ is decomposed into r endogenous and k predetermined variables, $p = r + k$, and the parameters are collected into the $r \times (r+k)$ matrix $\beta' = (\Gamma, B)$, where the $r \times r$ matrix Γ is assumed to have full rank. The errors are assumed to be independent Gaussian variables with mean zero and variance Ω ,

such that likelihood inference is available. Furthermore ϵ_t is assumed independent of z_t , such that (1) determines the joint distribution of the variables y_t conditional on the predetermined variables z_t .

In order to be able to recognize the individual relations we formulate linear restrictions R_i as full rank $p \times r_i$ matrices and assume that the coefficients in the i 'th equation, β_i , satisfy

$$(2) \quad R_i' \beta_i = 0, \quad i = 1, \dots, r.$$

We also work with a slightly different formulation by defining $H_i = R_{i\perp}$, that is, we define a $p \times (p-r_i) = p \times s_i$ matrix of full rank such that $H_i' R_i = 0$. In this case (2) is replaced by

$$\beta_i = H_i \varphi_i$$

for some s_i -dimensional vector φ_i . Thus a direct parameterization is

$$(3) \quad \beta = (H_1 \varphi_1, \dots, H_r \varphi_r).$$

We define a statistical model by the parameter space

$$\mathcal{L} = \{\beta_{p \times r}, \Omega \mid R_i' \beta_i = 0, \quad i = 1, \dots, r\}.$$

The model \mathcal{L} is linear in β and we shall call it a linear model. A parameter point (β, Ω) is identified if for any other value (β_1, Ω_1) the corresponding probability measures are distinct. The classical result on identification, see Sargan (1988, p.29), is given in

THEOREM 1 *A necessary and sufficient condition that a parameter value (β, Ω) is identified is that*

$$\text{rank}(R_i' \beta) = r-1, \quad i = 1, \dots, r.$$

Thus when applying the restrictions of one equation to the other $r-1$ equations we get a matrix of rank $r-1$. Hence it is not possible by taking linear combinations of for instance β_2, \dots, β_r to construct a vector and hence an equation which is restricted in the

same way as β_1 and in this sense could be confused with the equation defined by β_1 . Hence β_1 can be recognized among all linear combinations of β_1, \dots, β_r as the only one that is in $\text{sp}(H_1)$, the space spanned by the columns in H_1 , or the only one that satisfies the restrictions R_1 . In applications we often normalize the vectors β_1 on one of the variables. This will be discussed in section 4 where asymptotic distributions are derived. Theorem 1 gives rise to the definition of another model, namely the set of identified structures

$$\mathcal{M} = \{ \beta_{p \times r}, \Omega \mid R_i' \beta_i = 0, \text{ and } \text{rank}(R_i' \beta) = r-1, i = 1, \dots, r \}.$$

Clearly $\mathcal{M} \subset \mathcal{L}$ and our first result states that \mathcal{M} is a large subset of \mathcal{L} in the sense that \mathcal{M} is an open dense subset of \mathcal{L} or in other words that $\mathcal{L} \setminus \mathcal{M}$ is a nowhere dense subset:

THEOREM 2 *If \mathcal{L} contains an identified parameter value, that is, if \mathcal{M} is non-empty, then \mathcal{M} is an open dense subset of \mathcal{L} .*

The proof is given in the appendix. Note that \mathcal{L} is a linear model but \mathcal{M} is not linear since the rank condition is a non-linear restriction. The fact that \mathcal{M} is dense in \mathcal{L} means that the likelihood function cannot be used to distinguish between the models and the MLE derived from the linear model \mathcal{L} satisfies the rank conditions and hence corresponds to a point in \mathcal{M} with probability 1.

For a given set of restrictions either all structures in \mathcal{L} defined by the restrictions are unidentified, or "almost all" are identified. The paper by McManus (1992) contains a number of general results about identification of parametric models emphasizing the generic property of the subset of identified structures in a given model. The emphasis is on non-linear models showing that in some sense all the problems are created by focusing on linear models.

The purpose of this note is to discuss the linear models because we need them in the autoregressive formulation of the dynamic models and in the discussion of identification of cointegrating relations.

Theorem 2 gives rise to the definition

DEFINITION 1 *The restrictions R_1, \dots, R_r and the linear model \mathcal{L} are called identifying (generically identifying) if there exists a parameter value β such that the rank condition is satisfied, or equivalently such that \mathcal{M} is non-empty.*

The purpose of the note is to give a simple algebraic condition (see Theorem 3) for a linear model to be (generically) identifying, such that before a linear model is estimated a condition on the restrictions is tested.

