Identifiability of cointegrated systems

Boswijk, H.P.

Publication date
1995
Document Version
Submitted manuscript

Identifiability of Cointegrated Systems

H. Peter Boswijk*
Tinbergen Institute &
Department of Actuarial Science and Econometrics
University of Amsterdam
Roetersstraat 11
NL-1018 WB Amsterdam, The Netherlands
E-mail:peterb@butler.fee.uva.nl

March 16, 1995

Abstract

This paper stipulates conditions for identifiability of the parameters of a cointegrated VAR model under general linear restrictions, possibly including cross-equation restrictions. An algorithm is given to obtain the maximum likelihood estimators under such restrictions on both the cointegrating vectors and the adjustment parameters. Then the asymptotic distribution of the estimator and of the likelihood ratio statistic for the over-identifying restrictions is given. The importance of the observed information matrix for identification issues is emphasized.

*A previous version of this paper was presented at the Workshop on Multivariate Cointegration Analysis, June 1994, Mariefred, Sweden. Financial support from the Tinbergen Institute and the Royal Netherlands Academy of Arts and Sciences is gratefully acknowledged.

1 Introduction

Since the introduction of the notion of cointegration, cf. Engle and Granger (1987), a possible identification problem has been recognized. The fact that a number of stationary relationships exist between the components of a vector process which is integrated of order 1 is not sufficient to obtain unique estimates of the parameters characterizing these relationships. This has led researchers to impose identifying and normalizing restrictions on the cointegrating vectors, see, e.g., Stock (1987) and Boswijk (1994, 1995). In the cointegrating regression literature, cf. Phillips and Durlauf (1986) and Phillips (1991), the relationships are solved for a subset of the variables, leading to a reduced form system of equations. These can again be transformed into a structural form, as in Park (1990) and Saikkonen (1993). In all cases, consistent inference on the free parameters is only possible if some rank condition for identification is satisfied.

The identification problem in cointegrated systems is largely analogous to the simultaneous equations model, and this analogy has been explored extensively by Johansen (1995a). He shows that the classical rank condition for identifiability of a single equation in a system is somewhat impractical, since it involves the true parameter value. He therefore derives a condition for generic identification, i.e., identification for all parameter values except for a possible set of Lebesgue measure zero. In case generic identification is satisfied, Boswijk (1996) proposes a test for identifiability by testing whether the parameter vector lies in this measure-zero set of non-identifiable points. This is related to a test for identifiability of a single structural equation from a system of otherwise reduced form equations, see Koopmans and Hood (1953) and Cragg and Donald (1993).

In the present paper we analyze the identification of cointegrated vector autoregressions (VARs) under general linear restrictions, possibly including cross-equation restrictions [1] . This allows more involved economic theory restrictions, obtained e.g. from models under rational expectations, to be implemented in cointegration models. Moreover, it enables us to unify and at the same time simplify the earlier results obtained for the case of equation-by-equation restrictions only. Following Rothenberg (1971), we shall emphasize the central role of the information matrix in the analysis of identifiability of (functions of) parameters. We shall argue that in the present situation the *observed* information matrix provides a more natural and appropriate measure than the expected information matrix.

The outline of the paper is as follows. In Section 2 we derive a rank condition for identifiability of a whole system subject to linear restrictions. This condition involves the unknown parameters, but we discuss how this potential problem may be solved. In Section 3 we provide an algorithm to obtain the maximum likelihood (ML) estimators of the parameters of a vector error correction model subject to general linear restrictions on both the cointegrating

[1] After a first version of this paper was completed, I learned of related work by Pesaran and Shin (1994).

vectors and the error correction coefficients. The algorithm can also be applied to systems that are (partially) under-identified; in that case it provides only one of the many parameter values maximizing the likelihood. With this estimator, identifiability of (linear combinations of) the parameters is easily checked. In Section 4 the asymptotic distribution of the ML estimator and of the likelihood ratio statistic for the over-identifying restrictions is derived, and the consequences of lack of identifiability on these distributions are discussed. The fifth section concludes.

2 The Rank Condition Revisited

Consider the pth order vector autoregression (VAR(p)) in error correction form, for an observed n-vector time series $\{x_t, t = 1, \ldots, T\}$:

$$\Delta x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t, \tag{1}$$

where α and β are $n \times r$ parameter matrices, $0 < r < n$, where $\Gamma_i, i = 1, \ldots, p-1$, are $n \times n$ parameter matrices, and where $\{\varepsilon_t\}$ is an independent $N(0, \Omega)$ sequence, with Ω positive definite. The starting values $\{x_{1-p}, \ldots, x_0\}$ are considered fixed. We assume that the characteristic equation of (1) has exactly $n - r$ roots equal to one and all other roots outside the unit circle, so that x_t is cointegrated of order 1,1 (Engle and Granger, 1987; Johansen, 1991, 1995b). Then the columns of β are the cointegrating vectors, spanning the cointegrating space, r is the cointegrating rank, and α contains the adjustment coefficients (or error correction coefficients; factor loadings). The system can be extended by deterministic components such as a constant, seasonal dummies, or a linear trend. These are not considered explicitly here, but the analysis below can easily be adapted to such extensions.

Defining $w_t = (\Delta x'_{t-1}, \ldots, \Delta x'_{t-p+1})'$, and $\Gamma = (\Gamma_1, \ldots, \Gamma_{p-1})$, the log-likelihood of the model is given by:

$$\begin{aligned} \ln L(\alpha, \beta, \Gamma, \Omega) \;=\; & -\frac{Tn}{2} \ln 2\pi - \frac{T}{2} \ln |\Omega| \\ & -\frac{1}{2} \sum_{t=1}^{T} (\Delta x_t - \alpha\beta' x_{t-1} - \Gamma w_t)' \Omega^{-1} (\Delta x_t - \alpha\beta' x_{t-1} - \Gamma w_t). \end{aligned} \tag{2}$$

It is clear that any parameter point (α^0, β^0) is observationally equivalent to the set of all points $(\alpha^1, \beta^1) = (\alpha^0 A^{-1\prime}, \beta^0 A)$ with arbitrary non-singular $r \times r$ matrix A: all of these points result in the same likelihood value. Therefore, any unique estimator of β (and α) must, either explicitly or implicitly, satisfy a set of identifying restrictions. One popular choice is to normalize β with respect to the first r components of x_t, leading to $\beta' = [I_r : -B]$, where B is an $r \times (n-r)$ matrix, representing the long-run effect of the last $(n - r)$ components of x_t on the first r components

(without implying causality in one direction). Alternatively, Johansen's ML estimator $\hat{\beta}$ satisfies the property that the columns of $\hat{\beta}$ correspond to the ordered canonical correlations between Δx_t and x_{t-1} (conditional upon w_t); this also leads to a unique choice among all values of β that maximize the log-likelihood, provided that the canonical correlations are all distinct.

