Statistical analysis of hypotheses on the cointegrating relations in the I(2) model

Søren Johansen*
Department of Applied Mathematics and Statistics
University of Copenhagen

Abstract

The cointegrated vector autoregressive model for $I(2)$ variables is a non-linear parametric restriction on the linear $I(2)$ regression model for variables of order $I(0)$, $I(1)$ and $I(2)$. In this paper we discuss non-linear submodels given by smooth parametrizations. We give conditions on the parametrization which imply that the limit under local alternatives of the log likelihood ratio is quadratic, and show that the asymptotic distribution of the maximum likelihood estimator can be found by optimizing the limit function. This gives a reformulation of a condition by Boswijk (2000) and the reformulation is applied to show that some hypotheses on the cointegrating coefficients in the cointegrated $I(2)$ model give asymptotic χ^2 inference.

Keywords: Likelihood ratio test, $I(2)$ processes, cointegration, asymptotic theory
JEL Classification C32

*Søren Johansen, Department of Applied Mathematics and Statistics, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen Ø, Denmark, Tlf: +45 35320681, Fax +45 35320772, Email sjo@math.ku.dk

1 Introduction

The framework for this paper is the $I(2)$ regression model

$$Y_t = \theta_0' Z_{0t} + \theta_1' Z_{1t} + \theta_2' Z_{2t} + \varepsilon_t, \tag{1}$$

where Z_{it} is integrated of order i, $I(i),\ i = 0, 1, 2$. As a smooth submodel we consider the $I(2)$ cointegration model, see Johansen (1997), given by

$$\Delta^2 X_t = \alpha(\rho'\tau' X_{t-1} + \psi'\Delta X_{t-1}) + \Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\kappa'\tau'\Delta X_{t-1} + \varepsilon_t. \tag{2}$$

We show that (2) and the models we get under hypotheses on the cointegrating parameters $(\rho, \tau, \beta = \tau\rho, \psi)$ are smooth submodels of (1) defined by $\theta = \theta(\phi)$. We therefore find conditions on $\theta(\phi)$ under which the limit of the log likelihood function under local alternatives is quadratic, and show that these conditions are satisfied for the $I(2)$ cointegration model, building on Boswijk (2000). We thereby get a tool to discuss asymptotic inference on the cointegrating parameters. We find conditions for asymptotic χ^2 inference on β and ρ and show that in particular they are satisfied for hypotheses of the form $\beta = K\phi$ and $\beta = (b, b_\perp\phi)$. Similarly $\tau = K\phi$ leads to asymptotic χ^2, whereas $\tau = (b, b_\perp\phi)$ only leads to asymptotic χ^2 inference, when the overlap between b and β is minimal. A two step solution is suggested.

The thesis by Omtzigt (2002) contains a program for calculating the maximum likelihood estimator and likelihood ratio statistic under various restrictions on the parameters in the cointegrated $I(2)$ model, and the results of the present paper are therefore needed for understanding how to make inference in these models.

The paper is structured as follows: In section 2 we consider a general linear regression model with integrated regressors of order at most 2, of which the cointegrated model is a submodel given by a smooth hypothesis. We show that for suitable conditions on the smooth parameterization, the limit of the log likelihood function under local alternatives is quadratic (LAQ), which allows the derivation of the asymptotic distribution of the maximum likelihood estimator by optimizing the limit function.

In section 3 we discuss the cointegrated $I(2)$ model as a submodel of the $I(2)$ regression model and in section 4 we give the asymptotic theory as a consequence of the results in section 2 and find sufficient conditions for inference to be local asymptotic mixed normal (LAMN). Finally section 5 contains some examples of inference on the cointegrating coefficients.

2 Asymptotic inference in the $\mathbf{I(2)}$ regression model

In the regression model (1) for the $n-$dimensional process Y_t, we assume that the $I(d)$ process Z_{it} is of dimension k_i, and not cointegrated. Further Z_{it} is measurable with respect to $\sigma(\varepsilon_s, s \leq t-1)$ and hence independent of the i.i.d. errors ε_t, which in the likelihood derivations are assumed $N_n(0, \Omega)$, whereas for the asymptotics it is enough that they are i.i.d. with mean zero and variance Ω. Thus, the number of regressors $Z_t = (Z'_{0t}, Z'_{1t}, Z'_{2t})'$ is $k = k_0 + k_1 + k_2$, and the dimension of Y_t is n. We first want to make inference on the unrestricted parameters $\theta = (\theta'_0, \theta'_1, \theta'_2)'$, of dimension $n \times k$, and then analyse the model we get by a general set of restrictions expressed as $\theta = \theta(\phi)$ of the vector of parameters $\phi = (\phi'_0, \phi'_1, \phi'_2)'$ of dimension $q = q_0 + q_1 + q_2$. We make such assumptions on the processes Z_{0t}, Z_{1t}, Z_{2t} that the assumptions are satisfied in the $I(2)$ cointegration model formulated as a nonlinear $I(2)$ regression model.

The regression estimators of the unrestricted parameters $\theta = (\theta'_0, \theta'_1, \theta'_2)'$ are consistent of order $T^{-1/2}, T^{-1}$, and T^{-2} respectively. This follows from the integration properties of the regressors, see for instance Stock and Watson (1993), Kitamura (1995), and Boswijk (2000).

Here we find conditions, which guarantee that the parameters ϕ_0, ϕ_1, ϕ_2 are consistently estimated of order $T^{-1/2}, T^{-1}, T^{-2}$ respectively, and show that derivatives up to order four are needed to find the limit of the likelihood function under local alternatives. We then reformulate a result of Boswijk (2000) and find conditions on the parametrization which imply that the limit of the likelihood function is quadratic and show that the limit distribution of $\hat{\phi} - \phi$ suitably normalized can be found by maximizing the limit of the likelihood function.

2.1 Asymptotic distribution of the regression estimator $\hat{\theta}$

Equation (1) models Y_t given $\sigma(\varepsilon_i, i \leq t-1)$ and we get the likelihood function

$$\log L(\theta) = -\frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T}(Y_t - \theta' Z_t)'\Omega^{-1}(Y_t - \theta' Z_t), \tag{3}$$

which is maximized by regressing Y_t on Z_t. We do not discuss estimation of the variance as inference is standard and we assume that Ω is known in the following. In order to find the asymptotic distribution of the estimator of $\hat{\theta}$ we need assumptions on the processes $(\varepsilon_t, Z_{0t}, Z_{1t}, Z_{2t})$.

We define $n_j = \max(j, \frac{1}{2})$ and introduce the product moments

$$M_{\varepsilon j} = T^{-n_j}\sum_{t=1}^{T}\varepsilon_t Z'_{jt}, \quad M_{ij} = T^{-n_i-n_j}\sum_{t=1}^{T} Z_{it}Z'_{jt}, \tag{4}$$

which are normalized to converge.

Assumption 1 *a. The errors and regressors satisfy*

$$T^{-\frac{1}{2}}\begin{pmatrix}\sum_{t=1}^{[Tu]}\varepsilon_t \\ Z_{1[Tu]} \\ T^{-1}Z_{2[Tu]}\end{pmatrix} \xrightarrow{w} \begin{pmatrix} W(u) \\ H_1(u) \\ H_2(u)\end{pmatrix}, u \in [0,1] \tag{5}$$

where $W(u)$ is Brownian motion with variance Ω, and H_1 and H_2 are defined as the limits of $T^{-1/2}Z_{1[Tu]}$ and $T^{-3/2}Z_{2[Tu]}$ respectively.

b. With the notation $H_{ij} = \int_0^1 H_i H_j' du$ *and* $\Sigma = E(Z_{0t}Z_{0t}')$ *the product moments satisfy*

$$\begin{pmatrix} M_{\varepsilon 0} \\ M_{\varepsilon 1} \\ M_{\varepsilon 2} \end{pmatrix} \xrightarrow{w} \begin{pmatrix} N(0, \Sigma \otimes \Omega) \\ H_{\varepsilon 1} = \int_0^1 (dW) H_1' \\ H_{\varepsilon 2} = \int_0^1 (dW) H_2' \end{pmatrix} = \begin{pmatrix} N(0, \Sigma \otimes \Omega) \\ \int_0^1 (dW) H_*' \end{pmatrix} \tag{6}$$

$$\begin{pmatrix} M_{00} & M_{01} & M_{02} \\ M_{10} & M_{11} & M_{12} \\ M_{20} & M_{21} & M_{22} \end{pmatrix} \xrightarrow{w} \begin{pmatrix} \Sigma & 0 & 0 \\ 0 & H_{11} & H_{12} \\ 0 & H_{21} & H_{22} \end{pmatrix} = \begin{pmatrix} \Sigma & 0 \\ 0 & H_{**} \end{pmatrix}, \tag{7}$$

where $N(0, \Sigma \otimes \Omega)$ *indicates an* $n \times k_0$ *dimensional Gaussian distribution with mean zero and variance matrix* $\Sigma \otimes \Omega$*.*

We then find the asymptotic distribution of the estimator $\hat{\theta} - \theta$ normalized by

$$M_T = diag(T^{-1/2} I_{k_0}, T^{-1} I_{k_1}, T^{-2} I_{k_2}). \tag{8}$$

Theorem 1 *Under Assumption 1 the unrestricted estimators of the regression model (1) has an asymptotic distribution given by*

$$(\hat{\theta} - \theta)' M_T^{-1} \xrightarrow{w} (N(0, \Sigma^{-1} \otimes \Omega), \int_0^1 (dW) H_*' H_{**}^{-1}).$$

PROOF. This result follows from the expression

$$(\hat{\theta} - \theta)' M_T^{-1} = \sum_{t=1}^{T} \varepsilon_t Z_t' M_T \left(M_T \sum_{t=1}^{T} Z_t Z_t' M_T \right)^{-1},$$

applying Assumption 1. ■

2.2 Limit of the likelihood under local alternatives

We find from (3) the score and information about θ in the direction $\psi = (\psi_0', \psi_1', \psi_2')' \in \mathbb{R}^{k \times n}$ as

$$S_T(\psi) = \frac{d \log L(\theta + s\psi)}{ds}|_{s=0} = tr\{\Omega^{-1} \sum_{t=1}^{T} \varepsilon_t Z_t' \psi\}, \tag{9}$$

$$J_T(\psi, \psi) = -\frac{d^2 \log L(\theta + s\psi)}{ds^2}|_{s=0} = tr\{\Omega^{-1} \sum_{t=1}^{T} \psi' Z_t Z_t' \psi\}. \tag{10}$$

With this notation the likelihood function (3) satisfies

$$\log \frac{L(\theta + \psi)}{L(\theta)} = S_T(\psi) - \frac{1}{2} J_T(\psi, \psi).$$

The findings about the asymptotic distribution can be presented by finding the limit of the log likelihood ratio under local alternatives. It is convenient to introduce the notation $\psi' = (\psi_0', \psi_1', \psi_2') = (\psi_0', \psi_*')$. We find from (9)

$$S_T(M_T\psi) \xrightarrow{w} S(\psi) = tr\{\Omega^{-1}N(0, \Sigma \otimes \Omega)\psi_0\} + tr\{\Omega^{-1}\int_0^1 (dW)H_*'\psi_*\} = \psi^{v\prime}S^v,$$

with $\psi^v = vec(\psi')$ and the score vector S^v defined by

$$S^v = vec(N(0, \Sigma \otimes \Omega^{-1}), \Omega^{-1}\int_0^1 (dW)H_*'). \tag{11}$$

The limit of the information (10) under the local alternative $\theta + M_T\psi$ is

$$J_T(M_T\psi, M_T\psi) \xrightarrow{w} J(\psi, \psi) = tr\{\Omega^{-1}(\psi_0'\Sigma\psi_0 + \psi_*'H_{**}\psi_*)\} = \psi^{v\prime}J^m\psi^v, \tag{12}$$

where the $nk \times nk-$dimensional information matrix J^m is defined by

$$J^m = \Omega^{-1} \otimes \begin{pmatrix} \Sigma & 0 \\ 0 & H_{**} \end{pmatrix}. \tag{13}$$

Combining the above results and notation we find the limit of the likelihood function under local alternatives for the linear regression model

$$\log \frac{L(\theta + M_T\psi)}{L(\theta)} \xrightarrow{w} S(\psi) - \frac{1}{2}J(\psi, \psi) = \psi^{v\prime}S^v - \frac{1}{2}\psi^{v\prime}J^m\psi^v.$$

It is seen from Theorem 1 that the limit distribution of $((\hat{\theta} - \theta)'M_T^{-1})^v$ can be found by maximizing the limit under local alternatives and is given by the maximizer $(J^m)^{-1}S^v$. Note that we do not in general get a mixed Gaussian distribution, as H_* need not be independent of W. We shall use this type of result below to derive the asymptotic properties of the estimator in a smooth submodel given by $\theta = \theta(\phi)$.

In order to formulate the results we need some notation for derivatives. If $n = 1$, θ is a $k-$vector and ϕ a $q-$vector, the derivative $d\theta/d\phi'$ is a $k \times q$ matrix with elements $\partial\theta_i/\partial\phi_j$. The derivative in the direction of the $q-$vector ψ at the point ϕ^0 is denoted $d\theta/d\phi'(\psi)$ and defined as the $k-$vector with elements

$$\frac{d\theta_i(\phi^0 + s\psi)}{ds}|_{s=0} = \sum_k \frac{\partial\theta_i}{\partial\phi_k}\psi_k.$$

If θ is $k \times n$ and ϕ is $q \times r$, we can reduce the notation for derivatives to the case just above, by defining the vector function θ^v and the vector argument ϕ^v and define the $kn \times qr$ derivative matrix $d\theta^v/d\phi^{v\prime}$ so that

$$\frac{d\theta^v(\phi^0 + s\psi)}{ds}|_{s=0} = \frac{d\theta^v}{d\phi^{v\prime}}\psi^v.$$

In the following we also need higher order derivatives, where the above notation becomes more cumbersome if θ and ϕ are collections of matrices. We let $\psi^{(m)} =$

$(\psi,\ldots,\psi)$ denote m copies of ψ. If $\theta(\phi)$ is m times differentiable, we apply the notation

$$\frac{d^m\theta}{d\phi^m}(\psi^{(m)}) = \frac{d^m\theta}{d\phi^m}|_{\phi=\phi^0}(\psi^{(m)}) = \frac{d^m\theta(\phi^0+s\psi)}{ds^m}|_{s=0},$$

that is, the $m-$th derivative of the function $\theta(\phi)$ at the point ϕ^0 in the direction ψ. Note that $d\theta/d\phi(\psi)$ is linear in ψ, and that $d^2\theta/d\phi^2(\psi^{(2)}) = d^2\theta/d\phi^2(\psi,\psi)$ is quadratic in ψ. This notation allows ϕ and ψ to be matrices or collections of matrices. The main property used is that the argument ψ varies in a linear space. The notation is useful because if $\theta(\phi)$ is a matrix, then $d^m\theta/d\phi^m(\psi^{(m)})$ is a matrix of the same dimensions, which make the calculations easier to handle, and the only property we really need is the homogeneity

$$\frac{d^m\theta}{d\phi^m}(\lambda_1\psi,\ldots,\lambda_m\psi) = (\prod_{i=1}^m\lambda_i)\frac{d^m\theta}{d\phi^m}(\psi^{(m)})$$

and, using $\theta(\phi)=\theta(\phi_0,\phi_1,\phi_2)$, that we have the formula

$$\frac{d^m\theta}{d\phi^m}(\psi^{(m)}) = \sum_{i_1,\ldots,i_m=0}^{2}\frac{\partial^m\theta}{\partial\phi_{i_1}\ldots\partial\phi_{i_m}}(\psi_{i_1},\ldots,\psi_{i_m}). \tag{14}$$

In particular for $m=1$ we have

$$\frac{d\theta}{d\phi}(\psi) = \frac{\partial\theta}{\partial\phi_0}(\psi_0) + \frac{\partial\theta}{\partial\phi_1}(\psi_1) + \frac{\partial\theta}{\partial\phi_2}(\psi_2).$$

We also need a Taylor's formula with remainder term

$$\theta(\phi^0+\psi) = \sum_{m=0}^{l}\frac{1}{m!}\frac{d^m\theta}{d\phi^m}|_{\phi^0}(\psi^{(m)}) + R_T^{(l)},\ R_T^{(l)} = \frac{1}{l!}(\frac{d^l\theta}{d\phi^l}|_{\phi^*} - \frac{d^l\theta}{d\phi^l}|_{\phi^0})(\psi^{(l)}), \tag{15}$$

where we use the notation $d\theta/d\phi|_*$ to indicate the derivative taken at intermediate points, which depend on the coordinate of θ we expand. Finally we let $||\theta||^2 = \sum_{i=1}^k\sum_{j=1}^n\theta_{ij}^2$ denote the norm of an $k\times n$ matrix θ.