We conclude this section by pointing out a number of situations where the identification problem is met in the analysis of the vector autoregressive model for cointegration written in the structural error correction form:

$$(5) \quad A\Delta Y_t = a\beta' Y_{t-1} + G\Delta Y_{t-1} + m + \epsilon_t.$$

Here the ϵ_t are independent Gaussian with mean zero and variance Ω . For notational simplicity we have left out more lags.

The reduced form error correction model is

$$(6) \quad \Delta Y_t = \alpha\beta' Y_{t-1} + \Gamma\Delta Y_{t-1} + \mu + \delta_t,$$

where $\alpha = A^{-1}a$, $\Gamma = A^{-1}G$ etc. Notice that the β vectors are the same in the reduced form as in the structural form of the model.

The first identification problem is for the cointegrating relations $\beta' Y_t$. If it is assumed that Y_t is I(1) it follows from (6) that $\beta' Y_t$ is stationary. Thus

$$(7) \quad \beta' Y_t = \eta_t,$$

where η_t is stationary. Clearly any linear combination of these equations will produce new stationary relations, and hence there is a need for identifying restrictions.

The representation theorem of Granger, see for instance Johansen (1991), shows that the solution of the equations (6) under the assumption that X_t is I(1) is given by

$$(8) \quad Y_t = C(\Sigma_1^t \epsilon_i + \mu t) + U_t$$

where U_t is a stationary process and

$$C = \beta_{\perp} (\alpha'_{\perp} (I - \Gamma) \beta_{\perp})^{-1} \alpha'_{\perp},$$

such that the common trends are given by $\alpha'_{\perp} \Sigma_1^t \epsilon_i$. Any set of linear transformations of these $p-r$ variables can also serve as common trends, such that again restrictions are required to identify specific relations.

The third identification problem occurs in the short term dynamics, after the long-run relations have been identified. If in model (5) we define X_t by stacking ΔY_t , $\beta' Y_{t-1}$ and ΔY_{t-1} , then the equations have the form

$$(9) \quad (A, a, G)X_t = \epsilon_t,$$

such that the identification problem of the simultaneous effects and short-term dynamics involved in the matrices A , a and G can also be formulated as equation (1).

The main result in section 2 of this note does not involve any assumptions on the probability properties of the process, but rely entirely on the fact that we are interested in the linear relations, which we want to identify in the sense that we want to recognize the equations on the basis of suitable linear restrictions on the individual equations.

2. A necessary and sufficient condition for a set of restrictions to be generically identifying

We formulate the basic Theorem and show how it can be applied. The proof which is based on some classical results in mathematics by Hall (1935) and Rado (1942), will be discussed in the appendix.

THEOREM 3 *The linear statistical model \mathcal{L} defined by the restrictions R_1, \dots, R_r is identifying if and only if for $k = 1, \dots, r-1$ and any set of indices $1 \leq i_1 < \dots < i_k \leq r$ not containing i , it holds that*

$$(12) \quad \text{rank}(R'_i H_{i_1}, \dots, R'_i H_{i_k}) \geq k.$$

If (12) is satisfied for a particular value of i , then the restrictions are identifying that equation.

As an example of how to apply Theorem 3 we consider the situation with $r = 3$, where

$$(13) \quad \beta = (H_1 \varphi_1, H_2 \varphi_2, H_3 \varphi_3).$$

To see if R_1 , R_2 and R_3 identify the first equation we check the conditions

$$\text{rank}(R'_1 H_2) \geq 1, \text{rank}(R'_1 H_3) \geq 1, \text{rank}(R'_1 (H_2 : H_3)) \geq 2.$$

In terms of the matrices $M_{i,j,k} = H'_i H_j - H'_i H_k (H'_k H_k)^{-1} H'_k H_j$, $i, j = 1, 2, 3$ the conditions become

$$(14) \quad \text{rank}(M_{22.1}) \geq 1, \text{rank}(M_{33.1}) \geq 1, \text{rank} \begin{bmatrix} M_{22.1} & M_{23.1} \\ M_{32.1} & M_{33.1} \end{bmatrix} \geq 2.$$

Since the matrices M_{ij} are positive semidefinite an easy way of checking this is to calculate the eigenvalues of the matrices in (14).

The system is not identified if for instance $\text{sp}(H_2) \subset \text{sp}(H_1)$, since then $R'_1 H_2 = 0$ and condition (12) is violated by the choice $k = 1$ and $i_1 = 2$.

