Econometric Theory, **26**, 2010, 1565–1576.
doi:10.1017/S0266466610000095

NOTES AND PROBLEMS

MIXED NORMAL INFERENCE ON MULTICOINTEGRATION

H. PETER BOSWIJK
University of Amsterdam

Asymptotic likelihood analysis of cointegration in $I(2)$ models (see Johansen, 1997, 2006; Boswijk, 2000; Paruolo, 2000) has shown that inference on most parameters is mixed normal, implying hypothesis test statistics with an asymptotic χ^2 null distribution. The asymptotic distribution of the multicointegration parameter estimator so far has been characterized by a Brownian motion functional, which has been conjectured to have a mixed normal distribution, based on simulations. The present note proves this conjecture.

1. INTRODUCTION

The notion of multicointegration was introduced by Granger (1986) and Granger and Lee (1990). Although originally developed for processes integrated of order 1 ($I(1)$), it has subsequently become clear that the phenomenon occurs naturally in $I(2)$ cointegrated vector autoregressive (VAR) models (see Johansen, 1992; Engsted and Johansen, 1999). With $\{X_t\}_{t\geq 1}$ a p-vector time series process, the $I(2)$ VAR model of order k is expressed as

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

$$\bar{\alpha}_\perp \alpha'_\perp \Gamma \beta_\perp \bar{\beta}'_\perp = \alpha_1 \beta'_1, \tag{2}$$

where $\{\varepsilon_t\}_{t\geq 1}$ is assumed to be an independent and identically distributed (i.i.d.) $N(0, \Omega)$ sequence, and where α and β are $p \times r$ matrices ($0 \leq r < p$), α_1 and β_1 are $p \times s$ matrices ($0 \leq s < p - r$), and Γ, $\{\Psi_j\}_{j=1}^{k-2}$ and Ω are $p \times p$ matrices, with Ω positive definite. (For an $n \times m$ matrix A of rank $m < n$, $A_\perp$ denotes an $n \times (n-m)$ matrix of rank $n-m$ satisfying $A'_\perp A = 0$; and $\bar{A} = A(A'A)^{-1}$, so that $\bar{A}'A = I_m$.) The model can be extended to include deterministic components such as a constant and trend (see Rahbek, Kongsted, and Jørgensen, 1999) without qualitatively affecting the results to follow.

Helpful comments from Søren Johansen, Paolo Paruolo, Anders Rahbek, two anonymous referees, and the editor, Peter Phillips, are gratefully acknowledged. Address correspondence to H. Peter Boswijk, Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, NL-1018 WB Amsterdam, The Netherlands; e-mail: h.p.boswijk@uva.nl.

© Cambridge University Press 2010 0266-4666/10 $15.00

Paruolo and Rahbek (1999) show that the $I(2)$ restriction (2) implies

$$\Gamma = \alpha\delta\beta_2' + \zeta_1\beta' + \zeta_2\beta_1',$$

where $\delta = \bar{\alpha}'\Gamma\bar{\beta}_2$, $\beta_2 = (\beta, \beta_1)_\perp$, $\zeta_1 = \Gamma\bar{\beta}$, and $\zeta_2 = \Gamma\bar{\beta}_1$. Therefore, the model (1) under this restriction becomes

$$\Delta^2 X_t = \alpha(\beta' X_{t-1} + \delta\beta_2'\Delta X_{t-1}) + \zeta_1\beta'\Delta X_{t-1} + \zeta_2\beta_1'\Delta X_{t-1} + \sum_{j=1}^{k-2}\Psi_j\Delta^2 X_{t-j} + \varepsilon_t. \tag{3}$$

In this model, $(\beta, \beta_1)' X_t$ are $I(1)$ linear combinations of the $I(2)$ process X_t, and $\beta' X_t$ further cointegrates with the $I(1)$ process $\beta_2'\Delta X_t$ to the $I(0)$ linear combinations

$$\beta' X_t + \delta\beta_2'\Delta X_t = (\beta + \beta_2\delta'\Delta)' X_t. \tag{4}$$

This phenomenon is known as *multicointegration*, and also as *polynomial cointegration*, because the right-hand-side expression in (4) is a first-order matrix lag polynomial operating on X_t. Various alternative parametrizations of the $I(2)$ model have been proposed in the literature (see Johansen, 1997; Boswijk, 2000; Mosconi and Paruolo, 2010). However, they do not affect inference on the multicointegration parameter δ, which is the subject of this note.