The following result was proved by Boswijk (2000)

Theorem 2 *Let the $nk-$dimensional vector function $\theta^v=\theta^v(\phi)$ be a twice continuously differentiable function of the $q-$dimensional vector ϕ, with $nk\times q$ derivative matrix $H(\phi)=d\theta^v/d\phi'$. We let C_T be a $q\times q$ normalization matrix of ϕ, so that for some $nk\times q$ matrix K and M_T given by (8), it holds that*

$$\sup_{|\psi|\le\delta}||(I_n\otimes M_T^{-1})H(\phi^0+C_T\psi)C_T - K|| \to 0. \tag{16}$$

Note that the limit matrix K does not depend on ψ. Then the likelihood function is locally asymptotically quadratic (LAQ) and the limit is given by

$$\log\frac{L(\theta(\phi^0+C_T\psi))}{L(\theta(\phi^0))} \xrightarrow{w} \psi'K'S^v - \frac{1}{2}\psi'K'J^mK\psi, \tag{17}$$

with S^v and J^m given by (11) and (13).

Thus condition (16) implies that the derivative $d\theta/d\phi$ in the direction $C_T\eta$, normalized by $I_n \otimes M_T^{-1}$, and taken at the 'intermediate point' $\phi^0 + C_T\psi$, is convergent to a limit which is independent of the direction ψ of the 'intermediate point'. It is this property that Boswijk exploits to locally linearize the function $\theta(\phi)$, using the meanvalue theorem, and hence derive properties of the estimator of ϕ. This paper gives a reformulation of condition (16) expressed in terms of higher order derivatives.

We define the vector $D_T = (T^{-1/2}, T^{-1}, T^{-2})$, and the notation $D_T^{-1} \cdot (\hat{\theta} - \theta)$ to mean the normalized deviations:

$$D_T^{-1} \cdot (\hat{\theta} - \theta) = \left(T^{\frac{1}{2}}(\hat{\theta}_0 - \theta_0)', T(\hat{\theta}_1 - \theta_1)', T^2(\hat{\theta}_2 - \theta_2)' \right)'.$$

In the applications we have new (matrix) parameters $\phi = (\phi_0, \phi_1, \phi_2)$ and define

$$D_T^{-1} \cdot (\hat{\phi} - \phi) = (T^{1/2}(\hat{\phi}_0 - \phi_0)', T(\hat{\phi}_1 - \phi_1)', T^2(\hat{\phi}_2 - \phi_2)')'.$$

In order to prove results about the submodel defined by $\theta(\phi)$ we need to make some technical assumptions, which will be discussed below.

Assumption 2 *a. The function θ_0 is continuously differentiable, θ_1 twice continuously differentiable and θ_2 four times continuously differentiable at ϕ^0.*

b. The parameter ϕ is continuously identified

$$\theta(\phi_n) \to \theta(\phi^0) \text{ implies } \phi_n \to \phi^0. \tag{18}$$

c. The first order derivatives have full rank

$$rank(\frac{\partial \theta_i}{\partial \phi_i}|_{\phi=\phi^0}) = q_i, i = 0, 1, 2 \tag{19}$$

d. The derivatives

$$\frac{\partial \theta_1}{\partial \phi_0}, \frac{\partial \theta_2}{\partial \phi_1}, \frac{\partial^2 \theta_2}{\partial \phi_0^2}, \frac{\partial^2 \theta_2}{\partial \phi_0 \partial \phi_1}, \frac{\partial^3 \theta_2}{\partial \phi_0^3}, \tag{20}$$

are zero for $\phi = \phi^0$.

e. The derivatives

$$\frac{\partial^2 \theta_1}{\partial \phi_0^2}, \frac{\partial^2 \theta_2}{\partial \phi_1^2}, \frac{\partial^3 \theta_2}{\partial \phi_0^2 \partial \phi_1}, \frac{\partial^4 \theta_2}{\partial \phi_0^4} \tag{21}$$

are zero for $\phi = \phi^0$.

We can then prove the main result of this section using some technical results from the Appendix.

Theorem 3 *Under Assumptions 1 and 2 we have*

$$\log \frac{L(\theta(\phi^0 + D_T \cdot \psi))}{L(\theta(\phi^0))} \tag{22}$$

$$\xrightarrow{w} -\frac{1}{2} tr\{\Omega^{-1} \frac{\partial \theta_0'}{\partial \phi_0}(\psi_0) \Sigma \frac{\partial \theta_0}{\partial \phi_0}(\psi_0)\} + tr\{\frac{\partial \theta_0'}{\partial \phi_0}(\psi_0) N(0, \Omega^{-1} \otimes \Sigma)\}$$

$$-\frac{1}{2} tr \left\{ \Omega^{-1} \begin{pmatrix} \frac{\partial \theta_1}{\partial \phi_1}(\psi_1) \\ \frac{\partial \theta_2}{\partial \phi_2}(\psi_2) \end{pmatrix}' H_{**} \begin{pmatrix} \frac{\partial \theta_1}{\partial \phi_1}(\psi_1) \\ \frac{\partial \theta_2}{\partial \phi_2}(\psi_2) \end{pmatrix} \right\} + tr \left\{ \Omega^{-1} \begin{pmatrix} \frac{\partial \theta_1}{\partial \phi_1}(\psi_1) \\ \frac{\partial \theta_2}{\partial \phi_2}(\psi_2) \end{pmatrix}' H_{*\varepsilon} \right\}.$$

The maximum likelihood estimator exists with probability tending to one, and the asymptotic distribution of the maximum likelihood estimator $D_T^{-1} \cdot (\hat{\phi} - \phi^0)$ *can be found by maximizing the limiting quadratic likelihood function.*

PROOF. Let $\Delta\theta = \theta(\phi^0 + D_T \cdot \psi) - \theta(\phi^0)$. Then

$$\begin{aligned} l_T(\psi) &= \log \frac{L(\theta(\phi^0 + D_T \cdot \psi))}{L(\theta(\phi^0))} \qquad (23) \\ &= tr\{\Omega^{-1} \sum_{t=1}^{T} \varepsilon_t Z' \Delta\theta\} - \frac{1}{2} tr\{\Omega^{-1} \sum_{t=1}^{T} \Delta\theta' Z_t Z_t' \Delta\theta\} \\ &= tr\{\Omega^{-1} M_{\varepsilon z}(D_T^{-1} \cdot \Delta\theta)\} - \frac{1}{2} tr\{\Omega^{-1}(D_T^{-1} \cdot \Delta\theta)' M_{zz}(D_T^{-1} \cdot \Delta\theta)\}. \end{aligned}$$

Lemma 15 shows that

$$D_T^{-1} \cdot \Delta\theta \rightarrow (\frac{\partial \theta_0'}{\partial \phi_0}(\psi_0), \frac{\partial \theta_1'}{\partial \phi_1}(\psi_1), \frac{\partial \theta_2'}{\partial \phi_2}(\psi_2))' \qquad (24)$$

uniformly for bounded ψ, and (6) and (7) assume that the moments converge weakly

$$M_{z\varepsilon} \xrightarrow{w} H_{*\varepsilon},\ M_{zz} \xrightarrow{w} H_{**}. \qquad (25)$$

This shows that $l_T(\psi)$ converges weakly for fixed ψ to its limit $l(\psi)$ as given by (22). Lemma 16 shows that $\hat{\phi}$ exists with probability tending to one, and that $D_T^{-1} \cdot (\theta(\hat{\phi}) - \theta(\phi^0))$ is tight. We then find from Lemma 14 that also $D_T^{-1} \cdot (\hat{\phi} - \phi^0)$ is tight.

We want to conclude that the asymptotic distribution of $D_T^{-1} \cdot (\hat{\phi} - \phi^0)$ can be found from by maximizing $l(\psi)$, by applying the argmax Theorem, see van der Vaart (1998, Theorem 5.56). Three conditions are given there for this result to hold:

1. For any set A we need joint convergence of $\max_{\psi \in A, |\psi| \leq k} |l_T(\psi)|$ and $\max_{|\psi| \leq k} |l_T(\psi)|$ to $\max_{\psi \in A, |\psi| \leq k} |l(\psi)|$, and $\max_{|\psi| \leq k} |l(\psi)|$. This follows because the likelihood ratio $l_T(\psi)$ given in (22) is a simple combination of moments that converge weakly, see (25), and the function $D_T^{-1} \cdot \Delta\theta$ which converges uniformly for bounded ψ, see (24).

2. The limit function $l(\psi)$ possesses a well separated maximum, because the limit is quadratic and the information is positive definite.

3. The sequence $\hat{\psi}_T$ is tight, as discussed above.

Thus we can apply the argmax theorem and find that $D_T^{-1} \cdot (\hat{\phi} - \phi^0)$ converges weakly to the maximizer of the function derived in (22). ■

3 The I(2) cointegration model

In this section we briefly introduce the $I(2)$ cointegration model and discuss a convenient parametrization and show how the model is a non-linear regression model with integrated regressors. We start with the $I(1)$ model for cointegration as defined by the equation

$$\Delta^2 X_t = \alpha\beta' X_{t-1} - \Gamma\Delta X_{t-1} + \sum_{i=1}^{k-2} \Psi_i \Delta^2 X_{t-i} + \varepsilon_t, \tag{26}$$

where ε_t are i.i.d. $N_n(0, \Omega)$, and the initial values $X_0, \ldots, X_{-k+1}$ are fixed. When the $n \times r$ matrices α and β have full rank r, we define the $I(2)$ model by the non-linear restriction

$$\alpha'_\perp \Gamma \beta_\perp = \xi\eta', \tag{27}$$

where ξ and η are $(n-r) \times s$ matrices, see Johansen (1992).

We apply the reduced rank condition (27) to define the directions

$$\beta_1 = \bar{\beta}_\perp \eta,\ \beta_2 = \beta_\perp \eta_\perp,\ \alpha_1 = \bar{\alpha}_\perp \xi,\ \alpha_2 = \alpha_\perp \xi_\perp,$$

so that $(\beta, \beta_1, \beta_2)$ are orthogonal and span $\mathbb{R}^n$. Similarly for $(\alpha, \alpha_1, \alpha_2)$. For any $n \times m$ matrix a, of rank m, we have defined $a_\perp$ as an $n \times (n-m)$ matrix of full rank with $a' a_\perp = 0$, and $\bar{a} = a(a'a)^{-1}$. With this notation we find the $I(2)$ solution to equation (26), under condition (27), in the form

$$X_t = C_2 \sum_{s=1}^{t} \sum_{i=1}^{s} \varepsilon_i + C_1 \sum_{i=1}^{t} \varepsilon_i + \sum_{i=0}^{\infty} C_i^* \varepsilon_{t-i} + A + Bt, \tag{28}$$

see Johansen (1992) for the full set of conditions. The matrices C_2, C_1 and C_i^* can be expressed in terms of the parameters of the model using the relations

$$A(z) = (1-z)^2 I_n - \alpha\beta' z + \Gamma(1-z)z - \textstyle\sum_{i=1}^{k-2} \Psi_i (1-z)^2 z^i,$$

$$C(z) = A(z)^{-1} = C_2 \tfrac{1}{(1-z)^2} + C_1 \tfrac{1}{(1-z)} + \textstyle\sum_{i=0}^{\infty} C_i^* z^i,\ |z| < 1.$$

With the definition $\Theta = \Gamma\bar{\beta}\bar{\alpha}'\Gamma + I_n - \sum_{i=1}^{k-2} \Psi_i$ we find the relations

$$\begin{aligned} C_2 &= \beta_2(\alpha'_2 \Theta \beta_2)^{-1} \alpha'_2, \\ \beta' C_1 &= \bar{\alpha}' \Gamma C_2, \\ \beta'_1 C_1 &= (\bar{\alpha}'_1 \Gamma \bar{\beta}_1)^{-1} \bar{\alpha}'_1 (I_n - \Theta C_2). \end{aligned} \tag{29}$$

It follows from (28) and (29) that $\tau' X_t = (\beta, \beta_1)' X_t$ is $I(1)$. These relations are called the $CI(2,1)$ cointegrating relations. Moreover, it holds that $\beta' X_t - \bar{\alpha}' \Gamma \Delta X_t$ is stationary.

3.1 The $I(2)$ model as a non-linear regression model

The representation (26) is inconvenient because the parameters are restricted by the non-linear restriction (27), and for the asymptotic analysis we need another parametrization. In the new parameters the equation is

$$\Delta^2 X_t = \alpha(\rho'\tau' X_{t-1} + \psi' \Delta X_{t-1}) + \Omega\alpha_\perp (\alpha'_\perp \Omega \alpha_\perp)^{-1} \kappa' \tau' \Delta X_{t-1} + \varepsilon_t, \tag{30}$$

see Johansen (1997), where we let $\Psi_i = 0$ as they are not involved in the reparametrization. The new parameters are related to $(\alpha, \beta, \Gamma, \xi, \eta)$ by

$$\beta = \tau\rho, \Gamma = -(\alpha\psi' + \Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\kappa'\tau'),$$

so that

$$\alpha_\perp'\Gamma\beta_\perp = -\kappa'\tau'\beta_\perp = -(\kappa'\bar{\rho}_\perp)(\rho_\perp\tau'\beta_\perp) = \xi\eta',$$

since $\rho'\tau'\beta_\perp = \beta'\beta_\perp = 0$. The advantage of this parametrization is that the parameters

$$\alpha, \tau, \rho, \psi, \kappa, \Omega,$$

are variation free. It is seen that the $CI(2,1)$ relations are $\tau = (\beta, \beta_1)$ where ρ picks out the vectors β from τ, and that the multicointegration relation is $\rho'\tau'X_{t-1} + \psi'\Delta X_{t-1}$ or equivalently $\rho'\tau'X_{t-1} + \psi'\bar{\tau}_\perp\tau_\perp'\Delta X_{t-1}$, since $\tau'\Delta X_{t-1}$ is stationary. We define the multicointegration coefficient $\delta = \psi'\bar{\tau}_\perp$. We have here used ψ for a parameter, which should not be confused with the direction of derivatives, in expressions like $d\theta/d\phi(\psi)$ see above.

The parameters in the representation (30) are not identified and in order to investigate the asymptotic distribution, we need to normalize them. We introduce a reparametrization using the true value, which we indicate by the superscript ^{0}. The notation is a bit heavy, and will only be used when we discuss parametrizations.