Now suppose that we have an identifying statistical model \mathcal{L} as above. It is apparent that if we restrict the spaces H_i further, that is, formulate overidentifying restrictions, we may decrease the rank in condition (12) thus allowing the possibility that a linear submodel of the identifying model is not identifying.

An example of this is the hypothesis

$$(15) \quad \beta = (H_1 \varphi_1, (H_2 \cap H_1) \varphi_2, H_3 \varphi_3),$$

where $H_2 \cap H_1$ is a matrix such that $\text{sp}(H_2 \cap H_1) = \text{sp}(H_2) \cap \text{sp}(H_1)$. This is a

formulation of the hypothesis that the second vector is not only in $\text{sp}(H_2)$ but also in the $\text{sp}(H_1)$ which indicates that the first equation is not characterized by the restrictions R_1 , since evidently also β_2 satisfies the same restrictions.

Thus if we know where to look for the lack of identification we can formulate this as an overidentifying restriction on the model. The methods are illustrated and exemplified in Johansen and Juselius (1992) on money demand in Australia.

As a simple illustration consider the model defined by the linear structures

$$\beta = \begin{bmatrix} a & c \\ -a & 0 \\ b & -c \\ 0 & d \end{bmatrix}, (a,b,c,d) \in \mathbb{R}^4.$$

In this case we find

$$R_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, H_1 = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, R_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, H_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

such that

$$\begin{aligned} \text{rank}(R_1' H_2) &= \text{rank}\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = 2 \geq 1, \\ \text{rank}(R_2' H_1) &= \text{rank}\left(\begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}\right) = 2 \geq 1. \end{aligned}$$

Thus the conditions for generic identification (12) are satisfied. A given structure β in \mathcal{L} is identified if

$$\text{rank}(R_1' \beta_2) = \text{rank}(c,d) = 1$$

and

$$\text{rank}(R_2' \beta_1) = \text{rank}(a+b, -a) = 1.$$

This is, the structure β is identified if $(c,d) \neq (0,0)$ and $(a,b) \neq (0,0)$. Thus the linear model \mathcal{L} in this case contains the unidentified structures given by either $(a,b) = (0,0)$ or $(c,d) = (0,0)$, and this set is a small set of parameter values.

3. *An algorithm for estimating simultaneous equations under identifying linear restrictions and an expression for the asymptotic variance of the estimator*

We consider the situation described in (1). The likelihood function is given by

$$\begin{aligned}\log L(\beta, \Omega) &= -\frac{1}{2}T \log |\Omega| + \frac{1}{2}T \log |\Gamma \Gamma'| - \frac{1}{2} \sum_{t=1}^T X_t' \beta \Omega^{-1} \beta' X_t \\ &= -\frac{1}{2}T \log |\Omega| + \frac{1}{2}T \log |\Gamma \Gamma'| - \frac{1}{2} \text{tr} \{ \Omega^{-1} \beta' S_{xx} \beta \},\end{aligned}$$

where $S_{xx} = T^{-1} \sum_{t=1}^T X_t X_t'$, see Sargan (1988). Maximizing over Ω gives

$$(16) \quad L_{\max}^{-2/T}(\beta) = \frac{|\beta' S_{xx} \beta|}{|\Gamma \Gamma'|} = \frac{|\beta' S_{xx} \beta|}{|\beta' M \beta|},$$

where

$$M = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}.$$

We want to minimize (16) under the restrictions

$$\beta = (H_1 \varphi_1, \dots, H_r \varphi_r),$$

where we assume that the restrictions $R_i = H_{i\perp}$, $i = 1, \dots, r$ are identifying, that is, that they satisfy conditions (12).

We can rewrite (16) such that the dependence on φ_1 is more apparent and then apply a switching algorithm by cyclically minimizing with respect to one variable and fixing the others. In order to describe the successive steps we define matrices

$$\begin{aligned}\tau &= (\beta_2, \dots, \beta_r), \\ S_{xx, \tau} &= S_{xx} - S_{xx} \tau (\tau' S_{xx} \tau)^{-1} \tau' S_{xx}, \\ M_{\cdot, \tau} &= M - M \tau (\tau' M \tau)^{-1} \tau' M,\end{aligned}$$

and decompose

$$|\beta' S_{xx} \beta| = |\tau' S_{xx} \tau| |\beta_1' S_{xx, \tau} \beta_1| = |\tau' S_{xx} \tau| |\varphi_1' H_1' S_{xx, \tau} H_1 \varphi_1|$$

and

$$|\beta' M \beta| = |\tau' M \tau| |\varphi_1' H_1' M_{\cdot, \tau} H_1 \varphi_1|,$$