Asymptotic likelihood-based inference on the parameters of (3) was studied by Johansen (1997, 2006), Boswijk (2000), and Paruolo (2000). They showed that under suitable identifying restrictions, the asymptotic distributions of the maximum likelihood estimators $\hat{\beta}$ and $\hat{\beta}_1$ are scale mixtures of normals, where the random scaling matrix is the distributional limit of the inverse observed information matrix. This implies that likelihood ratio test statistics for smooth hypotheses on β and β_1 have an asymptotic χ^2 null distribution, at least under particular conditions on the hypotheses, derived by Boswijk (2000) and Johansen (2006). The asymptotic distribution of the multicointegration parameter estimator $\hat{\delta}$, however, at first sight does not appear to be mixed normal. It can be written as the distribution of the sum of two mixed normal random variables, but there is no common conditioning set such that both are conditionally normally distributed, which complicates deriving a valid inference procedure for δ. Yet, as noted by Paruolo (1995) and Johansen (2006), Monte Carlo simulation of the Brownian motion functionals that characterize the asymptotic distribution of $\hat{\delta}$ strongly suggests that $\hat{\delta}$ is in fact asymptotically mixed normal. The present note provides a proof of this conjecture, implying that likelihood-based inference on multicointegration can be conducted using χ^2 critical values.

The outline of the remainder of this note is as follows: Section 2 summarizes the asymptotic distributions of $\hat{\beta}$, $\hat{\beta}_1$, and $\hat{\delta}$, as obtained by Johansen (1997, 2006) and Paruolo (2000) (and in a mixture of their notation). In Section 3 the

main result is stated and proved. The final section discusses some extensions. The Appendix contains proofs of some auxiliary lemmas.

The following notation is used: The vec operator stacks the columns of a matrix into a column vector, and the (conditional) variance matrix of a random matrix X is always understood as the (conditional) variance matrix of $\mathrm{vec}X$. Integrals such as $\int_0^1 X(u)\,du$ and $\int_0^1 X(u)\,dW(u)$ are often abbreviated as $\int_0^1 X\,du$ and $\int_0^1 X\,dW$, respectively.

2. PRELIMINARY ASYMPTOTIC RESULTS

The starting point of the asymptotic analysis is the multivariate invariance principle: as $n \to \infty$,

$$n^{-1/2}\sum_{t=1}^{\lfloor un\rfloor}\varepsilon_t \xrightarrow{\mathcal{L}} W(u), \qquad u \in [0,1],$$

where W is a p-vector Brownian motion with variance matrix Ω. The i.i.d. normality of $\{\varepsilon_t\}_{t\geq 1}$ is sufficient but not necessary for this result to hold. From W, define

$$W_1 = (\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}W,$$

$$W_2 = \left(\bar{\alpha}_1' - \bar{\alpha}_1'\Omega\alpha_2(\alpha_2'\Omega\alpha_2)^{-1}\alpha_2'\right)W,$$

with $\alpha_2 = (\alpha, \alpha_1)_\perp$. These are two independent vector Brownian motions of dimensions r and s and with variance matrices denoted Ω_1 and Ω_2, respectively. Furthermore, (W_1, W_2) is independent of the $(p-r-s)$-vector Brownian motion $W_3 = \alpha_2' W$.

Johansen (2006) shows that

$$n^{-1/2}\begin{pmatrix}\beta_2'\Delta X_{\lfloor un\rfloor}\\ \beta_1' X_{\lfloor un\rfloor}\end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix}H_0(u)\\ H_1(u)\end{pmatrix} = \begin{pmatrix}A_{03}W_3(u)\\ A_{12}W_2(u) + A_{13}W_3(u)\end{pmatrix}, \qquad u \in [0,1],$$

where A_{03}, A_{12}, and A_{13} are conformable matrices, depending on the parameters, with A_{03} and A_{12} nonsingular. Define $H_2(u) = \int_0^u H_0(v)\,dv$,

$$H_*(u) = \begin{pmatrix}H_0(u)\\ H_1(u)\\ H_2(u)\end{pmatrix}, \qquad u \in [0,1],$$

and

$$H_{**} = \int_0^1 H_*(u)H_*(u)'du, \qquad H_{ij} = \int_0^1 H_i(u)H_j(u)'du, \qquad i,j = 0,1,2. \quad \textbf{(5)}$$

Let $\psi' = (\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}\Gamma$; it can be shown that $\psi'\bar{\beta}_2 = \bar{\alpha}'\Gamma\bar{\beta}_2 = \delta$. Johansen (2006, Thm. 4) and Paruolo (2000, Thms. 4.1 and 4.2) prove the following results for the maximum likelihood estimators[1] $\hat{\beta}$, $\hat{\beta}_1$, and $\hat{\psi}$ based on a sample $\{X_t\}_{t=1}^n$, with starting values $\{X_{1-k}, \ldots, X_0\}$:

$$\begin{pmatrix} n\bar{\beta}_2'(\hat{\psi}-\psi) \\ n\bar{\beta}_1'(\hat{\beta}-\beta) \\ n^2\bar{\beta}_2'(\hat{\beta}-\beta) \\ n\bar{\beta}_2'(\hat{\beta}_1-\beta) \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} B_0^\infty \\ B_1^\infty \\ B_2^\infty \\ C^\infty \end{pmatrix}, \tag{6}$$

where

$$B^\infty = \begin{pmatrix} B_0^\infty \\ B_1^\infty \\ B_2^\infty \end{pmatrix} = H_{**}^{-1}\int_0^1 H_* \, dW_1', \qquad C^\infty = H_{00}^{-1}\int_0^1 H_0 \, dW_2'. \tag{7}$$

