Econometric Theory, **16**, 2000, 878–904. Printed in the United States of America.

MIXED NORMALITY AND ANCILLARITY IN *I*(2) SYSTEMS

H. PETER BOSWIJK
University of Amsterdam

This paper studies asymptotic likelihood inference on cointegration parameters in systems integrated of order two. We start with so-called triangular systems and then extend the analysis to vector autoregressions. We show that even when all unit root restrictions have been imposed, the asymptotic observed information is not (locally) ancillary, which implies that the log-likelihood ratio is not locally asymptotically mixed normal. The results are applied to inference on polynomial cointegration. Some similarities and differences with $I(1)$ systems are also discussed.

1. INTRODUCTION

The theory of cointegration has gained a major role in the econometric analysis of economic time series. The Gaussian vector autoregressive (VAR) model has proved to be a fruitful basis for conducting likelihood inference on cointegration in systems integrated of order one (for a comprehensive overview, see Johansen, 1995c). In empirical applications however, it has appeared that some of the series involved, in particular prices and nominal variables, are more aptly described as $I(2)$ processes. Furthermore, the statistical analysis of multicointegrated $I(1)$ processes (see Granger and Lee, 1990) is facilitated by accumulating the original series, leading to a cointegrated $I(2)$ system (see, e.g., Haldrup, 1998). Representation of such systems is considered in Johansen (1992). Statistical inference in $I(2)$ models is analyzed, inter alia, by Stock and Watson (1993) and Kitamura (1995) in the context of so-called triangular systems (introduced for $I(1)$ processes by Phillips, 1991); and by Johansen (1995a, 1997) in VAR models. A recent overview is given by Haldrup (1998).

A central result in Gaussian cointegrated $I(1)$ systems, as shown by Phillips (1991), is that likelihood inference on the cointegrating vectors is asymptotically mixed normal, provided that the correct number of unit roots has been imposed. Jeganathan (1995, 1997) shows that essentially the same result holds in non-Gaussian systems. If, on the other hand, the unit root parameters are estimated, then the mixed normality is lost in general, and asymptotic distribu-

I thank Niels Haldrup, Søren Johansen, Paolo Paruolo, Thomas Rothenberg, Halbert White, two anonymous referees, and the editor, Peter Phillips, for helpful comments. Financial support from the Royal Netherlands Academy of Arts and Sciences is 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: peterb@fee.uva.nl.

 © 2000 Cambridge University Press 0266-4666/00 $9.50

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

tions fall into the more general class of Brownian motion functionals (see Sims, Stock, and Watson, 1990; Jeganathan, 1997). An interpretation of this result is given by Phillips (1989) and Jeganathan (1995), who show that log-likelihood ratios that admit a quadratic approximation are locally asymptotically mixed normal if and only if the observed information (i.e., minus the second derivative of the log-likelihood) is locally asymptotically ancillary, i.e., its asymptotic distribution is constant under a range of true values that shrinks to a single point at a suitable rate. An elegant proof of this result is given by Le Cam and Yang (1990, p. 81). This ancillarity is lost if unit root parameters are included.

The importance of imposing the correct number of unit roots is reemphasized for $I(2)$ systems by Kitamura (1995), who shows that inference on the cointegration parameters in three specific classes of triangular systems is mixed normal. In contrast, Johansen's (1997) analysis of asymptotic likelihood inference in $I(2)$ vector autoregressions shows that not all parameter estimators have asymptotic mixed normal distributions, even if the correct number of unit roots has been imposed.

The present paper seeks to resolve this apparent difference between inference in triangular systems and in vector autoregressions. Jeganathan's analysis implies that a lack of mixed normality should be due to a violation of asymptotic ancillarity, and we find this to be the case in cointegrated $I(2)$ systems even with the correct number of unit roots imposed. Therefore, inference in $I(2)$ is in general not mixed normal, whether one considers triangular systems or vector autoregressions. This apparent conflict with Kitamura's (1995) results is explained by the fact that we consider a more general class of models (closely related to the setup of Stock and Watson, 1993). Thus, unlike the $I(1)$ case considered by Johansen (1995b), imposing the unit roots does not imply ancillarity of the observed information.

The plan of the paper is as follows. In Section 2, we give a brief review of local asymptotic likelihood analysis, stressing the importance of ancillarity of the observed information for mixed normality. In Section 3, we apply this analysis to inference in triangular systems, showing that in a model more general than that of Kitamura (1995), the log-likelihood ratio is not locally asymptotically mixed normal. In Section 4, we show that the same result is obtained in vector autoregressions. In Section 5 we study inference on specific parameter restrictions, in particular the polynomial cointegration parameters. The final section discusses these results.

