Identifying, Estimating and Testing Restricted Cointegrated Systems: An Overview*

H. Peter Boswijk
UvA-Econometrics & Tinbergen Institute
Department of Quantitative Economics
Universiteit van Amsterdam †

Jurgen A. Doornik
Nuffield College
University of Oxford

January 2003

Abstract

The notion of cointegration has lead to a renewed interest in the identification and estimation of structural relations among economic time series, a field to which Henri Theil has made many pioneering contributions. This paper reviews the different approaches that have been put forward in the literature for identifying cointegrating relationships and imposing (possibly over-identifying) restrictions on them. Next, various algorithms to obtain (approximate) maximum likelihood estimates and likelihood ratio statistics are reviewed, with an emphasis on so-called switching algorithms. The implementation of these algorithms is discussed and illustrated using an empirical example.

1 Introduction

The need to analyse simultaneous structural relations between economic time series has been one of the main driving forces behind the development of econometrics as a separate discipline in the previous century. This lead to the influential work of the Cowles Commission on identification and likelihood-based estimation of simultaneous equations (see Hood and Koopmans, 1953), and subsequently to the computationally more attractive two-stage and three-stage least-squares methods of Theil (1953) and Zellner and Theil (1962).

In the following decades, it was realized that the dynamics of economic relations, and the statistical properties of economic time series, should be incorporated in these simultaneous equations models. Via the work of Theil and Boot (1962) and Zellner and Palm (1974) on dynamic simultaneous equations models, the concept of error correction mechanisms developed by Sargan (1964) and Davidson *et al.* (1978), and Sims's (1980) analysis of vector autoregressive models, this has lead to cointegration and

*A first version of this paper was presented at the Henri Theil Memorial Conference, Universiteit van Amsterdam, 16–18 August 2002. Helpful comments and suggestions from David Hendry and Søren Johansen are gratefully acknowledged.

†Address for correspondence: Department of Quantitative Economics, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands; e-mail: H.P.Boswijk@uva.nl.

vector error correction modelling (Engle and Granger, 1987) as the currently dominant approach to the econometric analysis of time series.

The problem of testing for cointegration and estimating the cointegrating relationships was solved by Johansen (1988, 1991), using reduced rank regression techniques developed by Anderson (1951). However, in the presence of multiple cointegrating relations, the resulting estimates are not unique and directly interpretable, unless some identifying restrictions are imposed. Therefore, the familiar identification problem of linear simultaneous structural relations reappears in vector error correction models, the main new element being that these relations are now embedded as error correction terms in a dynamic model.

Over the last decade, a number of approaches to identify and restrict multiple cointegration relations have been proposed in the literature, notably by Johansen (1988, 1991, 1995a,b), Johansen and Juselius (1990, 1992, 1994), Boswijk (1995), Doornik (1995), Elliott (2000), Hansen (2002), and Pesaran and Shin (2002). This paper reviews these approaches, and discusses the algorithms needed to apply these approaches in practice, as well as their implementation in some econometric software packages. An empirical example is used to illustrate the approaches.

The paper only considers identifying restrictions on the cointegrating vectors and adjustment coefficients. Identifying restrictions on short-run parameters, and the analysis of vector error correction models in structural form, is not discussed here; see, e.g., Johansen and Juselius (1994). Furthermore, the analysis is limited to processes integrated of order one. In $I(2)$ cointegration models similar identification issues arise, but the deviations from mixed normal inference in these models, as analysed in Boswijk (2000) and Johansen (2002a), lead to specific complications that go beyond the scope of the present paper.

The outline of the rest of the paper is as follows. In Section 2, the unrestricted cointegration model and the reduced rank regression procedure are discussed. Furthermore, an expression for the information matrix for the parameters of interest is obtained, and this is used to analyse the identification problem and the asymptotic properties of the unrestricted maximum likelihood estimators. Section 3 discusses identification and asymptotic results for a general class of restrictions on the cointegrating vectors and adjustment coefficients. In Section 4, various specific classes of restrictions are reviewed, together with the corresponding algorithms to maximize the likelihood function. The implementation of these algorithms in some econometric packages is also discussed. In Section 5, an empirical example of the various approaches discussed in the paper is considered.

2 The Model and Unrestricted Statistical Analysis

2.1 The Model

Consider the kth order vector error correction model (VECM) for a p-vector time series $\{X_t\}$:

$$\begin{aligned}\Delta X_t &= \Pi_x X_{t-1} + \Pi_d d_t + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Upsilon q_t + \varepsilon_t, \\ &= \Pi X^*_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Upsilon q_t + \varepsilon_t, \qquad t = 1, \ldots, T, \end{aligned} \tag{1}$$

where the starting values $(X_{1-k}, \ldots, X_0)$ are fixed, ε_t is i.i.d. $N(0, \Omega)$, and $X^*_{t-1} = (X'_{t-1}, d'_t)'$, where d_t and q_t are deterministic regressors, such as a constant and trend term, and (seasonal) dummy variables. The most common two specifications for these deterministic variables are $(d_t, q_t) = (1, \varnothing)$ (restricted constant, excluding a linear drift in X_t) and $(d_t, q_t) = (t, 1)$ (restricted linear trend, excluding a quadratic trend in X_t). The normality assumption on ε_t is made primarily for constructing the likelihood function; asymptotic results may be obtained under weaker conditions.

The dynamics of the process is determined by the roots of the characteristic equation

$$\phi(z) := \left| I_p(1-z) - \Pi_x z - \sum_{i=1}^{k-1} \Gamma_i z^i (1-z) \right| = 0. \tag{2}$$

If all roots of (2) are outside the unit circle, then Π_x has full row rank p and the process is (trend-) stationary. The process is integrated of order 1 ($I(1)$) if $\operatorname{rank} \Pi_x = r < p$, and (2) has $p - r$ roots equal to one and all other roots outside the unit circle (see Johansen, 1995a, Corollary 4.3). If this $I(1)$ condition holds with $r > 0$, and if $\operatorname{sp}(\Pi_d) \subseteq \operatorname{sp}(\Pi_x)$,[1] then we may write $\Pi = \alpha\beta'$, with α and β full column rank matrices of dimensions $p \times r$ and $p_1 \times r$, respectively, where $p_1 = \dim(X^*_t) = p + \dim(d_t)$. In that case the system is cointegrated, such that the r linear combinations $\beta' X^*_t$ are (trend-) stationary even though $X_t \sim I(1)$.

2.2 Reduced Rank Regression

The statistical analysis of the $I(1)$ cointegration model is described in detail in Johansen (1991, 1995a), the main elements of which are summarized here. Let $\Gamma = (\Gamma_1, \ldots, \Gamma_{k-1}, \Upsilon)$ and $W_t = (\Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1}, q'_t)'$, so that the model, with the reduced rank condition $\Pi = \alpha\beta'$ imposed, reads

$$\Delta X_t = \alpha\beta' X^*_{t-1} + \Gamma W_t + \varepsilon_t, \qquad t = 1, \ldots, T. \tag{3}$$

Since no restrictions will be imposed on Γ, we consider the concentrated log-likelihood function in terms of the remaining (freely varying) parameters (α, β, Ω), which up to a constant term is given by

$$\ell_c(\alpha, \beta, \Omega) = -\frac{T}{2} \log|\Omega| - \frac{T}{2} \operatorname{tr} \Omega^{-1} \left(S_{00} - S_{01}\beta\alpha' - \alpha\beta' S_{10} + \alpha\beta' S_{11} \beta\alpha' \right), \tag{4}$$

[1] If the columns of Π_d do not lie in $\operatorname{sp}(\Pi_x)$, then one should change the specification by moving some or all of the deterministic variables from d_t to q_t.

where

$$\begin{aligned} S_{00} &= \frac{1}{T}\left(\Delta X'\Delta X - \Delta X'W[W'W]^{-1}W'\Delta X\right), \\ S_{11} &= \frac{1}{T}\left(X_{-1}^{*\prime}X_{-1}^{*} - X_{-1}^{*\prime}W[W'W]^{-1}W'X_{-1}^{*}\right), \\ S_{01} = S_{10}' &= \frac{1}{T}\left(\Delta X'X_{-1}^{*} - \Delta X'W[W'W]^{-1}W'X_{-1}^{*}\right), \end{aligned}$$

with ΔX, X_{-1}^{*} and W the data matrices (consisting of T rows) of ΔX_t, X_{t-1}^{*} and W_t, respectively. For a given value of β, this log-likelihood function is maximized by $\hat{\alpha}(\beta) = S_{01}\beta(\beta'S_{11}\beta)^{-1}$ and $\hat{\Omega}(\beta) = S_{00} - S_{01}\beta(\beta'S_{11}\beta)^{-1}\beta'S_{10}$, which leads to the further concentrated log-likelihood function in terms of β only:

$$\begin{aligned} \ell_c(\beta) &= -\frac{T}{2}\log\left|S_{00} - S_{01}\beta(\beta'S_{11}\beta)^{-1}\beta'S_{10}\right| \\ &= -\frac{T}{2}\log|S_{00}|\,\frac{\left|\beta'\left(S_{11} - S_{10}S_{00}^{-1}S_{01}\right)\beta\right|}{\left|\beta'S_{11}\beta\right|}, \end{aligned} \tag{5}$$

where a further constant term has been omitted.

Recognizing (5) as a reduced rank regression problem, analysed originally by Anderson (1951), Johansen (1988, 1991) has shown that this concentrated log-likelihood function is maximized by $\hat{\beta} = (\hat{v}_1, \ldots, \hat{v}_r)$, where $\hat{v}_i$ are the eigenvalues corresponding to the eigenvalues $\hat{\lambda}_i$, in descending order, of the generalized eigenvalue problem

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

Since these eigenvectors satisfy the normalization $\hat{v}_i S_{11}\hat{v}_i = 1$ and $\hat{v}_i' S_{11}\hat{v}_j = 0, i \neq j$, it follows that $\hat{\beta}'S_{11}\hat{\beta} = I_r$, which leads to $\hat{\alpha} = S_{01}\hat{\beta}$ and $\hat{\Omega} = S_{00} - S_{01}\hat{\beta}\hat{\beta}'S_{10} = S_{00} - \hat{\alpha}\hat{\alpha}'$. It also implies that $\hat{\beta}'S_{10}S_{00}^{-1}S_{10}\hat{\beta} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \lambda_r)$, which in turn leads to

$$\ell_c(\hat{\beta}) = -\frac{T}{2}\left(\log|S_{00}| + \sum_{i=1}^{r}\log(1-\hat{\lambda}_i)\right).$$

Since these results also apply in case $r = p$, where the rank of Π is unrestricted, it follows that the likelihood ratio statistic for the hypothesis $\mathcal{H}_r : \mathrm{rank}\,\Pi \leq r$ in the general model $\mathcal{H}_p$ is given by

$$LR(\mathcal{H}_r|\mathcal{H}_p) = -T\sum_{i=r+1}^{p}\log(1-\hat{\lambda}_i). \tag{6}$$

The asymptotic null distribution of this statistic, expressed in terms of vector Brownian motion functionals, is derived and tabulated in, e.g., Johansen (1995a), and depends on the specification of the deterministic variables (d_t, q_t).