In this paper, we consider (generalizing the formulation of Stock, 1987, Engle and Granger, 1987, Boswijk, 1995, and Johansen, 1995a, *inter alia*), identifying restrictions of the form

$$\text{vec}\beta_\phi = H\phi + h, \tag{3}$$

where β_ϕ denotes the matrix of cointegrating vectors with the identifying restrictions imposed (which is distinguished from an arbitrary matrix β spanning the cointegrating space); furthermore, H is a known $nr \times k$ matrix of full column rank, h is a known $nr \times 1$ vector, and $\phi \in \mathbb{R}^k$ is an unrestricted parameter vector. Typically, both H and h contain zero and unit elements. The vector h corresponds to normalization restrictions on the vectors of β_ϕ. The restrictions can be rewritten in more conventional format as follows. For an arbitrary $n \times m$ matrix A of full column rank ($n > m$), we define $A_\perp$ to be an $m \times (n-m)$ matrix of full column rank such that $A'A_\perp = 0$. Letting $R = H_\perp$, an $nr \times (nr-k)$ matrix, and $s = R'h$, an $nr \times 1$ vector, the restrictions (3) correspond to

$$R'\text{vec}\beta_\phi = s. \tag{4}$$

A special case arises if there are no cross-equation restrictions. In that case, H and R satisfy:

Condition 1 *The matrices H and R are block-diagonal, i.e.,*

$$H = \begin{bmatrix} H_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & H_r \end{bmatrix}, \quad R = \begin{bmatrix} R_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R_r \end{bmatrix}.$$

Partition h, s and ϕ conformably with H and R, and let $\beta_\phi = (\beta_{\phi 1}, \ldots, \beta_{\phi r})$; then Condition 1 implies

$$\beta_{\phi i} = H_i\phi_i + h_i \quad \Leftrightarrow \quad R_i'\beta_{\phi i} = s_i, \quad i = 1, \ldots r.$$

Johansen (1995a) analyzes identification of the unrestricted parameters under Condition 1, and without any normalizations imposed (hence, with both h_i and s_i set to zero). He argues that the conventional formulation of the rank condition for identification of ϕ_i, viz.

$$\text{rank}(R_i'\beta_\phi) = \text{rank}(R_i'[H_1\phi_1, \ldots, H_r\phi_r]) = r - 1,$$

is rather impractical, since it requires knowledge of the other parameter vectors ϕ_j, $j \neq i$, which are still to be estimated, and which may in fact be under-identified themselves. Therefore, Johansen provides an operational criterion for generic identification, i.e., identification for all

$\phi \in \mathbb{R}^k$, except for a set of Lebesgue measure zero. However, for larger r his method requires quite a lot of computations; moreover, it is not applicable in the presence of cross-equation restrictions.

Therefore, we now analyze identification of in the general context (3)-(4). Let $\text{sp}(A)$ denote the column space of an arbitrary matrix A. Further, let $\mathcal{B}$ denote the parameter space of β, i.e., the space of all $n \times r$ matrices of full column rank that satisfy the (possibly over-) identifying restrictions (3)-(4), each together with all $n \times r$ matrices that span the same column space:

$$\mathcal{B} = \{\beta \in \mathbb{R}^{n \times r} : \text{sp}(\beta) = \text{sp}(\beta_\phi),\ \phi \in \mathbb{R}^k,\ \text{rank}(\beta_\phi) = r\}.$$

We now state the main result on identifiability of ϕ. Proofs of all theorems are given in the appendix.

Theorem 1 *Consider the model (1), together with the restrictions (3)-(4). For a particular parameter value $\beta \in \mathcal{B}$, the parameter vector ϕ is identifiable if and only if*

$$\text{rank}(H'[I_r \otimes \beta_\perp]) = k \quad \Leftrightarrow \quad \text{rank}(R'[I_r \otimes \beta]) = r^2. \tag{5}$$

For an arbitrary known $m \times k$ matrix F, the linear function $F\phi$ is identifiable if and only if

$$\text{sp}(F') \subseteq \text{sp}(H'[I_r \otimes \beta_\perp]). \tag{6}$$

Remarks (a) The rank condition (5) is a condition for *global identification*, since only linear restrictions on β are considered. This can be seen from the proof, where it is shown that under condition (5), each $\beta \in \mathcal{B}$ corresponds to a unique $\phi \in \mathbb{R}^k$. If instead of (3), a set of non-linear restrictions are considered, i.e., if $\text{vec}\beta_\phi = h(\phi)$, where $h : \mathbb{R}^k \mapsto \mathbb{R}^{rn}$ is a continuously differentiable function with Jacobian matrix $H(\phi) = \partial h / \partial \phi'$, then a similar rank condition holds, viz. $\text{rank}(H(\phi)'[I_r \otimes \beta_\perp]) = k$. This condition, however, is only sufficient for *local identification*: each $\beta \in \mathcal{B}$ now corresponds to a $\phi \in \mathbb{R}^k$ which is only unique in a neighbourhood of ϕ; see Rothenberg (1971).

(b) The second result of Theorem 1 determines the class of *estimable* [2] linear functions of ϕ. This is particularly useful if only $r_1 < r$ cointegrating vectors are subjected to identifying restrictions, with the remaining $r_2 = r - r_1$ vectors unrestricted. If $F = [I_{k_1} : 0]$ selects the k_1 components of ϕ corresponding to the first r_1 cointegrating vectors, then (6) requires the matrix consisting of the first k_1 rows of $H'[I_r \otimes \beta_\perp]$ to have full row rank k_1. This is essentially

[2] Formally, the notion of estimable functions in statistics refers to the possibility of unbiased estimation, whereas here we are concerned with consistent (and unique) ML estimation. However, as argued by Richmond (1974), the two notions are very similar, so that we may call a function estimable if the corresponding parameter is identifiable.

a *limited information* identifiability result. Note that if $F = I_k$, then $\text{sp}(F') = \mathbb{R}^k$, so that (6) implies to (5), as expected.

(c) The conditions (5) and (6) obviously suffers from the same problem as mentioned above, viz. that they cannot be checked without knowledge of β (or $\beta_\perp$). However, if the restrictions are just-identifying ($k = r(n-r)$), then the parameter space $\mathcal{B}$ is simply the space of all $n \times r$ matrices of full column rank (possibly except for a set of points of Lebesgue measure zero in this space). Thus, we could check generic identification simply by substituting an arbitrary $n \times r$ matrix of full column rank for β and check whether the rank condition is satisfied (e.g., by checking whether the smallest eigenvalue of $[I_r \otimes \beta]' RR' [I_r \otimes \beta]$ equals zero). More interestingly, we may substitute $\hat{\beta}$ (Johansen's ML estimator) for β, and check the rank condition for this choice. A violation of the rank condition may now occur for two reasons: either because ϕ is not generically identified, or because it is generically identified but happens to lie in the measure-zero set of non-identifiable parameter points (which, of course, happens with zero probability). However, this distinction is irrelevant for practical purposes, as in either case there is no unique ML estimator of ϕ.