Let β^0 and τ^0 be given and normalize $\beta = \tau\rho$ on $\bar{\beta}^{0\prime}\beta = I_r$ and τ on $\bar{\tau}^{0\prime}\tau = I_{r+s}$. We rewrite equation (30) as the non-linear regression model (1) with regressors $Z_t = (Z_{0t}', Z_{1t}', Z_{2t}')'$

$$\begin{aligned} Z_{0t}' &= (X_{t-1}'\beta^0 + \Delta X_{t-1}'\psi^0, \Delta X_{t-1}'\tau^0, \Delta^2 X_{t-1}', \cdots, \Delta^2 X_{t-k+2}'), \\ Z_{1t}' &= (\Delta X_{t-1}'\beta_2^0, X_{t-1}'\beta_1^0), \\ Z_{2t}' &= X_{t-1}'\beta_2^0, \end{aligned}$$

of dimensions $(2r + s + (k-2)n, n-r, n-r-s)$ respectively. The regressors are constructed so that the order of integration of Z_{it} is $I(i)$, $i = 0, 1, 2$, see (28), and Z_{it} is measurable with respect to $\sigma(\varepsilon_s, s \leq t-1)$, so that Assumption 1 is satisfied. Corresponding to this choice of regressors we find

$$\begin{aligned} \theta_0' &= (\alpha, \alpha(\psi - \psi^0)'\bar{\tau}^0 + \Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\kappa', \Psi_1, \cdots, \Psi_{k-2}), \\ \theta_1' &= (\alpha(\psi - \psi^0)'\bar{\beta}_2^0 + \Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\kappa'\tau'\bar{\beta}_2^0, \alpha\beta'\bar{\beta}_1^0), \\ \theta_2' &= \alpha\beta'\bar{\beta}_2^0. \end{aligned}$$

This leads to introducing the parameters

$$\begin{aligned} \Psi_* &= (\Psi_1, \cdots, \Psi_{k-2}), \\ A &= \alpha(\psi - \psi^0)'\bar{\tau}^0 + \Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\kappa', \\ B_0 &= \bar{\beta}_2^{0\prime}(\psi - \psi^0), \quad B_1 = \bar{\beta}_1^{0\prime}\beta, \quad B_2 = \bar{\beta}_2^{0\prime}\beta, \\ C &= \bar{\beta}_2^{0\prime}\tau\rho_\perp. \end{aligned} \tag{31}$$

so that

$$\begin{aligned} \theta_0' &= (\alpha, A, \Psi_*), \\ \theta_1' &= (\alpha B_0' + \Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}A(\bar{\rho}(B_1)B_2' + \bar{\rho}_\perp(B_1)C'), \alpha B_1'), \\ \theta_2' &= \alpha B_2'. \end{aligned} \tag{32}$$

Thus model (1), with the restriction (32), is a non-linear regression, where the coefficients θ_i are functions of the parameters

$$\alpha, A, B_0, B_1, B_2, C, \Psi_*, \Omega, \tag{33}$$

which are variation free and Z_{it} is integrated of order i, $I(i)$, $i = 0, 1, 2$.

Note that θ_0 is unrestricted in $\mathbb{R}^{(2r+s+(k-2)n)\times n}$, and that at the true value the cointegrating parameters B_0, B_1, B_2, C are normalized to zero.

4 Asymptotic inference in the cointegrated $I(2)$ model

In this section we first give the asymptotic distribution of the unrestricted maximum likelihood estimator in the cointegrated $I(2)$ model, following Johansen (1997). We then apply the formulation of the model as a non-linear regression model to show how Theorem 3 can be applied to prove a general result about estimators and test statistics under smooth restrictions on the cointegrating parameters, that is, where B_0, B_1, B_2 and C are functions of other parameters (ϕ_1, ϕ_2).

Before stating the result we need some notation. We let $W(u)$ denote the Brownian motion with variance Ω derived from the cumulated ε_t, see Assumption 1, and define the two independent Brownian motions

$$\begin{aligned} W_1(u) &= (\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}W(u), \\ W_2(u) &= (\bar{\rho}_\perp'\kappa(\alpha_\perp'\Omega\alpha_\perp)^{-1}\kappa'\bar{\rho}_\perp)^{-1}\bar{\rho}_\perp'\kappa(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'W(u), \end{aligned} \tag{34}$$

with variances

$$\begin{aligned} Var(W_1) &= (\alpha'\Omega^{-1}\alpha)^{-1} = \Omega_1, & r\times r, \\ Var(W_2) &= (\bar{\rho}_\perp'\kappa(\alpha_\perp'\Omega\alpha_\perp)^{-1}\kappa'\bar{\rho}_\perp)^{-1} = \Omega_2, & s\times s. \end{aligned} \tag{35}$$

The limits of the integrated processes Z_{1t} and Z_{2t} in this case are

$$T^{-1/2}Z_{1[Tu]} = T^{-1/2}\begin{pmatrix} \beta_2'\Delta X_{[Tu]} \\ \beta_1'X_{[Tu]} \end{pmatrix} \xrightarrow{w} \begin{pmatrix} \beta_2'C_2W(u) \\ \beta_1'C_1W(u) \end{pmatrix} = \begin{pmatrix} H_0(u) \\ H_1(u) \end{pmatrix},$$

$$T^{-3/2}Z_{2[Tu]} = T^{-3/2}\beta_2'X_{t-1} \xrightarrow{w} H_2(u) = \textstyle\int_0^u \beta_2'C_2W(s)ds = \int_0^u H_0(s)ds.$$

Note that the process $H_0(u)$ is not the limit of the normalized $I(0)$ process Z_{0t}, but the first components of the normalized $I(1)$ process Z_{1t}. We use the notation $H_* = (H_0', H_1', H_2')'$ as the limit of the normalized $I(1)$ and $I(2)$ regressors in order to conform with the notation in section 1.

Note also that W_1 is constructed from $\alpha'\Omega^{-1}W$ and is hence independent of all variables constructed from $\alpha_\perp'W$, and therefore H_*. Moreover, the process W_2, which is constructed from $\bar{\rho}_\perp'\kappa(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'W$, is independent of $H_0 = \beta_2'C_2W$, which is constructed from $\alpha_2' W(u) = (\kappa' \bar{\rho}_\perp)_\perp' \alpha_\perp' W$. We define the mixed Gaussian distributions

$$B^\infty = H_{**}^{-1} \int_0^1 H_*(dW_1)',$$

and

$$C^\infty = H_{00}^{-1} \int_0^1 H_0(dW_2)',$$

where the superscript refers to the fact that these quantities will turn out to be limit distributions of the corresponding estimators of the parameters (B_0, B_1, B_2, C), see Theorem 4. It is important to note that B^∞ and C^∞ are mixed Gaussian with conditional mean zero, but this does not hold for the joint distribution, since conditioning on $\alpha_\perp' W$, and hence H_*, we have also fixed C^∞. We quote here the result from Johansen (1997).

Theorem 4 *Provided the $I(2)$ process X_t is given by equation (30), and $\hat{\beta}$ is normalized on $\bar{\beta}$ and $\hat{\tau}$ is normalized on $\bar{\tau}$, the non-standard asymptotic distributions of the matrices $\hat{\psi}$, $\hat{\tau}$, $\hat{\beta}$, and $\hat{\rho}$ are given by*

$$\begin{pmatrix} T\hat{B}_0 = T\bar{\beta}_2'(\hat{\psi} - \psi) \\ T\hat{B}_1 = T\bar{\beta}_1'(\hat{\beta} - \beta) \\ T^2\hat{B}_2 = T^2\bar{\beta}_2'(\hat{\beta} - \beta) \\ T\hat{C} = T\bar{\beta}_2'(\hat{\tau} - \tau)\hat{\rho}_\perp \end{pmatrix} \xrightarrow{w} \begin{pmatrix} B^\infty \\ C^\infty \end{pmatrix},$$

The asymptotic distribution of the estimators of the remaining parameters

$$\theta_0' = (\alpha, A, \Psi_1, \cdots, \Psi_{k-2})$$

is given in Theorem 1.

We next give a result for a hypothesis on the cointegrating parameters of the $I(2)$ model. We assume throughout that τ is normalized on $\bar{\tau}^0$ and β on $\bar{\beta}^0$, and therefore express all conditions in the form that the matrices (B_0, B_1, B_2, C) are smooth functions of the (ϕ_1, ϕ_2) where we allow ϕ_1, ϕ_2 to be a matrices:

$$\begin{array}{lll} B_0 = \bar{\beta}_2^{0\prime}(\psi - \psi^0) = B_0(\phi_1, \phi_2), & (n-r-s) \times r, \\ B_1 = \bar{\beta}_1^{0\prime}\tau\rho = B_1(\phi_1, \phi_2), & s \times r, \\ B_2 = \bar{\beta}_2^{0\prime}\tau\rho = B_2(\phi_1, \phi_2), & (n-r-s) \times r, \\ C = \bar{\beta}_2^{0\prime}\tau\rho_\perp = C(\phi_1, \phi_2), & (n-r-s) \times s. \end{array} \tag{36}$$

We assume that the value $(\phi_1, \phi_2) = 0$ gives $(B_0, B_1, B_2, C) = 0$. Together with $(\alpha, A, \Psi_*, \Omega)$ we find the parameter $\theta = (\theta_0', \theta_1', \theta_2')'$ from (1) as function of the parameters $(\alpha, A, \Psi_*, \phi_1, \phi_2, \Omega)$, see (32). Let $n_C = s(n-r-s)$ and $n_B = (2(n-r-s)+s)r$ be the number of elements in C and (B_0, B_1, B_2) respectively.

Theorem 5 *Under the assumption that the parameters B_0, B_1, B_2, C are smoothly parametrized by the continuously identified parameters ϕ_1 and ϕ_2, of dimension q_1 and q_2, which for $\phi_1 = 0$ and $\phi_2 = 0$ satisfy*

$$\frac{\partial B_2}{\partial \phi_1} = 0, \frac{\partial^2 B_2}{\partial \phi_1^2} = 0, \tag{37}$$

the likelihood function is local asymptotic quadratic. Further assume that we can split $\phi_1 = (\phi_{1B}, \phi_{1C})$ of dimension q_{1B}, q_{1C} respectively $(q_{1B} + q_{1C} = q_1)$, so that at the point $\phi = 0$

$$rank(\frac{\partial(B_0, B_1)}{\partial \phi_{1B}}) = q_{1B},\ rank(\frac{\partial C}{\partial \phi_{1C}}) = q_{1C},\ rank(\frac{\partial B_2}{\partial \phi_2}) = q_2$$

$$\frac{\partial(B_0, B_1)}{\partial \phi_{1C}} = 0, \frac{\partial C}{\partial \phi_{1B}} = 0. \tag{38}$$

Then the asymptotic distributions of $T\hat{\phi}_{1C}$ and $(T\hat{\phi}_{1B}, T^2\hat{\phi}_2)$ are mixed Gaussian, so that with $K_C = \frac{\partial C^v}{\partial \phi_{1C}^{v}{}'}$, of dimension $n_C \times q_{1C}$, we have

$$T\hat{\phi}_{1C}^v \xrightarrow{w} \left(K_C'(H_{00} \otimes \Omega_2^{-1})K_C\right)^{-1} K_C'(H_{00} \otimes \Omega_2^{-1})(C^\infty)^v,$$

and with the $(n_B \times (q_{1B} + q_2))$ matrix K_B defined by

$$K_B = \begin{pmatrix} \frac{\partial(B_0^{v\prime}, B_1^{v\prime})'}{\partial \phi_{1B}^{v}{}'} & 0 \\ 0 & \frac{\partial B_2^v}{\partial \phi_2^{v}{}'} \end{pmatrix}$$

we have

$$(T\hat{\phi}_{1B}, T^2\hat{\phi}_2)^v \xrightarrow{w} \left(K_B' \left(H_{**} \otimes \Omega_1^{-1}\right) K_B\right)^{-1} K_B' \left(H_{**} \otimes \Omega_1^{-1}\right) (B^\infty)^v.$$

The asymptotic distribution of the estimates of the remaining parameters

$$\theta_0' = (\alpha, A, \Psi_1, \cdots, \Psi_{k-2})$$

is given in Theorem 1. Hence the asymptotic distribution of the likelihood ratio test for the hypothesis $\theta = \theta(\phi)$ is asymptotically $\chi^2(n_B + n_C - (q_1 + q_2))$.

Remark: The separation condition (38) is a formulation of the block diagonality of the matrix K, see (16) and Boswijk (2000, Theorem 5.1, equation (5.4)).

PROOF. *Local asymptotic quadratic*: We check that the conditions in (37) are enough for the conditions in Assumptions 1 and 2 to be satisfied so that Theorem 3 can be applied. We define $\phi_0 = \theta_0$. It is seen from (32) that $\partial\theta_1/\partial\phi_0$, $\partial\theta_2/\partial\phi_0$, and $\partial^2\theta_1/\partial\phi_0^2$ are zero at the true value $\phi_1 = 0$ and $\phi_2 = 0$, because we have assumed that $(B_0, B_1, B_2, C)(0, 0) = 0$. The derivatives $\partial^2\theta_2/\partial\phi_0^2, \partial^3\theta_2/\partial\phi_0^3, \partial^3\theta_2/\partial\phi_0^2\partial\phi_1$, and $\partial^4\theta_2/\partial\phi_0^4$ are zero because θ_2 is linear in ϕ_0. Finally, from $\theta_2 = \alpha B_2'$ we find

$$\begin{aligned} \frac{\partial\theta_2}{\partial\phi_1}(\psi_1) &= \alpha\frac{\partial B_2'}{\partial\phi_1}(\psi_1) = 0, \\ \frac{\partial^2\theta_2}{\partial\phi_0\partial\phi_1}(\psi_0, \psi_1) &= (\psi_{0\alpha})\frac{\partial B_2'}{\partial\phi_1}(\psi_1) = 0, \\ \frac{\partial^2\theta_2}{\partial\phi_1^2} &= \alpha\frac{\partial^2 B_2}{\partial\phi_1^2}(\psi_1, \psi_1) = 0, \end{aligned}$$

by assumption (37).

Thus most of the conditions (20) and (21) are satisfied automatically because of the formulation of the $I(2)$ model. Only (37) has to be checked to apply Theorem 3.

The conclusion of Theorem 3 is that the limit of the likelihood function is quadratic and that we can find the limit distribution of the maximum likelihood estimator by maximizing the limit function as given in (22). We therefore have to find the score and information and that reduces to finding the derivatives $\partial\theta_i/\phi_i$.

Score and information for ϕ_0 : Because we have no restriction on the parameter θ_0 we find $d\theta_0/d\phi_0(\psi_0) = \psi_0$, and from (11) and (12) we get that the terms depending on ψ_0 are

$$tr\{\Omega^{-1}N(0, \Sigma \otimes \Omega)\psi_0\} - \frac{1}{2}tr\{\Omega^{-1}\psi_0'\Sigma\psi_0\},$$

and the limit of the log likelihood function is maximized with respect to ψ_0 by the choice $\hat{\psi}_0' = N(0, \Sigma^{-1} \otimes \Omega)$, which is the limit of $T^{1/2}((\hat{\alpha}, \hat{A}, \hat{\Psi}_*) - (\alpha, A, \Psi_*))$.

Score for ϕ_1 *and* ϕ_2 : We next evaluate the first order terms of the score with respect to ϕ_1 and ϕ_2. We find that many of the derivatives are zero, and that the relevant non-zero derivatives, evaluated at the true value, see (32), are

$$\begin{aligned}
\tfrac{\partial\theta_1'}{\partial B_0}(\psi_{B_0}) &= (\alpha\psi_{B_0}', 0),\\
\tfrac{\partial\theta_1'}{\partial B_1}(\psi_{B_1}) &= (0, \alpha\psi_{B_1}'),\\
\tfrac{\partial\theta_1'}{\partial C}(\psi_C) &= (\Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\kappa'\bar{\rho}_\perp\psi_C', 0),\\
\tfrac{\partial\theta_2'}{\partial B_2}(\psi_{B_2}) &= \alpha\psi_{B_2}',
\end{aligned}$$

where we have left out the superscript ^{0}. This implies that

$$\begin{aligned}
&\tfrac{\partial\theta_1'}{\partial\phi_1}(\psi_1) = (\alpha\tfrac{\partial B_0'}{\partial\phi_1}(\psi_1) + \Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\kappa'\bar{\rho}_\perp\tfrac{\partial C'}{\partial\phi_1}(\psi_1), \alpha\tfrac{\partial B_1'}{\partial\phi_1}(\psi_1)),\\
&\tfrac{\partial\theta_2'}{\partial\phi_2}(\psi_2) = \alpha\tfrac{\partial B_2'}{\partial\phi_2}(\psi_2).
\end{aligned} \quad (39)$$

We find from (39) and the identity

$$A^{-1} = A^{-1}a(a'A^{-1}a)^{-1}a'A^{-1} + a_\perp(a_\perp'Aa_\perp)^{-1}a_\perp', \quad (40)$$

with $A = \Omega$ and $a = \alpha$, and the notation $\Omega_1 = (\alpha'\Omega^{-1}\alpha)^{-1}$ and $\Omega_2 = (\bar{\rho}_\perp'\kappa(\alpha_\perp'\Omega\alpha_\perp)^{-1}\kappa'\bar{\rho}_\perp)^{-1}$, see (35), that

$$(dW)'\Omega^{-1}\frac{\partial\theta_1'}{\partial\phi_1}(\psi_1) = \left((dW_1)'\Omega_1^{-1}\frac{\partial B_0'}{\partial\phi_1}(\psi_1) + (dW_2)'\Omega_2^{-1}\frac{\partial C'}{\partial\phi_1}(\psi_1), (dW_1)'\Omega_1^{-1}\frac{\partial B_1'}{\partial\phi_1}(\psi_1)\right)$$

and hence the first order term (22) becomes

$$\begin{aligned}
&tr\{[\tfrac{\partial B_0'}{\partial\phi_1}(\psi_1)\textstyle\int_0^1 H_0(dW_1)' + \tfrac{\partial B_1'}{\partial\phi_1}(\psi_1)\int_0^1 H_1(dW_1)']\Omega_1^{-1}\}\\
&\qquad + tr\{\tfrac{\partial B_2'}{\partial\phi_2}(\psi_2)\textstyle\int_0^1 H_2(dW_1)'\Omega_1^{-1}\} + tr\{\tfrac{\partial C'}{\partial\phi_1}(\psi_1)\int_0^1 H_0(dW_2)'\Omega_2^{-1}\},
\end{aligned}$$

where W_1 and W_2 are given in (34).