2.3 The Information Matrix

In order to characterize the asymptotic properties of $\hat{\alpha}$ and $\hat{\beta}$, and discuss the identification problem, it is useful to obtain the Fisher information matrix, which we derive from (4). Note first that, with $\Pi = \alpha\beta'$,

$$\begin{aligned} S_{00} - S_{01}\beta\alpha' - \alpha\beta'S_{10} + \alpha\beta'S_{11}\beta\alpha' &= S_{00} - S_{01}S_{11}^{-1}S_{10} + (\Pi - S_{01}S_{11}^{-1})S_{11}(\Pi - S_{11}^{-1}S_{10}) \\ &= \hat{\Omega}_{LS} + (\Pi - \hat{\Pi}_{LS})S_{11}(\Pi - \hat{\Pi}_{LS})', \end{aligned}$$

where $\hat{\Pi}_{LS} = S_{01}S_{11}^{-1}$ and $\hat{\Omega}_{LS} = S_{00} - S_{01}S_{11}^{-1}S_{10}$, the unrestricted maximum likelihood estimators of Π and Ω, which are obtained by applying least-squares to (1). This leads to

$$\begin{aligned}
\ell_c(\Pi,\Omega) &= -\frac{T}{2}\log|\Omega| - \frac{T}{2}\operatorname{tr}\Omega^{-1}\left[\hat{\Omega}_{LS} + (\Pi-\hat{\Pi}_{LS})S_{11}(\Pi-\hat{\Pi}_{LS})'\right] \\
&= -\frac{T}{2}\left(\log|\Omega| + \operatorname{tr}\Omega^{-1}\hat{\Omega}_{LS}\right) \\
&\quad -\frac{T}{2}\operatorname{vec}(\Pi'-\hat{\Pi}'_{LS})'\left(\Omega^{-1}\otimes S_{11}\right)\operatorname{vec}(\Pi'-\hat{\Pi}'_{LS}),
\end{aligned} \tag{7}$$

where $\operatorname{vec}(A)$ stacks the columns of A, and the Kronecker product $A \otimes B$ is defined by $(a_{ij}B)$. From (7) the usual block-diagonality of the information matrix between the regression parameters in $\Pi = \alpha\beta'$ and the variance-covariance parameters in Ω is obtained. An asymptotically valid expression for the observed (concentrated) information matrix[2] for $\gamma = (\operatorname{vec}(\alpha')', \operatorname{vec}(\beta')')'$ is given by

$$\begin{aligned}
\mathcal{I}_\gamma &= -\frac{\partial\operatorname{vec}(\Pi')'}{\partial\gamma}\frac{\partial^2\ell_c(\Pi,\Omega)}{\partial\operatorname{vec}(\Pi')\partial\operatorname{vec}(\Pi')'}\frac{\partial\operatorname{vec}(\Pi')}{\partial\gamma'} \\
&= T\begin{bmatrix} (I_p\otimes\beta') \\ (\alpha'\otimes I_{p_1}) \end{bmatrix}\left(\Omega^{-1}\otimes S_{11}\right)\begin{bmatrix} (I_p\otimes\beta) & : & (\alpha\otimes I_{p_1}) \end{bmatrix} \\
&= T\begin{bmatrix} (\Omega^{-1}\otimes\beta'S_{11}\beta) & (\Omega^{-1}\alpha\otimes\beta'S_{11}) \\ (\alpha'\Omega^{-1}\otimes S_{11}\beta) & (\alpha'\Omega^{-1}\alpha\otimes S_{11}) \end{bmatrix}.
\end{aligned} \tag{8}$$

The information matrix $\mathcal{I}_\gamma$ is of dimension $(p+p_1)r \times (p+p_1)r$, but has rank equal to $(p+p_1-r)r$, which is seen as follows (see also Johansen, 1995, Lemma 7.1). For any $m \times n$ matrix A of full column rank, let $A_\perp$ be an $m \times (m-n)$ matrix of full column rank such that $A'_\perp A = 0$. Then the left null space of the Jacobian matrix $J(\gamma) = \partial\operatorname{vec}(\Pi')/\partial\gamma' = [(I_p\otimes\beta) : (\alpha\otimes I_{p_1})]$ is spanned by the $pp_1 \times (p-r)(p_1-r)$ matrix $(\alpha_\perp\otimes\beta_\perp)$, which implies that $\operatorname{rank} J(\gamma) = pp_1 - (p-r)(p_1-r) = (p+p_1-r)r$. And since $\mathcal{I}_\gamma = TJ(\gamma)'\left(\Omega^{-1}\otimes S_{11}\right)J(\gamma)$, with $\Omega^{-1}\otimes S_{11}$ non-singular, this implies that $\operatorname{rank}\mathcal{I}_\gamma = (p+p_1-r)r$.

The fact that the rank of the Jacobian matrix J and the information matrix $\mathcal{I}_\gamma$ differs from the dimension of γ by a term r^2, implies that without further restrictions, γ (and hence α and β) is not identified; one might say that γ contains r^2 redundant parameters. This is also easily understood from the fact that α and β enter the likelihood function only via their product $\alpha\beta'$. And since $\alpha\beta' = \alpha Q^{-1}Q\beta' = \alpha^*\beta^{*\prime}$ for any non-singular $r \times r$ matrix Q, it follows that we may freely impose r^2 restrictions on α and/or β without affecting the maximum of the likelihood function.

A common set of identifying restrictions is $c'\beta = I_r$,for some known $p_1 \times r$ matrix c of full column rank. The maximum likelihood estimators of α and β under this restriction are obtained from the reduced rank regression estimators $(\hat{\alpha},\hat{\beta})$ via $\hat{\beta}_c = \hat{\beta}(c'\hat{\beta})^{-1}$ and $\hat{\alpha}_c = \hat{\alpha}\hat{\beta}'c$ (note that $\hat{\beta}'_c S_{11}\hat{\beta}_c \neq I_r$). Note that these restrictions may also be written as

$$\beta = c(c'c)^{-1} + c_\perp B,$$

[2]Often the observed information is defined as minus the second derivative of the log-likelihood, evaluated at the maximum likelihood estimate. Here we use the term to refer to minus the second derivative, evaluated at an arbitrary parameter point; also some terms with expectation zero (when evaluated at the true value) have been omitted.

where B is a $(p_1 - r) \times r$ freely varying (and identified) parameter matrix, with maximum likelihood estimator $\hat{B} = (c_\perp' c_\perp)^{-1} c_\perp' \hat{\beta}_c$. Letting $Y_t = (c'c)^{-1} c' X_t^*$ and $Z_t = -c_\perp' X_t^*$, it follows that in this new parametrization,

$$\beta' X_t^* = Y_t - B' Z_t,$$

which allows B to be interpreted as the equilibrium or long-run effect of Z_t on Y_t (see Johansen, 2002b, for a further discussion of such interpretations). Letting $\delta = (\mathrm{vec}(\alpha_c')', \mathrm{vec}(B)')'$, it follows that the observed information matrix corresponding to this parametrization becomes

$$\begin{aligned} \mathcal{I}_\delta &= T \begin{bmatrix} (I_p \otimes \beta') \\ (\alpha' \otimes c_\perp') \end{bmatrix} \left(\Omega^{-1} \otimes S_{11}\right) \begin{bmatrix} (I_p \otimes \beta) & : & (\alpha \otimes c_\perp) \end{bmatrix} \\ &= T \begin{bmatrix} (\Omega^{-1} \otimes \beta' S_{11} \beta) & (\Omega^{-1}\alpha \otimes \beta' S_{11} c_\perp) \\ (\alpha'\Omega^{-1} \otimes c_\perp' S_{11} \beta) & (\alpha'\Omega^{-1}\alpha \otimes c_\perp' S_{11} c_\perp) \end{bmatrix}. \end{aligned} \tag{9}$$

Provided that the true value of β satisfies $\left|\beta' c\right| \neq 0$, this matrix is non-singular.

The asymptotic properties of $\hat{\alpha}_c$ and $\hat{B}$ may now be characterized as follows, see Johansen (1991, 1995a), using a sequence of norming matrices $D_T = \mathrm{diag}(T^{-1/2} I_{pr}, D_{2T})$, where the form of D_{2T} depends on the specification of the deterministic components in d_t. In particular, when $d_t = 1$ then $D_{2T} = \mathrm{diag}(T^{-1} I_{(p-r)r}, T^{-1/2} I_r)$, and when $d_t = t$, we use $D_{2T} = \mathrm{diag}(T^{-1} I_{(p-r)r}, T^{-3/2} I_r)$. Then as $T \to \infty$,

$$\begin{aligned} D_T^{-1}(\hat{\delta} - \delta) &= \begin{pmatrix} T^{1/2} \,\mathrm{vec}(\hat{\alpha}_c' - \alpha') \\ D_{2T}^{-1} \,\mathrm{vec}(\hat{B} - B) \end{pmatrix} \\ &\xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Omega \otimes \Sigma_{\beta\beta}^{-1} & 0 \\ 0 & (\alpha'\Omega^{-1}\alpha)^{-1} \otimes V^{-1} \end{bmatrix} \right), \end{aligned} \tag{10}$$

$$D_T' \mathcal{I}_\delta D_T \xrightarrow{d} \begin{bmatrix} \Omega^{-1} \otimes \Sigma_{\beta\beta} & 0 \\ 0 & \alpha'\Omega^{-1}\alpha \otimes V \end{bmatrix}, \tag{11}$$

where $\Sigma_{\beta\beta} = \mathrm{plim}_{T\to\infty} \beta' S_{11} \beta$, and where V is a random matrix, which may be expressed as a functional of a vector Brownian motion. Thus the limiting distribution of $T^{1/2}(\hat{\alpha}_c - \alpha)$ is normal, whereas $D_{2T}^{-1}(\hat{B} - B)$ has a *mixed* normal limiting distribution. The limiting result for $\mathcal{I}_\delta$ implies that $(\hat{\delta} - \delta)' \mathcal{I}_\delta (\hat{\delta} - \delta) \xrightarrow{d} \chi^2((p + p_1 - r)r)$, and from the consistency of $\hat{\delta}$ it follows that the same result holds with $\mathcal{I}_\delta$ replaced by $\hat{\mathcal{I}}_\delta$, i.e., (9) evaluated at $\delta = \hat{\delta}$. Similarly it can be shown that Wald, likelihood ratio and Lagrange-multiplier tests for hypotheses on δ will have an asymptotic χ^2 distribution under the null hypothesis.

3 General Restrictions

3.1 Parametrization and Identification

Following Doornik (1995), we consider a general class of the restrictions on α and β of the following form, where the cointegrating rank r is given:

$$\mathcal{H}_g : \alpha' = f_\alpha(\theta), \quad \beta = f_\beta(\theta), \qquad \theta \in \Theta \subset \mathbb{R}^l, \tag{12}$$

where θ is an l-dimensional parameter vector with parameter space Θ, and where $f_\alpha(\cdot)$ and $f_\beta(\cdot)$ are matrix-valued functions on Θ of appropriate dimensions, satisfying the following conditions:

Condition 1 *The functions $f_\alpha(\cdot)$ and $f_\beta(\cdot)$ and the parameter space Θ satisfy:*

1. $\operatorname{rank} f_\alpha(\theta) = \operatorname{rank} f_\beta(\theta) = r$*, for all $\theta \in \Theta$ except a possible set of Lebesgue measure zero;*
2. *$f_\alpha(\cdot)$ and $f_\beta(\cdot)$ are continuously differentiable on Θ, with Jacobian matrices*

$$F_\alpha(\theta) = \frac{\partial \operatorname{vec} f_\alpha(\theta)}{\partial \theta'}, \qquad F_\beta(\theta) = \frac{\partial \operatorname{vec} f_\beta(\theta)}{\partial \theta'}.$$

The first condition is imposed to avoid restrictions that are in conflict with the cointegrating rank r. The second condition is standard, and allows the identification of θ to be analysed from the properties of the Jacobian matrices.