The following notation is used in the paper. A defining expression is signified by $:=$, and equality in distribution is denoted by $\equiv$. The Euclidean norm of a matrix A is denoted by $\|A\| := [\mathrm{tr}(A'A)]^{1/2}$. The operator $\mathrm{vec}(A)$ stacks the columns of the matrix A, and $\mathrm{vech}(A)$ stacks the lower-triangular part of a symmetric matrix A. The inequality > 0 means "is positive definite" when applied to matrices. Convergence in probability and convergence in distribution are denoted by $\xrightarrow{P}$ and $\xrightarrow{\mathcal{L}}$, respectively. The distribution of a random variable X under a probability measure P is denoted by $\mathcal{L}(X|P)$; if P is the distribution of

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

X itself, this is abbreviated as $\mathcal{L}(X)$. Vector Brownian motion with covariance matrix Ω is indicated by $\mathrm{BM}(\Omega)$. Riemann integrals such as $\int_0^1 W(u)\,du$ are abbreviated as $\int W$, whereas stochastic integrals such as $\int_0^1 W(u)\,dW(u)$ are denoted as $\int W dW$. For two (probability) measures P and Q defined on the same measurable space, dP/dQ denotes the Radon–Nikodym derivative of the absolutely continuous part of P with respect to Q. All limits are taken as the sample size n tends to infinity.

2. MIXED NORMALITY AND ANCILLARITY

Consider a p-vector stochastic process $\{X_t, t = 1,2,\ldots\}$. A statistical model of $(X_1,\ldots,X_n)$ can be represented by the experiment

$$\mathcal{E}_n = (\mathbb{R}^{pn}, \mathcal{B}_n, \{P_{\theta,n}, \theta \in \Theta\}),$$

where $\mathcal{B}_n$ is the Borel σ-field in $\mathbb{R}^{pn}$ and $\{P_{\theta,n}\}$ is a family of probability measures indexed by the k-dimensional parameter vector θ, with parameter space $\Theta \subseteq \mathbb{R}^k$. The likelihood function $L_n(\theta)$ is defined by the density $dP_{\theta,n}/d\mu$, with μ a dominating measure on $(\mathbb{R}^{pn}, \mathcal{B}_n)$, and the log-likelihood ratio is given by

$$\Lambda_n(\theta_1,\theta_2) := \log \frac{dP_{\theta_1,n}}{dP_{\theta_2,n}} = \log \frac{L_n(\theta_1)}{L_n(\theta_2)}.$$

Jeganathan (1995) defines the log-likelihood ratio to be locally asymptotically quadratic (LAQ) at $\theta \in \Theta$ if the following three conditions are satisfied. First, there exists a sequence of $k \times k$ nonsingular normalizing matrices D_n, satisfying $\|D_n\| \to 0$, a sequence of random k vectors S_n, and a sequence of random, almost surely positive definite $k \times k$ matrices J_n, all possibly depending on θ, such that under $P_{\theta,n}$,

$$\Lambda_n(\theta + D_n\tau_n, \theta) - (\tau_n' S_n - \tfrac{1}{2}\tau_n' J_n \tau_n) \xrightarrow{P} 0, \tag{2.1}$$

for all bounded sequences of k vectors $\{\tau_n\}$ satisfying $\theta + D_n\tau_n \in \Theta$. The second condition for LAQ is

$$(S_n, J_n) \xrightarrow{\mathcal{L}} (S, J), \tag{2.2}$$

under $P_{\theta,n}$, where S and J are a random vector and an almost surely positive definite random matrix, respectively. Combining (2.1) and (2.2) yields, for $\tau_n \to \tau$,

$$\Lambda_n(\theta + D_n\tau_n, \theta) \xrightarrow{\mathcal{L}} \Lambda(\tau) := \tau' S - \tfrac{1}{2}\tau' J \tau. \tag{2.3}$$

The third requirement for LAQ is that the sequences of measures $P_{\theta,n}$ and $P_{\theta+D_n\tau_n,n}$ be contiguous, i.e., $P_{\theta,n}(A_n) = 0 \Leftrightarrow P_{\theta+D_n\tau_n,n}(A_n) = 0, A_n \in \mathcal{B}_n$, which is equivalent to the requirement

$$E[\exp(\Lambda(\tau))] = 1. \tag{2.4}$$

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

In the applications to follow, the log-likelihood is twice continuously differentiable, so that, for some suitably chosen D_n,

$$\log L_n(\theta + D_n\tau_n) = \log L_n(\theta) + \tau_n' D_n' \frac{\partial \log L_n(\theta)}{\partial \theta} + \frac{1}{2}\tau_n' D_n' \frac{\partial^2 \log L_n(\theta)}{\partial \theta \partial \theta'} D_n \tau_n + o_p(1). \quad \textbf{(2.5)}$$

Hence S_n is the normalized score vector, and J_n is the normalized observed information matrix, both evaluated at the point θ. Similarly, S and J may be interpreted as the limiting score vector and observed information matrix, corresponding to the limiting quadratic log-likelihood ratio $\Lambda(\tau)$ in (2.3); condition (2.4) guarantees the existence of a family of probability measures $\{Q_\tau, \tau \in \mathbb{R}^k\}$ having $\Lambda(\tau) = \log dQ_\tau/dQ_0$ as its log-likelihood ratio (see Jeganathan, 1995). Note that τ may be seen as a local parameter, measuring the normalized deviation from θ.