Because W_1 is independent of (W_2, W_3) and H_* is defined from (W_2, W_3), it follows that W_1 is independent of H_*. Similarly, W_2 is independent of W_3 and hence H_0. This implies

$$B^\infty | H_* \sim N(0, \Omega_1 \otimes H_{**}^{-1}), \qquad C^\infty | H_0 \sim N(0, \Omega_2 \otimes H_{00}^{-1}). \tag{8}$$

Thus both B^∞ and C^∞ have a conditionally normal and hence mixed normal distribution, but with a different conditioning set. Therefore (B^∞, C^∞) is not *jointly* normal conditional on the same information: The distribution of $B^\infty | H_0$ is not normal, and $C^\infty | H_*$ has a degenerate distribution. This lack of joint mixed normality was analyzed in more detail by Boswijk (2000).

The conditional variances in (8) are estimated consistently by the estimated variance matrix based on the inverse observed information matrix, in the sense that

$$\widehat{\mathrm{var}} \begin{pmatrix} n\bar{\beta}_2'(\hat{\psi}-\psi) \\ n\bar{\beta}_1'(\hat{\beta}-\beta) \\ n^2\bar{\beta}_2'(\hat{\beta}-\beta) \\ n\bar{\beta}_2'(\hat{\beta}_1-\beta) \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} \Omega_1 \otimes H_{**}^{-1} & 0 \\ 0 & \Omega_2 \otimes H_{00}^{-1} \end{pmatrix}, \tag{9}$$

where in general we use the notation $\widehat{\mathrm{var}}(\hat{\theta}) = (-\partial^2\ell(\theta)/\partial\theta\partial\theta')^{-1}_{\theta=\hat{\theta}}$ for the inverse observed information matrix, with $\ell(\theta)$ the (concentrated) log-likelihood. Note that in (9) and in similar results below, $\widehat{\mathrm{var}}(\cdot)$ is not a consistent estimator of the unconditional asymptotic variance matrix of the argument (as would be the case in conventional, stationary asymptotic theory), but of its conditional asymptotic variance. In the remainder of this note, Wald and t-test statistics should always be taken as based on the maximum likelihood estimator and its conditional variance estimated by the inverse observed information.

Letting θ denote the full vector of cointegration parameters, these results imply the following sufficient conditions for likelihood ratio or Wald test statistics for smooth hypotheses $\mathcal{H}_0 : g(\theta) = 0$ to have an asymptotic χ^2 null distribution: First, $g(\hat{\theta})$ needs to be asymptotically linear in B^∞ and C^∞, in the sense that for some suitable sequence of norming matrices D_n,

$$D_n^{-1}\left(g(\hat{\theta}) - g(\theta)\right) \xrightarrow{\mathcal{L}} G \begin{pmatrix} \mathrm{vec}B^\infty \\ \mathrm{vec}C^\infty \end{pmatrix}, \tag{10}$$

with G a matrix of full row rank. Partitioning G conformably with $(\mathrm{vec}(B^\infty)', \mathrm{vec}(C^\infty)')'$, a second condition is that G is block-diagonal, i.e., $G = \mathrm{diag}(G_B, G_C)$ (for some choice of D_n) . As discussed by Boswijk (2000) and Johansen (2006), the block-diagonality condition is sufficient but possibly not necessary for mixed normal inference. Indeed, Theorem 1 in the next section implies another sufficient condition, which does not require $G = \mathrm{diag}(G_B, G_C)$.

The asymptotic distribution of the estimated multicointegration parameter $\hat{\delta} = \hat{\psi}'\hat{\beta}_2$ is obtained from (6), together with $n\beta_1'(\hat{\beta}_2 - \bar{\beta}_2) = -n(\hat{\beta}_1 - \beta_1)'\bar{\beta}_2 + o_p(1)$, which yields (Paruolo, 2000, Thm. 4.2)

$$n(\hat{\delta} - \delta)' \xrightarrow{\mathcal{L}} B_0^\infty - C^\infty A, \tag{11}$$

with $A = \bar{\beta}_1'\psi$. Its estimated conditional variance matrix, still based on the inverse observed information, satisfies

$$\widehat{\mathrm{var}}\left(n(\hat{\delta} - \delta)'\right) \xrightarrow{\mathcal{L}} \Omega_1 \otimes (H_{**}^{-1})_{00} + (A'\Omega_2 A) \otimes H_{00}^{-1} =: V_{B_0} + V_{CA}. \tag{12}$$

This implies that hypotheses on δ do not satisfy (10) unless the restriction $A = \bar{\beta}_1'\psi = 0$ is satisfied; in all other cases, the asymptotic distribution of $\hat{\delta}$ is characterized by the sum of two random variables that are marginally but not necessarily jointly mixed normal. As noted by Paruolo (1995) and Johansen (2006), however, Monte Carlo simulation of (11)–(12) suggests that inference on δ is asymptotically mixed normal even if $A \neq 0$. In the next section, this result will be proved.