Observe that α and β depend on the same parameter vector θ; hence restrictions linking α and β are possible in this general set-up. However, in many empirically relevant cases such restrictions are excluded, and we may partition θ as $(\psi', \phi')'$, with

$$\alpha' = f_\alpha(\theta) = g(\psi), \quad \beta = f_\beta(\theta) = h(\phi), \qquad (\psi, \phi) \in \Psi \times \Phi, \tag{13}$$

so that

$$F_\alpha(\theta) = \left[\begin{array}{ccc} \dfrac{\partial \operatorname{vec} g(\psi)}{\partial \psi'} & : & 0 \end{array} \right] = \left[\begin{array}{ccc} G(\psi) & : & 0 \end{array} \right], \tag{14}$$

$$F_\beta(\theta) = \left[\begin{array}{ccc} 0 & : & \dfrac{\partial \operatorname{vec} h(\phi)}{\partial \phi'} \end{array} \right] = \left[\begin{array}{ccc} 0 & : & H(\phi) \end{array} \right]. \tag{15}$$

The decomposition of the parameter space $\Theta = \Psi \times \Phi$ entails that the two parameter vectors ψ and ϕ are variation free, which simplifies the maximization of the log-likelihood considerably, as discussed below.

From the fact that $\Pi = \alpha\beta' = f_\alpha(\theta)' f_\beta(\theta)'$, and using the notation $\gamma = (\operatorname{vec}(\alpha')', \operatorname{vec}(\beta)')'$ and $J = \partial \operatorname{vec}(\Pi')/\partial \gamma' = [(I_p \otimes \beta) : (\alpha \otimes I_{p_1})]$ introduced in the previous section, we obtain from the chain rule,

$$\begin{aligned} \mathcal{J}(\theta) := \frac{\partial \operatorname{vec}(\Pi')}{\partial \theta'} &= \frac{\partial \operatorname{vec}(\Pi')}{\partial \gamma'} \frac{\partial \gamma}{\partial \theta'} \\ &= J(\gamma) F(\theta) \\ &= (I_p \otimes \beta)\, F_\alpha(\theta) + (\alpha \otimes I_{p_1})\, F_\beta(\theta), \end{aligned} \tag{16}$$

where $F(\theta) = [F_\alpha(\theta)' : F_\beta(\theta)]'$. Analogously to the derivations in the previous section, this leads to the following expression for the observed information matrix on θ:

$$\begin{aligned} \mathcal{I}_\theta &= T \mathcal{J}(\theta)' \left(\Omega^{-1} \otimes S_{11}\right) \mathcal{J}(\theta) \\ &= T F(\theta)' \left[\begin{array}{cc} (\Omega^{-1} \otimes \beta' S_{11} \beta) & (\Omega^{-1}\alpha \otimes \beta' S_{11}) \\ (\alpha'\Omega^{-1} \otimes S_{11}\beta) & (\alpha'\Omega^{-1}\alpha \otimes S_{11}) \end{array} \right] F(\theta). \end{aligned} \tag{17}$$

When restrictions linking α and β are excluded, such that (13)–(15) holds, then the expressions for $\mathcal{J}(\theta)$ and $\mathcal{I}_\theta$ simplify to

$$\mathcal{J}(\theta) = \left[\; (I_p \otimes \beta)\, G(\psi) \;\; : \;\; (\alpha \otimes I_{p_1})\, H(\phi) \; \right], \tag{18}$$

$$\mathcal{I}_\theta = T \begin{bmatrix} G(\psi)'(\Omega^{-1} \otimes \beta' S_{11}\beta) G(\psi) & G(\psi)'(\Omega^{-1}\alpha \otimes \beta' S_{11}) H(\phi) \\ H(\phi)'(\alpha'\Omega^{-1} \otimes S_{11}\beta) G(\psi) & H(\phi)'(\alpha'\Omega^{-1}\alpha \otimes S_{11}) H(\phi) \end{bmatrix}. \tag{19}$$

The following theorem, adapted from Doornik (1995, Proposition 1) and based on Rothenberg (1971), discusses local identification of θ. Let $N(\theta_0)$ denote a neighbourhood of θ_0.

Theorem 1 *Consider the model*

$$\Delta X_t = f_\alpha(\theta)' f_\beta(\theta)' X^*_{t-1} + \Gamma W_t + \varepsilon_t, \qquad \varepsilon_t \sim \text{i.i.d. } N(0, \Omega), \qquad t = 1, \ldots, T,$$

where $f_\alpha(\cdot)$ and $f_\beta(\cdot)$ satisfy Condition 1, and where $\theta \in \Theta$, an open subset of $\mathbb{R}^l$, with θ variation independent of the (unrestricted) parameters Γ and Ω. A sufficient condition for the parameter value $\theta_0 \in \Theta$ to be locally identified is that

$$\operatorname{rank} \mathcal{J}(\theta) = \operatorname{rank} \left\{ [I_p \otimes f_\beta(\theta)]\, F_\alpha(\theta) + [f_\alpha(\theta) \otimes I_{p_1}]\, F_\beta(\theta) \right\} = l, \qquad \theta \in N(\theta_0) \subset \Theta. \tag{20}$$

Proof. Let $\Pi_0 = f_\alpha(\theta_0)' f_\beta(\theta_0)'$. In the model $\Delta X_t = \Pi X^*_{t-1} + \Gamma W_t + \varepsilon_t$, with Π unrestricted, Π_0 is globally identified, since this is a regression model, where the regressors $(X^{*\prime}_{t-1}, W'_t)'$ are perfectly multicollinear with probability zero. The global identification of Π_0 still applies in the restricted model defined by $\Pi = f_\alpha(\theta)' f_\beta(\theta)', \theta \in \Theta$, simply because no matrices $\Pi \neq \Pi_0$ exist that are observationally equivalent to Π_0 (neither inside nor outside the restricted parameter space). Therefore, local identification of θ_0 requires that we can uniquely solve $\Pi_0 = f_\alpha(\theta)' f_\beta(\theta)'$ for θ, and a sufficient condition for this is that the Jacobian matrix $\mathcal{J}(\theta)$ has full row rank in a neighbourhood of θ_0. □

Note that the theorem only provides a sufficient condition for local identification, which is not necessary. In particular, a parameter value θ_0 can be identified even if $\operatorname{rank} \mathcal{J}(\theta_0) < l$, as long as θ_0 is a single point where rank deficiency occurs. For example, the equation $\pi = \theta^3$ can be uniquely solved for θ at $\pi_0 = \theta_0 = 0$, even though the derivative $\partial\pi/\partial\theta$ is zero at $\theta_0 = 0$. However, in such cases the asymptotic properties of the maximum likelihood estimator of θ will be different from the "regular" cases defined by Theorem 1.

Because the relationship between the Jacobian matrix $\mathcal{J}(\theta)$ and the observed information matrix $\mathcal{I}_\theta$, an equivalent requirement for local identification is that $\mathcal{I}_\theta$ is non-singular in a neighbourhood of θ_0. The non-singularity of the information matrix was proposed by Rothenberg (1971) as a condition for local identification, although he focussed on the usual *expected* information matrix, instead of the *observed* information matrix $\mathcal{I}_\theta$ which is more natural for non-ergodic models.

In the previous section we have shown that $\operatorname{rank} J(\gamma) = (p+p_1-r)r$, and since $\mathcal{J}(\theta) = J(\gamma)F(\theta)$, it follows that $(p + p_1 - r)r$ is an upper bound to the number of identified parameters in θ. Thus an *order condition* for identification is $l \leq (p + p_1 - r)r$.

Johansen (1991, Appendix C), Elliott (2000) and Pesaran and Shin (2002), consider a subclass of the type of restrictions here, where α is unrestricted, but general restriction are imposed on β. Hence this corresponds to (13)–(15) with $\psi = \mathrm{vec}(\alpha')$, such that $\mathrm{vec}\, g(\psi) = \psi$ and $G(\psi) = I_{pr}$. This leads to the rank condition

$$\mathrm{rank}\left[\ (I_p \otimes \beta)\ \ :\ \ (\alpha \otimes I_{p_1})\, H(\phi)\ \right] = l,$$

in a neighbourhood of ϕ_0. The matrix $(I_p \otimes \beta)$ has full column rank pr; if we eliminate these columns by pre-multiplying with $(I_p \otimes \beta'_\perp)$, then we arrive at the condition $\mathrm{rank}\left[(\alpha \otimes \beta'_\perp)H(\phi)\right] = \mathrm{rank}\left[(\alpha \otimes I_{p_1-r})(I_r \otimes \beta'_\perp)H(\phi)\right] = l - pr =: l_\phi$, where l_ϕ is the dimension of ϕ. And since $(\alpha \otimes I_{p_1-r})$ has full column rank, the rank condition becomes

$$\mathrm{rank}\left[(I_r \otimes \beta'_\perp)H(\phi)\right] = l_\phi, \qquad \phi \in N(\phi_0). \tag{21}$$

This rank condition was derived by Boswijk (1995) for the case where $h(\phi)$ is linear. It also corresponds to the condition $\mathrm{rank}\left\{[I_r \otimes C(1)']H(\phi)\right\} = l_\phi$ considered by Pesaran and Shin (2002), where $C(1) = \beta_\perp \left[\alpha'_\perp(I_p - \sum_{i=1}^{k-1}\Gamma_i)\beta_\perp\right]^{-1}\alpha'_\perp$ (in this case β only contains coefficients of X_{t-1}, not of any restricted deterministic components d_t).

Both rank conditions (20) and (21) involve unknown parameters. One can check whether these conditions are generically satisfied by drawing random elements from Θ or Φ, and numerically determining the rank for those parameter values. If the rank condition is satisfied outside a set of Lebesgue measure zero, then the probability of drawing an element from this set is zero (assuming that a continuous distribution on Θ or Φ is used). Alternatively, one can check the rank condition by evaluating (20) in $\tilde{\theta}$, which is any (possibly non-unique) value that maximizes the likelihood function. In that case one does not investigate generic identification, but rather the possibility of identification problems at the maximum likelihood estimate (MLE); this clearly requires an algorithm to maximize the likelihood that does not require a full rank information matrix.

Instead of investigating the rank of $\mathcal{J}(\theta)$ at a random element Θ or the MLE $\tilde{\theta}$, one could investigate the rank of the observed information matrix $\mathcal{I}_\theta$ at such points, as proposed by Boswijk (1995). However, as argued by Doornik (1995), we may expect numerical evaluation of the rank of $\mathcal{J}(\theta)$ to be somewhat more reliable than that of $\mathcal{I}_\theta$, because the latter will be contaminated by the possible near-singularities in Ω^{-1} or S_{11}.