Information for ϕ_1 and ϕ_2: The second order term of (22) consists of three terms

$$N_{22} = -\tfrac{1}{2}tr\{\Omega^{-1}\tfrac{\partial\theta_2}{\partial\phi_2}(\psi_2)'H_{22}\tfrac{\partial\theta_2}{\partial\phi_2}(\psi_2)\},$$

$$N_{12} = -tr\{\Omega^{-1}\tfrac{\partial\theta_1}{\partial\phi_1}(\psi_1)'\begin{pmatrix} H_{02} \\ H_{12} \end{pmatrix}\tfrac{\partial\theta_2}{\partial\phi_2}(\psi_2)\},$$

$$N_{11} = -\tfrac{1}{2}tr\{\Omega^{-1}\tfrac{\partial\theta_1}{\partial\phi_1}(\psi_1)'\begin{pmatrix} H_{00} & H_{01} \\ H_{10} & H_{11} \end{pmatrix}\tfrac{\partial\theta_1}{\partial\phi_1}(\psi_1)\}.$$

The first is

$$N_{22} = -\frac{1}{2}tr\{\Omega_1^{-1}\frac{\partial B_2'}{\partial\phi_2}(\psi_2)H_{22}\frac{\partial B_2'}{\partial\phi_2}(\psi_2)\}.$$

Next we find

$$N_{12} = -tr\{\Omega_1^{-1}\begin{pmatrix} \frac{\partial B_0}{\partial\phi_1}(\psi_1) \\ \frac{\partial B_1}{\partial\phi_1}(\psi_1) \end{pmatrix}'\begin{pmatrix} H_{02} \\ H_{12} \end{pmatrix}\frac{\partial B_2}{\partial\phi_2}(\psi_2)\},$$

and finally from (40) we get

$$\frac{\partial\theta_1}{\partial\phi_1}(\psi_1)\Omega^{-1}\frac{\partial\theta_1}{\partial\phi_1}(\psi_1)' = \begin{pmatrix} \frac{\partial B_0}{\partial\phi_1}(\psi_1) & \frac{\partial C}{\partial\phi_1}(\psi_1) \\ \frac{\partial B_1}{\partial\phi_1}(\psi_1) & 0 \end{pmatrix}\begin{pmatrix} \Omega_1^{-1} & 0 \\ 0 & \Omega_2^{-1} \end{pmatrix}\begin{pmatrix} \frac{\partial B_0'}{\partial\phi_1}(\psi_1) & \frac{\partial B_1'}{\partial\phi_1}(\psi_1) \\ \frac{\partial C'}{\partial\phi_1}(\psi_1) & 0 \end{pmatrix}.$$

We then find

$$N_{11} = -\tfrac{1}{2}tr\{\Omega_1^{-1}\begin{pmatrix} \frac{\partial B_0}{\partial\phi_1}(\psi_1) \\ \frac{\partial B_1}{\partial\phi_1}(\psi_1) \end{pmatrix}'\begin{pmatrix} H_{00} & H_{01} \\ H_{10} & H_{11} \end{pmatrix}\begin{pmatrix} \frac{\partial B_0}{\partial\phi_1}(\psi_1) \\ \frac{\partial B_1}{\partial\phi_1}(\psi_1) \end{pmatrix}\}$$
$$-\tfrac{1}{2}tr\{\Omega_2^{-1}\tfrac{\partial C'}{\partial\phi_1}(\psi_1)H_{00}\tfrac{\partial C}{\partial\phi_1}(\psi_1)\}.$$

Optimization of limiting likelihood function: Let $\psi_1 = (\psi_{1B}, \psi_{1C})$. From condition (38) we find that $\frac{\partial C'}{\partial\phi_1}(\psi_1)$ depends on ψ_{1C} whereas $(\frac{\partial B_0}{\partial\phi_1}(\psi_1)', \frac{\partial B_1}{\partial\phi_1}(\psi_1)')'$ depends on ψ_{1B}. We then find the limiting likelihood function in terms of the parameters

$$\tilde{\psi}_B = (\frac{\partial B_0}{\partial\phi_1}(\psi_{1B})', \frac{\partial B_1}{\partial\phi_1}(\psi_{1B})', \frac{\partial B_2}{\partial\phi_2}(\psi_2)')', \qquad \tilde{\psi}_C = \frac{\partial C}{\partial\phi_1}(\psi_{1C}),$$

is given by

$$tr\{\tilde{\psi}_B'\int_0^1 H_*(dW_1)'\Omega_1^{-1}\} - \frac{1}{2}tr\{\Omega_1^{-1}\tilde{\psi}_B'H_{**}\tilde{\psi}_B\}$$
$$+ tr\{\tilde{\psi}_C'\int_0^1 H_0(dW_2)'\Omega_2^{-1}\} - \frac{1}{2}tr\{\Omega_2^{-1}\tilde{\psi}_C'H_{00}\tilde{\psi}_C\}.$$

This function has to be optimized with respect to ψ_{1B}, ψ_{1C} and ψ_2. With $K_C = \partial C^v/\partial\phi_{1C}^v{}'$ of dimension $(n_C \times q_{1C})$, and the notation $\int_0^1 H_0(dW_2)' = H_{00}C^\infty$, and $\tilde{\psi}_C^v = K_C\psi_{1C}^v$, the contribution from ψ_{1C} takes the form

$$\psi_{1C}^v{}'K_C'(H_{00}\otimes\Omega_2^{-1})(C^\infty)^v - \frac{1}{2}\psi_{1C}^v{}'K_C'(H_{00}\otimes\Omega_2^{-1})K_C\psi_{1C}^v,$$

which is optimized for

$$\psi_{1C}^v = \left(K_C'(H_{00} \otimes \Omega_2^{-1})K_C\right)^{-1} K_C'(H_{00} \otimes \Omega_2^{-1})(C^\infty)^v.$$

This shows that the limit $T\hat{\psi}_{1C}$ is mixed Gaussian. The optimal value is

$$(C^\infty)^{v\prime}(H_{00} \otimes \Omega_2^{-1})K_C \left(K_C'(H_{00} \otimes \Omega_2^{-1})K_C\right)^{-1} K_C'(H_{00} \otimes \Omega_2^{-1})(C^\infty)^v.$$

The corresponding contribution for the likelihood ratio test for the hypothesis (36) is found by comparing with the expression for the unrestricted C, that is, by replacing K_C by $I_{(n-r-s)s}$:

$$(C^\infty)^{v\prime}(H_{00} \otimes \Omega_2^{-1})(C^\infty)^v,$$

and the difference is

$$LR_C = (C^\infty)^{v\prime} K_{C\perp} \left(K_{C_\perp}'(H_{00}^{-1} \otimes \Omega_2)K_{C\perp}\right)^{-1} K_{C\perp}'(C^\infty)^v,$$

using the identity (40) with $A = (H_{00} \otimes \Omega_2^{-1})$ and $a = K_C$.

The optimization with respect to ψ_{1B} and ψ_2 is handled the same way, and the contribution to the likelihood ratio test is

$$LR_B = (B^\infty)^{v\prime} K_{B_\perp} \left(K_{B\perp}'(H_{**}^{-1} \otimes \Omega_1)K_{B\perp}\right)^{-1} K_{B\perp}'(B^\infty)^v.$$

The usual conditioning argument then goes as follows. If we condition on $\alpha_\perp' W$, then LR_B is distributed as $\chi^2(n_B - q_{1B} - q_2)$. As this does not depend on the conditioning variable it is independent of $\alpha_\perp' W$, and hence of LR_C, which by a similar conditioning argument is distributed as $\chi^2(n_C - q_{1C})$. Thus the sum is distributed as

$$\chi^2(n_B + n_C - (q_1 + q_2)),$$

as was to be proved. ■

5 Asymptotic inference for the cointegrated $I(2)$ model

In (31) the cointegration parameters B_0, B_1, B_2, C are defined as functions of τ, ρ, ψ and the hypotheses we test below on β, τ, ρ and ψ are reformulated as hypotheses on B_1, B_2, and C in terms of new parameter ϕ_{1B}, ϕ_{1C}, and ϕ_2. Generally we define ϕ_2, as the parameters that we can recover from B_2, which are then T^2 consistent, and ϕ_{1B} and ϕ_{1C} are determined so that together with ϕ_2 they determine B_0, B_1 and C. In Theorem 5, the conditions we need to get a quadratic likelihood function in the limit are

$$\frac{\partial B_2}{\partial \phi_1} = 0, \frac{\partial^2 B_2}{\partial \phi_1^2} = 0, \tag{41}$$

and the conditions for asymptotic χ^2 inference are

$$\frac{\partial B_1}{\partial \phi_{1C}} = 0, \frac{\partial C}{\partial \phi_{1B}} = 0. \tag{42}$$

Throughout we denote by τ and $\beta = \tau\rho$ the parameters that are normalized on $\bar{\tau}^0$ and $\bar{\beta}^0$ respectively. The normalization implies that

$$\tau = \tau^0\bar{\tau}^{0\prime}\tau + \tau_\perp^0\bar{\tau}_\perp^{0\prime}\tau = \tau^0 + \tau_\perp^0\bar{\tau}_\perp^{0\prime}\tau,$$
$$\beta = \beta^0\bar{\beta}^{0\prime}\beta + \beta_\perp^0\bar{\beta}_\perp^{0\prime}\beta = \beta^0 + \beta_\perp^0\bar{\beta}_\perp^{0\prime}\beta.$$

From $\bar{\tau}^{0\prime}\beta = \bar{\tau}^{0\prime}\tau\rho = \rho$ and $\beta_\perp^0 = (\tau_\perp^0, \bar{\tau}^0\rho_\perp^0)$ we find that

$$\beta_\perp^0\bar{\beta}_\perp^{0\prime} = \tau_\perp^0(\tau_\perp^{0\prime}\tau_\perp^0)^{-1}\tau_\perp^{0\prime} + \bar{\tau}^0\rho_\perp^0(\rho_\perp^{0\prime}\bar{\tau}^{0\prime}\bar{\tau}^0\rho_\perp^0)^{-1}\rho_\perp^{0\prime}\bar{\tau}^{0\prime},$$

so that

$$\rho = \bar{\tau}^{0\prime}\beta = \bar{\tau}^{0\prime}(\beta^0\bar{\beta}^{0\prime} + \beta_\perp^0\bar{\beta}_\perp^{0\prime})\beta = \rho^0 + \bar{\tau}^{0\prime}\bar{\beta}_1^0\rho_\perp^{0\prime}\rho, \tag{43}$$

and

$$\rho_\perp = \rho_\perp^0 - \tau^{0\prime}\bar{\beta}^0\rho'\rho_\perp^0. \tag{44}$$

This shows that under the normalizations of τ and β the parameter $B = \bar{\beta}_1^{0\prime}\beta = (\beta_1^{0\prime}\beta_1^0)^{-1}\rho_\perp^{0\prime}\rho$ depends on ρ through $\rho_\perp^{0\prime}\rho$ only.

5.1 Hypotheses on β

We start by testing hypotheses on

$$\beta = \beta^0 + \beta_1^0 B_1 + \beta_2^0 B_2,$$

see (36). Hypotheses on β can be expressed as restrictions on $B_1 = B_1(\phi_1, \phi_2)$ and $B_2 = B_2(\phi_1, \phi_2)$, and we define $\phi_{1C} = C$ and $\phi_{1B_0} = B_0$, so that (42) is satisfied, and the only conditions we need to check are $\partial B_2/\partial\phi_1 = 0$ and $\partial^2 B_2/\partial\phi_1^2 = 0$ at $\phi = 0$, see (41). We consider the hypotheses $\beta = K\phi$, and $\beta = (b, b_\perp\phi)$ in detail.

Proposition 6 *The test that $\beta = K\phi$, $K(n\times m)$ in model (30) satisfies*

$$-2\log LR(\beta = K\phi) \xrightarrow{w} \chi^2((n-m)r).$$

PROOF. If $\beta^0 = K\phi^0$, then $\beta_\perp^0 = (K_\perp, \bar{K}\phi_\perp^0)$, and therefore $\bar{\beta}_1^0 = K_\perp a_1 + \bar{K}\phi_\perp^0 a_2$, $\bar{\beta}_2^0 = K_\perp b_1 + \bar{K}\phi_\perp^0 b_2$ for some $(n-r)\times(n-r)$ matrix

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}$$

of full rank. Then $K'\bar{\beta}_1^0 = \phi_\perp^0 a_2, K'\bar{\beta}_2^0 = \phi_\perp^0 b_2$ and from (36)

$$B_1 = \bar{\beta}_1^{0\prime}K\phi = a_2'\phi_\perp^{0\prime}\phi, \quad B_2 = \bar{\beta}_2^{0\prime}K\phi = b_2'\phi_\perp^{0\prime}\phi.$$

Thus the $(m-r)r$ parameters of $\phi_{\perp}^{0\prime}\phi$ parametrize B_1 and B_2 under the null hypothesis. The number of parameters without the hypothesis is $(n-r)r$, leaving $(n-m)r$ degrees of freedom if the asymptotic distribution is χ^2.

The $(m-r)\times(n-r)$ matrix (a_2, b_2) has rank $m-r$ so that $\phi_{\perp}^{0\prime}\phi$ can be recovered from (B_1, B_2). In order to define ϕ_2 we let f =rank(b_2), so that $b_2 = uv'$, for u and v of rank f. We define $\phi_2 = u'\phi_{\perp}^{0\prime}\phi$ as the parameters that determine B_2, and let $\phi_{1B_1} = u_{\perp}'\phi_{\perp}^{0\prime}\phi$ be the remaining combinations of $\phi_{\perp}^{0\prime}\phi$. We then find

$$B_1 = a_2'\bar{u}\phi_2 + a_2'\bar{u}_{\perp}\phi_{1B_1}, \quad B_2 = v\phi_2.$$

The parameters $\phi_2, \phi_{1B}, \phi_{1C}$ are variation free and the conditions of Theorem 5 are satisfied. We find

$$\frac{\partial B_2}{\partial \phi_2} = v, \frac{\partial B_1}{\partial \phi_{1B_1}} = a_2'\bar{u}_{\perp},$$

where rank$(v) = f$, and rank$(a_2'\bar{u}_{\perp}) = m-r-f$, because rank$(a_2, b_2) = m-r$. Thus ϕ_2, ϕ_{1B_1} can be found from B_1, B_2, and inference on the hypothesis is asymptotically χ^2. In case $b_2 = 0$, we modify the definitions and take $\phi_2 = 0$ and $\phi_{1B_1} = \phi_{\perp}^{0\prime}\phi$. The argument is then as before. ■

Next consider $\beta = (b, b_{\perp}\phi)$, where b is of dimension $n\times m, m \leq r$.