(d) If the system is over-identified, then the above method is no longer valid, because $\hat{\beta}$ will lie in $\mathcal{B}$ with probability zero, and the rank condition may be satisfied for almost all $n \times r$ matrices *outside* $\mathcal{B}$, but at the same time violated for almost all matrices *within* $\mathcal{B}$. Consider the following example with $n = 3, r = 2, k = 1$, and with $\beta'_{\phi 1} = (1, \phi, 0)'$ and $\beta'_{\phi 2} = (0, 1, 0)$. It is immediately clear that ϕ is not generically identified. In this case we have $R = \text{diag}(R_1, R_2)$, with

$$R_1' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad R_2' = I_3.$$

Now if β is restricted to lie in $\mathcal{B}$, it can easily be checked that $\text{rank}(R_1'\beta) = 1$ and $\text{rank}(R_2'\beta) = 2$, which implies $\text{rank}(R'[I_2 \otimes \beta]) = 3 < r^2$. However, the unrestricted estimator $\hat{\beta}$ satisfies $\text{rank}(R_1'\hat{\beta}) = 2$ with probability 1, so that the lack of identification would not be detected. Therefore, in the presence of over-identifying restrictions generic identification should be checked by substituting an arbitrary element of $\mathcal{B}$. This could be done by selecting an arbitrary k-vector ϕ and constructing the corresponding β_ϕ from it. The analogue of the above method is to use an ML estimate of β under the restrictions. Note that the cointegration space, and hence an arbitrary matrix β spanning it, is always identified, even if ϕ is not identified. This follows since the model (1) is linear in $\Pi = \alpha\beta'$, and $\text{sp}(\beta) = \text{sp}(\Pi')$. An algorithm to obtain the restricted ML estimate of β (which is also valid if ϕ is not identified) is given in Section 3. This is again more useful, because we want to check whether a unique ML estimate is available for the specific case under consideration, rather than for an arbitrary point in the parameter

space. Alternatively, if Condition 1 is satisfied, Johansen's (1995a) method to check generic identification can be used; however, this technique does not seem to be easily generalized to the case of cross-equation restrictions.

(e) Theorem 1 provides a sufficient condition for identifiability of β_ϕ, where the matrix of adjustment coefficients is unrestricted. If restrictions are imposed on α, these may be identifying too, in which case less restrictions on β are required. As an extreme example, consider the restriction $\alpha' = [\alpha_1' : I_r]$, where α_1 is an $(n-r) \times r$ matrix. This restriction is considered by Breitung (1994), who uses it to derive a particular simultaneous equations representation of the model (1). Under this restriction, no further restrictions on β are required: it is uniquely determined by $\beta' = [0 : I_r]\Pi$. However, although such restrictions lead to a unique estimator of β (just like Johansen's reduced rank regression estimator), there is no reason to interpret the individual vectors of $\hat{\beta}$ as specific long-run relationships. Thus, identification is obtained only in a purely mathematical sense, and not in any interpretational sense. Moreover, as we shall see in Section 4, such identifying restrictions do not lead to a *super-consistent* estimator.

(f) The results of Theorem 1 can be straightforwardly applied to the classical simultaneous equations model

$$By_t + Cz_t = u_t \quad \Leftrightarrow \quad y_t = \Pi z_t + v_t,$$

where y_t is a g-vector of endogenous variables, z_t is a m-vector of exogenous variables, u_t is a g-vector of structural disturbances, B and C are structural coefficient matrices of appropriate order, which satisfy certain identifying restrictions, and finally $\Pi = -B^{-1}C$ and $v_t = B^{-1}u_t$ are the reduced-form coefficient matrix and disturbance vector, respectively. Defining $x_t' = (y_t', z_t')$ and $\beta_\phi' = [B : C]$, the model in structural form reads $\beta_\phi' x_t = u_t$. Notice that the column space of β_ϕ is spanned by $\beta = [I_g : -\Pi]'$, corresponding to the reduced from $\beta' x_t = v_t$, and the orthogonal complement is spanned by $\beta_\perp = [\Pi' : I_m]'$. If the identifying restrictions on β_ϕ are again formulated as (3)-(4), then the rank condition for identifiability reads

$$\operatorname{rank}\left(H'\left[I_g \otimes \begin{bmatrix} \Pi \\ I_m \end{bmatrix}\right]\right) = k,$$

or

$$\operatorname{rank}\left(R'\left[I_g \otimes \begin{bmatrix} B' \\ C' \end{bmatrix}\right]\right) = \operatorname{rank}\left(R'\left[I_g \otimes \begin{bmatrix} I_g \\ -\Pi' \end{bmatrix}\right]\right) = g^2,$$

with k the dimension of ϕ. The second formulation of the rank condition is well-known, see, e.g., Richmond (1974). Notice that this condition can be checked with the estimated reduced form matrix, but in the presence of over-identifying restrictions, this should be the *restricted* reduced form estimator (which, again, is identified even if ϕ is not identified).

3 Maximum Likelihood Estimation

Consider again the log-likelihood (2). Since no restrictions will be imposed on the parameter matrix Γ, it will be useful to concentrate the log-likelihood. Define the observation matrices $X_0 = (\Delta x_1, \ldots, \Delta x_T)'$, $X_1 = (x_0, \ldots, x_{T-1})'$, and $W = (w_1, \ldots, w_T)'$. Next, let $M = I_T - W(W'W)^{-1}W'$ and define the moment matrices

$$S_{ij} = \frac{1}{T} X_i' M X_j, \quad i, j = 0, 1.$$

Using well-established results from partitioned regression, it is easily checked that the concentrated log-likelihood equals (apart from a constant term):

$$\ln L(\alpha, \beta, \Omega) = -\frac{T}{2} \ln |\Omega| - \frac{T}{2} \mathrm{tr} \Omega^{-1} (S_{00} - S_{01}\beta\alpha' - \alpha\beta' S_{10} + \alpha\beta' S_{11} \beta\alpha'). \tag{7}$$

Johansen (1988,1991) further concentrates (7) with respect to α and Ω, and derives from that a maximum likelihood estimator of β, based on reduced rank regression methods. However, for our purposes the expression (7) will be most convenient.

As in the previous section, the identifying restrictions on are expressed as (the distinction between β and β_ϕ is no longer important in the sequel):

$$\mathrm{vec}\beta = H\phi + h. \tag{8}$$

Similarly, we impose homogeneous linear restrictions on α:

$$\mathrm{vec}\alpha' = G\gamma, \tag{9}$$

where G is a known $nr \times l$ matrix of full column rank, and $\gamma \in \mathbb{R}^l$ is an unrestricted parameter vector. Although non-homogeneous restrictions could be imposed on α as well, in practice these are less relevant, since most restrictions on α will be exclusion restrictions, corresponding to weak exogeneity of some of the variables for some of the cointegrating vectors, see Johansen (1992) and Boswijk (1995). Note that if $G = I_{nr}$, then α is unrestricted.

An obvious requirement of H, h and G is that they are not in conflict with the cointegrating rank r. For example, if we would set G to 0, then $\alpha = 0$ for all $\gamma \in \mathbb{R}^l$, which violates the assumption $\mathrm{rank}\alpha = r$. More formally, we require:

Condition 2 *The matrices H and G and the vector h in (8)-(9) are such that* $\mathrm{rank}\alpha = \mathrm{rank}\beta = r$ *for all $(\phi', \gamma')' \in \mathbb{R}^{k+l}$, except for a possible set of Lebesgue measure zero.*

If Condition 2 is not satisfied, then the parameter space is void, and the maximum likelihood estimators of ϕ and γ are not defined. Note that the condition can easily be checked by substituting an arbitrary k-vector for ϕ and an l-vector for γ.