An important subclass of the LAQ likelihood ratios is given by the locally asymptotically mixed normal (LAMN) likelihood ratios. They are characterized by

$$(S, J) \equiv (J^{1/2}Z, J), \quad \textbf{(2.6)}$$

where Z is a standard normal vector, independent of J. Thus the conditional distribution of S given J is $N(0, J)$, or in other words, the distribution of S is a mixture of normal distributions, with the observed information matrix J acting as the mixing variate. As shown by Jeganathan (1995), a necessary and sufficient condition for (2.6) is

$$\mathcal{L}(J|Q_\tau) = \mathcal{L}(J|Q_0), \quad \textbf{(2.7)}$$

where Q_τ is the weak limit of $\mathcal{L}(S_n, J_n|P_{\theta+D_n\tau_n,n})$ (still assuming $\tau_n \to \tau$). Thus, (2.7) requires the limiting distribution of the observed information not to vary locally with the parameters, so that it acts as an ancillary statistic; hence we say that the observed information is locally asymptotically ancillary.

Jeganathan (1995) also shows that for LAQ likelihood ratios, (2.7) and hence (2.6) hold for almost every $\theta \in \Theta$, i.e., outside a set of Lebesgue measure zero, provided that $D_n(\theta)$ is a sufficiently regular function of θ; in particular, this requires that for all $\theta \in \Theta$ and $\alpha > 0$,

$$\sup_{\|\tau\| \le \alpha} \|D_n(\theta)^{-1} D_n(\theta + D_n(\theta)\tau) - I_k\| \to 0, \quad \textbf{(2.8)}$$

where the supremum is taken only over those τ that satisfy $\theta + D_n(\theta)\tau \in \Theta$ (Condition C.5 in Jeganathan, 1995). Subsequently we shall encounter a case where this condition is violated, so that LAQ does not imply LAMN almost everywhere.

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

Establishing the LAQ and LAMN property of a model serves two purposes. First, it may be used to obtain the asymptotic distribution of likelihood-based statistics, such as the maximum likelihood estimator (MLE) and likelihood ratio test statistics. Under LAQ and sufficient smoothness conditions on the likelihood function, the normalized MLE $\hat{\tau}_n = D_n^{-1}(\hat{\theta}_n - \theta)$ converges in distribution to $\hat{\tau} = J^{-1}S$, the maximizer of $\Lambda(\tau)$ in (2.3), which defines the MLE to be asymptotically centering. Under LAMN, $J^{-1}S$ has an $N(0, J^{-1})$ distribution conditional on J, which in turn implies that likelihood ratio test statistics for a large class of hypotheses will have an asymptotic χ^2 null distribution. For cases where the likelihood function is not sufficiently smooth, asymptotically centering estimators and corresponding test statistics may be constructed by the methods outlined in Jeganathan (1995).

A second motivation for this analysis is that the LAMN property allows one to determine certain optimality properties of likelihood-based inference procedures. In particular, the usual optimality results associated with exponential families will apply to LAMN models in a local asymptotic and conditional form. (For details, see Jeganathan, 1995.) The emphasis in this paper will be on the first issue (pertaining to asymptotic distributions), although the results will have consequences for the second issue also.

3. LIKELIHOOD ANALYSIS IN TRIANGULAR SYSTEMS

Consider the following triangular $I(2)$ system, obtained as a special case ($d = 2$, no deterministic components) of Stock and Watson's (1993) representation:

$$X_{0t} = A_1 X_{1t} + A_2 X_{2t} + A_3 \Delta X_{2t} + \varepsilon_{0t}, \tag{3.1}$$

$$\Delta X_{1t} = A_4 \Delta X_{2t} + \varepsilon_{1t}, \tag{3.2}$$

$$\Delta^2 X_{2t} = \varepsilon_{2t}, \tag{3.3}$$

$t = 1, \ldots, n$, where the vectors X_{it} are $p_i \times 1$, so that $X_t := (X_{0t}', X_{1t}', X_{2t}')'$ is of dimension $p := p_0 + p_1 + p_2$. We assume that $X_0 = \Delta X_0 = 0$ (a.s.), which excludes the possibility that the level of X_t contains a linear trend. The innovations $\{\varepsilon_t := (\varepsilon_{0t}', \varepsilon_{1t}', \varepsilon_{2t}')'\}$ are assumed to be independent and identically distributed (i.i.d.) $N(0, \Omega)$, with Ω known; some relaxation of this assumption is possible and will be discussed subsequently. The parameter vector of the model is $\theta := (\mathrm{vec}(A_1, A_2, A_3)', \mathrm{vec}(A_4)')'$, with parameter space $\Theta = \mathbb{R}^k, k = p_0(p_1 + 2p_2) + p_1 p_2$.