3. MAIN RESULT

This section studies asymptotic inference on the multicointegration parameter δ, based on the limit in distribution of the standardized estimator, as implied by (11)–(12):

$$\widehat{\mathrm{var}}(\hat{\delta}')^{-1/2}\mathrm{vec}(\hat{\delta} - \delta)' \xrightarrow{\mathcal{L}} \left(V_{B_0} + V_{CA}\right)^{-1/2}\left(\mathrm{vec}B_0^\infty - \mathrm{vec}(C^\infty A)\right) =: Z. \tag{13}$$

When δ is a scalar parameter, Z may be interpreted as the limit in distribution of a t-statistic of $\hat{\delta}$. More generally, a likelihood ratio or Wald test statistic for a simple hypothesis on δ will converge in distribution, under the null hypothesis, to $Z'Z$.

The main result is Theorem 1, which states that the triplet $(B_0^\infty, B_1^\infty, C^\infty)$ is jointly mixed normal. This directly implies that the distribution of Z is standard normal, conditionally on (V_{B_0}, V_{CA}), and hence also unconditionally, so that standard inference applies to $\hat{\delta}$. A generalization of this result implied by Theorem 1 will be discussed in the next section.

In order to prove Theorem 1, we first need some auxiliary lemmas, proved in the Appendix. We introduce the subscript coupling notations $a = \{0, 1\}$ and $b = \{0, 2\}$ so that $B_a^\infty = (B_0^{\infty\prime}, B_1^{\infty\prime})'$, $H_b = (H_0', H_2')'$, and $H_{bb} = \int_0^1 H_b H_b' \, du$. This leads to

$$V_{B_a} = \operatorname{var}(B_a^\infty | H_*) = \Omega_1 \otimes \begin{pmatrix} (H_{**}^{-1})_{00} & (H_{**}^{-1})_{01} \\ (H_{**}^{-1})_{10} & (H_{**}^{-1})_{11} \end{pmatrix} = \Omega_1 \otimes (H_{**}^{-1})_{aa}. \tag{14}$$

The first lemma provides a convenient expression for the blocks of $(H_{**}^{-1})_{aa}$.

LEMMA 1. *Let H_{**} and $H_{ij}, i, j = 0, 1, 2, a, b$, be as defined in (5), and define*

$$H_{ij|k} = H_{ij} - H_{ik} H_{kk}^{-1} H_{kj}, \qquad i, j, k = 0, 1, 2, a, b.$$

Then

$$(H_{**}^{-1})_{aa} = \begin{pmatrix} H_{00|2}^{-1} + H_{00|2}^{-1} H_{01|2} H_{11|b}^{-1} H_{10|2} H_{00|2}^{-1} & -H_{00|2}^{-1} H_{01|2} H_{11|b}^{-1} \\ -H_{11|b}^{-1} H_{10|2} H_{00|2}^{-1} & H_{11|b}^{-1} \end{pmatrix}. \tag{15}$$

A simpler expression for the first diagonal block $(H_{**}^{-1})_{00}$ is available, but the expression in Lemma 1 is most convenient for our purposes. In particular, using the fact that $H_1 = A_{12} W_2 + A_{13} W_3$, and H_0 and H_2 are defined from W_3, the lemma implies that W_2 appears in $(H_{**}^{-1})_{aa}$ and hence V_{B_a} in the linear functional $H_{01|2}$, and in the quadratic functional $H_{11|b}$. The next lemma characterizes conditions for conditional independence between stochastic integrals and such functionals of a vector Brownian motion.

A function or *kernel* K on $[0, 1]^2$ is said to be symmetric if $K(u, v) = K(v, u)$ for all $(u, v) \in [0, 1]^2$, and positive semidefinite if $\int_0^1 \int_0^1 K(u, v) g(u) g(v) \, du \, dv \geq 0$ for all continuous functions g on $[0, 1]$.