Theorem 2 *Consider the log-likelihood (7), subject to (8) and (9), satisfying Condition 2. The maximum likelihood estimators of ϕ, γ and Ω, each conditional upon the remaining parameters, are given by*

$$\begin{aligned}\hat{\phi}(\gamma,\Omega) &= \left[H'(\alpha'\Omega^{-1}\alpha\otimes S_{11})H\right]^{-1} \\ &\quad\times\left(H'(\alpha'\Omega^{-1}\otimes I_n)\mathrm{vec}S_{10}-H'(\alpha'\Omega^{-1}\alpha\otimes S_{11})h\right),\end{aligned} \tag{10}$$

$$\hat{\gamma}(\phi,\Omega)=\left[G'(\Omega^{-1}\otimes\beta'S_{11}\beta)G\right]^{-1}G'[\Omega^{-1}\otimes\beta']\mathrm{vec}S_{10}, \tag{11}$$

$$\hat{\Omega}(\phi,\gamma)=S_{00}-\alpha\beta'S_{10}-S_{01}\beta\alpha'+\alpha\beta'S_{11}\beta\alpha'. \tag{12}$$

Fisher's expected information matrix is block-diagonal with respect to $\theta=(\phi',\gamma')'$ and Ω; the observed information matrix for θ (excluding terms with zero expectation) is given by

$$J_\theta=T\begin{bmatrix}H'(\alpha'\Omega^{-1}\alpha\otimes S_{11})H & H'(\alpha'\Omega^{-1}\otimes S_{11}\beta)G\\ G'(\Omega^{-1}\alpha\otimes\beta'S_{11})H & G'(\Omega^{-1}\otimes\beta'S_{11}\beta)G\end{bmatrix}. \tag{13}$$

Corollary 1 *The maximum likelihood estimators of $\hat{\phi}$, $\hat{\gamma}$ and $\hat{\Omega}$ may be obtained from the following iterative procedure, starting from a set of initial values $\{\hat{\phi}_0,\hat{\gamma}_0,\hat{\Omega}_0\}$:*

$$\hat{\phi}_j=\hat{\phi}(\hat{\gamma}_{j-1},\hat{\Omega}_{j-1}),\quad \hat{\gamma}_j=\hat{\gamma}(\hat{\phi}_j,\hat{\Omega}_{j-1}),\quad \hat{\Omega}_j=\hat{\Omega}(\hat{\phi}_j,\hat{\gamma}_j),\quad j=1,2,\ldots$$

Remarks (a) Rothenberg (1971) analyzes the connection between (local) identifiability of a parameter vector and non-singularity of Fisher's information matrix. He focuses on the expected information matrix, but in cointegration models, which are non-ergodic (in certain directions), the observed information matrix should be used instead. From (13), we observe that the blocks on the diagonal are non-singular, provided that α, β, G and H all have full column rank. This means that given α, ϕ is always identified; similarly, γ is always identified given β. This is reflected in the fact that (10) and (11) provide unique estimators of ϕ and γ, each given the other (and given Ω). However, if both ϕ and γ are unknown, then what is required for identification is that the full information matrix is non-singular. This matrix can be expressed as

$$J_\theta=T\left[(\alpha\otimes I_n)H:(I_n\otimes\beta)G\right]'(\Omega^{-1}\otimes S_{11})\left[(\alpha\otimes I_n)H:(I_n\otimes\beta)G\right], \tag{14}$$

so that the rank condition for local identifiability of θ becomes

$$\mathrm{rank}\left[(\alpha\otimes I_n)H:(I_n\otimes\beta)G\right]=k+l. \tag{15}$$

If $G=I_{nr}$, then this condition reduces to $\mathrm{rank}[(\alpha\otimes I_n)H:(I_n\otimes\beta)]=k+nr$. The matrix $(I_n\otimes\beta)$ has full column rank nr; if we eliminate these columns by pre-multiplying with $(I_n\otimes\beta_\perp)'$, then (15) requires

$$\mathrm{rank}\left[(\alpha\otimes\beta_\perp)H\right]=\mathrm{rank}\left[(\alpha\otimes I_{n-r})(I_r\otimes\beta_\perp)'H\right]=k. \tag{16}$$

Since $(\alpha \otimes I_{n-r})$ has full column rank, (14)-(16) provide an alternative derivation of the rank condition (5).

(b) Following Johansen and Juselius (1992) and Johansen (1995a), the procedure in Corollary 1 may be called a *switching algorithm*, since we switch between optimization over ϕ, γ and Ω. This idea has a long-standing tradition in econometrics; examples based on the same idea are the Cochrane- Orcutt method (if iterated until convergence), or Zellner's seemingly unrelated regression equation estimator. Johansen (1995a) discusses the pros and cons of a switching algorithm in comparison with Newton-type algorithms. The most important advantage is that in each step the optimization is explicit; no special care has to be taken of singularity problems or step lengths. Since the likelihood increases with each step, the procedure will always converge eventually. On the other hand, this convergence may take quite some time, since the direction in which we move may be far from optimal. For example, if the log-likelihood is a quadratic function of two scalar parameters a and b, then a switching algorithm will be very slow if the contour ellipses of the log-likelihood lie in the direction of the diagonal ($a = b$) line; in this case Newton-Raphson will of course yield the optimum in one step.

(c) The switching algorithm does not require ϕ and γ to be identified: as long as both H and G have full column rank and Condition 2 is satisfied, the procedure will never break down due to singularity problems, and will eventually lead to some values of ϕ and γ that maximize the likelihood; of course, these are not unique if ϕ and γ are not identified. However, the corresponding estimated cointegrating space *is* unique, since this space is always identified. Thus the algorithm provides us with the ML estimate of the cointegrating space under the (possibly) over-identifying restrictions; since this estimate $\hat{\beta}$ lies in the parameter space $\mathcal{B}$, we can use it to check the rank condition (5). Note that if α is restricted ($G \neq I_{nr}$), it is possible that this condition is violated, whereas (14) is satisfied. If that is the case, ϕ and γ are identified, but $\hat{\phi}$ is not super-consistent; we shall return to this in the next Section. Observe that because identification is not required, the algorithm can also be applied to limited information systems, where only a subset of the vectors is restricted, see Remark (b), Section 2.

(d) Given that the switching algorithm may be relatively slow, the choice of starting values is an important one. As shown in the proof of Theorem 1, the defining equation for ϕ, for a given β spanning the cointegrating space, is

$$(I_r \otimes \beta_\perp)' H\phi = -(I_r \otimes \beta_\perp)' h. \tag{17}$$

From Johansen's ML estimate $\hat{\beta}$, we can calculate the unrestricted ML estimate of $\beta_\perp$ as the eigenvectors corresponding to the $(n-r)$ zero eigenvalues of the matrix $\hat{\beta}\hat{\beta}'$. Substituting this estimate in (17), we can try to solve the equation for ϕ. If the system is exactly identified

$(k = r(n-r))$, so that $(I_r \otimes \hat{\beta}_\perp)H$ is a non-singular $k \times k$ matrix, then the starting value $\hat{\phi}_0 = -[(I_r \otimes \hat{\beta}_\perp)'H]^{-1}(I_r \otimes \hat{\beta}_\perp)'h$ is the ML estimate, and no further iterations are required. If, on the other hand, there are over-identifying restrictions so that $r(n-r) > k$, then a possible starting value is the one that minimizes the length of the vector $(I_r \otimes \hat{\beta}_\perp)'(H\phi + h)$, i.e., the least-squares estimate

$$\hat{\phi}_0 = -[(I_r \otimes \hat{\beta}_\perp)'H]^+(I_r \otimes \hat{\beta}_\perp)'h,$$

where A^+ denotes the Moore-Penrose inverse of A. The same formula may be used if the rank condition is violated, although in that case $\hat{\phi}_0$ is only one of the many possible choices for ϕ that minimize the vector $(I_r \otimes \hat{\beta}_\perp)'(H\phi + h)$. For that case, it should be stressed that the choice of H and h should be such that the system is suitably normalized. If, e.g., $h = 0$, then $\hat{\phi}_0 = 0$, so that the starting value for the cointegrating matrix would be zero, in which case the algorithm breaks down. More generally, one should check whether $\hat{\phi}_0$ corresponds to a full rank matrix; if not, an alternative starting value should be chosen. However, this problem only arises in systems that are under-identified or violate Condition 2.