and insert into (16). This shows that the likelihood function depends on φ_1 in such a way that the maximization can be performed by solving the eigenvalue problem

$$|\lambda H_1' S_{xx, \tau} H_1 - H_1' M_{\cdot, \tau} H_1| = 0$$

for eigenvalues $\lambda_1 > \dots > \lambda_{s_1}$ and eigenvectors v_1, \dots, v_{s_1} . We get the maximum

likelihood estimator for φ_1 , for fixed values of the other parameters τ , as the eigenvector corresponding to the largest eigenvalue λ_1 , and hence, $\hat{\beta}_1 = H_1 v_1$. Next we fix $\beta_1, \beta_2, \dots, \beta_r$ and maximize over φ_2 using a similar eigenvalue procedure. We then continue until convergence.

The algorithm is probably slower than Newton-Raphson iteration, but has the advantage that it is explicit in each step. Once the eigenvalue problem has been programmed each step is easy to perform, the likelihood function is increasing in each step, and finally the algorithm works for general linear restrictions on the individual equations, not just zero restrictions.

It is interesting to note that a good set of initial values is found by first solving (16) without restrictions on β , that is, by solving

$$|\lambda S_{xx} - M| = 0$$

for eigenvalues $\lambda_1 > \dots > \lambda_p$ and eigenvectors w_1, \dots, w_p giving $\hat{\beta} = (w_1, \dots, w_p)$. Next we find a linear combination of the β 's which is closest to the space H_i , that is, we solve the problem

$$\max_{w, \varphi} \frac{(w' \hat{\beta}' H_i \varphi)^2}{w' \hat{\beta}' \hat{\beta} w \varphi' H_i' H_i \varphi},$$

or equivalently the eigenvalue problem

$$|\lambda \hat{\beta}' \hat{\beta} - \hat{\beta}' H_i (H_i' H_i)^{-1} H_i' \hat{\beta}'| = 0.$$

Let u_1 be the eigenvector corresponding to the largest eigenvalue, then $\hat{\beta}_i = \hat{\beta} u_1$ is the starting value for β_i in the iteration.

This choice of initial value also has the advantage that if H_1, \dots, H_r is a just identifying set of restrictions, then no iteration is needed.

In order to discuss the asymptotic distribution and the expression for the variance it is necessary to normalize the vectors, so that the length of the eigenvectors is determined. This is often done by normalizing on one of the coefficients, and a general formulation is

$$\beta_i = h_i + H^i \psi_i,$$

for some $s_i - 1$ dimensional vector ψ_i . Thus the vectors in h_i and H^i span the same space as the vectors in H_i , and h_i is linearly independent of the vectors in H^i . By differentiating the concentrated likelihood function (16) we derive the expressions for the derivative of $\log L$ with respect to ψ_i

$$\begin{aligned} \partial \log L / \partial \psi_i &= e_i' (\beta' S_{xx} \beta)^{-1} \beta' S_{xx} H^i - e_i' (\beta' M \beta)^{-1} \beta' M H^i \\ &= e_i' S_{\epsilon\epsilon}^{-1} S_{\epsilon Z}^{-1} \beta' H^i, \end{aligned}$$

where $\beta' = (-B' \Gamma^{-1}, I)$, and e_i is the i 'th unit vector. The second derivative is given by

$$\partial^2 \log L / \partial \psi_i \partial \psi_j = e_i' S_{\epsilon\epsilon}^{-1} e_j H^i \beta' S_{zz} \beta' H^j + o_P(1).$$

Now $S_{\epsilon\epsilon} \xrightarrow{P} \Omega$, and assuming that $S_{zz} \xrightarrow{W} \Sigma > 0$, we find by an expansion of the likelihood function that

$$T^{\frac{1}{2}} \text{vec}(\hat{\psi} - \psi) \xrightarrow{W} N(0, \{e_i' \Omega^{-1} e_j H^i \beta' \Sigma \beta' H^j\}^{-1}).$$

and since $\hat{\beta}_i = h_i + H^i \hat{\psi}_i$, so that $\text{vec}(\hat{\beta}' - \beta') = \{H^i\} \text{vec}(\hat{\psi}_i - \psi_i)$, we find that

$$(17) \quad T^{\frac{1}{2}} \text{vec}(\hat{\beta} - \beta) \xrightarrow{W} N_{p \times r}(0, \{H^i\} \{e_i' \Omega^{-1} e_j H^i A' \Sigma A' H^j\}^{-1} \{H^j\}).$$