LEMMA 2. *Let W be a vector Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, independent of $\mathcal{G} \subset \mathcal{F}$. Let X and Y be $\mathcal{G}$-measurable vector processes satisfying $\mathrm{E}\big(\int_0^1 (XX' + YY') \, du\big) < \infty$, and let K be a positive semidefinite $\mathcal{G}$-measurable kernel on $[0, 1]^2$. Then, conditionally on $\mathcal{G}$, $\int_0^1 X \, dW'$ is independent of $\int_0^1 Y W' \, du$ and $\int_0^1 \int_0^1 K(u, v) \, dW(u) \, dW(v)'$ if and only if, with probability one,*

$$\int_0^1 \left(\int_0^u X(v) \, dv \right) Y(u)' \, du = 0, \qquad \int_0^1 K(u, v) X(u) \, du = 0, \qquad v \in [0, 1].$$

We are now in a position to prove the main result.

THEOREM 1. *Let* $B_a^\infty = (B_0^{\infty\prime}, B_1^{\infty\prime})'$, C^∞, V_{B_a} *and* Z *be as defined in (7), (14), and (13), and let* $V_C = \Omega_2 \otimes H_{00}^{-1}$. *Then we have*

$$\begin{pmatrix} B_a^\infty \\ C^\infty \end{pmatrix} \Bigg| \left(V_{B_a}, V_C\right) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_{B_a} & 0 \\ 0 & V_C \end{pmatrix}\right), \tag{16}$$

so that inference on δ *is asymptotically mixed normal; i.e.,*

$$Z \left| \left(V_{B_a}, V_C\right) \sim N(0, I_{r(p-r-s)}). \right. \tag{17}$$

Proof. Define

$$Z_{B_a} = V_{B_a}^{-1/2} \mathrm{vec} B_a^\infty, \qquad Z_C = V_C^{-1/2} \mathrm{vec} C^\infty.$$

From (8), it directly follows that $Z_{B_a}|H_* \sim N(0, I_{r(p-r)})$ and $Z_C|H_0 \sim N(0, I_{s(p-r-s)})$. Note that conditioning on a process X in fact means conditioning on the σ-field generated by $\{X(u)\}_{u\in[0,1]}$. We will use the notation $X \equiv Y$ if both processes or random variables generate the same σ-field, and $X \subset Y$ if the σ-field generated by X is contained in the σ-field generated by Y.

The result $Z_{B_a}|H_* \sim N(0, I_{r(p-r)})$ implies that Z_{B_a} is independent of H_*, and hence also of $(Z_C, V_{B_a}, H_0) \subset (H_0, W_1) \equiv H_*$, so that

$$Z_{B_a}|(Z_C, V_{B_a}, H_0) \sim N(0, I_{r(p-r)}). \tag{18}$$

We will show that conditionally on H_0, Z_C is independent of V_{B_a}. This implies $Z_C|(V_{B_a}, H_0) \sim N(0, I_{s(p-r-s)})$, and together with (18), this implies

$$\begin{pmatrix} Z_{B_a} \\ Z_C \end{pmatrix} \Bigg| \left(V_{B_a}, H_0\right) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_{r(p-r)} & 0 \\ 0 & I_{s(p-r-s)} \end{pmatrix}\right).$$

Because this conditional distribution of (Z_{B_a}, Z_C) does not depend on $\left(V_{B_a}, H_0\right)$, the same joint $N(0, I_{(r+s)(p-r)-s^2})$ distribution applies conditionally on $\left(V_{B_a}, V_C\right) \subset \left(V_{B_a}, H_0\right)$. This directly implies (16) and (17), noting that $V_{B_0} = EV_{B_a}E'$ with $E = I_r \otimes (I_{p-r-s}, 0)$, and $V_{CA} = (A' \otimes I_{p-r-s})V_C(A \otimes I_{p-r-s})$.

Recall that $Z_C = \mathrm{vec}\left(H_{00}^{-1/2} \int_0^1 H_0\, dW_2' \Omega_2^{-1/2}\right)$ and $V_{B_a} = \Omega_1 \otimes (H_{**}^{-1})_{aa}$, where $(H_{**}^{-1})_{aa}$ is given by (15). This means that, conditionally on H_0, Z_C is independent of V_{B_a} if $\int_0^1 H_0\, dW_2'$ is independent of $H_{01|2}$ and $H_{11|b}$; the other

ingredients of $(H_{**}^{-1})_{aa}$ are fixed conditional on H_0. Using $H_1 = A_{12}W_2 + A_{13}W_3$ for fixed matrices A_{12} and A_{13}, with $|A_{12}| \neq 0$, it follows that

$$\begin{aligned} H_{01|2} &= \int_0^1 (H_0 - H_{02}H_{22}^{-1}H_2)H_1'\,du \\ &= \int_0^1 H_{0|2}H_1'\,du \\ &= \int_0^1 H_{0|2}W_2'\,du A_{12}' + \int_0^1 H_{0|2}W_3'\,du A_{13}', \end{aligned}$$

where $H_{0|2}(u) = H_0(u) - H_{02}H_{22}^{-1}H_2(u)$. By Lemma 2, this implies that $\int_0^1 H_0\,dW_2'$ is conditionally independent of $H_{01|2}$, because

$$\int_0^1 \left(\int_0^u H_0\,dv\right) H_{0|2}(u)'\,du = \int_0^1 H_2(u)H_{0|2}(u)'\,du = 0.$$