4 Asymptotic Properties

We now consider the asymptotic properties of the ML estimator defined in Theorem 2. Let $m = \text{rank}J_\theta$, and let C denote a semi-orthogonal $(k+l) \times m$ matrix (hence $C'C = I_m$), partitioned as $C = [C_1 : C_2]$, where

$$\text{sp}(C) = \text{sp}\begin{pmatrix} H'(\alpha' \otimes I_n) \\ G'(I_n \otimes \beta') \end{pmatrix}, \quad \text{sp}(C_1) = \text{sp}\begin{pmatrix} H'(I_r \otimes \beta_\perp) \\ 0 \end{pmatrix}.$$

To check that the column space of the second right-hand side matrix is contained in the column space of the first right-hand side matrix, post-multiply the first matrix by $(I_n \otimes \beta_\perp)$, yielding

$$\begin{pmatrix} H'(\alpha' \otimes \beta_\perp) \\ 0 \end{pmatrix} = \begin{pmatrix} H'(I_r \otimes \beta_\perp) \\ 0 \end{pmatrix}(\alpha' \otimes I_{n-r}).$$

Note that C spans the column space of J_θ in (14), whereas C_1 determines the linear functions of ϕ that satisfy (6). If the information matrix is non-singular, then $m = k + l$, and $C_\perp$ is void; otherwise, it is easily seen that the information on $C_\perp'\theta$ is zero, so that this function is not identifiable. Therefore, Theorem 3 only provides the asymptotic distribution of $C'\hat{\theta}$. First, we make the assumption of cointegration (of order 1,1) explicit, as well as a further assumption that will be required below:

Assumption 1 *The vector time series $\{x_t, t = 1, \ldots, T\}$ is generated by (1) with $\{\varepsilon_t\} \sim IN(0, \Omega)$, for particular parameter matrices α, β (of order $n \times r$), $\Gamma_i, i = 1, \ldots, p-1$, and positive definite Ω, such that*

1. $\text{rank}\alpha = \text{rank}\beta = r$, *and*

2. *the characteristic equation* $|\Gamma(z)(1-z) - \alpha\beta' z| = 0$ *has exactly* $(n-r)$ *roots equal to 1 and all other roots outside the unit circle, where* $\Gamma(z) = I_n - \sum_{i=1}^{p-1} \Gamma_i z^i$.

Assumption 2 *The true values of* β *and* α *can be expressed as (8) and (9) for some (possibly non-unique) value of* θ^0 *which is a regular point of* J_θ, *i.e., there exists an open neighbourhood* $N(\theta^0)$ *such that*

$$\forall \theta \in N(\theta^0) : \text{rank}(J_\theta) = m.$$

Let $\int_{V>0} N(0,V)\mathrm{d}P(V)$ denote a mixed normal distribution, i.e., a distribution that is $N(0,V)$ conditional upon the random matrix V; the integral is over all positive definite matrices V (denoted by $V > 0$).

Theorem 3 *Under Assumptions 1 and 2, and as* $T \to \infty$,

$$\begin{array}{rcl} TC_1'(\hat{\theta} - \theta) & \xrightarrow{D} & \int_{V_1>0} N(0,V_1)\mathrm{d}P(V_1), \\ \sqrt{T}C_2'(\hat{\theta} - \theta) & \xrightarrow{D} & N(0,V_2), \end{array} \tag{18}$$

independently, where V_1 *and* V_2 *are a random and a fixed matrix, respectively, defined by*

$$\left(T^{-2}C_1'J_\theta C_1\right)^{-1} \xrightarrow{D} V_1, \qquad \left(T^{-1}C_2'J_\theta C_2\right)^{-1} \xrightarrow{P} V_2. \tag{19}$$

The likelihood ratio statistic LR for the over-identifying restrictions satisfies

$$LR \xrightarrow{D} \chi^2(d), \tag{20}$$

where d, *the degree of over-identification, is given by* $d = r(n-r) + rn - m$.

Corollary 2 *For an arbitrary known* $m \times k$ *matrix* F,

1. $F(\hat{\phi} - \phi) = O_p(T^{-1})$ *if and only if condition (6) is satisfied, so that* $\text{sp}([F:0]') \subseteq \text{sp}(C_1)$;

2. $F(\hat{\phi} - \phi) = O_p(T^{-1/2})$ *if and only if* $\text{sp}([F:0]') \subseteq \text{sp}(C) = \text{sp}(J_\theta)$.

Remarks (a) Theorem 3, together with Corollary 2, determines whether a certain linear combination of ϕ (and γ) can be estimated consistently, i.e., whether it is identifiable. Letting the matrix F equal the identify matrix, we observe from the corollary that $\hat{\phi}$ may be consistent even if the rank condition (5)-(6) is not satisfied, provided that there are sufficient restrictions on α, so that $\text{sp}([F:0]') \subseteq \text{sp}(J_\theta)$; see also Remark (e), Section 2. However, $\hat{\phi}$ is only super-consistent if the rank condition (5) is satisfied.

(b) In order to analyze the role of Assumption 2, consider the following example where $n = 4, r = 2, k = 2$ and $l = 8$ (α is unrestricted), and where

$$\beta_\phi = \begin{bmatrix} 1 & 0 \\ \phi_1 & 1 \\ 0 & \phi_{21} \\ 0 & \phi_{22} \end{bmatrix}.$$

It is easily seen that ϕ and hence θ is generically identified, which implies that J_θ has rank 11 almost everywhere in the parameter space; the corresponding degree of over-identification equals 1. If $\phi_2 = (\phi_{21}, \phi_{22})' = 0$ however, then ϕ_1 is not identified, and this is reflected in the rank of J_θ reducing to 10. A parameter point θ^0 with $\phi_2 = 0$ is not a regular point of J_θ: any open neighbourhood of θ^0 contains parameter points with $\phi_2 \neq 0$. This implies that the LR statistic for the over-identifying restrictions does not have an asymptotic $\chi^2(1)$ distribution, which would be the case if θ^0 were a regular point. In fact, from the analysis of Boswijk (1996) it follows that the asymptotic distribution of LR is stochastically dominated by a $\chi^2(1)$ distribution.