Proposition 7 *The test that $\beta = (b, b_{\perp}\phi)$, $b(n\times m)$ in model (30) satisfies*

$$-2\log LR(\beta = (b, b_{\perp}\phi)) \overset{w}{\to} \chi^2((n-r)m).$$

PROOF. If $\beta^0 = (b, b_{\perp}\phi^0)$ then $\beta_{\perp}^0 = \bar{b}_{\perp}\phi_{\perp}^0$, and $\bar{\beta}_1^0 = \bar{b}_{\perp}\phi_{\perp}^0 a_1, \bar{\beta}_2^0 = \bar{b}_{\perp}\phi_{\perp}^0 a_2$, where (a_1, a_2) has rank $n-r$. Then

$$B_1 = \bar{\beta}_1^{0\prime}(b, b_{\perp}\phi) = (0, a_1'\phi_{\perp}^{0\prime}\phi) = (0, \phi_{1B_1}),$$
$$B_2 = \bar{\beta}_2^{0\prime}(b, b_{\perp}\phi) = (0, a_2'\phi_{\perp}^{0\prime}\phi) = (0, \phi_2),$$

and we define again $\phi_{1C} = C$ and $\phi_{1B_0} = B_0$. This defines variation free parameters $(\phi_{1B_0}, \phi_{1B_1}, \phi_{1C}, \phi_2)$, so that condition (42) are satisfied. Note that $\hat{\phi}_2$ becomes T^2 consistent and $\hat{\phi}_1$ is T consistent, so that the asymptotic distribution is χ^2. The total number of parameters under the null is $(n-r)(r-m)$ and outside the null it is $(n-r)r$, giving $(n-r)m$ degrees of freedom for the test. ■

5.2 Test on ρ

We want here to test the hypotheses $\rho = K\phi$, and $\rho = (b, b_{\perp}\phi)$, where K is $(r\times s)\times m$ and b is $(r+s)\times m$. Because $\beta = \tau\rho$, and $\bar{\tau}^{0\prime}\beta = \rho$, these are really hypotheses on β, when τ is normalized on $\bar{\tau}^0$. The hypotheses only involve the parameters B_1, and the conditions (37) and (41) are satisfied. Hence we only have to define ϕ_{1B_1}, so that $dB_1/d\phi_{1B1}$ has full rank.

Proposition 8 *The test that $\rho = K\phi$, where K is $((r+s)\times m)$ in model (30) satisfies*

$$-2\log LR(\rho = K\phi) \overset{w}{\to} \chi^2((n-m)r).$$

PROOF. For $\rho = K\phi$, we find $\rho_\perp = (K_\perp, \bar{K}\phi_\perp)$, so that with $\beta_1^0 = \bar{\tau}^{0\prime}\rho_\perp^0$ we find

$$B_1 = \bar{\beta}_1^{0\prime}\tau\rho = (\beta_1^{0\prime}\beta_1^0)^{-1}(K_\perp, \bar{K}\phi_\perp^0)'K\rho = \begin{pmatrix} 0 \\ (\beta_1^{0\prime}\beta_1^0)^{-1}\phi_\perp^{0\prime}\phi \end{pmatrix} = \begin{pmatrix} 0 \\ \phi_{1B} \end{pmatrix}$$

while C, B_0 are unrestricted and the conditions for asymptotic χ^2 are satisfied. We find that the number of parameters under the null in ϕ_{1B} are $(m-r)r$, and hence the degrees of freedom are given by $(n-m)r$. ■

Proposition 9 *The test that $\rho = (b, b_\perp\phi)$, $b((r+s)\times m)$ in model (30) satisfies*

$$-2\log LR(\rho = (b, b_\perp\phi)) \xrightarrow{w} \chi^2(s(n-m)).$$

PROOF. Under the hypothesis $\rho = (b, b_\perp\phi)$, we have $\rho_\perp = \bar{b}_\perp\phi_\perp$, so that

$$B_1 = (\beta_1^{0\prime}\beta_1^0)^{-1}\rho_\perp^{0\prime}\rho = (0, (\beta_1^{0\prime}\beta_1^0)^{-1}\phi_\perp^{0\prime}\phi) = (0, \phi_{1B_1}),$$

whereas B_0, B_2, C are unrestricted. Thus the hypothesis $\rho = (b, b_\perp\phi)$ is a linear hypothesis on ϕ_{1B} only and hence the conditions for Theorem 5 are satisfied. The number of parameters under the null hypothesis in ϕ_{1B} are $s(r+s-m)$, giving $s(n-m)$ degrees of freedom for the test. ■

5.3 Hypotheses on τ

In this subsection we investigate hypotheses on τ. We consider the usual hypotheses $\tau = K\phi$, and $\tau = (b, b_\perp\phi)$, which are invariant to normalization of τ and β on $\bar{\tau}^0$ and $\bar{\beta}^0$ respectively. We also investigate the test on individual elements of τ, a hypothesis that requires that τ has been identified.

5.3.1 The same linear restrictions on all of τ

Let $K(n \times m)$ be given, where $r + s \leq m \leq n$. We want to test the hypothesis $\tau = K\phi$ or $K_\perp'\tau = 0$.

Proposition 10 *The test that $\tau = K\phi$, $K(n\times m)$ in model (30) satisfies*

$$-2\log LR(\tau = K\phi) \xrightarrow{w} \chi^2((r+s)(n-m)).$$

PROOF. We find $\tau_\perp = (K_\perp, \bar{K}\phi_\perp)$, so that from

$$\tau = \tau^0 + \tau_\perp^0\bar{\tau}_\perp^{0\prime}\tau = K\phi^0 + \tau_\perp^0\bar{\tau}_\perp^{0\prime}\tau$$

we find by multiplying by $\bar{K}'$ that

$$\phi = \phi^0 + (K'K)^{-1}\phi_\perp^0(\phi_\perp^{0\prime}(K'K)^{-1}\phi_\perp^0)^{-1}\phi_\perp^{0\prime}\phi,$$

which shows that ϕ can be recovered from $\phi_\perp^{0\prime}\phi$. The restriction is clearly only a restriction on τ and not on ρ, so we define $\phi_{1B_1} = B_1$, and

$$B_2 = \bar{\beta}_2^{0\prime} K\phi\rho = \begin{pmatrix} 0 \\ (\phi_\perp^{0\prime}\bar{K}'\bar{K}\phi_\perp^0)^{-1}\phi_\perp^{0\prime}\phi\rho \end{pmatrix} = \begin{pmatrix} 0 \\ \phi_2 \end{pmatrix},$$
$$C = \bar{\beta}_2^{0\prime} K\phi\rho_\perp = \begin{pmatrix} 0 \\ (\phi_\perp^{0\prime}\bar{K}'\bar{K}\phi_\perp^0)^{-1}\phi_\perp^{0\prime}\phi\rho_\perp \end{pmatrix} = \begin{pmatrix} 0 \\ \phi_{1C} \end{pmatrix}.$$

Under the normalization the parameters $(\phi_\perp^{0\prime}\phi\rho, \phi_\perp^{0\prime}\phi\rho_\perp)$ of dimensions $(m-r-s)\times r$ and $(m-r-s)\times s$ respectively, are variation free, in the sense that for any values of these parameters we can recover ϕ from $\phi_\perp^{0\prime}\phi\rho$ and $\phi_\perp^{0\prime}\phi\rho_\perp$ for given ρ. The degrees of freedom for the test is therefore

$$(n-r-s)(r+s) - (m-r-s)(r+s) = (n-m)(r+s).$$

It is seen that (B_1, B_2, C) are variation free and linear in the parameters $(\phi_{1B}, \phi_{1C}, \phi_2)$ respectively. Thus the conditions of Theorem 5 are satisfied, and therefore the test of $\tau = K\phi$ is asymptotically $\chi^2((r+s)(n-m))$. ■

5.3.2 The hypothesis of some vectors in τ known

Let b $(n \times m)$ be given, where $1 \leq m < r+s$. We want to test the restriction that $\tau = (b, b_\perp\phi)$.

Proposition 11 *If the dimension of sp(β)$\cap$sp(b) is max$(0, m-s)$, then the test that $\tau = (b, b_\perp\phi)$ in model (30) satisfies*

$$-2\log LR(\tau = (b, b_\perp\phi)) \xrightarrow{w} \chi^2((n-r-s)m)).$$

COMMENT: The result depends on how the spaces spanned by β and b are related. The condition states that the intersection should be as small as possible. That is, if $m \leq s$, the spaces should intersect only at $\{0\}$ and if $m > s$, so that they have to overlap, the intersection should be as small as possible, that is, of dimension $m-s$. In practice on can check the condition by testing if the individual vectors of b are contained in the space spanned by β, see Proposition 7.

We first give a reformulation of the overlap condition, which will be used in the proof of Proposition 11. Corresponding to the decomposition $\tau = (b, b_\perp\phi)$ we decompose $\rho = (\rho_1', \rho_2')'$, and for more complicated expressions like $\rho_\perp$ we use the notation $\rho_\perp = ((\rho_\perp)_1', (\rho_\perp)_2')'$.

Lemma 12 *Let $\tau = (b, b_\perp\phi)$ so that $\beta = \tau\rho = b\rho_1 + b_\perp\phi\rho_2$. The dimension of sp($\beta$)$\cap$sp(b) is max$(0, m-s)$ if and only if the dimension of sp($b_\perp'\beta$) is $\min((r+s-m), r)$, which is equivalent to*

$$rank(\rho_2) = \min((r+s-m), r). \qquad (45)$$

The main observation is that if v is a vector so that $\rho_2 v = 0$, then $\beta v = b\rho_1 v + b_\perp \phi \rho_2 v = b\rho_1 v$, gives a vector in both sp(β) and in sp(b). Thus if ρ_2 has maximal rank, which is $\min((r+s-m), r)$, then the number of linearly independent vectors in sp$(\beta)\cap$sp(b) is as small as possible.

PROOF. Case 1: Assume that $m < s$, so that $r+s-m > r$. In this case (45) implies $rank(\rho_2) = r$, which means that no non-zero vector satisfies $\rho_2 v = 0$, which is the same as sp$(\beta)\cap$sp$(b) = \{0\}$.

Case 2: Assume that $m \geq s$, so that $(r+s-m) \leq r$. In this case (45) implies $rank(\rho_2) = r+s-m$, so that there are exactly $m-s$ linearly independent vectors orthogonal to the rows of ρ_2. This means that there are exactly $m-s$ linearly independent vectors in *sp*$(\beta)\cap$*sp*(b). ■

PROOF. OF PROPOSITION **11:** The restriction implies that $\tau_\perp = \bar{b}_\perp \phi_\perp$, or $b'\tau_\perp = 0$. The normalization $\bar{\tau}^{0\prime}\tau = I_{r+s}$ implies that

$$\phi = \phi^0 + (b_\perp' b_\perp)^{-1}\phi_\perp^0(\phi_\perp^{0\prime}(b_\perp' b_\perp)^{-1}\phi_\perp^0)^{-1}\phi_\perp^{0\prime}\phi,$$

which shows that ϕ can be recovered from $\phi_\perp^{0\prime}\phi$, and from (43) we see that ρ can be recovered from $\rho_\perp^{0\prime}\rho$. The parameter τ (normalized on τ^0) has $(n-r-s)(r+s)$ parameters whereas $\phi_\perp^{0\prime}\phi$ only has $(n-r-s)(r+s-m)$ and the difference is $(n-r-s)m$ corresponding to the degrees of freedom, if the asymptotic distribution is χ^2.

We find from $\beta = \tau\rho = (b, b_\perp\phi)\rho = b\rho_1 + b_\perp\phi\rho_2$ that with $\beta_1^0 = \bar{\tau}^0\rho_\perp^0$, and $\beta_2^0 = \tau_\perp^0$ we get

$$\begin{aligned}
&B_1 = \bar{\beta}_1^{0\prime}\tau\rho = (\beta_1^{0\prime}\beta_1^0)^{-1}\rho_\perp^{0\prime}\rho = \phi_{1B}\\
&B_2 = \bar{\beta}_2^{0\prime}\tau\rho = (\phi_\perp^{0\prime}(b_\perp' b_\perp)^{-1}\phi_\perp^0)^{-1}\phi_\perp^{0\prime}\bar{b}_\perp'(b, b_\perp\phi)\rho = (0, \tilde{\phi}\rho_2)\\
&C = \bar{\beta}_2^{0\prime}\tau\rho_\perp = (\phi_\perp^{0\prime}(b_\perp' b_\perp)^{-1}\phi_\perp^0)^{-1}\phi_\perp^{0\prime}\bar{b}_\perp'(b, b_\perp\phi)(\rho_\perp)_2 = (0, \tilde{\phi}(\rho_\perp)_2)
\end{aligned}$$

where

$$(\rho, \rho_\perp) = \begin{pmatrix} \rho_1 & (\rho_\perp)_1 \\ \rho_2 & (\rho_\perp)_2 \end{pmatrix}$$

and where $\tilde{\phi} = (\phi_\perp^{0\prime}(b_\perp' b_\perp)^{-1}\phi_\perp^0)^{-1}\phi_\perp^{0\prime}\phi$ is of dimension $(n-r-s)\times(r+s-m)$. The matrix $(\rho_2, (\rho_\perp)_2)$ has dimension $(r+s-m)\times(r+s)$ and rank $(r+s-m)$, so that $\tilde{\phi}$ can be recovered from (B_2, C, ρ).

We define $\phi_2 = \tilde{\phi}\rho_2$. What remains to define is ϕ_{1C} and discuss C as a function of ϕ_{1B}, ϕ_{1C}, and ϕ_2. Note, however, that condition (41) is satisfied so we get asymptotic quadratic likelihood function, whereas the problem is (42) which does not in general hold.

Case 1. Assume that $m < s$, so that from (45) ρ_2^0 has rank $r < r+s-m$. For ρ close to ρ^0 the matrix ρ_2 will have rank r and we define in this case $\phi_{1C} = \tilde{\phi}(\rho_2)_\perp$, that is, some components of $\tilde{\phi}$ are in ϕ_2 and the remaining ones in ϕ_{1C}, then

$$\begin{aligned}
C &= \tilde{\phi}\,(\rho_\perp)_2 = \tilde{\phi}\rho_2\bar{\rho}_2'\,(\rho_\perp)_2 + \tilde{\phi}\,(\rho_2)_\perp\,\overline{(\rho_2)}_\perp'\,(\rho_\perp)_2\\
&= \phi_2\bar{\rho}_2'\,(\rho_\perp)_2 + \phi_{1C}\overline{(\rho_2)}_\perp'\,(\rho_\perp)_2\,.
\end{aligned}$$

It is seen that C is a non-linear function of $(\phi_2, \phi_{1B}, \phi_{1C})$. The functions

$$\begin{aligned} &\rho \mapsto \bar{\rho}_2' = (\rho_2'\rho_2)^{-1}\rho_2' \\ &\rho \mapsto \overline{(\rho_2)}_\perp' = \left[(\rho_2)_\perp' (\rho_2)_\perp\right]^{-1} (\rho_2)_\perp' \\ &\rho \mapsto (\rho_\perp)_2 = ((I_{r+s} - \rho^0(\rho'\rho^0)^{-1}\rho')\bar{\rho}_\perp^0)_2 \end{aligned}$$

are differentiable at the point $\rho = \rho^0$, because ρ_2^0 and hence $\rho_2^{0\prime}\rho_2^0$ have full rank $r < r+s-m$ and $(\rho_2^0)_\perp$ and hence $(\rho_2^0)_\perp' (\rho_2^0)_\perp$ have full rank $r+s-m-r = s-m \geq 1$.

Thus we have $\partial C/\partial\phi_{1B}|_{\phi_2=0,\phi_{1C}=0} = 0$, which by Theorem 5 implies that inference is asymptotic χ^2.