Next,

$$\begin{aligned} H_{11|b} &= \int_0^1 H_1H_1'\,du - \int_0^1 H_1H_b'\,du H_{bb}^{-1}\int_0^1 H_bH_1'\,du \\ &= A_{12}\left(\int_0^1 W_2W_2'\,du - \int_0^1 W_2H_b'\,du H_{bb}^{-1}\int_0^1 H_bW_2'\,du\right)A_{12}', \end{aligned}$$

where the final equality follows from $H_1 = A_{12}W_2 + A_{13}W_3 = A_{12}W_2 + A_{13}A_{03}^{-1}H_0$ and the fact that H_0 is contained in H_b. From this, we find

$$\begin{aligned} A_{12}^{-1}H_{11|b}A_{12}'^{-1} &= \int_0^1 \left(\int_0^u dW_2\right)\left(\int_0^u dW_2\right)'du \\ &\quad - \int_0^1 \left(\int_0^u dW_2\right)H_b'\,du H_{bb}^{-1}\int_0^1 H_b\left(\int_0^u dW_2\right)'du \\ &= \int_0^1\int_0^1 K(u,v)\,dW_2(u)\,dW_2(v), \end{aligned}$$

where

$$K(u,v) = 1 - u\vee v - \tilde{H}_b(u)'H_{bb}^{-1}\tilde{H}_b(v),$$

with $\tilde{H}_i(u) = \int_u^1 H_i\,dv$, $i = 0,1,2,b$. Applying again Lemma 2, we find that conditionally on H_0, $\int_0^1 H_0\,dW_2'$ is independent of $H_{11|b}$, because

$$\begin{aligned} \int_0^1 K(u,v)H_0(u)\,du &= \int_v^1\left(\int_0^u H_0\,dw\right)du - \int_0^1\left(\int_0^u H_0\,dw\right)H_b(u)'\,du H_{bb}^{-1}\tilde{H}_b(v) \\ &= \tilde{H}_2(v) - \int_0^1 H_2H_b(u)'\,du H_{bb}^{-1}\tilde{H}_b(v) = 0. \end{aligned}$$

The final equality follows from $H_2 = (0, I_{p-r-s})H_b$, and hence $\tilde{H}_2 = (0, I_{p-r-s})\tilde{H}_b$. Thus we have shown that $\int_0^1 H_0\, dW_2'$ is independent of both $H_{01|2}$ and $H_{11|b}$, and hence of V_{B_a}. ■

4. DISCUSSION

Theorem 1 states that the maximum likelihood estimator of the multicointegration parameter δ has an asymptotically mixed normal distribution. This means that a likelihood ratio or Wald test statistic of a simple hypothesis $\mathcal{H}_0 : \delta = \delta_0$ will have an asymptotic $\chi^2_{r(p-r-s)}$ null distribution, arising as the distribution of $Z'Z$. More generally, it is not hard to prove that test statistics of smooth hypotheses $\mathcal{H}_0 : g(\delta) = 0$, with g a continuously differentiable function with derivative $G(\delta)$ of full row rank, will have an asymptotic χ^2 null distribution.

A further extension is to consider hypotheses on β, β_1, and δ together. For example, Mosconi and Paruolo (2010) consider possibly overidentifying restrictions of the form $(\beta', \delta)' = h(\phi)$, where h is a linear function of a parameter vector ϕ. Extending Johansen's (2006) Theorem 5.1, we may obtain conditions on h such that the restricted log-likelihood is locally asymptotically quadratic. As indicated by Johansen (2006), these conditions entail that ϕ can be partitioned as (ϕ_1, ϕ_2), with $n\hat{\phi}_1$ and $n^2\hat{\phi}_2$ converging in distribution to linear functions of $(B_0^\infty, B_1^\infty, C^\infty)$ and B_2^∞, respectively. From Theorem 1 we know that $(B_0^\infty, B_1^\infty, C^\infty)$ is jointly mixed normal, but it can be shown that $(B_0^\infty, B_1^\infty, C^\infty)$ is not independent of B_2^∞ and its conditional variance $\Omega_1 \otimes \left(H_{**}^{-1}\right)_{22}$. This implies that hypotheses that only restrict $B_0 = \bar{\beta}_2^{0\prime}(\psi - \psi^0)$, $B_1 = \bar{\beta}_1^{0\prime}\beta$, and $C = \bar{\beta}_2^{0\prime}\beta_1$ but leave $B_2 = \bar{\beta}_2^{0\prime}\beta$ unrestricted (where θ^0 denotes the true value of θ) can be tested based on asymptotically χ^2 likelihood ratio statistics. In other words, hypotheses that involve β and δ only allow for mixed normal inference if they do not restrict B_2.