(c) In a generically identified model, Assumption 2 is violated only in the measure-zero set of non-identifiable parameter points. For the specific case where α is unrestricted and H is block-diagonal (Condition 1) a test for this hypothesis is analyzed by Boswijk (1996) . A generalization of such a test to the present context would entail a test for the null hypothesis

$$\mathcal{H}_0 : \text{rank} \begin{pmatrix} H'(\alpha' \otimes I_n) \\ G'(I_n \otimes \beta') \end{pmatrix} \leq k + l - 1.$$

If α is unrestricted, this reduces to

$$\mathcal{H}_0 : \text{rank}\left(H'(I_r \otimes \beta_\perp)\right) \leq k - 1 \quad \Leftrightarrow \quad \text{rank}\left(R'(I_r \otimes \beta)\right) \leq r^2 - 1.$$

In principle, such restrictions can be tested in generically identified models. In practice however, it is not clear how to construct likelihood ratio or even Wald statistics for these hypotheses, because they are not formulated as explicit smooth restrictions $g(\alpha, \beta) = 0$.

(d) The results of Theorem 3 are not directly applicable, because the matrices C_1 and C_2 depend upon the true value. The best we can do is evaluate the observed information matrix in the ML estimate, and determine its rank and column space. First, the rank can be used to obtain the correct number of degrees of freedom of the LR test for the over-identifying restrictions, provided that both $\hat{\theta}$ and θ are regular points of J_θ. Second, the column space will indicate whether a linear function of $\hat{\theta}$ is unique, in the sense that it is the only one maximizing the likelihood. If it is, then an estimate of its covariance matrix may be obtained from the

information matrix. In particular, letting $A\theta$ denote a linear function of interest, identifiability requires that $A = BC'$ for some matrix B of full row rank. In that case AJ_θ^+A' may be used as an asymptotic covariance matrix of $A\hat{\theta}$, where

$$J_\theta^+ = C(C'J_\theta C)^{-1}C',$$

the Moore-Penrose inverse of J_θ. Note that $AJ_\theta^+A' = B(C'J_\theta C)^{-1}B'$, which is the inverse of the concentrated information on $BC'\theta = A\theta$. Letting $\hat{J}_\theta$ denote the estimated observed information matrix, $A\hat{J}_\theta^+A'$ is a consistent estimate of the covariance matrix of $A\hat{\theta}$, in the sense that

$$(A\hat{J}_\theta^+A')^{-1/2}A(\hat{\theta}-\theta) \xrightarrow{D} N(0,I),$$

still provided that $A\theta$ is identifiable.

5 Discussion

In this paper we have analyzed conditions for identifiability of the parameters characterizing the cointegrating vectors and adjustment coefficients in cointegrated VAR models. The conditions that have been derived are always expressed in terms of the parameters, which has the disadvantage that it does not enable the researcher to check generic identifiability, cf. Johansen (1995a). As a solution to this problem we may substitute the restricted ML estimates in the conditions, and thus check whether the ML estimate is unique. If not, then this may be caused either by the fact that the model is not generically identified, or because the ML estimate is a singular point. In both cases neither a unique ML estimate is available, nor an estimate of its variance, since the estimated information matrix will be singular. Note that in such cases it is still possible that some functions of the parameter vector are identifiable; conditions for this have also been derived. A second and complementary way to investigate identifiability is to test the hypothesis that the parameter lies in the set of non-identifiable structures; this is analyzed in Boswijk (1996).

It should be stressed that identification has only been considered here in the mathematical sense, indicating uniqueness, and related to the traditional statistical problem of estimable functions of parameters. Hendry (1993) lists two more attributes of identification, viz. "'correspondence to the desired entity', and 'satisfying the assumed interpretation (usually of a theory model)'" (Hendry, 1993, p.11). Although we haven't paid any explicit attention to these connotations of identification, we have noted that some classes of restrictions to obtain uniqueness may not provide interpretability, whereas others will in fact be inspired by such considerations. For example, one could argue that Johansen's (1988) reduced rank regression estimator is unique, provided that we restrict the ith vector in β to correspond to the ith canonical correlation between Δx_t and x_{t-1} (in addition to some normalization restriction). However, there is no guarantee that

these vectors are interpretable or correspond the "the desired entity" (e.g., a money demand function or a purchasing power parity relationship). Therefore, explicit identifying restrictions are preferable if the individual cointegrating vectors are to be interpreted.

Appendix: Proofs

Proof of Theorem 1: Regardless of any identifying restrictions, the cointegrating space, and hence an arbitrary basis for this space, is identifiable; Johansen's maximum likelihood estimator, which is unique up to rotations which span the same space, provides a consistent estimator of this space. Thus the question is whether a unique value of ϕ can be obtained from β (or $\beta_\perp$, which can be derived from β) and the restrictions (3)-(4). In other words, the question is whether ϕ is an estimable function of β.

Since the restricted matrix of cointegrating vectors β_ϕ should span the cointegrating space, it should satisfy $\beta_\perp'\beta_\phi = 0$. Vectorizing this condition yields

$$(I_r \otimes \beta_\perp)'\text{vec}\beta_\phi = (I_r \otimes \beta_\perp)'(H\phi + h) = 0,$$

so that the defining equation for ϕ is

$$(I_r \otimes \beta_\perp)'H\phi = -(I_r \otimes \beta_\perp)'h. \tag{A.1}$$

This system only has a unique solution if the $r(n-r) \times k$ matrix $(I_r \otimes \beta_\perp)'H$ is of full column rank, i.e., if rank $((I_r \otimes \beta_\perp)'H) = k$.

Suppose that $(I_r \otimes \beta_\perp)'H$ has a deficient column rank; then there exists a non-null vector x such that $(I_r \otimes \beta_\perp)'Hx = 0$. Because H has full column rank, $Hx \neq 0$ for all $x \neq 0$; since this vector lies in the null space of $(I_r \otimes \beta_\perp)'$, it can be expressed as $(I_r \otimes \beta)y$ for some non-null vector y. For that same y, we have $R'(I_r \otimes \beta)y = R'Hx = 0$, so that $R'(I_r \otimes \beta)$ has a deficient column rank. The same argument can be reversed, so that a deficient column rank of $(I_r \otimes \beta_\perp)'H$ implies and is implied by a deficient column rank of $R'(I_r \otimes \beta)$. This in turn implies that a full column rank (k) of the former matrix is equivalent to a full column rank (r^2) of the latter matrix.