A method to establish the numerical rank of $\mathcal{J}(\theta)$ or $\mathcal{I}_\theta$ is proposed by Doornik (1995, Definition 1): the rank of an $m \times n$ matrix A is determined by the number of singular values w_i satisfying $w_i > 10^4 \epsilon_m \max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|$, where ϵ_m is the machine accuracy for double precision ($\epsilon_m \approx 2 \times 10^{-16}$). The singular values are obtained from the singular value decomposition, available in matrix programming languages such as Ox (Doornik, 2001).

3.2 Estimation and Testing

Consider again the log-likelihood function (4). The maximum likelihood estimator of Ω for a fixed value of θ is

$$\tilde{\Omega}(\theta) = S_{00} - S_{01}\Pi(\theta)' - \Pi(\theta)S_{10} + \Pi(\theta)S_{11}\Pi(\theta)', \qquad \Pi(\theta) = g(\theta)'h(\theta)',$$

which leads to the concentrated log-likelihood in terms of θ only:

$$\ell_c^*(\theta) = -\frac{T}{2}\log\left|\tilde{\Omega}(\theta)\right|, \tag{22}$$

omitting an additive constant. Using the rules of matrix differential calculus, see Magnus and Neudecker (1988), and in particular $d\log|A| = \mathrm{tr}\, A^{-1}dA$, the (concentrated) score vector is

$$\begin{aligned} q(\theta) := \frac{\partial \ell_c^*(\theta)}{\partial \theta} &= T\mathcal{J}(\theta)'\left[\tilde{\Omega}(\theta)^{-1}\otimes I_{p_1}\right]\mathrm{vec}\left[S_{10} - S_{11}\Pi(\theta)'\right] \\ &= T\mathcal{J}(\theta)'\left[\tilde{\Omega}(\theta)^{-1}\otimes S_{11}\right]\mathrm{vec}\left[\hat{\Pi}'_{LS} - \Pi(\theta)'\right]. \end{aligned} \tag{23}$$

A (local) maximum of the likelihood function is obtained by any root of $q(\theta) = 0$ (provided that the Hessian is negative semi-definite at this root). When the Jacobian matrix $\mathcal{J}(\theta)$ is of full column rank for all $\theta \in \Theta$, then this maximum may be found by Newton-type methods, where minus the Hessian matrix might be approximated by $\mathcal{I}_\theta = T\mathcal{J}(\theta)'\left[\tilde{\Omega}(\theta)^{-1}\otimes S_{11}\right]\mathcal{J}(\theta)$. In practice the BFGS (Broyden-Fletcher-Goldfarb-Shanno; see, e.g., Fletcher, 1987) method is known to be more robust, and furthermore this method does not require an invertible Hessian matrix. Clearly, if $\mathcal{J}(\theta)$ has a deficient column rank throughout the entire parameter space, such that θ is not fully identified, then the iterations can only be expected to converge to one out of a continuum of roots.

Of particular interest is the case where there are no restrictions linking α and β, see (13)–(15). In that case the score vector may be partitioned into

$$q_\psi(\psi,\phi) = TG(\psi)'\left[\tilde{\Omega}(\theta)^{-1}\otimes \beta' S_{11}\right]\mathrm{vec}\left[\hat{\Pi}'_{LS} - \Pi(\theta)'\right], \tag{24}$$

$$q_\phi(\psi,\phi) = TH(\phi)'\left[\alpha'\tilde{\Omega}(\theta)^{-1}\otimes S_{11}\right]\mathrm{vec}\left[\hat{\Pi}'_{LS} - \Pi(\theta)'\right], \tag{25}$$

where of course $\alpha' = g(\psi)$ and $\beta' = h(\phi)'$. Now the equations $q_\psi(\psi,\phi) = 0$ may be solved to obtain $\tilde{\psi}(\phi)$, i.e., the maximum likelihood estimator of ψ for a given value of ϕ (and hence β), and similarly $q_\phi(\psi,\phi) = 0$ may be solved to obtain $\tilde{\phi}(\psi)$. This suggests a so-called *switching algorithm*, which means alternating between optimizing over ψ for a given value of ϕ, and optimizing over ϕ for a given value of ψ. Thus, for a starting value $(\tilde{\psi}_0, \tilde{\phi}_0)$, the iterations involve evaluating

$$\tilde{\psi}_1 = \tilde{\psi}(\tilde{\phi}_0), \quad \tilde{\phi}_1 = \tilde{\phi}(\tilde{\psi}_1), \quad \ldots, \tilde{\psi}_j = \tilde{\psi}(\tilde{\phi}_{j-1}), \quad \tilde{\phi}_j = \tilde{\phi}(\tilde{\psi}_j), \quad \ldots$$

until the value of the likelihood function converges. This algorithm was proposed by Doornik (1995). In a different context, such switching algorithms were considered earlier by Sargan (1964) and Oberhofer and Kmenta (1974). As discussed in the next section, Johansen and Juselius (1994) introduced this idea into cointegration modelling. Because the value of the log-likelihood function is non-decreasing in each step, the algorithm will eventually converge to a point where no further improvements are possible in the directions of ϕ and ψ (provided that there are no numerical problems preventing us from reaching this point). Such a point may be a global maximum, but could also be a local maximum or even a saddle point (when likelihood improvements are possible in directions associated with combinations of ϕ and ψ). Therefore, careful selection of starting values is required as always, and positive semi-definiteness of the observed information matrix needs to be checked. Unless g and h are linear (considered in the next

section), each step involves a BFGS or Newton optimization, which might suggest that this procedure will be computationally more intensive than direct maximization over the full parameter vector. The advantage of this procedure however, is that for all reasonable specifications of g and h, ψ is fully identified given a known value of ϕ, and conversely ϕ is identified given ψ. In other words, $G(\psi)$ and $H(\phi)$ will have full column rank even if $\mathcal{J}(\theta)$ does not have full column rank, which makes the algorithms particularly suited for partially identified models.

Once the MLE $\tilde{\theta}$ has been obtained, its covariance matrix may be estimated by

$$\tilde{\mathcal{I}}_\theta^{-1} = \left[T\mathcal{J}(\tilde{\theta})'\left(\tilde{\Omega}^{-1} \otimes S_{11}\right)\mathcal{J}(\tilde{\theta})\right]^{-1}, \tag{26}$$

and the restrictions implied by $\mathcal{H}_g$ in (12) relative to the unrestricted cointegration model $\mathcal{H}_r$ may be tested using the likelihood ratio statistic

$$\begin{aligned} LR(\mathcal{H}_g|\mathcal{H}_r) &= 2\left[\ell_c(\hat{\beta}) - \ell_c^*(\tilde{\theta})\right] \\ &= T\left[\log\left|\tilde{\Omega}\right| - \log|S_{00}| - \sum_{i=1}^{r}\log(1-\hat{\lambda}_i)\right]. \end{aligned} \tag{27}$$

The asymptotic justification of this is provided in the next theorem:

Theorem 2 *Consider the model (3) under the restrictions (12), and assume that the parameter space Θ is compact, that the true value θ_0 lies in the interior of Θ, and that $\operatorname{rank}\mathcal{J}(\theta) = s, \theta \in N(\theta_0)$. Then, as $T \to \infty$,*

$$LR(\mathcal{H}_g|\mathcal{H}_r) \xrightarrow{d} \chi^2((p+p_1-r)r - s). \tag{28}$$

If, in addition, $s = l$, then $\tilde{\theta}$ is consistent and asymptotically mixed normal, i.e.,

$$\tilde{\theta} \xrightarrow{p} \theta_0, \tag{29}$$

and for any vector $a \neq 0$,

$$\left(a'\tilde{\mathcal{I}}_\theta^{-1}a\right)^{-1/2} a'(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0,1). \tag{30}$$

The proof of this theorem is given in the appendix. Note that the theorem does not explicitly specify the rate of consistency. In the absence of restrictions linking α and β, see (13)–(15), we may in general expect $\tilde{\psi}$ and hence $\tilde{\alpha} = f_\alpha(\tilde{\psi})'$ to be $O_p(T^{-1/2})$-consistent and asymptotically normal, whereas $\tilde{\phi}$ and hence $\tilde{\beta} = f_\beta(\tilde{\phi})$ is expected to have a faster rate of convergence and to be asymptotically mixed normal. However, whether this is indeed the case depends on whether the parameters are fully identified by restrictions on β, which corresponds to the condition (21). As an example of a case where this is violated, suppose that β is unrestricted and $\alpha' = (I_r, \alpha_2')$, with α_2 an unrestricted $(p-r)\times r$ matrix. It can be checked that this corresponds to a just-identified model, but the corresponding estimator of β will not be super-consistent, since the identifying restrictions on α imply that part of the information on the adjustment toward equilibrium is contained in β.

The result (28), which does not require a fully identified model, will be particularly useful in the next section, where we consider various specific classes of restrictions on α and β which are not necessarily

fully identifying. It may be emphasized however, that all results of Theorem 2 break down when θ_0 is a singular point of $\mathcal{J}(\theta_0)$, i.e., when the rank of the Jacobian matrix is lower at θ_0 than in a neighbourhood of θ_0. Such cases are associated with local non-identifiability. Although the asymptotic distributions of maximum likelihood estimators and likelihood ratio test statistics may be derived for such cases in specific models, these distributions are typically non-normal and non-chi-squared, respectively.

4 Specific Classes of Restrictions

4.1 Weak Exogeneity Restrictions

Johansen and Juselius (1990) considered the hypothesis

$$\mathcal{H}_a : \alpha = A\zeta, \tag{31}$$

where A is a known $p \times m_a$ matrix of full column rank, with $r \leq m_a < p$, and ζ is a $m_a \times r$ freely varying parameter matrix. Letting $Y_t = \bar{A}' X_t$ and $Z_t = A'_\perp X_t$ (with $\bar{A} = A(A'A)^{-1}$), this hypothesis implies that the model (3) may be written as

$$\begin{aligned} \Delta Y_t &= \psi\beta' X^*_{t-1} + \Gamma_y W_t + \varepsilon_{yt}, \\ \Delta Z_t &= \Gamma_z W_t + \varepsilon_{zt}, \end{aligned}$$

where $\left[\Gamma'_y : \Gamma'_z\right] = \Gamma'[\bar{A} : A_\perp]$ and $(\varepsilon'_{yt}, \varepsilon'_{zt}) = \varepsilon'_t[\bar{A} : A_\perp]$. Thus there is no adjustment toward equilibrium in the equations describing Z_t, which implies that Z_t is *weakly exogenous* for the parameter β, see Johansen (1995a, Chapter 8).