NOTE

1. The cointegration parameters have been identified by $c'\beta = I_r$ and $c_1'\beta_1 = I_s$, where c and c_1 are known conformable matrices. The results given here are for $c = \bar{\beta}$ and $c_1 = \bar{\beta}_1$, from which the results for general (c, c_1) can be derived.

REFERENCES

Boswijk, H.P. (2000) Mixed normality and ancillarity in $I(2)$ systems. *Econometric Theory* 16, 878–904.

Engsted, T. & S. Johansen (1999) Granger's representation theorem and multicointegration. In R.F. Engle & H. White (eds.), *Cointegration, Causality and Forecasting: Festschrift in Honour of Clive Granger*, pp. 200–212. Oxford University Press.

Granger, C.W.J. (1986) Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics* 48, 213–228.

Granger, C.W.J. & T.-H. Lee (1990) Multicointegration. In G.F. Rhodes Jr. & T.B. Fomby (eds.), *Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots*. pp. 71–84. JAI Press.

Johansen, S. (1992) A representation of vector autoregressive processes integrated of order 2. *Econometric Theory* 8, 188–202.

Johansen, S. (1997) Likelihood analysis of the $I(2)$ model. *Scandinavian Journal of Statistics* 24, 433–462.

Johansen, S. (2006) Statistical analysis of hypotheses on the cointegrating relations in the $I(2)$ model. *Journal of Econometrics* 132, 81–115.

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

Mosconi, R. & P. Paruolo (2010) Identification of cointegrating relations in $I(2)$ vector autoregressive models. Working paper, University of Insubria.

Paruolo, P. (1995) Topics in Statistical Inference for Vector Autoregressive Processes Integrated of Order 1 and 2. PhD thesis, University of Copenhagen.

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

Paruolo, P. & A. Rahbek (1999) Weak exogeneity in $I(2)$ VAR systems. *Journal of Econometrics* 93, 281–308.

Rahbek, A., H.C. Kongsted, & C. Jørgensen (1999) Trend-stationarity in the $I(2)$ cointegration model. *Journal of Econometrics* 90, 265–289.

Tanaka, K. (1996) *Time Series Analysis. Nonstationary and Noninvertible Distribution Theory*. Wiley.

APPENDIX: Proofs of Lemmas

Proof of Lemma 1. We use the following well-known result for partitioned inverses; see, e.g., Magnus and Neudecker (1988, p. 11):

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A_{11|2}^{-1} & -A_{11|2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11|2}^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}A_{11|2}^{-1}A_{12}A_{22}^{-1} \end{pmatrix}, \tag{A.1}$$

where $A_{11|2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$, and where A_{11}, A_{22}, and $A_{11|2}$ are assumed to be nonsingular. It is convenient to define a reordered version of H_{**}:

$$H_{\dagger\dagger} = \begin{pmatrix} H_{11} & \begin{pmatrix} H_{10} & H_{12} \end{pmatrix} \\ \begin{pmatrix} H_{01} \\ H_{21} \end{pmatrix} & \begin{pmatrix} H_{00} & H_{02} \\ H_{20} & H_{22} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} H_{11} & H_{1b} \\ H_{b1} & H_{bb} \end{pmatrix},$$

so that $(H_{**}^{-1})_{00}$ is identical to the middle diagonal block of $H_{\dagger\dagger}^{-1}$, $(H_{**}^{-1})_{11}$ is the first diagonal block, and $(H_{**}^{-1})_{01}$ is the part of $H_{\dagger\dagger}^{-1}$ corresponding to the second row and first column of blocks. Applying (A.1) to $H_{\dagger\dagger}$ with $A_{11} = H_{11}$, $A_{12} = H_{1b}$, and $H_{22} = H_{bb}$ directly leads to

$$(H_{**}^{-1})_{11} = H_{11|b}^{-1},$$

and

$$(H_{**}^{-1})_{bb} = \begin{pmatrix} H_{00} & H_{02} \\ H_{20} & H_{22} \end{pmatrix}^{-1} + \begin{pmatrix} H_{00} & H_{02} \\ H_{20} & H_{22} \end{pmatrix}^{-1} \begin{pmatrix} H_{01} \\ H_{21} \end{pmatrix} H_{11|b}^{-1} (H_{10}, H_{12}) \begin{pmatrix} H_{00} & H_{02} \\ H_{20} & H_{22} \end{pmatrix}^{-1}.$$