Note that X_{2t} is $I(2)$; if $A_4 \neq 0$, $A_1 \neq 0$, or $A_2 \neq 0$, then X_{0t} and X_{1t} may also be $I(2)$, but the common $I(2)$ trends in the system are represented by $X_{2t} = \sum_{s=1}^{t} \sum_{j=1}^{s} \varepsilon_{2j}$. Similarly, the common $I(1)$ trends in the system are $(X_{1t} - A_4 X_{2t}) = \sum_{s=1}^{t} \varepsilon_{1s}$ and $\Delta X_{2t} = \sum_{s=1}^{t} \varepsilon_{2s}$, so that X_{1t} and X_{2t} together represent the common $I(1)$ and $I(2)$ trends in the system. The linear combination $(I_{p_0}, -A_1, -A_2) X_t$ is (in general) $I(1)$ and cointegrates with ΔX_{2t} to $I(0)$. Thus the polynomial $(I_{p_0}, -A_1, -A_2 - A_3 \Delta)$ transforms the $I(2)$ process X_t into

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

the $I(0)$ process ε_{0t}, a phenomenon called polynomial cointegration or multicointegration. If $A_3 = 0$, then the polynomial part of the cointegrating relation vanishes, and $(I_{p_0}, -A_1, -A_2)X_t$ is already stationary. We shall return to this hypothesis later. Finally, the linear combination $(0, I_{p_1}, -A_4)$ transforms X_t into the $I(1)$ process $\sum_{s=1}^{t} \varepsilon_{1s}$, which does not cointegrate with any other process to $I(0)$.

The fact that $\{\varepsilon_t\} \equiv$ i.i.d. $N(0, \Omega)$ implies the following functional central limit theorem:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{[un]} \varepsilon_i \xrightarrow{\mathcal{L}} W(u) \equiv \mathrm{BM}(\Omega), \qquad u \in [0,1]. \tag{3.4}$$

Let W and Ω be partitioned conformably with ε_t. We shall make use of the subscript coupling notation $b = 1,2$, so that, e.g., $\varepsilon_{bt} := (\varepsilon_{1t}', \varepsilon_{2t}')'$. Next, define

$$W_{0\cdot b} := W_0 - \Omega_{0b}\Omega_{bb}^{-1} W_b \equiv \mathrm{BM}(\Omega_{00\cdot b}), \quad \Omega_{00\cdot b} := \Omega_{00} - \Omega_{0b}\Omega_{bb}^{-1}\Omega_{b0},$$

$$W_{1\cdot 2} := W_1 - \Omega_{12}\Omega_{22}^{-1} W_2 \equiv \mathrm{BM}(\Omega_{11\cdot 2}), \quad \Omega_{11\cdot 2} := \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}.$$

Note that $W_{0\cdot b}$ is independent of W_b and $W_{1\cdot 2}$ is independent of W_2. Finally, define the integrated Brownian motion $\overline{W}_2(u) = \int_0^u W_2(s)ds$ and the normalizing matrix

$$D_n = \begin{pmatrix} n^{-1}I_{p_0p_1} & 0 & 0 & 0 \\ -n^{-1}(A_4' \otimes I_{p_0}) & n^{-2}I_{p_0p_2} & 0 & 0 \\ 0 & 0 & n^{-1}I_{p_0p_2} & -n^{-1}(I_{p_2} \otimes \Omega^{10}(\Omega^{00})^{-1}) \\ 0 & 0 & 0 & n^{-1}I_{p_1p_2} \end{pmatrix}, \tag{3.5}$$

where $\Omega^{ij} := [\Omega^{-1}]_{ij}$.

THEOREM 3.1. *In the model* (3.1)–(3.3), *as* $n \to \infty$ *and under* $P_{\theta,n}$,

$$\begin{aligned} \Lambda_n(\theta + D_n\tau, \theta) \xrightarrow{\mathcal{L}} \Lambda(\tau) &= \tau'S - \tfrac{1}{2}\tau' J\tau \\ &= \Lambda_0(\tau_0) + \Lambda_1(\tau_1) \\ &= (\tau_0'S_0 - \tfrac{1}{2}\tau_0'J_0\tau_0) + (\tau_1'S_1 - \tfrac{1}{2}\tau_1'J_1\tau_1), \end{aligned} \tag{3.6}$$

where $\tau = (\tau_0', \tau_1')'$,

$$J = \begin{pmatrix} J_0 & 0 \\ 0 & J_1 \end{pmatrix} \equiv \begin{pmatrix} \int FF' \otimes \Omega_{00\cdot b}^{-1} & 0 \\ 0 & \int W_2W_2' \otimes \Omega_{11\cdot 2}^{-1} \end{pmatrix}, \tag{3.7}$$

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

and

$$S = \begin{pmatrix} S_0 \\ S_1 \end{pmatrix} \equiv \begin{pmatrix} \text{vec} \int (d\Omega_{00\cdot b}^{-1} W_{0\cdot b}) F' \\ \text{vec} \int (d\Omega_{11\cdot 2}^{-1} W_{1\cdot 2}) W_2' \end{pmatrix} \equiv \begin{pmatrix} J_0^{1/2} Z_0 \\ J_1^{1/2} Z_1 \end{pmatrix}, \tag{3.8}$$

where $\mathcal{L}(F|Q_0) = \mathcal{L}((W_1', \overline{W}_2', W_2')')$ *and where* Z_0 *and* Z_1 *are independent standard normal vectors, with* Z_0 *independent of* (J_0, J_1) *and* Z_1 *independent of* J_1.