We have here applied the notation $\{M_{ij}\}$ for a block matrix with blocks M_{ij} , and apply the notation $\{M_i\}$ for the block diagonal matrix with blocks M_i along the diagonal. A consistent estimator of the asymptotic variance is given by replacing Σ , Ω and β by their estimators. Note that β_{\perp} involves the estimated reduced form coefficients. We conclude this section by summarizing the results

THEOREM 4 *The asymptotic distribution of the maximum likelihood estimators of the simultaneous equations (1) under the identifying restrictions R_1, \dots, R_r and the normalization $\beta_i = h_i + H^i \psi_i$ is given by*

$$(18) \quad T^{\frac{1}{2}} \text{vec}\{\hat{\beta} - \beta\} \xrightarrow{W} N_{p \times r}(0, \{H^i\} \{e_i' \Omega^{-1} e_j H^i \beta' \Sigma \beta' H^j\}^{-1} \{H^j\}).$$

Note that under the identification condition the asymptotic covariance matrix must have full rank at the true value of β since if

$$0 = \sum_{i,j} e_i' \Omega^{-1} e_j \psi_i' H^i \beta_{\perp} \Sigma \beta_{\perp}' H^j \psi_j = \text{tr} \{ e_i' \Omega^{-1} e_j \} \{ \psi_i' H^i \beta_{\perp} \Sigma \beta_{\perp}' H^j \psi_j \},$$

then

$$\psi_i' H^i \beta_{\perp} \Sigma \beta_{\perp}' H^i \psi_i = 0.$$

This implies that $\beta_{\perp}' H^i \psi_i = 0$, such that the vector $H^i \psi_i$, which is clearly in H_i , is a linear combination of the vectors in β . Since the parameter is identified there is only one such combination which is in the space spanned by H_i , and that is $\beta_i = h_i + H^i \psi_i^0$, say, hence $H^i \psi_i = c(h_i + H^i \psi_i^0)$ or $H^i(\psi_i - \psi_i^0 c) = c h_i$, which is clearly impossible unless $\psi_i = 0$, and $c = 0$, since h_i is linearly independent of H^i .

4. *Estimation of cointegration parameters and common trends under identifying restrictions*

We give here a discussion of the estimation problem for identified cointegrating relations in a vector autoregressive model. It turns out that the same estimation procedure as discussed in section 3 can be applied. Let Y_t be generated by the reduced form equations (6)

$$(19) \quad \Delta Y_t = \alpha \beta' Y_{t-1} + \Gamma \Delta Y_{t-1} + \mu + \epsilon_t,$$

let α and β be $p \times r$ and assume that the cointegrating relations are restricted by

$$(20) \quad \beta = (H_1 \varphi_1, \dots, H_r \varphi_r).$$

For ease of notation we have left out more lags. The cointegrating relations are estimated without restrictions by reduced rank regression of ΔY_t on Y_{t-1} corrected for ΔY_{t-1} , and 1, see Johansen (1988). This procedure gives residuals R_{0t} and R_{1t} and product moment matrices

$$S_{ij} = T^{-1} \sum_1^T R_{it} R_{jt}'.$$

The estimation problem is solved by the eigenvalue problem

$$(21) \quad |\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0$$

for eigenvalues $1 > \lambda_1 > \dots > \lambda_p > 0$ and eigenvectors v_1, \dots, v_p , and the estimator of β is given by (v_1, \dots, v_r) . When the restrictions (20) are imposed the estimation procedure is no longer explicit, and the switching algorithm described in section 3 can be applied:

For fixed values of $\varphi_2, \dots, \varphi_r$ or equivalently β_2, \dots, β_r , we define $\tau = (\beta_2, \dots, \beta_r)$ and decompose α into (α_1, α_2) where α_1 is $p \times 1$. Then

$$\Delta Y_t = \alpha_1 \varphi_1' H_1' Y_{t-1} + \alpha_2 \tau' Y_{t-1} + \Gamma \Delta Y_{t-1} + \mu + \epsilon_t.$$

The solution for fixed τ is given by reduced rank regression of ΔY_t on $H_1' Y_{t-1}$ corrected for ΔY_{t-1} , $\tau' Y_{t-1}$ and 1. Hence we solve the eigenvalue problem

$$|\lambda H_1' S_{11.\tau} H_1 - H_1' S_{10.\tau} S_{00.\tau}^{-1} S_{01.\tau} H_1| = 0$$

for eigenvalues λ_i and eigenvectors φ_i and choose $\hat{\beta}_1 = H_1 \varphi_1$. Here

$$S_{11.\tau} = S_{11} - S_{11} \tau (\tau' S_{11} \tau)^{-1} \tau' S_{11}$$

with similar definitions of $S_{10.\tau}$ and $S_{00.\tau}$.