Let $\mathcal{K}$ denote the null space of $(I_r \otimes \beta_\perp)'H$ (possibly void). If ϕ is a solution of (A.1), then so is $(\phi + x)$ for any $x \in \mathcal{K}$. Thus $F\phi$ is observationally equivalent to the set of all points $\{F(\phi + x), x \in \mathcal{K}\}$. If (and only if) condition (6) is satisfied, then $Fx = 0, \forall x \in \mathcal{K}$, so that this set contains only one point; thus in that case F is identifiable. □

Proof of Theorem 2: Using $\text{tr}AB = \text{tr}BA = (\text{vec}A')'\text{vec}B$ for matrices A and B of conformable orders, the derivatives of the log-likelihood with respect to ϕ and γ are:

$$\frac{\partial \ln L}{\partial \phi} = TH'\text{vec}(S_{10}\Omega^{-1}\alpha) - TH'\text{vec}(S_{11}\beta\alpha'\Omega^{-1}\alpha), \tag{A.2}$$

$$\frac{\partial \ln L}{\partial \gamma} = TG'\text{vec}(\beta' S_{10}\Omega^{-1}) - TG'\text{vec}(\beta' S_{11}\beta\alpha'\Omega^{-1}). \tag{A.3}$$

Substituting (8) in (A.2) and solving for ϕ yields (10); similarly, (9) and (A.3) lead to (11). The expression for $\hat{\Omega}(\phi, \gamma)$, as well as the block-diagonality of the information matrix are standard

results for ML estimation of multivariate regression models with Gaussian disturbances, see, e.g., Magnus and Neudecker (1988) (note that (12) equals the average outer product of the residual vector). The diagonal blocks of the information matrix simply follow from differentiation. For the off-diagonal block, we have

$$\frac{\partial \ln L}{\partial \phi \partial \gamma'} = -TH'(\Omega^{-1}\alpha \otimes S_{11}\beta)G + TH'(I_r \otimes [S_{10} - S_{11}\beta\alpha']\Omega^{-1})KG, \tag{A.4}$$

with K the commutation matrix of appropriate order, so that $\text{vec}\alpha = K\text{vec}\alpha'$ (cf. Magnus and Neudecker, 1988). The matrix in square brackets in (A.4) has expectation zero if evaluated in the true value; it also vanishes asymptotically if the observed information matrix is suitably standardized, which is why it is omitted from (13). □

Proof of Theorem 3: Define the orthogonal matrix $A = [A_1 : A_2]$, where $A_1 = [I_n \otimes \beta_\perp]$, and $A_2 = [I_n \otimes \beta(\beta'\beta)^{-1/2}]$; note that the matrix $\beta_\perp$ may be taken semi-orthogonal without loss of generality. Next, write the observed information matrix as

$$J_\theta = \begin{bmatrix} H'(\alpha' \otimes I_n) \\ G'(I_n \otimes \beta') \end{bmatrix} AA'(\Omega^{-1} \otimes S_{11})AA' \begin{bmatrix} H'(\alpha' \otimes I_n) \\ G'(I_n \otimes \beta') \end{bmatrix}' = B\tilde{J}_\theta B', \tag{A.5}$$

where

$$B = \begin{bmatrix} H'(I_r \otimes \beta_\perp) & H'(\alpha' \otimes \beta(\beta'\beta)^{-1}) \\ 0 & G' \end{bmatrix},$$

$$\tilde{J}_\theta = \begin{bmatrix} (\alpha'\Omega^{-1}\alpha) \otimes (\beta_\perp' S_{11}\beta_\perp) & (\alpha'\Omega^{-1}) \otimes (\beta_\perp' S_{11}\beta) \\ (\Omega^{-1}\alpha) \otimes (\beta' S_{11}\beta_\perp) & \Omega^{-1} \otimes (\beta' S_{11}\beta) \end{bmatrix}. \tag{A.6}$$

In a similar fashion, the score vector in (A.2)-(A.3) can be summarized and rewritten as

$$q_\theta = \frac{\partial \ln L}{\partial \theta} = T \begin{bmatrix} H'(\alpha' \otimes I_n) \\ G'(I_n \otimes \beta') \end{bmatrix} AA'\text{vec}\left((S_{10} - S_{11}\beta\alpha')\Omega^{-1}\right) = B\tilde{q}_\theta,$$

where

$$\tilde{q}_\theta = T \begin{pmatrix} \text{vec}((\beta_\perp' S_{10} - \beta_\perp' S_{11}\beta\alpha')\Omega^{-1}) \\ \text{vec}((\beta' S_{10} - \beta' S_{11}\beta\alpha')\Omega^{-1}) \end{pmatrix}.$$

Let $N(\theta^0)$ denote a neighbourhood of the true value where $\text{rank}J_\theta = m$. For any $\theta \in N(\theta^0)$, let C_θ denote a semi-orthogonal matrix of full column rank spanning $\text{sp}(J_\theta)$, such that in the true value, $C_\theta = C$. Choose $C_{\theta\perp}$ such that $[C_\theta : C_{\theta\perp}]$ is orthogonal. Now consider the reparametrization $\lambda = f(\theta)$, where (for $\theta \in N(\theta^0)$) $\partial f/\partial\theta' = [C_\theta : C_{\theta\perp}]'$; partition $\lambda = (\lambda_1', \lambda_2')'$ conformably with $[C_\theta : C_{\theta\perp}]$. Note that for $\theta \in N(\theta^0)$, this reparametrization is one to one, because the derivative is non-singular; and because this derivative matrix is orthogonal, the inverse transformation $\theta = f^{-1}(\lambda)$ has derivative $[C_\theta : C_{\theta\perp}]$. Therefore,

$$q_\lambda = \frac{\partial \ln L}{\partial \lambda} = \begin{pmatrix} C_\theta' q_\theta \\ C_{\theta\perp}' q_\theta \end{pmatrix} = \begin{pmatrix} q_{\lambda_1} \\ 0 \end{pmatrix},$$

$$J_\lambda = \frac{\partial^2 \ln L}{\partial\lambda\partial\lambda'} = [C_\theta : C_{\theta\perp}]' J_\theta [C_\theta : C_{\theta\perp}] = \begin{bmatrix} J_{\lambda_1} & 0 \\ 0 & 0 \end{bmatrix},$$

i.e., the parts of the score vector and the information matrix that correspond to λ_2 are all equal to zero. This in turn implies that all point λ (corresponding to $N(\theta^0)$) that have the same value of λ_1 are observationally equivalent, i.e., λ_2 is not identifiable. Thus we may restrict λ_2 to an arbitrary value (e.g. zero), without restricting the log-likelihood. The information matrix for the remaining free parameters, $J_{\lambda_1} = C_\theta' J_\theta C_\theta$, is non-singular by construction, so that λ_1 is identified. From this, consistency of the (hypothetical) ML estimator $\hat{\lambda}_1$ can be proved, so that we can use a Taylor series expansion around the true value (so that $C_\theta = C$):

$$\Upsilon_T(\hat{\lambda}_1 - \lambda_1) = \Upsilon_T C'(\hat{\theta} - \theta) + o_p(1) = \left[\Upsilon_T^{-1} C' J_\theta C \Upsilon_T^{-1}\right]^{-1} \Upsilon_T^{-1} C' q_\theta + o_p(1),$$

where Υ_T is a scaling matrix chosen such that the normalized sore vector and observed information matrix converge (recall that J_θ is equal to minus the Hessian matrix, except for an asymptotically negligible term with expectation zero). Letting $m_1 = \text{rank}C_1$ and $m_2 = \text{rank}C_2$ ($m = m_1 + m_2$), it will follow from the subsequent derivations that $\Upsilon_T = \text{diag}(T \cdot I_{m_1}, \sqrt{T} \cdot I_{m_2})$ is a suitable choice.