Under $P_{\theta + D_n \tau, n}$, (3.6) *and* (3.7) *still hold but with*

$$\mathcal{L}(F|Q_\tau) = \mathcal{L}\begin{pmatrix} W_1 + \mathcal{T}_1 \overline{W}_2 \\ \overline{W}_2 \\ W_2 \end{pmatrix}, \tag{3.9}$$

where $\mathcal{T}_1$ *is a* $p_1 \times p_2$ *matrix such that* $\tau_1 = \text{vec}(\mathcal{T}_1)$. *Hence*

$$\mathcal{L}(J|Q_\tau) = \mathcal{L}(H_\tau' J H_\tau | Q_0), \tag{3.10}$$

where

$$H_\tau = \begin{pmatrix} I_{p_0 p_1} & 0 & 0 \\ \mathcal{T}_1' \otimes I_{p_0} & I_{p_0 p_2} & 0 \\ 0 & 0 & I_{p_0 p_2 + p_1 p_2} \end{pmatrix}. \tag{3.11}$$

Proofs are given in the Appendix.

Remark 3.1. Theorem 3.1 shows that the log-likelihood ratio for the triangular system (3.1)–(3.3) is LAQ. In fact, given that S and J may be expressed as functionals of Brownian motion processes, the LAQ approximation is locally asymptotically Brownian functional (LABF) (see Jeganathan, 1995). The fact that J is the quadratic variation of the stochastic integral S implies that the contiguity condition (2.4) is satisfied.

Remark 3.2. The normalizing matrix D_n in (3.5) is not diagonal, which means that the ith component of τ does not in general correspond to the ith component of θ. The reason for the off-diagonal block involving A_4 is that the regressors X_{1t} and X_{2t} in (3.1) are in general both $I(2)$, but the linear combination $X_{1t} - A_4 X_{2t}$ is $I(1)$ because of (3.2). Because the observed information involves the second moment matrix of the regressors of (3.1), some rotation is required for the normalized observed information matrix to converge in distribution in all directions. On the other hand, the off-diagonal block in D_n involving $\Omega^{10}(\Omega^{00})^{-1}$ is not strictly required and could be set to zero; however, in that case the block-diagonality of J in (3.7) would be lost.

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

Remark 3.3. Theorem 3.1 implies that the limiting log-likelihood ratio $\Lambda(\tau)$ may be decomposed into two quadratic forms $\Lambda_0(\tau_0)$ and $\Lambda_1(\tau_1)$. The former stems from the conditional distribution of ε_{0t} given ε_{bt}, whereas the latter corresponds to the marginal distribution of ε_{bt} (or, equivalently, the conditional distribution of ε_{1t} given ε_{2t}; the marginal distribution of ε_{2t} does not involve θ). In fact, the normality assumption implies that these two parts of the log-likelihood ratio are exactly quadratic in finite samples, so that the MLE $\hat{\theta}_n$ satisfies

$$D_n^{-1}(\hat{\theta}_n - \theta) = \begin{pmatrix} \hat{\tau}_{0n} \\ \hat{\tau}_{1n} \end{pmatrix} = \begin{pmatrix} J_{0n}^{-1} S_{0n} \\ J_{1n}^{-1} S_{1n} \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} J_0^{-1} S_0 \\ J_1^{-1} S_1 \end{pmatrix}.$$

Because the distribution of S_i given J_i is $N(0, J_i)$, $i = 0,1$, we have obtained what might be called a partial LAMN result: the two parts of the log-likelihood ratio, and the two corresponding parts of the normalized MLE, have an asymptotic mixed normal distribution when taken in isolation.

Remark 3.4. To conclude from the previous remark that the full log-likelihood ratio and MLE are LAMN, we would have to show that the joint distribution of $(Z_0', Z_1')'$ is standard normal and independent of $J = \operatorname{diag}(J_0, J_1)$. However, there is no reason to expect this to hold, because J_0 depends on $W_b = (W_1', W_2')'$ and $Z_1 = J_1^{-1/2} S_1$ is a functional of this vector Brownian motion. In fact, the result (3.10) shows that the ancillarity condition (2.7) is violated, so that the full limiting log-likelihood ratio $\Lambda(\tau)$ is not mixed normal. Because this holds for all $\theta \in \Theta$, we see that in this case, LAQ does not imply LAMN almost everywhere. The explanation of this result is that the normalizing matrix $D_n(\theta)$ does not satisfy the regularity condition (2.8); in particular, we find $D_n(\theta)^{-1} D_n(\theta + D_n(\theta)\tau) \to H_\tau^{-1}$.