The algorithm then continues by fixing $\beta_1, \beta_3, \dots, \beta_r$ and solve for β_2 by reduced rank regression. The process is repeated until convergence by cyclically maximizing over one of the parameters φ_i keeping all others fixed.

If further submodels are formulated in the form (15) then the likelihood ratio test is asymptotically χ^2 distributed. This follows from the general result that for error correction models, hypotheses on β can be tested by the likelihood ratio procedure by the χ^2 distribution.

Without identifying restrictions the coefficients of the common trends can be estimated by the eigenvalue problem

$$|\lambda S_{00} - S_{01} S_{11}^{-1} S_{10}| = 0,$$

which is solved for eigenvalues $1 > \lambda_1 > \dots > \lambda_p > 0$ and eigenvectors w_1, \dots, w_p . The unrestricted maximum likelihood estimator of α_{\perp} is given by $\hat{\alpha}_{\perp} = (w_{p-r+1}, \dots, w_p)$,

and under identifying restrictions the switching algorithm described above can be applied.

We formulate in the next section the asymptotic distribution of the maximum likelihood estimator of β when it is estimated under the identifying restrictions.

4. *The asymptotic distribution of the estimated identified cointegrating relations*

The general result about the limit distribution is given in Johansen (1991), see also Phillips (1991) and Ahn and Reinsel (1990). It is proved that the limit distribution is mixed Gaussian and that inference concerning the cointegrating vectors can be performed by the χ^2 distribution if likelihood ratio tests are applied.

The estimation is most conveniently performed with the individual vectors unnormalized, but for the asymptotic distribution we normalize them, usually on one of the variables. A general formulation is of the form

$$\beta_i = h_i + H^i \psi_i$$

for some $(s_i - 1)$ -dimensional vector ψ_i . The asymptotic distribution is most conveniently derived from the likelihood function concentrated with respect to μ . We also use the general result that inference on β can be conducted as if the other parameters were known.

Under the identifying restrictions and the normalization the derivatives with respect to ψ_i are given by

$$\begin{aligned} \partial \log L / \partial \psi_i &= \alpha_i' \Omega^{-1} \Sigma_1^T \epsilon_t (Y_{t-1} - Y_{-1})' H^i = T \alpha_i' \Omega^{-1} S_{\epsilon y} H^i, \\ \partial^2 \log L / \partial \psi_i \partial \psi_j &= \alpha_i' \Omega^{-1} \alpha_j H^j \Sigma_1^T (Y_{t-1} - Y_{-1}) (Y_{t-1} - Y_{-1})' H^i \\ &= T \alpha_i' \Omega^{-1} \alpha_j H^j S_{yy} H^i. \end{aligned}$$

An expansion of the likelihood function gives the expression

$$\text{vec}(\hat{\psi} - \psi) \approx \{ \alpha_i' \Omega^{-1} \alpha_j H^j S_{yy} H^i \}^{-1} \{ H^i S_{y\epsilon} \Omega^{-1} \alpha_i \}$$

The representation of Y_t under the assumption that the process is I(1), is given

by (8). It follows that

$$H^i \cdot Y_t = H^i \cdot C \sum_{i=1}^t \epsilon_i + \tau_i t + H^i \cdot U_t,$$

for $\tau_i = H^i \cdot C\mu$. Thus the process is asymptotically linear in the direction τ_i and orthogonally to this direction behaves like a random walk. Let W_t be the Brownian motion defined on the unit interval as the limit of the normalized sums of the ϵ 's, that is,

$$T^{-\frac{1}{2}} \sum_1^{[Tt]} \epsilon_i \xrightarrow{W} W_t,$$

and let $W = \int_0^1 W_t dt$. Let $\gamma_i = \tau_{i\perp}$, then it follows from the representation (8) that

$$(22) \quad T^{-1} \tau_i' H^i \cdot (Y_{[Tt]} - Y) \xrightarrow{W} \tau_i' \tau_i (t - \frac{1}{2}) = \tau_i' \tau_i G_1(t),$$

$$(23) \quad T^{-\frac{1}{2}} \gamma_i' H^i \cdot (Y_{[Tt]} - Y) \xrightarrow{W} \gamma_i' H^i \cdot C(W_t - W) = \gamma_i' H^i \cdot G_2(t),$$