Case 2. Assume that $m \geq s$, so that from (45), ρ_2^0 has rank $r+s-m \leq r$. In a neighbourhood of ρ^0, the matrix ρ_2 will also have rank $r+s-m$ and in this case we can recover all components of $\tilde{\phi}$ from $\tilde{\phi}\rho_2$, because $\tilde{\phi}\rho_2\rho_2'(\rho_2\rho_2')^{-1} = \tilde{\phi}$, and we therefore define $\phi_2 = \tilde{\phi}\rho_2$ and $\phi_{1C} = 0$. We find

$$C = \tilde{\phi}(\rho_\perp)_2 = \tilde{\phi}\rho_2\rho_2'(\rho_2\rho_2')^{-1}(\rho_\perp)_2 = \phi_2\rho_2'(\rho_2\rho_2')^{-1}(\rho_\perp)_2,$$

which is a function of ϕ_2 and ϕ_{1B}, through ρ. The functions

$$\begin{aligned} &\rho \mapsto (\rho_2\rho_2')^{-1}\rho_2' \\ &\rho \mapsto (\rho_\perp)_2 = ((I_{r+s} - \rho^0(\rho'\rho^0)^{-1}\rho')\bar{\rho}_\perp^0)_2 \end{aligned}$$

are differentiable at a point $\rho = \rho^0$, because ρ_2^0 is of rank $r+s-m$. We therefore find that $\partial C/\partial\phi_{1B}|_{\phi_2=0} = 0$, so that the conditions of Theorem 5 for asymptotic χ^2 inference are satisfied. ∎

5.3.3 Test of individual coefficients in τ

In order for this hypothesis to make sense, τ need to be normalized. We take the simplest situation where $\tau = \tau_k$ is normalized on a matrix k, so that $\tau_k'k = I_{r+s}$. Let $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^{r+s}$ be vectors. We want to test that $a'\tau_k b = 0$. This is in particular a formulation of the hypothesis that a single element of τ_k is zero.

Proposition 13 *Let τ be normalized on $k'\tau = I_{r+s}$, and let $a \in \mathbb{R}^n, b \in \mathbb{R}^{r+s}$. We want to test that $a'\tau b = 0$, and assume that $a'k_\perp \neq 0$ and $\bar{\rho}_\perp^{0\prime} b \neq 0$. Then in model (30) the test satisfies*

$$-2\log LR(a'\tau b = 0) \xrightarrow{w} \chi^2(1).$$

Comment. If $a'k_\perp = 0$, then $a = kc$, for some c, and then, because $k'\tau = I_{r+s}$, we would have $a'\tau b = c'k'\tau b = c'b$, is automatically satisfied for all τ due to the normalization, and hence not testable.

If $\bar{\rho}_\perp^{0\prime} b = 0$, then inference is probably not asymptotic χ^2, as can be seen from the proof below. The assumption can be tested as a hypothesis on $\rho : \rho_\perp' b = 0$, or $\rho = (b, b_\perp\phi)$, see Proposition 9.

Proof. Let τ^0 be the true value, also normalized on k. For simplicity we also assume that $\tau_\perp^0$ is normalized so that $\bar{\tau}_\perp^{0\prime} k_\perp = I_{n-r-s}$. The results above are all formulated for $\tau_{\bar\tau^0}$ normalized on $\bar\tau^0$, for which we have $\tau_{\bar\tau^0} = \tau^0 + \tau^0_\perp(B_2\bar\rho' + C\bar\rho'_\perp)$ so that

$$\begin{aligned}\tau_k &= \tau_{\bar\tau^0}(k'\tau_{\bar\tau^0})^{-1} = (\tau^0 + \tau^0_\perp(B_2\bar\rho' + C\bar\rho'_\perp))(k'\tau^0 + k'\tau^0_\perp(B_2\bar\rho' + C\bar\rho'_\perp))^{-1} \\ &= \tau^0 + k_\perp(B_2\bar\rho' + C\bar\rho'_\perp) + R(B_2, C, \phi_{1B}),\end{aligned}$$

where the remainder term $R(B_2, C, \phi_{1B})$ has at least two factors C or a factor B_2, so that $\partial R/\partial B_1 = 0$ and $\partial R/\partial C = 0$ for $C = 0$ or $B_2 = 0$.

From $a'\tau^0_k b = 0$ we find that the restriction is

$$\begin{aligned}0 = a'\tau_k b &= a'k_\perp(B_2\bar\rho' + C\bar\rho'_\perp)b + a'R(B_2, C, \phi_{1B})b \\ &= a'k_\perp(B_2\bar\rho' + C\bar\rho^{0\prime}_\perp)b + a'R(B_2, C, \phi_{1B})b,\end{aligned}$$

which is a non-linear restriction on the parameters in (36). We define $v' = a'k_\perp \neq 0$ and $u = \bar\rho^{0\prime}_\perp b \neq 0$, and the restriction becomes

$$\begin{aligned}0 &= v'B_2u + v'Cu + a'k_\perp(B_2(\bar\rho - \bar\rho^0)' + C(\bar\rho_\perp - \bar\rho^{0\prime}_\perp)b + a'R(B_2, C, \phi_{1B})b \\ &= v'B_2u + v'Cu + v'C(\bar\rho_\perp - \bar\rho^{0\prime}_\perp)b + a'R_1(B_2, C, \phi_{1B})b,\end{aligned}$$

where again $R_1(B_2, C, B_1)$ has at least two factors C or a factor B_2. We to solve the relation as

$$v'Cu = -v'B_2u - v'C(\bar\rho_\perp - \bar\rho^{0\prime}_\perp)b - a'R_1(B_2, C, \phi_{1B})b.$$

We define C_{ij} by the decomposition

$$C = \bar v C_{11}\bar u' + \bar v_\perp C_{21}\bar u' + \bar v C_{12}\bar u'_\perp + \bar v_\perp C_{22}\bar u'_\perp,$$

so that $v'Cu = C_{11}$ and define $\phi_{1C} = (C_{21}, C_{22}, C_{12})$. The restriction then becomes

$$C_{11} = -v'B_2u - (C_{11}\bar u' + C_{12}\bar u'_\perp)(\bar\rho_\perp - \bar\rho^{0\prime}_\perp)b - a'R_1(C_{11}, \phi_{1B}, \phi_{1C}, \phi_2)b,$$

so that

$$\frac{\partial C_{11}}{\partial\phi_{1B}} = -\bar u'(\bar\rho_\perp - \bar\rho^{0\prime}_\perp)b - (C_{11}\bar u' + C_{12}\bar u'_\perp)\frac{\partial(\bar\rho_\perp - \bar\rho^{0\prime}_\perp)b}{\partial\phi_{1B}} - \frac{\partial a'R_1 b}{\partial\phi_{1B}} - \frac{\partial a'R_1 b}{\partial C_{11}}\left(\frac{\partial C_{11}}{\partial\phi_{1B}}\right),$$

which is zero for $B_1 = 0, C = 0$ and $B_2 = 0$, because $\partial R_1/\partial\phi_{1B} = 0$, and $\partial R_1/\partial C_{11} = 0$. This shows that condition (38) is satisfied and Theorem 5 can be applied, so that inference is asymptotically $\chi^2(1)$. ■

5.4 Test for no multicointegration $\psi'\tau_\perp = 0$.

A more complicated hypothesis is to test for the absence of multicointegration, that is that $\delta = \psi'\bar\tau_\perp = 0$. One could also test that some entries of the matrix δ were zero, but such a hypothesis would depend on how β was identified, whereas the hypothesis that they are all zero is invariant to normalizations of β. We consider only $\delta = 0$. From

$$\tau_\perp - \tau_\perp^0 = -\bar{\tau}^0(\bar{\rho}(B_1)B_2' + \bar{\rho}_\perp(B_1)C')(\tau_\perp^{0\prime}\tau_\perp^0),$$
$$\psi - \psi^0 = \tau^0 A'\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1} + \tau_\perp^0 B_0,$$

we find

$$\psi'\bar{\tau}_\perp = [\psi^{0\prime} + (\psi - \psi^0)'\bar{\tau}^0\tau^{0\prime} + B_0'\tau_\perp^{0\prime}][\tau_\perp^0 - \bar{\tau}^0(\bar{\rho}(B_1)B_2' + \bar{\rho}_\perp(B_1)C')(\tau_\perp^{0\prime}\tau_\perp^0)](\tau_\perp'\tau_\perp)^{-1},$$

from which it follows that

$$T(\hat{\delta} - \delta) = -\psi^{0\prime}\bar{\tau}^0\bar{\rho}_\perp^0 T\hat{C}' + T\hat{B}_0' + O_P(T^{-1/2})$$

a result due to Paruolo (2000). The result shows that the asymptotic distribution is a combination of the distributions of B_0^∞ and C^∞, see Theorem 4. The restriction that $\delta = \psi'\bar{\tau}_\perp = 0$, can be reduced, using that $\delta^0 = \psi^{0\prime}\bar{\tau}_\perp^0 = 0$, to

$$B_0' = \psi'\bar{\tau}^0(\bar{\rho}(B_1)B_2' + \bar{\rho}_\perp(B_1)C'),$$

that is, B_0 is expressed as a function of the variation free parameters $\psi'\bar{\tau}^0$, B_1, B_2, and C, and we see that the condition $\partial B_0/\partial C|_{\phi=0} = 0$, see (38), is not satisfied. This means that although the likelihood function is local asymptotic quadratic we do not get asymptotic χ^2 because the limit is not LAMN. Hence we are not guaranteed asymptotic χ^2 inference by the results in Theorem 5.

It turns out, see Table 1, that simulations show that in fact inference is probably asymptotic χ^2. This means that there is an interesting problem to be solved in the future and that the separation condition (38) is probably not necessary.

If we manage to determine τ or $\tau_\perp$ completely, that is, accept the hypothesis that $B_2 = 0$ and $C = 0$, then the situation is different, since the test for no multi-cointegration becomes the test that $\psi'\tau_\perp^0 = 0$ which is equivalent to $B_0 = 0$, which gives asymptotic χ^2 inference. Similarly if only $C = 0$, which is a bit difficult to interpret, we get asymptotic χ^2 inference.

5.5 Comparison with results of Boswijk

The results obtained can be compared to the results of Boswijk (2000). Instead of normalizing as done here on the true value of the parameters, matrices c and d are introduced for normalization of β and η respectively, see (27), so that $c'\beta = I_r$ and $d'\eta = I_s$. We use the notation $\alpha_c, \beta_c, \eta_d$ to distinguish them from the parameters α, β, η when β and τ normalized on $\bar{\beta}^0$ and $\bar{\tau}^0$ respectively.

Boswijk introduced the parameters

$$A_1 = -\beta_c' c_1,\ A_2 = -\beta_c' c_2,\ A_3 = -\bar{\alpha}_c'\Gamma\beta_{c2},\ A_4 = -\eta_d' d_\perp$$

and developed the asymptotic theory for this parametrization. A tedious calculation, see Johansen (2004), gives the connection between these parameters and the parameters used in the present paper. The local linear relations are given by

$$\begin{aligned} B_0 &\sim \left(\beta_2^{0\prime}\beta_2^0\right)^{-1}[(A_3' - A_3^{0\prime}) - (A_4' - A_4^{0\prime})(A_1^{0\prime}c' + c_1')\psi^0], \\ B_1 &\sim -A_1' + A_1^{0\prime}, \\ B_2 &\sim -(\beta_2^{0\prime}\beta_2^0)^{-1}(A_4^{0\prime}(A_1' - A_1^{0\prime}) + (A_2' - A_2^{0\prime})), \\ C &\sim -(\beta_2^{0\prime}\beta_2^0)^{-1}(A_4' - A_4^{0\prime})\rho_\perp^{0\prime}\rho_\perp^0. \end{aligned}$$

In Boswijk (2000, Remarks 5.3, 5.4, 5.5) it is pointed out that a number of different tests give asymptotic χ^2 inference. We shall show here that the same results are covered by the results of Theorem 3.

Thus for instance, under the hypothesis $A_4 = A_4^0$, we have

$$\begin{aligned} B_0 &\sim \left(\beta_2^{0\prime}\beta_2^0\right)^{-1}(A_3' - A_3^{0\prime}), \\ B_1 &\sim -A_1' + A_1^{0\prime}, \\ B_2 &= -(\beta_2^{0\prime}\beta_2^0)^{-1}(A_4^{0\prime}(A_1' - A_1^{0\prime}) + (A_2' - A_2^{0\prime}))(\bar{\beta}^{0\prime}\beta_c)^{-1}, \\ C &\sim 0. \end{aligned}$$

With A_1, A_2, and A_3 varying unrestricted the parameters B_0, B_1, and B_2 are unrestricted and the hypothesis corresponds to $C = 0$. All conditions of Theorem 5 are satisfied, so that inference is asymptotic χ^2. The hypothesis $A_1 = A_1^0$ corresponds to $B_1 = 0$, and (B_0, B_2, C) unrestricted and by Theorem 5 we get asymptotic χ^2 inference. Finally the multicointegration parameter is $\delta = \psi'\bar{\tau}_\perp = A_3$, and the test for $A_3 = A_3^0$ results in the parameters

$$\begin{aligned} B_0 &\sim -\left(\beta_2^{0\prime}\beta_2^0\right)^{-1}(A_4' - A_4^{0\prime})(A_1^{0\prime}c' + c_1')\psi^0, \\ B_1 &\sim -A_1' + A_1^{0\prime}, \\ B_2 &= -(\beta_2^{0\prime}\beta_2^0)^{-1}(A_4^{0\prime}(A_1' - A_1^{0\prime}) + (A_2' - A_2^{0\prime}))(\bar{\beta}^{0\prime}\beta_c)^{-1}, \\ C &\sim -(\beta_2^{0\prime}\beta_2^0)^{-1}(A_4' - A_4^{0\prime})\rho_\perp^{0\prime}\rho_\perp^0, \end{aligned}$$

which shows that $B_0 \sim C(\rho_\perp^{0\prime}\rho_\perp^0)^{-1}(A_1^{0\prime}c' + c_1')\psi^0$, so that the result of Theorem 5 does not give asymptotic χ^2 inference. However, as pointed out in Boswijk (2000, Remark 5.5), one can sometimes test this hypothesis in two stages: First test that $A_4 = A_4^0$ using an asymptotic χ^2 test, and then, if that is accepted, continue with testing $A_3 = A_3^0$, which again is asymptotic χ^2, see section 5.4.

5.6 A simple example with a simulation

In order to illustrate the findings we consider a three variable system

$$\begin{aligned} \Delta^2 X_{1t} &= -0.5[X_{1t-1} + \rho_2 X_{2t-1} + (\tau_1 + \rho\tau_2)X_{3t-1} \\ &\quad +2\Delta X_{1t-1} + \Delta X_{2t-1} + (\psi_3 + 2\tau_1)\Delta X_{3t-1}] + \varepsilon_{1t}, \\ \Delta^2 X_{2t} &= -\Delta X_{2t-1} - \tau_2\Delta X_{3t-1} + \varepsilon_{2t}, \\ \Delta^2 X_{3t} &= \varepsilon_{3t}. \end{aligned}$$

The example is so constructed that the cointegration parameters are

$$\tau = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \tau_1 & \tau_2 \end{pmatrix}, \rho = \begin{pmatrix} 1 \\ \rho_2 \end{pmatrix}, \beta = \begin{pmatrix} 1 \\ \rho_2 \\ \tau_1 + \rho_2\tau_2 \end{pmatrix}, \psi = \begin{pmatrix} 2 \\ 1 \\ \psi_3 + 2\tau_1 \end{pmatrix},$$

and the true value is taken to be $\tau_1 = \tau_2 = \rho_2 = \psi_3 = 0$, so that τ and β are normalized on $\bar{\tau}^0$ and $\bar{\beta}^0$. We have fixed the parameters $\alpha = (-0.5, 0, 0)', \Omega_3 = I_3$, and $\psi'\bar{\tau}^0 = (2, 1)$, as these parameters do not enter the cointegrating relations. The new parameters (31) become

$$B_0 = \psi_3 + 2\tau_1,\ B_1 = \rho_2,\ B_2 = \tau_1 + \rho_2\tau_2,\ C = -\rho_2\tau_1 + \tau_2,$$

which are variation free. This shows that the model can be tested by an asymptotic χ^2 test since the conditions of Theorem 5 are satisfied.

We now consider four hypotheses and for each check the conditions of Theorem 5. At the end of this section we have conducted a small simulation experiment to illustrate what may happen when we cannot prove asymptotic χ^2 inference.

5.6.1 The hypothesis $\tau_2 = 0$

The hypothesis $\tau_2 = 0$ is equivalent to $C = -B_1B_2$. We find $\partial C/\partial B_1 = -B_2$, which is zero at the true value so that inference is asymptotic $\chi^2(1)$ by Theorem 5. The hypothesis is a special case of test on a coefficient in τ, see section 5.3.3, but it is also a special case of testing that a given vector is contained in $\mathrm{sp}(\tau)$, see section 5.3.2.