The maximum likelihood estimator under this restriction may again be obtained by reduced rank regression, in a conditional model of ΔY_t given ΔZ_t. For details, see Johansen (1995a, Section 8.2.1). This results in a new set of eigenvalues $\tilde{\lambda}_i$, with corresponding eigenvectors $\tilde{v}_i$ defining the restricted estimator $\tilde{\beta} = (\tilde{v}_1, \ldots, \tilde{v}_r)$, and the concentrated restricted log-likelihood $\ell_c(\tilde{\beta}) = -0.5T \times \left(\log|S_{00}| + \sum_{i=1}^r \log(1-\tilde{\lambda}_i)\right)$, so that the likelihood ratio statistic becomes

$$LR(\mathcal{H}_a|\mathcal{H}_r) = T\sum_{i=1}^r \log\frac{1-\tilde{\lambda}_i}{1-\hat{\lambda}_i}. \tag{32}$$

Note that the null hypothesis may be reformulated as $\mathrm{vec}(\alpha') = (A \otimes I_r)\,\mathrm{vec}(\zeta') = G\psi$, which implies that the Jacobian matrix $\mathcal{J}(\theta)$ in this case is a special case of (18), given by

$$\mathcal{J}(\theta) = \left[\begin{array}{c:c} (A \otimes \beta) & (\alpha \otimes I_{p_1}) \end{array}\right].$$

The left null space of this matrix is now spanned by $[\alpha_\perp \otimes \beta_\perp : A_\perp \otimes \beta]$, which means that $\mathrm{rank}\,\mathcal{J}(\theta) = pp_1 - (p-r)(p_1-r) - (p-m_a)r = (m_a+p_1-r)r = s$, whereas the number of columns is $(m_a+p_1)r = l$. Thus the restrictions are not identifying: indeed, we still have that (ζ, β) is observationally equivalent to $(\zeta^*, \beta^*) = (\zeta Q^{-1}, \beta Q')$ for arbitrary non-singular Q. From Theorem 2, we find that the degrees of freedom for the likelihood ratio test is $(p + p_1 - r)r - s = (p - m_a)r$. In fact, the null hypothesis is equivalent to $A'_\perp \alpha = 0$, which amounts to exactly $(p - m_a)r$ restrictions on α.

4.2 Linear Restrictions on the Cointegrating Space

Johansen (1988) and Johansen and Juselius (1990) considered hypotheses of the following form:

$$\mathcal{H}_b : \beta = H_b\varphi, \tag{33}$$

where H_b is a known $p_1 \times m_b$ matrix of full column rank, with $r \leq m_b < p_1$, and φ is a freely varying $m_b \times r$ parameter matrix. This hypothesis restricts the column space of β to lie in $\mathrm{sp}(H_b)$. Again, the restrictions are not identifying, since $\beta = H_b\varphi$ is equivalent to $\beta^* = H_b\varphi^* = H_b\varphi Q'$ for arbitrary non-singular Q, (accommodated by changing α to $\alpha^* = \alpha Q^{-1}$).

Writing the model (3) under this restriction as

$$\Delta X_t = \alpha\varphi' H_b' X_{t-1}^* + \Gamma W_t + \varepsilon_t, \tag{34}$$

it is easily seen that the statistical analysis of this restricted model is entirely analogous to the reduced rank regression procedure discussed in Section 3, with β and X_{t-1}^* replaced by φ and $H_b'X_{t-1}^*$, respectively, and similarly with (S_{11}, S_{10}, S_{01}) replaced by $(H_b'S_{11}H_b, H_b'S_{10}, S_{01}H_b)$. This again results in restricted eigenvalues $\tilde{\lambda}_i$, and a likelihood ratio statistic $LR(\mathcal{H}_b|\mathcal{H}_r) = T\sum_{i=1}^{r}\log(1-\tilde{\lambda}_i)/(1-\hat{\lambda}_i)$.

When $s = r$, then $\mathrm{sp}(H_b)$ is an r-dimensional subspace, which therefore fully specifies the column space of β. Given that β is only identified up to its column space, this means that in this case β is fully specified (after appropriate normalization). Now φ may be set to I_r in (34) without loss of generality, and the remaining parameters α, Γ and Ω may be estimated simply by least-squares.

The Jacobian matrix in this case becomes $\mathcal{J}(\theta) = [(I_p \otimes \beta) : (\alpha \otimes H_c)]$. Analogously to the previous sub-section, it can be shown that the rank of $\mathcal{J}(\theta)$ is $(p + m_b - r)r = s$, so that the degrees of freedom for the likelihood ratio test is $(p + p_1 - r)r - s = (p_1 - m_b)r$.

Johansen and Juselius (1992) considered:

$$\mathcal{H}_c : \beta = (H_c : \varphi), \tag{35}$$

where H_c is a known $p_1 \times r_1$ matrix of full column rank, and where φ is a freely varying $p_1 \times (r - r_1)$ parameter matrix, with $0 \leq r_1 \leq r$. This corresponds to the case where r_1 cointegrating vectors are fully known, and the remaining $p_1 - r_1$ vectors are unrestricted. Partitioning α conformably with β, we may write the model under this restriction as

$$\Delta X_t = \alpha_1 H_c' X_{t-1}^* + \alpha_2\phi' X_{t-1}^* + \Gamma W_t + \varepsilon_t. \tag{36}$$

By adding the term $H_c'X_{t-1}^*$ to the stationary regressors W_t, this is again recognized as a reduced rank regression problem. The resulting eigenvalues may again be used to construct the likelihood ratio statistic, and the degrees of freedom, following from the rank of the Jacobian matrix, is given by $(p_1 - r)r_1$.

4.3 Linear Identifying Restrictions on Separate Cointegrating Vectors

The restrictions on β discussed in the previous sub-section are testable, but not identifying. Johansen and Juselius (1994) and Johansen (1995) considered linear identifying restrictions on each of the r cointegrating vector β_i separately, of the form $\beta_i = H_i\varphi_i$, and hence

$$\mathcal{H}_d : \beta = (H_1\varphi_1, \ldots, H_r\varphi_r), \tag{37}$$

where H_i are $p_1 \times m_i$ matrices of full column rank, and φ_i are m_i-vectors. These restrictions can only be identifying up to a scale factor; the scale can be fixed by restricting the first component of φ_i to 1, i.e., $\varphi_i' = (1, \phi_i')$ and writing $H_i = [h_i : H_i^*]$, so that $\beta_i = h_i + H_i^*\phi_i$.

Johansen (1995) developed conditions under which such restrictions are generically identifying. The appropriate rank condition for identification of the ith cointegrating vector is that

$$\operatorname{rank}\left[H_{i\perp}'(H_1\varphi_i, \ldots, H_r\varphi_r)\right] = r - 1, \tag{38}$$

for all parameter values $(\varphi_1, \ldots, \varphi_r)$ except a possible set of measure zero. Johansen showed how this condition may be checked from $\operatorname{sp}(H_1), \ldots, \operatorname{sp}(H_r)$, without having to evaluate (38) at an arbitrary point $(\varphi_1, \ldots, \varphi_r)$.

Johansen and Juselius (1994) considered the case where the same H matrix applies to a number of cointegrating vectors, leading to $\mathcal{H}_e : \beta = (H_d\varphi_d, H_e\varphi_e)$, where φ_d and φ_e are matrices, collecting the φ_i vectors corresponding to H_d and H_e, respectively. In such cases these restrictions clearly cannot be fully identifying. They proposed a switching algorithm to maximize the likelihood, alternating between maximization over φ_d for a given value of φ_e, and the converse maximization problem; both maximization problems are solved by reduced rank regression. This algorithm was subsequently generalized to (37) by Johansen (1995), where the algorithm cycles through the different vectors φ_i.

An expression for the asymptotic covariance matrix of the MLE of the normalized parameters $(\phi_1', \ldots, \phi_r')'$, assuming that the rank condition (38) holds, may be obtained from the general expression (19), as discussed in the next sub-section. The degrees of freedom for the likelihood ratio test for $\mathcal{H}_d$, in case of full identification, is $(p_1 - r)r - \sum_{i=1}^{r}(m_i - 1)$, where $m_i = \dim(\varphi_i) = \dim(\phi_1) + 1$.

4.4 General Linear Restrictions

Boswijk (1995) considered the case where $\alpha' = g(\psi)$ and $\beta = h(\phi)$ are linear and affine, respectively, i.e.,

$$\mathcal{H}_l : \operatorname{vec}(\alpha') = G\psi, \qquad \operatorname{vec}\beta = H\phi + h_0, \tag{39}$$

where G and H are constant matrices of full column rank, and h_0 is a constant vector. This class of hypotheses encompasses all hypotheses considered in the previous subsections. For this case, Boswijk (1995) proposed a switching algorithm that is explicit in each step, replacing $\tilde{\Omega}(\theta)^{-1}$ in (24)–(25) by $\tilde{\Omega}_{j-1} = \tilde{\Omega}(\tilde{\theta}_{j-1})$; this is justified by starting from the log-likelihood (4) instead of the concentrated log-likelihood (22). Then the switching algorithm involves alternating over the following three explicit

steps:

$$\begin{aligned}
\tilde{\psi}(\phi,\Omega) &= \left[G'\left(\Omega^{-1}\otimes\beta' S_{11}\beta\right)G\right]^{-1} G'\left(\Omega^{-1}\otimes\beta' S_{11}\right)\mathrm{vec}\left(\hat{\Pi}'_{LS}\right), \\
\tilde{\phi}(\psi,\Omega) &= \left[H'\left(\alpha'\Omega^{-1}\alpha\otimes S_{11}\right)H\right]^{-1} H'\left(\alpha'\Omega^{-1}\otimes S_{11}\right)\left[\mathrm{vec}\left(\hat{\Pi}'_{LS}\right)-(\alpha\otimes I_{p_1})h_0\right], \\
\tilde{\Omega}(\psi,\phi) &= S_{00}-S_{01}\beta\alpha'-\alpha\beta' S_{10}+\alpha\beta' S_{11}\beta\alpha'.
\end{aligned}$$

Starting from a set of initial values $(\tilde{\psi}_0,\tilde{\phi}_0,\tilde{\Omega}_0)$, the iterations then become

$$\tilde{\phi}_j=\tilde{\phi}(\tilde{\psi}_{j-1},\tilde{\Omega}_{j-1}),\quad \tilde{\psi}_j=\tilde{\psi}(\tilde{\phi}_j,\tilde{\Omega}_{j-1}),\quad \tilde{\Omega}_j=\tilde{\Omega}(\tilde{\psi}_j,\tilde{\phi}_j),\quad \ldots.$$

(The order of the evaluation of $\tilde{\phi}(\psi,\Omega)$ and $\tilde{\psi}(\phi,\Omega)$ could also be reversed.) Recently Hansen (2002) provided a generalization of this algorithm, allowing for non-homogeneous linear restrictions on (α,Γ), and a possibly time-varying covariance matrix Ω, labelling this *generalized reduced rank regression*. The same approach may also be applied in the $I(2)$ cointegration model, see Boswijk (2000), and in the seasonal cointegration model, see Johansen and Schaumburg (1999).

The expressions (18)-(19) for the Jacobian matrix $\mathcal{J}(\theta)$ and the observed information matrix $\mathcal{I}_\theta$ apply to this case, with G and H constant matrices instead of functions of ϕ and ψ:

$$\mathcal{J}(\theta) = \left[\begin{array}{ccc}(I_p\otimes\beta)G & : & (\alpha\otimes I_{p_1})H\end{array}\right], \tag{40}$$

$$\mathcal{I}_\theta = T\left[\begin{array}{cc} G'(\Omega^{-1}\otimes\beta' S_{11}\beta)G & G'(\Omega^{-1}\alpha\otimes\beta' S_{11})H \\ H'(\alpha'\Omega^{-1}\otimes S_{11}\beta)G & H'(\alpha'\Omega^{-1}\alpha\otimes S_{11})H\end{array}\right]. \tag{41}$$

Generic identification may be investigated by checking whether $\mathcal{J}(\theta)$ has full column rank for randomly chosen θ. If this is satisfied, the asymptotic covariance matrix of the MLE $\tilde{\theta}$, following from Theorem 2, is given by $\tilde{\mathcal{I}}_\theta^{-1}$. The degrees of freedom for the likelihood ratio test again follows from Theorem 2.