say, and we define $G' = (G_1', G_2')$, and

$$G_{2.1}(t) = G_2(t) - [\int_0^1 G_2 G_1 dt] [\int_0^1 G_1 G_1 dt]^{-1} G_1(t),$$

that is, the Brownian motion corrected for a trend. For $B_T = (T^{-\frac{1}{2}} \gamma_i, T^{-1} \tau)$ we find

$$(24) \quad B_T' H^i \cdot S_{yy} H^j B_T \xrightarrow{W} (H^i \gamma_i, \tau_i' \tau_i)' \int_0^1 G G' dt (H^i \gamma_i, \tau_i' \tau_i),$$

$$(25) \quad S_{ey} B_T \xrightarrow{W} \int_0^1 (dW) G' (H^i \gamma_i, \tau_i' \tau_i)$$

Applying the results (22), ..., (25) we find that

$$\text{Tvec}(\hat{\psi} - \psi) \xrightarrow{W} \{\gamma_i\} \{ \gamma_i' H^i \cdot \int_0^1 G_{2.1} G_{2.1}' dt H^j \gamma_j \alpha_j' \Omega^{-1} \alpha_j \}^{-1} \{ \gamma_j' H^j \cdot \int_0^1 G_{2.1} (dW) \cdot \Omega^{-1} \alpha_j \}.$$

Thus the asymptotic distribution is asymptotically mixed Gaussian with an asymptotic quadratic variation process given by

$$\{\gamma_i\} \{ \gamma_i' H^i \cdot \int_0^1 G_{2.1} G_{2.1}' dt H^j \gamma_j \alpha_j' \Omega^{-1} \alpha_j \}^{-1} \{\gamma_j\},$$

which is consistently estimated by

$$T \{ \hat{\alpha}_1 \hat{\Omega}^{-1} \hat{\alpha}_j H^j \cdot S_{yy} H^i \}^{-1} = T^2 \{ \hat{\alpha}_1 \hat{\Omega}^{-1} \hat{\alpha}_j H^j \cdot \Sigma_1^T R_{yt} R_{yt}' H^i \}^{-1}.$$

Thus in practice, ignoring finite sample problems, we can act as if the asymptotic distribution of $\hat{\beta} - \beta$ is asymptotically Gaussian with a variance matrix given by (26):

$$(26) \quad \{ \hat{\alpha}_1 \hat{\Omega}^{-1} \hat{\alpha}_j H^j \cdot \Sigma_1^T R_{yt} R_{yt}' H^i \}^{-1}.$$

The square root of the diagonal elements of (26) provide the "standard

deviations" which are used for constructing Wald tests for specific values of individual coefficients, if the likelihood ratio test is not worked out.

Note that the distribution of $\hat{\beta}$ is mixed Gaussian such that the distribution has heavy tails. Since the "asymptotic variance" or the observed Fisher information is random even in the limit it may happen that we get a very small value of the information and hence a large value of the variance. In this case it may happen that value of the likelihood estimator appears far away from the true value, but in reality we only have a situation, where the data has little information on the true parameter.

It is thus of more interest to simulate the distribution of the t-ratios, that is the estimated coefficient minus the coefficient divided by its standard deviation, since it is this ratio that is being used in the testing, rather than the distribution of $\hat{\beta}$ itself. The limit distribution of the t-ratio is asymptotically Gaussian such that one would expect much smaller tails in finite samples than is found in the simulations of the distribution of $\hat{\beta}$.

Appendix

Proof of Theorem 2

We can formulate the rank condition by defining the matrix $A_k(\varphi)$ with entries

$$A_k(\varphi)_{ij} = \varphi_i' H_i' R_k R_k' H_j \varphi_j, \quad i, j \neq k.$$

Then

$$\text{rank}(R_k' \beta) = \text{rank}(A_k(\varphi)),$$

such that the rank condition is equivalent to

$$|A_k(\varphi)| \neq 0.$$

Now the function $\varphi \rightarrow |A_k(\varphi)|$ is a polynomial, and it is well known that a polynomial which is zero on an open set is constant ($= 0$) and hence if there exists a vector φ such that $|A_k(\varphi)| \neq 0$, then the set of zeros $\{\varphi \mid |A_k(\varphi)| = 0\}$ contains no open set. Under the assumptions of Theorem 2 \mathcal{L} contains an identified parameter value φ such that

$|A_k(\varphi)| \neq 0$ for all k . In this case the set of unidentified structures is equal to the closed set

$$\bigcup_{k=1}^r \{\varphi \mid |A_k(\varphi)| = 0\}$$

which does not contain any interior points. Thus the complement \mathcal{M} is an open dense set.