Remark 3.5. It may be useful to compare the preceding analysis to the $I(1)$ triangular system as analyzed by Phillips (1991) and Jeganathan (1997):

$$X_{0t} = AX_{1t} + \varepsilon_{0t},$$

$$\Delta X_{1t} = \Pi X_{1,t-1} + \varepsilon_{1t},$$

with $\Pi = 0$. Using the same notation as before, this model is LAQ with $D_n = n^{-1} I_{p_0 p}$ and $J \equiv \int FF' \otimes \operatorname{diag}(\Omega_{00\cdot 1}, \Omega_{11})^{-1}$, where F is the Brownian motion process W_1 under $\Pi = 0$. However, under contiguous alternatives $\Pi_n = n^{-1}\mathcal{T}_1$, $F(u)$ is the Ornstein–Uhlenbeck process $\int_0^u \exp(\mathcal{T}_1(u - s)) dW_1(s)$, so that J does not satisfy (2.7). In this case, imposing the unit root restriction $\Pi = 0$ (and hence $\mathcal{T}_1 = 0$) leads to mixed normality with respect to the remaining parameter A. We have seen that in the $I(2)$ model, imposing these unit root restrictions is not sufficient for LAMN. Although the distribution of the limiting information matrix only involves (integrated) Brownian motions, which do

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

not depend locally upon unknown parameters themselves, the form in which they appear in J does vary locally with θ. Just as in the $I(1)$ model, one may obtain full LAMN by imposing a restriction, namely, that A_4 be known (and hence not treated as a parameter). Of course, when $p_1 = 0$, so that X_{1t} and hence (3.2) vanish from the system, full mixed normality is obtained also. The $I(2)$ triangular systems considered by Kitamura (1995) have similar restrictions imposed, which explains his conclusion that LAMN obtains when all unit roots have been imposed. We shall return to this in Section 5.

Remark 3.6. The breakdown of LAMN in the present $I(2)$ model might be seen as a particular instance of the general lack of mixed normality in systems with an arbitrary number of roots on or in the vicinity of the unit circle, as established recently by Jeganathan (1997). However, in Jeganathan's inferential procedure the unit roots are not assumed known but are treated as parameters. Therefore, his results can be seen as a generalization of the well-known loss of mixed normality due to estimating the unit roots, as discussed in the previous remark. The striking feature of Theorem 3.1 is that in $I(2)$ systems LAMN breaks down despite imposing the unit root restrictions. In this sense our result is not an obvious corollary of Jeganathan's (1997) theorems.

Remark 3.7. The distributional assumption on ε_t may be loosened in different directions without changing the main points of Theorem 3.1. In the next section, we shall analyze a dynamic version of the model (3.1)–(3.3), which is equivalent to replacing ε_t by a stationary Gaussian VAR process. When the Gaussian i.i.d. assumption is maintained, we may allow the covariance matrix Ω to be unknown and hence treated as a parameter. Letting $\omega := \text{vech}(\Omega)$, this entails replacing θ by $(\theta', \omega')'$, Θ by $\Theta \times \{\omega \in \mathbb{R}^{p(p+1)/2} : \Omega > 0\}$, and D_n by $\text{diag}(D_n, n^{-1/2} I_{p(p+1)/2})$ and finally adding $\Lambda_2(\tau_2)$ to $\Lambda(\tau)$, where

$$\Lambda_2(\tau_2) = \tau_2 J_2^{1/2} Z_2 - \tfrac{1}{2}\tau_2' J_2 \tau_2, \qquad J_2 = \tfrac{1}{2} D'(\Omega^{-1} \otimes \Omega^{-1}) D,$$

with Z_2 a standard normal vector, independent of (Z_0, Z_1) and (J_0, J_1), and where D is the duplication matrix (cf. Magnus and Neudecker, 1988). Thus, this extension leads to an additional locally asymptotically normal (LAN) part $\Lambda_2(\tau_2)$ in the log-likelihood ratio, without influencing the properties of the original part $\Lambda_0(\tau_0) + \Lambda_1(\tau_1)$.