5.6.2 The hypothesis $\tau_1 = 0$

The hypothesis $\tau_1 = 0$ is equivalent to $B_2 = B_1C$. In this case we have $\partial B_2/\partial B_1 = C$ and $\partial B_2/\partial C = B_1$, which are both zero at the true value. We also find, however, that $\partial^2 B_2/\partial C\partial B_1 = 1$, so that condition (37) is not satisfied. The simulation in Table 1 indicates that indeed the asymptotic distribution is not $\chi^2(1)$ as the mean, variance and 95% quantile are all too small. Thus we get a situation of LAQ but not LAMN.

The hypothesis is a special case of test on a coefficient in τ, see section 5.3.3, but it is also a special case of testing that a given vector b is contained in $\mathrm{sp}(\tau)$, see section 5.3.2, and the example has been constructed so that $b = \beta^0$, so that the condition of Lemma 12 is not satisfied.

5.6.3 The test that $\psi_3 = 0$,

The test that $\psi_3 = 0$, is equivalent to $B_0 = 2(B_2 - CB_1)/(1 + B_1^2)$. In this case we have to check the derivative $\partial B_0/\partial C = -2B_1/(1 + B_1^2)$ which is zero at the true value. Thus we get asymptotic χ^2 inference.

5.6.4 The test that $\delta = 0$,

Finally we consider the test that $\delta = \psi'\tau_\perp = \psi_3 - \tau_2 = 0$. The test is equivalent to $B_0 = (2(B_2 - CB_1) + B_1B_2 + C)/(1 + B_1^2)$. The conditions of Theorem 5 are not satisfied as we find $\partial B_0/\partial C = (-2B_1 + 1)/(1 + B_1^2)$ which is non-zero at the true value. Note, however, that the simulations in Table 1 indicates that we nevertheless get an asymptotic distribution that is very close to that of a $\chi^2(1)$. Thus the conditions given in Theorem 5 are probably not necessary. The possibility of asymptotic χ^2 inference, based on simulations was also pointed out by Paruolo, see Paruolo (1995), in connection with the derivation of the distribution.

Table 1 here

6 Conclusion

Likelihood based inference in the $I(2)$ cointegrated model is now feasible since the programs have been developed. It is therefore of importance to investigate when we can use asymptotic χ^2 inference. A reformulation of the result of Boswijk (2000) is given with the purpose of applying it to a number of situations of practical importance. A number of hypotheses is then investigated involving the cointegrating parameters and it is found that in many cases inference on the cointegrating parameters is asymptotic χ^2. Some hypotheses on the $C(2,1)$ parameters τ do not give asymptotic χ^2 inference, but a procedure is suggested to solve the problem. The conditions do not cover the situation of the multicointegrating parameter even if simulation experiments indicate that inference could be asymptotic χ^2.

7 Appendix

In this appendix we discuss the function $\theta = \theta(\phi)$ from section 2. We formulate results for a vector function of a vector variable and we can then apply the results to $\theta^v = \theta^v(\phi^v)$. This saves some notation. Thus we consider three vector valued functions $\theta = (\theta_0, \theta_1, \theta_2)$ of dimension k_0, k_1, k_2 respectively, which depend on three vector valued parameters $\phi_i = (\phi_0, \phi_1, \phi_2)$ of dimension q_0, q_1, q_2 respectively with $q_i \leq k_i$. In the $I(2)$ regression model we show that with $n_i = \max(\frac{1}{2}, i)$, the estimators $T^{n_i}(\hat{\theta}_i - \theta_i^0)$ converge weakly and hence are tight. We want to find conditions under which the same holds for $T^{n_i}(\hat{\phi}_i - \phi_i^0)$.

We therefore prove two results. First we show under some conditions on $\theta(\phi)$ that if the sequence $T^{n_i}(\theta_i(\phi_T) - \theta(\phi^0))$ is bounded then $T^{n_i}(\phi_{Ti} - \phi_i^0)$ is bounded too. This can obviously be use to prove tightness of $T^{n_i}(\hat{\phi}_i - \phi_i^0)$. Next we find conditions under which $T^{n_i} d\theta_i/d\phi|_{\phi=\phi^0+D_T\cdot\eta}(D_T\cdot\psi) \to \frac{\partial\theta_i}{\partial\phi_i}(\psi_i)$ and $T^{n_i}(\theta_i(\phi^0 + D_T\cdot\psi) - \theta(\phi^0)) \to \partial\theta_i/\partial\phi_i(\psi_i)$ uniformly in ψ bounded, so that the function $\theta(\phi)$ becomes locally linear in such a way that θ_i depends only on ϕ_i in the local linear approximation. These results can be applied to find the limit distribution of $T^{n_i}(\hat{\phi}_i - \phi_i^0)$. We let $\theta_{i|i}$ denote the $k_i \times q_i$ matrix $\partial\theta_i/\partial\phi_i$ for $\phi = \phi^0$, and define $\bar{\theta}_{i|i} = \theta_{i|i}(\theta'_{i|i}\theta_{i|i})'$.

We shall use in the proofs that condition (20) of Assumption 2 can be expressed as

$$\frac{\partial^m \theta_i}{\partial \phi_{i_1} \dots \partial \phi_{i_m}} = 0, \text{ for } n_{i_1} + \dots + n_{i_m} < n_i, \; m = 1, 2, 3, \tag{46}$$

whereas condition (21) of Assumption 2 is

$$\frac{\partial^m \theta_i}{\partial \phi_{i_1} \dots \partial \phi_{i_m}} = 0, \text{ for } n_{i_1} + \dots + n_{i_m} = n_i, \; m = 2, 3, 4. \tag{47}$$

Lemma 14 *Under Assumption 2(a,b,c,d), it holds that if ϕ_T is a sequence of parameters for which*

$$\left(T^{1/2}(\theta_0(\phi_T) - \theta_0^0), T(\theta_1(\phi_T) - \theta_1^0), T^2(\theta_2(\phi_T) - \theta_2^0)\right) \tag{48}$$

is bounded, then so is

$$\left(T^{1/2}(\phi_{T0} - \phi_0^0), T(\phi_{T1} - \phi_1^0), T^2(\phi_{T2} - \phi_2^0)\right). \tag{49}$$

Proof. Note that (48) implies that $\theta(\phi_T) \to \theta(\phi^0)$, so that Assumption 2b implies that $\phi_T \to \phi^0$. We assume without loss of generality that ϕ^0 and $\theta^0 = \theta(\phi^0)$ are zero.

$T^{1/2}\phi_{T0}$ is bounded

We expand as follows

$$T^{1/2}\theta_0(\phi_T) = \frac{\partial \theta_0}{\partial \phi_0}|_*(T^{1/2}\phi_{T0}) + \frac{\partial \theta_0}{\partial \phi_1}|_*(T^{1/2}\phi_{T1}) + \frac{\partial \theta_0}{\partial \phi_2}|_*(T^{1/2}\phi_{T2}). \tag{50}$$

Note that $\theta_{0|0} = \partial\theta_0/\partial\phi_0$ has full rank q_0 and that $\bar{\theta}'_{0|0}\partial\theta_0/\partial\phi_0|_* \to I_{q_0}$, so that if we multiply (50) by $\bar{\theta}'_{0|0}$ we can solve for $T^{1/2}\phi_{T2}$. Thus if we can prove that $T^{1/2}\phi_{T1}$ and $T^{1/2}\phi_{T2}$ are both $o(T^{1/2}\phi_{T0}) + o(1)$ we have proved the first result because $T^{1/2}\theta_0(\phi_T)$ is bounded. We therefore expand $T^{1/2}\theta_2(\phi_T)$ and multiply by $T^{1/2}\bar{\theta}'_{2|2}$, and find using $\bar{\theta}'_{2|2}\frac{\partial\theta_2}{\partial\phi_2}|_* \to I_{q_2}$ that

$$T^{1/2}\bar{\theta}'_{2|2}\theta_2(\phi_T) = \bar{\theta}'_{2|2}\frac{\partial \theta_2}{\partial \phi_0}|_*(T^{1/2}\phi_{T0}) + \bar{\theta}'_{2|2}\frac{\partial \theta_2}{\partial \phi_1}|_*(T^{1/2}\phi_{T1}) + (I_{q_2} + o(1))(T^{1/2}\phi_{T2}).$$

The left hand side is $o(1)$, see (48), and the first two terms on the right hand side are $o(T^{1/2}\phi_{T0})$ and $o(T^{1/2}\phi_{T1})$ respectively, because $\partial\theta_2/\partial\phi_0$ and $\partial\theta_2/\partial\phi_1$ are zero for $\phi = 0$, see (20). We solve for $T^{1/2}\phi_{T2}$ and find

$$T^{1/2}\phi_{T2} = o(1) + o(T^{1/2}\phi_{T0}) + o(T^{1/2}\phi_{T1}). \tag{51}$$

Now expand $T^{1/2}\theta_1(\phi_T)$ multiplied by $\bar{\theta}'_{1|1}$, and we find as before

$$T^{1/2}\bar{\theta}'_{1|1}\theta_1(\phi_T) = \bar{\theta}'_{1|1}\frac{\partial \theta_1}{\partial \phi_0}|_*(T^{1/2}\phi_{T0}) + (I_{q_1} + o(1))(T^{1/2}\phi_{T1}) + \bar{\theta}'_{1|1}\frac{\partial \theta_1}{\partial \phi_2}|_*(T^{1/2}\phi_{T2}) \tag{52}$$

The left hand side is $o(1)$, and $\partial\theta_1/\partial\phi_0|_*(T^{1/2}\phi_{T0})$ is $o(T^{1/2}\phi_{T0})$ because $\partial\theta_1/\partial\phi_0$ is zero. Finally, by (51), we have that $\partial\theta_1/\partial\phi_2|_*(T^{1/2}\phi_{T2})$ is $o(1)+o(T^{1/2}\phi_{T0})+o(T^{1/2}\phi_{T1})$, so solving (52) for $T^{1/2}\phi_{T1}$, we find

$$T^{1/2}\phi_{T1} = o(1) + o(T^{1/2}\phi_{T0}). \tag{53}$$

This implies that $T^{1/2}\phi_{T2} = o(1)+o(T^{1/2}\phi_{T0})$, see (51). Hence from (50) we get that $T^{1/2}\phi_{T0}$ is bounded and finally that

$$T^{1/2}\phi_{T1} = o(1), \quad T^{1/2}\phi_{T2} = o(1). \tag{54}$$

$T\phi_{T1}$ is bounded

Before expanding $T\theta(\phi_{T1})$ we want to prove that $T\phi_{T2}$ is $o(1)$. In order to do this we expand $T\theta_2(\phi_T)$ by applying Taylor's formula (15) with $l=2$ and find because $T\theta_2(0)=0$, that the first term with $m=1$ is

$$T\frac{d\theta_2}{d\phi}(\phi_T) = T\frac{\partial\theta_2}{\partial\phi_0}(\phi_{T0}) + T\frac{\partial\theta_2}{\partial\phi_1}(\phi_{T1}) + T\frac{\partial\theta_2}{\partial\phi_2}(\phi_{T2}) = \frac{\partial\theta_2}{\partial\phi_2}(T\phi_{T2}), \tag{55}$$

because $\partial\theta_2/\partial\phi_0$ and $\partial\theta_2/\partial\phi_1$ are zero. The left hand side is $o(1)$, see (48), and hence solving (55) for $T\phi_{T2}$, by multiplying by $\bar{\theta}'_{2|2}$, shows that $T\phi_{T2}$ is $o(1)$, provided we only get $o(1)$ contributions from the remaining terms.

We find for $m=2$

$$T\frac{1}{2}\frac{d^2\theta_2}{d\phi^2}(\phi_T^{(2)}) = \frac{1}{2}\sum_{i,j=0}^{2}\frac{\partial^2\theta_2}{\partial\phi_i\partial\phi_j}(T^{1/2}\phi_{Ti}, T^{1/2}\phi_{Tj}).$$

If $(i,j)=(0,0)$ the term $\partial^2\theta_2/\partial\phi_0^2$ is zero, see (20). If $n_i+n_j>1$, we have that either i or j is non-zero, which means that the term is $o(1)$, see (54).

Finally the remainder term is

$$R_T^{(2)} = \frac{1}{2}\sum_{i,j=0}^{2}\left(\frac{\partial^2\theta_2}{\partial\phi_i\partial\phi_j}|_* - \frac{\partial^2\theta_2}{\partial\phi_i\partial\phi_j}\right)(T^{1/2}\phi_{Ti}, T^{1/2}\phi_{Tj}).$$

The same type of argument shows that $R_T^{(2)}$ is $o(1)$, because $\partial^2\theta_2/\partial\phi_0^2|_\phi$ is continuous in ϕ.

Thus (55) implies that $T\phi_{T2}$ is $o(1)$. Expanding $T\theta_1(\phi_T)$ using Taylor's formula (15) with $l=2$ and find because $T\theta_1(0)=0$, that the first term with $m=1$ is

$$T\frac{d\theta_1}{d\phi}(\phi_T) = T\frac{\partial\theta_1}{\partial\phi_0}(\phi_{T0}) + T\frac{\partial\theta_1}{\partial\phi_1}(\phi_{T1}) + T\frac{\partial\theta_1}{\partial\phi_2}(\phi_{T2}) = \frac{\partial\theta_1}{\partial\phi_1}(T\phi_{T1}) + o(1). \tag{56}$$

Multiplying by $\bar{\theta}'_{1|1}$ we see that $T\phi_{T1}$ is bounded provided the other terms of the Taylor's expansion are either $o(T\phi_{T1})$ or bounded.

The next term is

$$T\frac{1}{2}\frac{d^2\theta_1}{d\phi^2}(\phi_T^{(2)}) = \frac{1}{2}\sum_{i,j=0}^{2}\frac{\partial^2\theta_1}{\partial\phi_i\partial\phi_j}(T^{1/2}\phi_{Ti}, T^{1/2}\phi_{Tj}).$$

The term with $(i,j) = (0,0)$ is clearly bounded, because $T^{1/2}\phi_{T0}$ is bounded, and if either i or j is non-zero, then the term is $o(1)$, see (54). Finally the remainder term is

$$R_T^{(2)} = \frac{1}{2}\sum_{i,j=0}^{2}(\frac{\partial^2\theta_1}{\partial\phi_i\partial\phi_j}|_* - \frac{\partial^2\theta_1}{\partial\phi_i\partial\phi_j})(T^{1/2}\phi_{Ti}, T^{1/2}\phi_{Tj}).$$

The same type of argument shows that $R_T^{(2)}$ is $o(1)$, because $\partial^2\theta_2/\partial\phi_0^2|_\phi$ is continuous in ϕ. This completes the proof that $T\phi_{T1}$ is bounded.

$T^2\phi_{T2}$ is bounded

Finally we apply the results found so far to investigate $T^2\theta_2(\phi_T)$. We apply Taylor's formula (15) to $T\theta_2(\phi_T)$ with $l = 3$, and find the first term

$$T^2\frac{d\theta_2}{d\phi}(\phi_T) = T^{3/2}\frac{\partial\theta_2}{\partial\phi_0}(T^{1/2}\phi_{T0}) + T\frac{\partial\theta_2}{\partial\phi_1}(T\phi_{T1}) + \frac{\partial\theta_2}{\partial\phi_2}(T^2\phi_{T2}) = \frac{\partial\theta_2}{\partial\phi_2}(T^2\phi_{T2}), \quad (57)$$

which we can invert, see Assumption 2b, by multiplying by $\bar{\theta}_{2|2}'$ to prove that $T^2\phi_{T2}$ is bounded, provided the remaining terms in the expansion are all bounded or $o(T^2\phi_{T2})$.