Proof of Theorem 3

1. *Necessity of condition (12)*

It is easily seen that if the set of restrictions identify the 1'st equation, say, then one can take vectors $\beta_j \in \text{sp}(H_j)$, $j \neq 1$, such that they satisfy the rank condition (4) and hence such that they are linearly independent. This means that the matrix

$$(R_1' H_{i_1}, \dots, R_1' H_{i_k})$$

contains at least k linearly independent columns. Thus condition (12) is necessary, but the sufficiency is not so simple.

2. *Sufficiency of condition (12) for zero restrictions*

The result follows for zero restrictions from Hall's Theorem, a classical result from combinatorics, see Hall (1935).

THEOREM 5 (Hall (1935))

Let M_i , $i = 1, \dots, n$ be finite sets, then a necessary and sufficient condition that we can take distinct elements $m_i \in M_i$ is that for all $k = 1, \dots, n$ and all sets of indices $1 \leq i_1 < \dots < i_k \leq n$

$$(A.1) \quad \text{card}(M_{i_1} \cup \dots \cup M_{i_k}) \geq k$$

What is characteristic for zero restrictions is that the vector spaces spanned by H_j are defined by the subset of the unit vectors that they contain. Hence the vector spaces $\text{sp}(R_1' H_j)$, $j = 2, \dots, r$, are characterized by subsets M_j of $\{1, \dots, p\}$, and condition

(12) translates into the condition (A.1).

Under this condition Hall's Theorem states that there exists distinct representative elements $m_j \in M_j$, $j = 2, \dots, r$. The unit vectors corresponding to these are linearly independent, and hence guarantee that the rank condition is satisfied for this choice of parameter values.

3. Sufficiency in the general case

The proof follows from a generalization of Hall's theorem due to Rado (1942).

THEOREM 6 (Rado (1942)) *Let M_i , $i = 1, \dots, n$ be finite sets of vectors from a given vector space, then a necessary and sufficient condition that one can take linearly independent vectors $v_i \in M_i$, $i = 1, \dots, n$ is that for all $k = 1, \dots, n$ and all sets of indices $1 \leq i_1 < \dots < i_k \leq n$*

$$(A.2) \quad \dim\{sp(M_{i_1} \cup \dots \cup M_{i_k})\} \geq k.$$

With this result the proof of Theorem 1 goes as follows: We define M_i as the columns of $R_1' H_{i+1}$, $i = 1, \dots, r-1$ and let $n = r-1$. Then (12) implies by Rado's theorem that we can choose vectors $v_i \in M_i$ linearly independent. These have the form $v_i = R_1' H_{i+1} \varphi_{i+1}$ for some unit vector φ_{i+1} . The vectors $\beta_i = H_i \varphi_i$, $i = 2, \dots, r$ satisfy the rank condition and hence the first equation is identified by R_1 .

The reason for proving the result of Theorem 3 first for zero restrictions and then in the general case is to point out the connection with the two important results from operations research. A systematic exposition of the theory of matroids derived from Rado's theorem is given by Welsh (1976), and the applications of Hall's theorem in the theory of flows in networks is given by Ford and Fulkerson (1962).

Acknowledgements

I am grateful for discussions with many colleagues in the preparation of this paper. In particular I have benefited from discussions with Katarina Juselius on the topic of identification, and Jon Faust helped with some of the econometric formulations. The fact that the main result could be proved by the theorems by Hall and Rado was pointed out to me by Lawrence Woolsey, Steffen Lauritzen and Erik Kjær Jørgensen.

References

- Ahn, S.K. & Reinsel, C G. (1990). Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association*, 85, 813–823.
- Ford, L.R. & Fulkerson, D.R. (1962). *Flows in Networks*. Princeton University Press.
- Hall, P. (1935). On representatives of subsets. *Journal of the London Mathematical Society*, 10, 26–30.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12, 231–254.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59, 1551–1580.
- Johansen, S. & Juselius, K. (1992). Identification of the long–run and the short–run structure. An application to the ISLM model. University of Copenhagen.
- McManus, D.A. (1992). How common is identification in parametric models? To appear in *Journal of Econometrics*.
- Phillips, P.C.B. (1991). Optimal inference in cointegrated systems. *Econometrica*, 59, 283–306.
- Rado, R. (1942). A theorem on independence relations. *Quarterly Journal of Mathematics Oxford*, 13, 83–89.
- Sargan, J.D. (1988). *Lectures on advanced econometric theory*. Oxford : Basil Blackwell.
- Welsh, D.J.A. (1976). *Matroid Theory*. London: Academic Press.