Next, applying (A.1) again to

$$\begin{pmatrix} H_{00} & H_{02} \\ H_{20} & H_{22} \end{pmatrix}^{-1} = \begin{pmatrix} H_{00|2}^{-1} & -H_{00|2}^{-1} H_{02} H_{22}^{-1} \\ -H_{22}^{-1} H_{20} H_{00|2}^{-1} & H_{22}^{-1} + H_{22}^{-1} H_{20} H_{00|2}^{-1} H_{02} H_{22}^{-1} \end{pmatrix}$$

yields

$$\begin{aligned}(H_{**}^{-1})_{00} &= H_{00|2}^{-1} \\ &\quad + H_{00|2}^{-1} \left(I_{p-r-s}, -H_{02} H_{22}^{-1} \right) \begin{pmatrix} H_{01} \\ H_{21} \end{pmatrix} H_{11|b}^{-1} (H_{10}, H_{12}) \begin{pmatrix} I_{p-r-s} \\ -H_{22}^{-1} H_{20} \end{pmatrix} H_{00|2}^{-1} \\ &= H_{00|2}^{-1} + H_{00|2}^{-1} H_{01|2} H_{11|b}^{-1} H_{10|2} H_{00|2}^{-1}.\end{aligned}$$

Finally, $(H_{**}^{-1})_{01}$ equals the first block of $-H_{bb}^{-1} H_{b1} H_{11|b}^{-1}$, which from the partitioned expression of H_{bb}^{-1} leads to $(H_{**}^{-1})_{01} = -H_{00|2}^{-1} H_{01|2} H_{11|b}^{-1}$. ■

Proof of Lemma 2. Let $\tilde{Y}(u) = \int_u^1 Y(v)\,dv$. Because $\tilde{Y}(1) = 0$ and $W(0) = 0$, integration by parts yields

$$\int_0^1 Y(u) W(u)'\,du = \int_0^1 Y(u) \left(\int_0^u dW(v)' \right) du = \int_0^1 \tilde{Y}(u)\,dW(u)'.$$

Next, for a positive definite K, Mercer's theorem (see Tanaka, 1996) states that

$$K(u,v) = \sum_{i=1}^{\infty} \lambda_i f_i(u) f_i(v), \tag{A.2}$$

where $\{\lambda_i\}_{i\geq 1}$ and $\{f_i\}_{i\geq 1}$ are the eigenvalues (Tanaka, 1996, refers to $1/\lambda_i$ as the eigenvalues) and orthonormal eigenfunctions of K, solving the integral equation

$$\int_0^1 K(u,v) f(u)\,du = \lambda f(v).$$

This implies that

$$\begin{aligned}\int_0^1 \int_0^1 K(u,v)\,dW(u)\,dW(v)' &= \int_0^1 \int_0^1 \sum_{i=1}^{\infty} \lambda_i f_i(u) f_i(v)\,dW(u)\,dW(v)' \\ &= \sum_{i=1}^{\infty} \lambda_i \left(\int_0^1 f_i\,dW \right) \left(\int_0^1 f_i\,dW \right)',\end{aligned}$$

where the second equality (term by term integration) follows from uniform convergence of the series representation in (A.2), together with the fact that the eigenfunctions are orthonormal and hence bounded in L^2. The basic properties of the Itô integral imply

$$\begin{pmatrix} \int_0^1 X\,dW' \\ \int_0^1 \tilde{Y}\,dW' \\ \int_0^1 f_i\,dW' \end{pmatrix} \Bigg| \mathcal{G} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \int_0^1 \begin{pmatrix} XX' & X\tilde{Y}' & Xf_i' \\ \tilde{Y}X' & \tilde{Y}\tilde{Y}' & \tilde{Y}f_i' \\ f_iX' & f_i\tilde{Y}' & f_if_i' \end{pmatrix} du \right).$$

Therefore $\int_0^1 X\,dW'$ is conditionally independent of $\int_0^1 YW'\,du$ and $\int_0^1 f_i\,dW'$ if and only if

$$\int_0^1 X(u)\tilde{Y}(u)\,du = \int_0^1 \left(\int_0^u X(v)\,dv \right) Y'\,du = 0, \tag{A.3}$$

(the first equality follows from integration by parts), and

$$\int_0^1 X(u) f_i(u)\,du = 0. \tag{A.4}$$

This in turn implies that $\int_0^1 X\,dW'$ is conditionally independent of $\int_0^1 YW'\,du$ and $\int_0^1 \int_0^1 K\,dW\,dW'$ if and only if both (A.3) holds and (A.4) holds for all eigenfunctions f_i corresponding to nonzero eigenvalues. The latter condition is equivalent to

$$\int_0^1 K(u,v)X(u)\,du = \sum_{i=1}^{\infty} \lambda_i f_i(v) \int_0^1 f_i(u)X(u)\,du = 0, \qquad v \in [0,1].$$

Hence the components of X are eigenfunctions of K corresponding to zero eigenvalues. ■