Remark 3.8. A further generalization of Theorem 3.1 is obtained by assuming $\{\varepsilon_t\}$ to be not necessarily Gaussian i.i.d. with mean zero and finite positive definite covariance matrix Ω. If the analysis is still based on the Gaussian (quasi-) likelihood function, then the results of Theorem 3.1 still apply, because the invariance principle and weak convergence to matrix stochastic integral result continue to hold for general i.i.d. $(0, \Omega)$ innovations. If the analysis is based on the true (non-Gaussian) likelihood, then we require ε_1 to have an absolutely continuous Lebesgue density $f(\varepsilon)$ with score $\psi(\varepsilon) := -\partial \log f(\varepsilon)/\partial \varepsilon$ and with

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

finite and positive definite information matrix $\mathcal{I} := E(\psi(\varepsilon_1)\psi(\varepsilon_1)')$, which implies $E(\psi(\varepsilon_1)) = 0$. Then (3.4) can be extended to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{[un]}\begin{pmatrix}\psi(\varepsilon_i)\\ \varepsilon_i\end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix}B(u)\\ W(u)\end{pmatrix} \equiv \mathrm{BM}\begin{pmatrix}\mathcal{I} & \Phi\\ \Phi' & \Omega\end{pmatrix}, \qquad u \in [0,1], \tag{3.12}$$

where $\Phi := E(\psi(\varepsilon_1)\varepsilon_1')$. In the Gaussian case, $\psi(\varepsilon) = \Omega^{-1}\varepsilon$, so that $B = \Omega^{-1}W$, $\mathcal{I} = \Omega^{-1}$, and $\Phi = I_p$. In more general cases, B will not be perfectly correlated with W. Replacing Ω^{ij} in (3.5) by $\mathcal{I}_{ij}$, Theorem 3.1 still applies, but with $(W_{0\cdot b}, \Omega^{-1}_{00\cdot b})$ replaced by $(B_0, \mathcal{I}_{00})$ and with $(W_{1\cdot 2}, \Omega^{-1}_{11\cdot 2})$ replaced by $([B_1 - \mathcal{I}_{10}\mathcal{I}_{00}^{-1}B_0], [\mathcal{I}_{11} - \mathcal{I}_{10}\mathcal{I}_{00}^{-1}\mathcal{I}_{01}])$. This may be proved analogously to Proposition 2 and Theorem 3 of Jeganathan (1997). In particular, the lack of asymptotic ancillarity (3.10) remains unaltered. The same result essentially applies when $f(\varepsilon)$ depends upon some nuisance parameters (such as scale parameters, analogous to ω in the previous remark) and even when the entire density function is unknown (so that the nuisance parameter is infinite dimensional); in that case adaptive estimation may be used to recover the log-likelihood ratio, provided that f is symmetric (see Jeganathan, 1995).

4. LIKELIHOOD ANALYSIS IN VECTOR AUTOREGRESSIONS

Consider now the VAR model of order l, written in error-correction form:

$$\Delta^2 X_t = \Gamma\Delta X_{t-1} + \Pi X_{t-1} + \sum_{i=1}^{l-2}\Psi_i\Delta^2 X_{t-i} + \varepsilon_t, \tag{4.1}$$

where X_t is a p vector time series and ε_t is i.i.d. $N(0,\Omega)$. As in the previous section, we assume $X_0 = \Delta X_0 = 0$ for simplicity. Let $\Psi(z) = I_p - \sum_{i=1}^{l-2}\Psi_i z^i$ and $\Psi = \Psi(1)$. The following assumption guarantees that the process X_t is $I(2)$ (see Johansen, 1992):

Assumption 4.1. In the model (4.1),

(i) The characteristic equation $|\Psi(z)(1-z)^2 - \Gamma z(1-z) - \Pi z| = 0$ has all roots equal to one or outside the unit circle;
(ii) $\Pi = \alpha\beta'$, where α and β are $p \times p_0$ matrices of full column rank;
(iii) $\alpha_\perp'\Gamma\beta_\perp = \xi\eta'$, where ξ and η are $(p - p_0) \times p_1$ matrices of full column rank;
(iv) $\mathrm{rank}(\alpha_2'[\Gamma\bar{\beta}\bar{\alpha}'\Gamma + \Psi]\beta_2) = p_2 := (p - p_0 - p_1)$, where $\alpha_2 := \alpha_\perp\xi_\perp$ and $\beta_2 := \beta_\perp\eta_\perp$.

Here and subsequently, we use for any $n \times m$ matrix of rank $m < n$ the notation $A_\perp$ to signify an $n \times (n - m)$ matrix of rank $(n - m)$ satisfying $A_\perp'A = 0$, and we use $\bar{A}$ to denote $A(A'A)^{-1}$. Subsequently we shall also use $\alpha_1 = \bar{\alpha}_\perp\xi$ and $\beta_1 = \bar{\beta}_\perp\eta$. Johansen (1997) considers two alternative parametri-

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

zations of (4.1), both of which satisfy conditions (ii) and (iii) almost everywhere in the parameter space (in which case (iv) is also generically satisfied). Here we consider a slightly different parametrization, with the same property.

Let c and d be known matrices, of order $p \times p_0$ and $(p_1 + p_2) \times p_1$, respectively, both of full column rank. We shall need the following assumption.

Assumption 4.2. In the model (4.1) under Assumption 4.1, $|c'\beta| \neq 0$ and $|d'\eta| \neq 0$.