4.5 Non-Causality Restrictions

An important class of non-linear restrictions that does not fit easily in the framework of Theorems 1 and 2 of this paper, is given by Ganger-non-causality restrictions in cointegrated models, see Toda and Phillips (1993). Let X_t be partitioned as $X_t=(X'_{1t},X'_{2t})'$, and let α, β, Π and $\Gamma_i, i=1,\ldots,k-1$, be partitioned conformably (assuming that d_t is void, so that $X_t^*=X_t$). Then the hypothesis that X_{2t} does not Granger-cause X_{1t} corresponds to the null hypothesis

$$\mathcal{H}_{gc}:\Pi_{12}=\alpha_1\beta_2'=0,\qquad \Gamma_{12,i}=0,\quad i=1,\ldots,k-1. \tag{42}$$

The restrictions on Γ_i do not lead to statistical complications, so we will focus on the restriction $\alpha_1\beta_2'=0$. Furthermore, for ease of exposition we concentrate on the case $p=2$, $r=1$, so that α_1 and β_2 are both scalars. We could formulate this hypothesis as $\alpha=f_\alpha(\theta)', \beta=f_\beta(\theta)$, where $\theta\in\Theta\subset\mathbb{R}^4$, and

$$f_\alpha(\theta)'=\left(\begin{array}{c}\theta_1\\ \theta_2\end{array}\right),\qquad f_\beta(\theta)=\left(\begin{array}{c}\theta_3\\ \theta_4 1_{\{\theta_1=0\}}\end{array}\right);$$

however, it is clear that f_β is not continuously differentiable, and furthermore the dimension of Θ does not reflect the number of restrictions.

Alternatively, we may write the hypothesis as $g(\gamma) = \alpha_1\beta_2 = 0$ (recall that $\gamma = (\alpha', \beta')'$), and use a Wald test statistic

$$W = g(\hat{\gamma})[G(\hat{\gamma})\hat{\mathcal{I}}_\gamma^{-1}G(\hat{\gamma})']^{-1}g(\hat{\gamma}), \tag{43}$$

where $G(\gamma)$ is the Jacobian matrix

$$G(\gamma) = \frac{\partial g(\gamma)}{\partial \gamma'} = \begin{pmatrix} \beta_2 & 0 & 0 & \alpha_1 \end{pmatrix}.$$

If either $\beta_2 \neq 0$ or $\alpha_1 \neq 0$, this Wald statistic can be shown to have a limiting χ^2 null distribution, and to be asymptotically equivalent to the LR statistic. However, the null hypothesis also contains parameter values with $\alpha_1 = \beta_2 = 0$, so that $G(\gamma)$ is zero; in such cases the asymptotic distribution of W and LR will be non-standard, and these tests will no longer be asymptotically equivalent. For more details we refer to Phillips and Toda (1993).

4.6 Implementation in Econometric Software

Many of the algorithms to maximize the likelihood function under various parameters restrictions, discussed in the previous sub-sections, have been implemented in econometric software packages. Clearly, as long as the likelihood function under the restriction can be programmed, it may be maximized in any matrix programming language containing a good numerical optimization routine (such as Ox, see Doornik, 2001). Many programs and modules written in such languages have becomes available over the years. However, most practitioners will prefer an interactive econometric package with built-in routines, that require only a minimal amount of programming. Here we discuss the most widely used candidates to fill this need that are currently available.

The most flexible software within this class is PcGive (see Doornik and Hendry, 2001). Over the past fifteen years, PcGive (and its companion PcFiml) have regularly been updated to include the most recent classes of cointegration restrictions as they became available. The switching algorithm for the class of general restrictions discussed in Section 3, and the associated method to check the rank condition and compute the degrees of freedom for the likelihood ratio test via the numerical rank of the Jacobian matrix, were developed by Doornik (1995) with the purpose of implementing them in PcGive. The most recent version (PcGive 10.1) includes a refined version of those algorithms, but also allows the hypotheses discussed in Sections 4.1 and 4.2 to be analysed using the reduced rank regression algorithms mentioned in those sections. The empirical results in the next section have all been obtained using PcGive 10.1.

A close competitor is EViews (Quantitative Micro Software, Irvine (CA); http://www.eviews.com). The most recent version 4.1 allows for the class of general linear restrictions discussed in Section 4.4, implementing the linear switching algorithm of Boswijk (1995). Thus non-linear restrictions, and linear restrictions linking α and β are not allowed. The program does check the rank condition, and allows for partially identified systems.

The Cats in Rats package was developed by Hansen and Juselius (1995), and was used to empirically implement the estimation and testing procedures discussed in Sections 4.1–4.3 as they were developed. The more general hypotheses and algorithms discussed in Sections 3 and 4.4 are not implemented. Unlike the competitors discussed here, which are all general-purpose time-series econometrics packages, Cats in Rats focuses exclusively on cointegration analysis. It is essentially a module within Rats (see http://www.estima.com) which does allow for more general econometric analyses of time series.

General (non-linear) restrictions on the cointegrating vectors, as analysed by Pesaran and Shin (2002), may be empirically implemented using Microfit 4.0, see Pesaran and Pesaran (1997). The program does not allow for restrictions on α. Also, it does not allow for partially identified models: first the user should impose a set of just-identifying linear restrictions, after which general over-identifying restrictions on the remaining parameters may be imposed and tested. The program also allows for the classes of restrictions on β discussed in Sections 4.2 and 4.3.

Finally, it may be noted that none of these packages have a built-in option to test non-causality hypotheses of the type discussed in Section 4.5.

5 An Empirical Application: UK M1

As an application of the various approaches to identifying and restricting cointegrated systems, we consider a model of the demand for narrow money in the UK. We use quarterly seasonally adjusted data over the sample period 1963(1)– 1989(2) that was originally analysed by Hendry and Ericsson (1991), and subsequently by many others, including Doornik *et al.* (1998). Following these authors, we consider the following variables:

- $m - p$: log of real M1, deflated by the total final expenditure deflator;
- y: real total final expenditure;
- Δp: rate of inflation;
- R: interest rate differential, i.e., the three-month local authority interest rate minus the learning-adjusted own interest rate;
- $dout$: dummy variable for output shifts, zero except unity in 1972(4), 1973(1), and 1979(2);
- $doil$: dummy variable for price shocks, zero except unity in 1973(3), 1973(4), and 1979(3).

See Hendry and Ericsson (1991) for details on data sources and transformations.

Following Doornik *et al.* (1998), we specify a VAR(2) model for $X_t = [(m-p)_t, y_t, \Delta p_t, R_t]'$ with deterministic variables $d_t = t$ (restricted trend) and $q_t = (1, dout_t, doil_t)'$ (unrestricted constant and dummy variables); the estimation sample is 1964(3)-1989(2). Furthermore, we assume that the cointegrating rank is given by $r = 2$, although the likelihood ratio tests within this specification formally only

support $r = 1$. (In a model without the dummy variables, there is empirical support for a second cointegrating vector at the 10% significance level.) The unrestricted estimates of α and β (with normalizations $\beta_{11} = \beta_{22} = 1$ imposed) are

$$\hat{\alpha} = \begin{pmatrix} -0.088 & -0.010 \\ -0.022 & -0.097 \\ -0.001 & 0.078 \\ -0.002 & -0.062 \end{pmatrix}, \qquad \hat{\beta} = \begin{pmatrix} 1 & -0.064 \\ -0.997 & 1 \\ 7.343 & -3.383 \\ 7.651 & 0.858 \\ -0.051 & -0.593 \end{pmatrix}. \begin{matrix} m-p \\ y \\ \Delta p \\ R \\ t/100 \end{matrix}$$

All numerical results have been obtained using PcGive version 10.1, see Doornik and Hendry (2001). No standard errors are given for the estimates of α and β, since these parameters are not identified yet. The rank of the Jacobian matrix $\mathcal{J}(\theta)$ is $s = 14$, whereas the number of estimated parameters in α and β is $l = 16$.

As a first sub-model, we test the hypothesis that the rate of inflation and the interest rate differential are weakly exogenous. This is a hypothesis of the form $\mathcal{H}_a$ discussed in Section 4.1, with $A = [I_2 : 0]'$. Using the algorithm referred to in Section 4.1, we obtain the following restricted estimates (imposing the same normalization as before) and likelihood ratio statistic:

$$\tilde{\alpha} = \begin{pmatrix} -0.090 & 0.065 \\ -0.022 & -0.151 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \tilde{\beta} = \begin{pmatrix} 1 & -0.164 \\ -0.952 & 1 \\ 7.234 & -1.764 \\ 7.692 & 0.131 \\ -0.082 & -0.555 \end{pmatrix}, \qquad LR = 4.886\ [0.30].$$

The same restriction may also be imposed using the switching algorithm for general restrictions discussed in Section 3. The resulting likelihood ratio statistic is identical, whereas the estimates of α and β are rotations of the ones given above. Given the p-value of 0.3, the weak exogeneity hypothesis is not rejected.

Suppose now that we wish to identify the long-run money demand relation (corresponding to the first cointegrating vector, β_1) by the restriction of a unit long-run income elasticity, ($\beta_{12} = -\beta_{11}$) and the exclusion of a trend term ($\beta_{15} = 0$), with no other restrictions (other than the normalizations) imposed. Using the switching algorithm once more, this leads to

$$\tilde{\alpha} = \begin{pmatrix} -0.095 & 0.004 \\ -0.002 & -0.070 \\ -0.020 & 0.059 \\ 0.012 & -0.047 \end{pmatrix}, \qquad \tilde{\beta} = \begin{pmatrix} 1 & 0.235 \\ -1 & 1 \\ 6.643 & -2.395 \\ 7.444 & 3.499 \\ 0 & -0.780 \end{pmatrix}, \qquad LR = 0.333\ [0.56].$$

Note that these restrictions generically (over-) identify β_1. The rank of the Jacobian matrix in this model is $s = 13$, showing that there is a single over-identifying restriction, which is not rejected by the likelihood ratio test. Note that because β_1 is identified, so is α_2, but not α_1 (since $\alpha_1\beta_1' + \alpha_2\beta_2'$

is observationally equivalent to $(\alpha_1 + c\alpha_2)\beta_1' + \alpha_2(\beta_2' - c\beta_1') = \alpha_1^*\beta_1' + \alpha_2\beta_2^{*\prime}$, for arbitrary c). A possible local identification problem arises if the restrictions imposed on β_1 are also satisfied by β_2. As discussed in Boswijk (1996), this hypothesis of local non-identifiability may be tested in the form $\mathcal{H}_b : \beta = H_b\varphi$, where

$$H_b = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \tag{44}$$

The likelihood ratio statistic for this hypothesis, obtained from the algorithm referred to in Section 4.2, is given by $LR = 5.889\ [0.21]$, which shows that we cannot reject this hypothesis, and identification based on the restrictions on β_1 only is fragile. A possible explanation for this unexpected result is as follows. The first column $\hat{\beta}_1$ of the unrestricted estimator $\hat{\beta}$, which would be the MLE of the cointegrating vector under the hypothesis $r = 1$, almost satisfies the restriction $\hat{\beta}_1 = H_b\varphi_1$ for some vector φ_1. Therefore, the evidence against the hypothesis $\beta = H_b\varphi$ should come from the second cointegrating vector. However, since the empirical support for this second cointegrating vector is rather weak, it might not be estimated very precisely, which would imply that the resulting test will not be very powerful.[3]