The term for $m = 2$ is

$$T^2\frac{1}{2}\frac{d^2\theta_2}{d\phi^2}(\phi_T^{(2)}) = \frac{1}{2}\sum_{i,j=0}^{2}T^2\frac{\partial^2\theta_2}{\partial\phi_i\partial\phi_j}(\phi_{Ti}, \phi_{Tj}).$$

For $(i,j) = (0,0)$ or $(0,1)$ we have assumed that $\partial^2\theta_2/\partial\phi_i\partial\phi_j = 0$, see (46). For $(i,j) = (1,1)$ the term is bounded because $T\phi_{T1}$ is bounded, and for either i or j $= 2$, the term is $o(T^2\phi_{T2})$.

The term for $m = 3$ is

$$T^2\frac{1}{3!}\frac{d^3\theta_2}{d\phi^3}(\phi_T^{(3)}) = \frac{1}{2}\sum_{i,j,k=0}^{2}T^2\frac{\partial^3\theta_2}{\partial\phi_i\partial\phi_j\partial\phi_k}(\phi_{Ti}, \phi_{Tj}, \phi_{Tk}).$$

If $(i,j,k) = (0,0,0)$ we have assumed that $\partial^3\theta_2/\partial\phi_0^3 = 0$, see (46). If $(i,j,k) = (0,0,1)$, we find the term $\partial^3\theta_2/\partial\phi_0^2\partial\phi_1(T^{1/2}\phi_{T0}, T^{1/2}\phi_{T0}, T\phi_{T1})$ which is bounded, because $T^{1/2}\phi_{T0}$ and $T\phi_{T1}$ are bounded. The remaining terms are either $o(1)$ or $o(T^2\phi_{T2})$.

The remainder term is

$$R_T^{(3)} = \frac{1}{3!}\sum_{i,j,k=0}^{2}T^2(\frac{\partial^3\theta_2}{\partial\phi_i\partial\phi_j\partial\phi_k}|_* - \frac{\partial^3\theta_2}{\partial\phi_i\partial\phi_j\partial\phi_k})(\phi_{Ti}, \phi_{Tj}, \phi_{Tk}),$$

and the same arguments shows that this is $o(1)$ or $o(T^2\phi_{T2})$. Returning to (57) we have seen that $T^2\phi_{T2}$ is also bounded, which completes the proof of the Lemma 14. ∎

Lemma 15 *Under Assumption 2a and with* $D_T = (T^{-1/2}, T^{-1}, T^{-2})$ *it holds that (20) and (21) are necessary and sufficient for the convergence*

$$T^{n_i}(\frac{d\theta_i}{d\phi}|_{\phi=\phi^0+D_T\cdot\psi}(D_T\cdot\eta)) \to \frac{\partial\theta_i}{\partial\phi_i}(\eta_i), \tag{58}$$

uniformly for ψ *bounded. In this case*

$$T^{n_i}(\theta_i(\phi^0 + D_T\cdot\psi) - \theta_i(\phi^0)) \to \frac{\partial\theta_i}{\partial\phi_i}(\psi_i), \tag{59}$$

uniformly for ψ *bounded.*

Remark. It follows from the proof that only condition (20) is needed to find a limit and condition (21) is needed to get a limit that does not depend on ψ. If this is the case the limit only involves the derivatives $\partial\theta_i/\partial\phi_i$. Thus the first statement shows that the conditions given in (20) and (21) give a reformulation of the condition of Boswijk, see Theorem 16.

PROOF. *$i = 0$:*

We let $\phi_T = \phi^0 + D_T\cdot\psi$ and assume $\phi^0 = 0$. To prove (58) for $i = 0$, we find

$$T^{1/2}\frac{d\theta_0}{d\phi}|_{\phi_T}(D_T\cdot\eta) = \frac{\partial\theta_0}{\partial\phi_0}|_{\phi_T}(\eta_0) + T^{-1/2}\frac{\partial\theta_0}{\partial\phi_1}|_{\phi_T}(\eta_1) + T^{-3/2}\frac{\partial\theta_0}{\partial\phi_2}|_{\phi_T}(\eta_2),$$

which converges, uniformly in ψ bounded, to $\partial\theta_0/\partial\phi_0(\eta_0)$ because $\partial\theta_0/\partial\phi_0|_\phi(\eta_0)$ is continuous in ϕ.

$i = 1$:

To prove (58) for $i = 1$ we consider

$$T\frac{d\theta_1}{d\phi}|_{\phi_T}(D_T\cdot\eta) = T^{1/2}\frac{\partial\theta_1}{\partial\phi_0}|_{\phi_T}(\eta_0) + \frac{\partial\theta_1}{\partial\phi_1}|_{\phi_T}(\eta_1) + T^{-1}\frac{\partial\theta_1}{\partial\phi_2}|_{\phi_T}(\eta_2).$$

The last term tends to zero and the second to $\partial\theta_1/\partial\phi_1(\eta_1)$. For the first term to converge we need $\partial\theta_1/\partial\phi_0 = 0$. We then expand and find

$$T^{1/2}\frac{\partial\theta_1}{\partial\phi_0}|_{\phi_T}(\eta_0) = \frac{\partial^2\theta_1}{\partial\phi_0^2}|_*(\eta_0,\psi_0) + T^{-1/2}\frac{\partial^2\theta_1}{\partial\phi_0\partial\phi_1}|_*(\eta_0,\psi_1) + T^{-3/2}\frac{\partial^2\theta_1}{\partial\phi_0\partial\phi_2}|_*(\eta_0,\psi_2).$$

This tends, uniformly in ψ bounded, to $\partial^2\theta_1/\partial\phi_0^2(\eta_0,\psi_0)$ which is independent of ψ if and only if $\partial^2\theta_1/\partial\phi_0^2 = 0$, see (21) in Assumption 2.

$i = 2$:

Finally we take $i = 2$ in (58) and find, using Taylor's Formula (15) for $T^2 d\theta_2/d\phi|_{\phi_T}(D_T\cdot\eta)$ with $l = 2$, and find the term with $m = 0$

$$T^2\frac{d\theta_2}{d\phi}(D_T\cdot\eta) = T^{3/2}\frac{\partial\theta_2}{\partial\phi_0}(\eta_0) + T\frac{\partial\theta_2}{\partial\phi_1}(\eta_1) + \frac{\partial\theta_2}{\partial\phi_2}(\eta_2).$$

We have convergence if and only if $\partial\theta_2/\partial\phi_0$ and $\partial\theta_2/\partial\phi_0$ are zero and the limit is the main term $\partial\theta_2/\partial\phi_2(\eta_2)$.

For $m = 1$ we get

$$T^2 \frac{d^2\theta_2}{d\phi^2}(D_T \cdot \eta, D_T \cdot \psi) = \sum_{j,k=0}^{2} T^{2-n_j-n_k} \frac{\partial^2 \theta_2}{\partial \phi_j \partial \phi_k}(\eta_j, \psi_k).$$

For $n_j + n_k < 2$, $T^{2-n_j-n_k} \to \infty$, and the limit exists if and only if $\partial^2\theta_2/\partial\phi_0^2$ and $\partial^2\theta_2/\partial\phi_0\partial\phi_1$ are zero, see (20). For $n_j + n_k = 2$, the limit is $\partial^2\theta_2/\partial\phi_1^2(\eta_1, \psi_1)$ which is independent of ψ if and only if $\partial^2\theta_2/\partial\phi_1^2$ is zero. Finally for $n_j + n_k > 2$, $T^{2-n_j-n_k} \to 0$, and the terms tends to zero.

For $m = 2$ we get

$$T^2 \frac{1}{2} \frac{d^3\theta_2}{d\phi^3}(D_T \cdot \eta, (D_T \cdot \psi)^{(2)}) = \frac{1}{2} \sum_{j,k,l=0}^{2} T^{2-n_j-n_k-n_l} \frac{\partial^3 \theta_2}{\partial \phi_j \partial \phi_k \partial \phi_l}(\eta_j, \psi_k, \psi_l).$$

For $n_j + n_k + n_l < 2$, $T^{2-n_j-n_k-n_l} \to \infty$, and the limit exists if and only if $\partial^3\theta_2/\partial\phi_0^3$ is zero. For $n_j + n_k + n_l = 2$, we have the limit $\partial^3\theta_2/\partial\phi_0^2\partial\phi_1(\eta_0, \psi_0, \psi_1)$ or $\partial^3\theta_2/\partial\phi_0^2\partial\phi_1(\eta_1, \psi_0, \psi_0)$ which are independent of ψ if and only if $\partial^3\theta_2/\partial\phi_0^2\partial\phi_1$ is zero, see (21). Finally for $n_j + n_k + n_l > 2$, $T^{2-n_j-n_k-n_l} \to 0$.

For $m = 3$ we get

$$T^2 \frac{1}{2} \frac{d^4\theta_2}{d\phi^4}(D_T \cdot \eta, (D_T \cdot \psi)^{(3)}) = \frac{1}{3!} \sum_{j,k,l,h=0}^{2} T^{2-n_j-n_k-n_l-n_h} \frac{\partial^4 \theta_2}{\partial \phi_j \partial \phi_k \partial \phi_l \partial \phi_h}(\eta_j, \psi_k, \psi_l, \psi_h).$$

For $n_j + n_k + n_l + \psi_h = 2$, the limit is $\partial^4\theta_2/\partial\phi_0^4(\eta_0, \psi_0^{(3)})$ which is independent of ψ if and only if $\partial^4\theta_2/\partial\phi_0^4$ is zero, see (21). Finally for $n_j + n_k + n_l + n_h > 2$, $T^{2-n_j-n_k-n_l-n_h} \to 0$ and the terms tend to zero. The remainder term is

$$R_T^{(2)} = \frac{1}{3!} \sum_{j,k,l,h=0}^{2} T^{2-n_j-n_k-n_l} \left(\frac{\partial^4 \theta_2}{\partial \phi_j \partial \phi_k \partial \phi_l \partial \phi_h}|_* - \frac{\partial^4 \theta_2}{\partial \phi_j \partial \phi_k \partial \phi_l \partial \phi_h} \right)(\eta_j, \psi_k, \psi_l, \psi_h).$$

For $(i, j, k, h) = (0, 0, 0, 0)$ we find $(\frac{\partial^4\theta_2}{\partial\phi_0^4}|_* - \frac{\partial^4\theta_2}{\partial\phi_0^4})(\eta_0, \psi_0^{(3)}) \to 0$, and the remaining terms tend to zero because of the factor $T^{2-n_j-n_k-n_l}$.

To prove the result (59) we apply the meanvalue theorem

$$T^{n_i}\theta_i(D_T \cdot \psi) = T^{n_i} \frac{d\theta_i}{d\phi}|_*(D_T \cdot \psi)$$

and the result follows from (58). ■

Lemma 16 *If $\theta(\phi)$ is continuous, it holds that for all $\varepsilon > 0$, there exists a $K > 0$ so that for all T we have that $\hat{\phi}$ exists with probability greater than $1 - \varepsilon$ and satisfies*

$$P_0\{(M_T^{-1}(\theta(\hat{\phi}) - \theta(\phi^0)))' M_{zz} M_T^{-1}(\theta(\hat{\phi}) - \theta(\phi^0)) \leq K\} \geq 1 - \varepsilon.$$

Thus with probability approaching one, the maximum likelihood estimator of ϕ exists and $M_T^{-1}(\theta(\hat{\phi}) - \theta(\phi^0))$ is tight, so that $\hat{\phi}$ is consistent for ϕ^0.

PROOF. Because the unrestricted estimator $\hat{\theta}$ satisfies that $(M_T^{-1}(\hat{\theta}-\theta^0), M_{zz})$ is weakly convergent it holds that $||\hat{\theta}-\theta^0||_T^2 = (M_T^{-1}(\hat{\theta}-\theta^0))' M_{zz} M_T^{-1}(\hat{\theta}-\theta^0)$ is tight and hence for any $\varepsilon > 0$, there exists a $k > 0$ so that for all T

$$1-\varepsilon \leq P_0\{||\hat{\theta}-\theta^0||_T^2 \leq k\}. \tag{60}$$

Now introduce the notation for the open ellipsoid with center θ^0 and radius $k+1$:

$$S(\theta^0, k+1) = \{\theta : ||\theta-\theta^0||_T^2 < k+1\}.$$

Obviously on the set defined in (60), we have

$$\theta^0 \in S(\hat{\theta}, k+1).$$

Because $\theta(\phi^0) = \theta^0$, $\theta(\phi)$ is continuous, and $S(\hat{\theta}, k+1)$ is open, it follows that $\theta(\phi) \in S(\hat{\theta}, k+1)$ for ϕ close to ϕ^0. The open ellipsoid $S(\hat{\theta}, k+1)$ is defined by a level curve of the likelihood function and it follows that all values of θ outside $S(\hat{\theta}, k+1)$ will give values of the likelihood function less than or equal to the values inside $S(\hat{\theta}, k+1)$. This proves that a maximum likelihood estimator exists and satisfies $\hat{\phi} \in S(\hat{\theta}, k+1)$, that is,

$$||\hat{\theta}-\theta(\hat{\phi})||_T^2 < k+1$$

so that

$$||\theta(\hat{\phi})-\theta^0||_T^2 \leq ||\theta(\hat{\phi})-\hat{\theta}||_T^2 + ||\hat{\theta}-\theta^0||_T^2 < 2k+1 = K.$$

Thus we have proved that for all T

$$1-\varepsilon \leq P_0\{||\theta(\hat{\phi})-\theta^0||_T^2 \leq K\},$$

which shows that $||\theta(\hat{\phi})-\theta^0||_T^2$, and hence $M_T^{-1}(\theta(\hat{\phi})-\theta(\phi^0))$ is tight. This immediately implies that $\theta(\hat{\phi}) \xrightarrow{P} \theta(\phi^0)$, and hence by Assumption 2b, that $\hat{\phi}$ is consistent for ϕ^0. ■

8 Acknowledgments

The author wishes to thank the Danish Social Sciences Research Council for support and to thank Peter Boswijk for very helpful discussions. I am grateful to two anonymous referees, whose informed and constructive comments made a uniform improvement to the paper. Finally I would like to thank the members of the ESF-EMM network for useful discussions.

References

[1] Boswijk, H. P., 2000, Mixed Normality and Ancillarity in $I(2)$ Systems, Econometric Theory 16, 878-904.

[2] Jeganathan, P., 1995, Some aspects of asymptotic theory with applications to time series models, Econometric Theory 11, 818-887.

[3] Johansen, S., 1992, A representation of vector autoregressive processes integrated of order 2, Econometric Theory 8, 188-202.

[4] Johansen, S., 1996, Likelihood based inference in cointegrated vector autoregressive models (Oxford University Press, Oxford).

[5] Johansen S., 2004, The equivalence of two parametrizations for the $I(2)$ model and an example of the conditions for asymptotic χ^2 inference, www.math.ku.dk/~sjo.

[6] Johansen, S., 1997, Likelihood analysis of the $I(2)$ model, Scandinavian Journal of Statistics 24, 433-462.

[7] Kitamura, Y., 1995, Estimation of cointegrated systems with $I(2)$ processes, Econometric Theory 11, 1-24.

[8] Omtzigt, P., 2002, Essays on cointegration, Ph.D. Thesis, European University Institute, Florence.

[9] Paruolo, P., 1995, Topics in statistical inference for vector autoregressive processes integrated of order 1 and 2, Ph.D. thesis, University of Copenhagen.

[10] Paruolo, P., 2000, Asymptotic efficiency of the two stage estimator in $I(2)$ systems, Econometric Theory 16, 524-550.

[11] Stock, J. H and M. W. Watson, 1993, A simple estimator of cointegrating vectors in higher order integrated systems, Econometrica 61, 783-820.

[12] van der Vaart, A. W., 1998, Asymptotic Statistics (Cambridge University Press, Cambridge).

Hypothesis	$E(-2\log LR)$	$Var(-2\log LR)$	95% *quantile*
$\tau_1 = 0$	0.731	1.119	2.84
$\delta = 0$	0.993	1.926	3.85

Table 1: The simulation has $T = 500$, and $10,000$ simulations. The table shows the estimated mean, variance, and 95% quantile of the log likelihood ratio test statistic for the hypotheses $\tau_1 = 0$ and $\delta = 0$. For the $\chi^2(1)$ we have $E(\chi^2(1)) = 1, Var(\chi^2(1)) = 2$, and $\chi^2(1)_{0.95} = 3.84$.