Partition $B = [B_1 : B_2]$ in an obvious fashion, and note that $\text{sp}(B) = \text{sp}(C)$ and $\text{sp}(B_1) = \text{sp}(C_1)$. This implies that we may express B as $B = CD$, and therefore $C'B = D$, where D is a matrix of appropriate order, with full row rank, partitioned as $D = (D_{ij}), i, j = 1, 2$ (note that $D_{21} = 0$). This leads to

$$\begin{aligned} \Upsilon_T C'(\hat{\theta} - \theta) &= \left[\Upsilon_T^{-1} D \tilde{J}_\theta D' \Upsilon_T^{-1}\right]^{-1} \Upsilon_T^{-1} D \tilde{q}_\theta \\ &= \left[(\Upsilon_T^{-1} D \tilde{\Upsilon}_T)(\tilde{\Upsilon}_T^{-1} \tilde{J}_\theta \tilde{\Upsilon}_T^{-1})(\tilde{\Upsilon}_T^{-1} D' \Upsilon_T^{-1})\right]^{-1} (\Upsilon_T^{-1} D \tilde{\Upsilon}_T)(\tilde{\Upsilon}_T^{-1} \tilde{q}_\theta), \end{aligned}$$

where we have defined the scaling matrix $\tilde{\Upsilon}_T = \text{diag}(T \cdot I_{r(n-r)}, \sqrt{T} \cdot I_{nr})$. From the general results of Johansen (1991), it can be deduced that

$$\tilde{\Upsilon}_T^{-1} \tilde{J}_\theta \tilde{\Upsilon}_T^{-1} \xrightarrow{D} \begin{bmatrix} (\alpha'\Omega^{-1}\alpha) \otimes \int_0^1 FF' \mathrm{d}t & 0 \\ 0 & \Omega^{-1} \otimes \Sigma_{\beta\beta} \end{bmatrix},$$

$$\tilde{\Upsilon}_T^{-1} \tilde{q}_\theta \xrightarrow{D} \begin{pmatrix} \text{vec} \int_0^1 F \mathrm{d}B' \\ N(0, \Omega^{-1} \otimes \Sigma_{\beta\beta}) \end{pmatrix},$$

where $F(t)$ is an $(n-r)$-vector Brownian motion process on $[0, 1]$ with positive definite covariance matrix, and $B(t)$ is an r-vector Brownian motion process, independent of $F(t)$, with covariance matrix $(\alpha'\Omega^{-1}\alpha)$. Furthermore, $\Sigma_{\beta\beta} = \text{Var}[\beta' x_t | w_t]$. Because $\Upsilon_T^{-1} D \tilde{\Upsilon}_T \to \text{diag}(D_{11}, D_{22})$, it follows that

$$\left[\Upsilon_T^{-1} C' J_\theta C \Upsilon_T^{-1}\right]^{-1} \xrightarrow{D} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix}, \tag{A.7}$$

with

$$V_1 = \left(D_{11} \left(\alpha' \Omega^{-1} \alpha \otimes \int_0^1 FF' \mathrm{d}t \right) D_{11}' \right)^{-1}, \quad V_2 = \left(D_{22} (\Omega^{-1} \otimes \Sigma_{\beta\beta}) D_{22}' \right)^{-1},$$

and

$$\Upsilon_T C'(\hat{\theta} - \theta) \quad \xrightarrow{D} \quad \begin{pmatrix} V_1 D_{11} \mathrm{vec} \int_0^1 F \mathrm{d}B' \\ V_2 D_{22} N(0, \Omega^{-1} \otimes \Sigma_{\beta\beta}) \end{pmatrix} = \begin{pmatrix} \int_{V_1 > 0} N(0, V_1) \mathrm{d}P(V_1) \\ N(0, V_2) \end{pmatrix}, \qquad \text{(A.8)}$$

which proves (18)-(19). The independence of the two components in (A.8) follows from the block-diagonality of the limiting variance in (A.7).

Finally, the limiting χ^2 distribution of the LR statistic follows from (mixed) normality of the estimators of the identified components of α and β in the unrestricted model, see Johansen (1991, Appendix C). It follows from (A.5)-(A.6) that the number of identified parameters in the unrestricted model is $r(n-r) + nr$. This can also be seen from the rank condition in (5): if $G = I_{nr}$, then α contains nr unrestricted parameters, and the rank condition limits the number of identified parameters in β to $r(n-r)$. Since the number of identified parameters in the restricted model is m (in a neighbourhood of the true value), the number of effective over-identifying restrictions is $d = r(n-r) + nr - m$. These can be expressed as smooth restrictions $g(\alpha, \beta) = 0$, with a derivative $\mathsf{D}g(\alpha, \beta)$ which is of full column rank, provided that the true value is not a singular point of J_θ. This proves (20). $\square$

References

Boswijk, H.P. (1994), "Testing for an Unstable Root in Conditional and Structural Error Correction Models", *Journal of Econometrics*, 63, 37-60.

Boswijk, H.P. (1995), "Efficient Inference on Cointegration Parameters in Structural Error Correction Models", *Journal of Econometrics*, 69, 133-158.

Boswijk, H.P. (1996), "Testing Identifiability of Cointegrating Vectors", *Journal of Business and Economics Statistics*, 14,153-160.

Breitung, J. (1994), "A Simultaneous Equations Approach to Cointegrated Systems", Working paper, Humboldt University of Berlin.

Cragg, J.D. and S.G. Donald (1993), "Testing Identifiability and Specification in Instrumental Variable Models", *Econometric Theory*, 9, 222-240.

Engle, R.F. and C.W.J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 251-276.

Hendry, D.F. (1993), "The Roles of Economic Theory and Econometrics in Time Series Economics", Invited paper presented at the Econometric Society European Meeting, Uppsala, Sweden.

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. (1992), "Cointegration in Partial Systems and the Efficiency of Single-Equation Analysis", *Journal of Econometrics*, 52, 389-402.

Johansen, S. (1995a), "Identifying Restrictions of Linear Equations – With Applications to Simultaneous Equations and Cointegration", *Journal of Econometrics*, 69, 111-132.

Johansen, S. (1995b), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Johansen, S. and K. Juselius (1992), "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK", *Journal of Econometrics*, 53, 211-244.

Koopmans, T.C. and W.C. Hood (1953), "The Estimation of Simultaneous Linear Economic Relationships", in W.C. Hood and T.C. Koopmans (eds.), *Studies in Econometric Method*. New York: John Wiley.

Magnus, J.R. and H. Neudecker (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester: John Wiley.

Park, J.Y. (1990), "Maximum Likelihood Estimation of Simultaneous Cointegrated Models", Memo 1990-18, University of Aarhus.

Pesaran, M.H. and Y. Shin (1994), "Long-Run Structural Modelling", Working paper, Univer-

sity of Cambridge.

Phillips, P.C.B. (1991), "Optimal Inference in Cointegrated Systems", *Econometrica*, 59, 283-306.

Phillips, P.C.B. and S.N. Durlauf (1986), "Multiple Time Series Regression with Integrated Processes", *Review of Economic Studies*, 53, 473-495.

Richmond, J. (1974), "Identifiability in Linear Models", *Econometrica*, 42, 731-736.

Rothenberg, T.J. (1971), "Identification in Parametric Models", *Econometrica*, 39, 577-591.

Saikkonen, P. (1993), "Estimation of Cointegration Vectors with Linear Restrictions", *Econometric Theory*, 9, 19-35.

Stock, J.H. (1987), "Asymptotic Properties of Least-Squares Estimators of Cointegrating Vectors", *Econometrica*, 55, 1035-1056.