As a final model, we consider the case where in addition to the restrictions $\beta_{12} = \beta_{11}, \beta_{15} = 0$ and the normalizations, we also impose the restriction that adjustment towards money demand equilibrium occurs only in the equation for $\Delta(m-p)_t$, i.e., $\alpha_{12} = \alpha_{13} = \alpha_{14} = 0$. The Jacobian matrix for this model has rank $s = 11$, and since the number of unrestricted parameters l also equals 11, this is an identified model. The likelihood ratio statistic (3 over-identifying restrictions) is $LR = 4.240\ [0.24]$, so that the restrictions cannot be rejected. The estimates, with their standard errors in parentheses, are given by

$$\tilde{\alpha} = \begin{pmatrix} \underset{(0.013)}{-0.107} & \underset{(0.066)}{0.059} \\ 0 & \underset{(0.033)}{-0.141} \\ 0 & \underset{(0.023)}{0.018} \\ 0 & \underset{(0.043)}{-0.040} \end{pmatrix}, \qquad \tilde{\beta} = \begin{pmatrix} 1 & \underset{(0.075)}{0.006} \\ -1 & 1 \\ \underset{(1.360)}{6.401} & \underset{(0.765)}{-1.739} \\ \underset{(0.528)}{7.409} & \underset{(0.502)}{1.650} \\ 0 & \underset{(0.037)}{-0.651} \end{pmatrix}. \tag{45}$$

The striking result is that we may obtain full identification by only imposing restrictions on β_1 and α_1, the cointegrating vector and adjustment coefficients associated with the money demand relation. The second relation, for which we have less identifying information, is left unrestricted, but still is identified.

Although the final model is generically identified, local identification problems could still arise if β_2 would satisfy the same identifying restrictions as β_1. In other words, the final model, corresponding

[3] Another explanation might be that the asymptotic χ^2 distribution of the LR test provides an inaccurate approximation to its actual null distribution, such that the reported p-value of 0.21 deviates from the actual marginal significance level. We have investigated this possibility using a bootstrap analysis (following the approach of Omtzigt and Fachin, 2002), but this leads to a larger p-value, and hence even weaker evidence against the hypothesis of local non-identifiability.

to the hypothesis $\mathcal{H}_g : \alpha_{12} = \alpha_{13} = \alpha_{14} = 0, \beta_1 = H_b\varphi_1$, with H_b as given in (44), contains a further sub-model $\mathcal{H}_0 : \alpha_{12} = \alpha_{13} = \alpha_{14} = 0, \beta = H_b\varphi$, which is not identified. The likelihood ratio test for $\mathcal{H}_0$ in $\mathcal{H}_g$ is $LR = 4.489\ [0.21]$, so that this hypothesis is not rejected (note that the rank of the Jacobian matrix for the null model is $s = 8$, so this is a $\chi^2(3)$ test). Again this result is surprising, since in (45) it appears that $\tilde{\beta}_{21}$ and $\tilde{\beta}_{25}$ differ significantly from -1 and 0, respectively. Indeed, the LR statistics for the individual restrictions $\beta_{21} = -1$ and $\beta_{25} = 0$ in model $\mathcal{H}_g$ both have a p-value of about 0.04. Apparently the estimators $\tilde{\beta}_{21}$ and $\tilde{\beta}_{25}$ are strongly correlated, leading to the non-rejection of the joint hypothesis in conjunction with rejection of the individual hypotheses.

To avoid such problems, we may follow Doornik *et al.* (1998) in imposing the additional restrictions $\alpha_{21} = 0$ and $\beta_{21} = 0$ within the model $\mathcal{H}_g$. The likelihood ratio statistic for the resulting model against the unrestricted model (5 over-identifying restrictions) is $LR = 4.763\ [0.45]$, which shows that the likelihood has hardly decreased by imposing these two additional restrictions, and indeed the estimates are very close to those reported above. The advantage of this further restriction is that is does not contain any non-identified sub-models; i.e., no problems of local non-identifiability can occur.

Appendix: Proof of Theorem 2

Consider first the case where $s = l$, such that θ_0 is identified and $\mathcal{J}(\theta_0)$ has full column rank. To prove consistency of $\tilde{\theta}$, we follow the approach of Saikkonen (1995). Let $\lambda = (\mathrm{vec}(\Pi'), \mathrm{vec}(\Gamma')', \mathrm{vech}(\Omega)')'$, and let the unrestricted parameter space of λ be

$$\Lambda_u = \{\lambda \in \mathbb{R}^{p(p_1+p_2+(p+1)/2)} : \Pi \in \mathbb{R}^{p\times p_1}, \Gamma \in \mathbb{R}^{p\times p_2}, \Omega \in \mathbb{R}^{p\times p}, \Omega > 0\},$$

where $p_2 = p\dim(W_t) = p\,[(k-1)p + \dim(q_t)]$. Analogously, define the restricted parameter space Λ_r as

$$\Lambda_r = \{\lambda \in \mathbb{R}^{p(p_1+p_2+(p+1)/2)} : \Pi = f_\alpha(\theta)' f_\beta(\theta)', \theta \in \Theta, \Gamma \in \mathbb{R}^{p\times p_2}, \Omega \in \mathbb{R}^{p\times p}, \Omega > 0\}.$$

Letting $\ell(\lambda)$ denote the unrestricted log-likelihood function, it is well known the unrestricted MLE $\hat{\lambda} = \mathrm{argmax}_{\lambda\in\Lambda_u}\, \ell(\lambda)$ is obtained from a least-squares regression. Furthermore, when the true value λ_0 is such that the system is $I(1)$, it is known that $\hat{\lambda}$ is consistent, and that there exists a sequence of norming matrices B_T with $\|B_T\| \to 0$ such that $B_T^{-1}(\hat{\lambda} - \lambda) = O_p(1)$. Finally, from the simple form of the log-likelihood $\ell(\lambda)$, it may easily be established that for a true value λ_0 satisfying the $I(1)$ condition,

$$\forall \delta > 0 : \lim_{T\to\infty} P_{\lambda_0}\left\{\sup_{\lambda\in\bar{N}(\lambda_0,\delta)} \ell(\lambda) < \ell(\lambda_0)\right\} = 1, \tag{A.1}$$

where $\bar{N}(\lambda_0, \delta)$ is the complement of the neighbourhood $N(\lambda_0, \delta) = \{\lambda \in \Lambda_u : \|\lambda - \lambda_0\| < \delta\}$. The property (A.1) is a sufficient condition for (weak) consistency of $\hat{\lambda}$, see Wu (1981). However, since $\sup_{\lambda\in\bar{N}(\lambda_0,\delta)\cap\Lambda_r} \ell(\lambda) \leq \sup_{\lambda\in\bar{N}(\lambda_0,\delta)} \ell(\lambda)$, it follows that

$$\forall \delta > 0 : \lim_{T\to\infty} P_{\lambda_0}\left\{\sup_{\lambda\in\bar{N}(\lambda_0,\delta)\cap\Lambda_r} \ell(\lambda) < \ell(\lambda_0)\right\} = 1,$$

which in turn is sufficient for weak consistency of the restricted MLE $\tilde{\lambda} = \operatorname{argmax}_{\lambda \in \Lambda_r} \ell(\lambda)$. And since $\Pi(\theta)$ is a continuously differentiable function with Jacobian matrix $\mathcal{J}(\theta)$ of full column rank in a neighbourhood of θ_0, consistency of $\tilde{\Pi} = \Pi(\tilde{\theta})$ implies consistency of $\tilde{\theta}$.

To obtain the asymptotic distribution of $\tilde{\theta}$, it is useful to write the function $\Pi(\theta)$ as $\Pi(\delta(\theta))$, where

$$\begin{aligned}\delta(\theta) = \begin{pmatrix} \operatorname{vec}(\alpha_c') \\ \operatorname{vec}(B) \end{pmatrix} &= \begin{pmatrix} \operatorname{vec}(c'\beta\alpha) \\ \operatorname{vec}\left[(c_\perp' c_\perp)^{-1} c_\perp' \beta (c'\beta)^{-1}\right] \end{pmatrix} \\ &= \begin{pmatrix} \operatorname{vec}(c' f_\beta(\theta) f_\alpha(\theta)) \\ \operatorname{vec}\left\{(c_\perp' c_\perp)^{-1} c_\perp' f_\beta(\theta)[c' f_\beta(\theta)]^{-1}\right\} \end{pmatrix},\end{aligned}$$

see Section 3. Recall that (α_c, B) is a just-identified parameterization, so that any $\Pi(\theta) = f_\alpha(\theta)' f_\beta(\theta)'$ may be written as $\Pi(\delta(\theta))$, provided that $|c' f_\beta(\theta_0)| \neq 0$. Let $K(\theta) = \partial\delta(\theta)/\partial\theta'$; since

$$\begin{aligned}\mathcal{J}(\theta) &= \frac{\partial \operatorname{vec}(\Pi')}{\partial \theta} \\ &= \frac{\partial \operatorname{vec}(\Pi')}{\partial \delta'} \frac{\partial \delta}{\partial \theta'} \\ &= \left[\begin{array}{ccc} (I_p \otimes \beta_c & : & \alpha_c \otimes c_\perp \end{array}\right] K(\theta),\end{aligned}$$

the concentrated score vector and observed information matrix for θ may be written as

$$q(\theta) = TK(\theta)' \begin{bmatrix} (\Omega^{-1} \otimes \beta_c' S_{11}) \\ (\alpha_c' \Omega^{-1} \otimes c_\perp' S_{11}) \end{bmatrix} \operatorname{vec}\left[\hat{\Pi}_{LS}' - \Pi(\theta)'\right] \tag{A.2}$$

$$\mathcal{I}_\theta = TK(\theta)' \begin{bmatrix} (\Omega^{-1} \otimes \beta_c' S_{11} \beta_c) & (\Omega^{-1}\alpha_c \otimes \beta_c' S_{11} c_\perp) \\ (\alpha_c' \Omega^{-1} \otimes c_\perp' S_{11} \beta_c) & (\alpha_c' \Omega^{-1} \alpha_c \otimes c_\perp' S_{11} c_\perp) \end{bmatrix} K(\theta). \tag{A.3}$$

Now let C_T be any sequence of non-singular norming matrices such that $D_T^{-1} K(\theta_0) C_T \to \bar{K}$, where $\bar{K}$ is a matrix of full column rank. Then it follows that

$$\begin{aligned} C_T' \mathcal{I}_\theta C_T &= C_T' K(\theta_0)' D_T'^{-1} \\ &\quad \times T D_T' \begin{bmatrix} (\Omega^{-1} \otimes \beta_c' S_{11} \beta_c) & (\Omega^{-1}\alpha_c \otimes \beta_c' S_{11} c_\perp) \\ (\alpha_c' \Omega^{-1} \otimes c_\perp' S_{11} \beta_c) & (\alpha_c' \Omega^{-1} \alpha_c \otimes c_\perp' S_{11} c_\perp) \end{bmatrix} D_T \times D_T^{-1} K(\theta_0) C_T \\ &\xrightarrow{d} \bar{K}' \begin{bmatrix} \Omega^{-1} \otimes \Sigma_{\beta\beta} & 0 \\ 0 & \alpha_c' \Omega^{-1} \alpha_c \otimes V \end{bmatrix} \bar{K} \\ &= \bar{K}' \bar{\mathcal{I}} \bar{K}, \end{aligned} \tag{A.4}$$

where $\bar{\mathcal{I}} = \operatorname{diag}([\Omega^{-1} \otimes \Sigma_{\beta\beta}], \left[\alpha_c' \Omega^{-1} \alpha_c \otimes V\right])$, and furthermore

$$C_T' q(\theta_0) \xrightarrow{d} \bar{K}' \bar{\mathcal{I}}^{1/2} Z, \tag{A.5}$$

where Z is a standard normal vector.

Following Saikkonen (1995), the asymptotic distribution of $\tilde{\theta}$ may now be obtained from the usual first-order Taylor series approximation of the score vector, leading to

$$\begin{aligned} C_T^{-1}(\tilde{\theta} - \theta_0) &= \left[C_T' \mathcal{I}_\theta C_T\right]^{-1} C_T' q(\theta_0) + o_p(1) \\ &\xrightarrow{d} (\bar{K}' \bar{\mathcal{I}} \bar{K})^{-1} \bar{K}' \bar{\mathcal{I}}^{1/2} Z \\ &\sim N\left(0, [\bar{K}' \bar{\mathcal{I}} \bar{K}]^{-1}\right). \end{aligned} \tag{A.6}$$

Consistency of $\tilde{\theta}$ at the rate determined by C_T, and stochastic equicontinuity of $\mathcal{I}_\theta$ (see Saikkonen, 1995) implies that $C_T'\tilde{\mathcal{I}}_\theta C_T$ has the same limit as $C_T'\mathcal{I}_\theta C_T$. For any vector a, let T^d be that power of T such that $T^d a' C_T$ converges to a non-zero row vector b'. Then

$$\begin{aligned}\left(a'\tilde{\mathcal{I}}_\theta^{-1}a\right)^{-1/2}a'(\tilde{\theta}-\theta_0) &= \left(T^d a' C_T[C_T'\tilde{\mathcal{I}}_\theta C_T]^{-1}C_T'aT^d\right)^{-1/2}T^d a' C_T C_T^{-1}(\tilde{\theta}-\theta_0)\\ &\xrightarrow{d} (b'\left[\bar{K}'\bar{\mathcal{I}}\bar{K}\right]^{-1}b)^{-1/2}b'[\bar{K}'\bar{\mathcal{I}}\bar{K}]^{-1}\bar{K}'\bar{\mathcal{I}}^{1/2}Z\\ &\sim N(0,1). \end{aligned} \tag{A.7}$$

The distribution of the likelihood ratio statistic follows from a quadratic approximation of the likelihood function, leading to

$$\begin{aligned}LR(\mathcal{H}_g|\mathcal{H}_r) &= [\hat{\delta}-\delta(\tilde{\theta})]'\mathcal{I}_\delta[\hat{\delta}-\delta(\tilde{\theta})]+o_p(1)\\ &= [\hat{\delta}-\delta(\tilde{\theta})]'D_T^{-1}D_T'\mathcal{I}_\delta D_T D_T^{-1}[\hat{\delta}-\delta(\tilde{\theta})]. \end{aligned} \tag{A.8}$$

Now it can be shown that

$$\begin{aligned}D_T^{-1}[\hat{\delta}-\delta(\tilde{\theta})] &\xrightarrow{d} \left(\bar{\mathcal{I}}^{-1/2}-\bar{K}\left[\bar{K}'\bar{\mathcal{I}}\bar{K}\right]^{-1}\bar{K}'\bar{\mathcal{I}}^{1/2}\right)Z\\ &\sim N(0,\bar{\mathcal{I}}^{-1}-\bar{K}\left[\bar{K}'\bar{\mathcal{I}}\bar{K}\right]^{-1}\bar{K}'). \end{aligned}$$

The covariance matrix has rank $(p+p_1-r)r-l$, and $\bar{\mathcal{I}}$ is a generalized inverse of this covariance matrix, which implies that

$$LR(\mathcal{H}_g|\mathcal{H}_r) \xrightarrow{d} \chi^2([p+p_1-r]r-l). \tag{A.9}$$

Finally, consider the case where $s<l$, such that θ_0 is not fully identified. Since $\mathcal{J}(\theta)$ has constant rank s in a neighbourhood of θ_0, it follows that the rank deficiency and the associated identification problem does not occur at an isolated point, but occurs generically, i.e., everywhere in Θ except for a possible set of measure zero where the rank is lower than s. This in turn may be interpreted as θ containing $l-s$ redundant parameters. In such cases we may reparametrize θ as $\theta(\eta)$, where $\eta\in \mathrm{H}\subset\mathbb{R}^s$, such that $\Theta=\{\theta(\eta),\eta\in\mathrm{H}\}$, and where $\theta(\eta)$ is a continuously differentiable function. The model then becomes $\Pi(\theta)=\Pi(\theta(\eta))=\Pi^*(\eta),\eta\in\mathrm{H}$, where Π^* is a continuously differentiable function, with Jacobian matrix of full column rank s in a neighbourhood of the true value η_0. This means that the above result apply with θ replaced by η, and in particular, $LR(\mathcal{H}_g|\mathcal{H}_r)\xrightarrow{d}\chi^2([p+p_1-r]r-s)$. □

References

Anderson, T. W. (1951), "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions", *Annals of Mathematical Statistics*, 22, 327–351.

Boswijk, H. P. (1995), "Identifiability of Cointegrated Systems", Tinbergen Institute Discussion Paper TI 95-78, Tinbergen Institute, http://www.fee.uva.nl/ke/boswijk/ident.pdf.

Boswijk, H. P. (1996), "Testing Identifiability of Cointegrating Vectors", *Journal of Business & Economic Statistics*, 14, 153–160.

Boswijk, H. P. (2000), "Mixed Normality and Ancillarity in I(2) Systems", *Econometric Theory*, 16, 878–904.

Davidson, J. E. H., D. F. Hendry, F. Srba and S. Yeo (1978), "Econometric Modelling of the Aggregate Time-Series Relationship between Consumers' Expenditure and Income in the United Kingdom", *Economic Journal*, 88, 661–692.

Doornik, J. A. (1995), "Testing General Restrictions on the Cointegrating Space", Discussion Paper, Nuffield college, http://hicks.nuff.ox.ac.uk/Users/Doornik/papers/coigen.pdf.

Doornik, J. A. (2001), *Object-Oriented Matrix Programming using Ox 3*. London: Timberlake Consultants Ltd. and http://www.nuff.ox.ac.uk/Users/Doornik.

Doornik, J. A. and D. F. Hendry (2001), *Modelling Dynamic Systems using PcGive 10*. London: Timberlake Consultants Ltd. and http://www.pcgive.com.

Doornik, J. A., D. F. Hendry and B. Nielsen (1998), "Inference in Cointegrated Models: UK M1 Revisited", *Journal of Economic Surveys*, 12, 533–572.

Elliott, G. (2000), "Estimating Restricted Cointegrating Vectors", *Journal of Business and Economic Statistics*, 18, 91–99.

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

Fletcher, R. (1987), *Practical Methods of Optimization*, 2nd edition. New York: John Wiley & Sons.

Hansen, H. and K. Juselius (1995), *CATS in RATS. Manual to Cointegration Analysis of Time Series*. Evanston (IL): Estima and http://www.estima.com/catsinfo.htm.

Hansen, P. R. (2002), "Generalized Reduced Rank Regression", Economics Working Paper 2002-02, Brown University, http://www.econ.brown.edu/fac/Peter_Hansen/Papers/grrr.pdf.

Hendry, D. F. and N. R. Ericsson (1991), "Modelling the Demand for Narrow Money in the United Kingdom and the United States", *European Economic Review*, 35, 833–886.

Hood, W. C. and T. C. Koopmans (eds.) (1953), *Studies in Econometric Method*. New York: John Wiley & Sons.

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

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

Johansen, S. (2002a), "The Statistical Analysis of Hypotheses on the Cointegrating Relations in the I(2) Model", Working paper, Department of Statistics and Operations Research, University of Copenhagen.

Johansen, S. (2002b), "The Interpretation of Cointegrating Coefficients in the Cointegrated Vector Autoregressive Model", Working paper, Department of Statistics and Operations Research, University

of Copenhagen.

Johansen, S. and K. Juselius (1990), "Maximum Likelihood Estimation and Inference on Cointegration – With Application to the Demand for Money", *Oxford Bulletin of Economics and Statistics*, 52, 169–210.

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

Johansen, S. and K. Juselius (1994), "Identification of the Long-Run and the Short-Run Structure. An Application to the ISLM Model", *Journal of Econometrics*, 63, 7–36.

Johansen, S. and E. Schaumburg (1999), "Likelihood Analysis of Seasonal Cointegration", *Journal of Econometrics*, 88, 301–339.

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

Oberhofer, W. and J. Kmenta (1974), "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models", *Econometrica*, 42, 579–590.

Omtzigt, P. and S. Fachin (2002), "Bootstrapping and Bartlett Corrections in the Cointegrated VAR Model", Working paper 2002/024, Faculty of Economics, University of Insubria, http://eco.uninsubria.it/dipeco/Quaderni/files/QF2002_24.pdf.

Pesaran, M. H. and B. Pesaran (1997), *Working with Microfit 4.0: Interactive Econometric Analysis*. Oxford: Oxford University Press and http://www.intecc.co.uk/camfit.

Pesaran, M. H. and Y. Shin (2002), "Long-Run Structural Modelling", *Econometric Reviews*, 21, 49–87.

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

Saikkonen, P. (1995), "Problems with the Asymptotic Theory of Maximum Likelihood Estimation in Integrated and Cointegrated Systems", *Econometric Theory*, 11, 888–911.

Sargan, J. D. (1964), "Wages and Prices in the United Kingdom: A Study in Econometric Methodology", in Hart, P. E., G. Mills and J. K. Whitaker (eds.), *Econometric Analysis for National Economic Planning*, Vol. 16 of *Colston Papers*, pp. 25–63. London: Butterworth Co.

Sims, C. A. (1980), "Macroeconomics and Reality", *Econometrica*, 48, 1–48.

Theil, H. (1953), "Estimation and Simultaneous Correlation in Complete Equation Systems", Central Planning Bureau Mimeograph, The Hague.

Theil, H. and J. C. G. Boot (1962), "The Final Form of Econometric Equation Systems", *Review of the International Statistical Institute*, 30, 136–152.

Toda, H. and P. C. B. Phillips (1993), "Vector Autoregressions and Causality", *Econometrica*, 61, 1367–1394.

Wu, C.-F. (1981), "Asymptotic Theory of Nonlinear Least Squares Estimation", *Annals of Statistics*, 9, 501–513.

Zellner, A. and F. C. Palm (1974), "Time Series Analysis and Simultaneous Equation Econometric Models", *Journal of Econometrics*, 2, 17–54.

Zellner, A. and H. Theil (1962), "Three-Stage Least-Squares: Simultaneous Estimation of Simultaneous Equations", *Econometrica*, 30, 54–78.