The conditions of Assumption 4.2 will be satisfied by almost all parameter values of (α, β, Γ) satisfying conditions (ii) and (iii) of Assumption 4.1. We shall restrict the parameter space such that all points where these conditions and Assumption 4.2 are not satisfied are excluded. This allows the identifying restrictions (or normalizations) $c'\beta = I_{p_0}$ and $d'\eta = I_{p_1}$ to be imposed. Note that under Assumption 4.2, these restrictions are without loss of generality, because, e.g., $\alpha\beta' = \alpha\beta' c(\beta' c)^{-1}\beta' = \alpha_c \beta_c'$, where $\alpha_c = \alpha(\beta' c)$ is unrestricted and $\beta_c = \beta(c'\beta)^{-1}$ satisfies $c'\beta_c = I_{p_0}$. Thus, these restrictions are exactly identifying.

Define $c_1 = c_\perp d, c_2 = c_\perp d_\perp$ and note that $\bar{c}_1 = \bar{c}_\perp \bar{d}$ and $\bar{c}_2 = \bar{c}_\perp \bar{d}_\perp$. Next, let

$$X_{0t} = \bar{c}' X_t, \qquad X_{1t} = \bar{c}_1' X_t, \qquad X_{2t} = \bar{c}_2' X_t.$$

In the simplest example with $c = (I_{p_0}, 0)'$ and $d = (I_{p_1}, 0)'$, we have $(c, c_1, c_2) = (\bar{c}, \bar{c}_1, \bar{c}_2) = I_p$ and hence $X_t' = (X_{0t}', X_{1t}', X_{2t}')$.

THEOREM 4.1. *The model* (4.1) *under Assumptions* 4.1(ii) and (iii) *and* 4.2 *may be reparametrized as*

$$\begin{aligned}\Delta^2 X_t = {} & \alpha(X_{0,t-1} - A_1 X_{1,t-1} - A_2 X_{2,t-1} - A_3 \Delta X_{2,t-1}) \\ & + \gamma(\Delta X_{1,t-1} - A_4 \Delta X_{2,t-1}) + \lambda(\Delta X_{0,t-1} - A_1 \Delta X_{1,t-1} - A_2 \Delta X_{2,t-1}) \\ & + \sum_{i=1}^{l-2} \Psi_i \Delta^2 X_{t-i} + \varepsilon_t. \end{aligned} \tag{4.2}$$

The full parameter vector is $\theta := (\mathrm{vec}(A_1, A_2, A_3)', \mathrm{vec}(A_4)', \mathrm{vec}(\alpha, \gamma, \lambda, \Psi_1, \ldots, \Psi_{l-2})', \mathrm{vech}(\Omega)')'$, *and its parameter space is* $\Theta := \{\theta \in \mathbb{R}^k : \mathrm{rank}(\alpha, \gamma) = p_0 + p_1, \Omega > 0\}$, *where* $k := p_0(p_1 + 2p_2) + p_1 p_2 + p(2p_0 + p_1) + (l-2)p^2 + p(p+1)/2$.

Remark 4.1. The formulation (4.2) is a natural dynamic generalization of (3.1)–(3.3). It has three error-correction terms: the p_0 vector of lagged $CI(2,2)$ polynomial combinations, with adjustment matrix α, and two lagged differenced $CI(2,1)$ linear combinations, of dimension p_1 and p_0 and with adjustment matrices γ and λ, respectively. The model implies that, similarly to the previous section, we may take X_{2t} as the $I(2)$ common trends and $X_{bt} = (X_{1t}', X_{2t}')'$ to

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

represent the full vector of common trends. Formulated differently, Assumption 4.2 corresponds to the assumption that all (nonzero) linear combinations of X_{2t} are $I(2)$ and that all polynomial combinations of X_{bt} that include levels are $I(2)$ or $I(1)$.

Remark 4.2. The main attraction of the parametrization (4.2) over alternative forms found in the literature is that the parameters enter in an explicit but simple bilinear form. In fact, this immediately suggests a switching algorithm to maximize the log-likelihood, analogous to the one proposed for $I(1)$ systems in Boswijk (1995). This would amount to alternating between maximization over $(\alpha, \gamma, \lambda, \Psi_i, \Omega)$ given (A_1, A_2, A_3, A_4), which is solved by linear least-squares methods, and the converse maximization problem (which can also be shown to have a closed form solution). Observe that this algorithm does not impose the condition $\text{rank}(\alpha, \gamma) = p_0 + p_1$, but this condition will be satisfied by all matrices α and γ except for a set of measure zero, and hence it will be satisfied almost surely by the maximizer of the switching algorithm. Because the emphasis of this paper is on asymptotic inference, we will not pursue this avenue any further here. A possible disadvantage is that we seem to need the extra Assumption 4.2, but one can always find matrices c and d satisfying this assumption. In practice, one has to make sure that the corresponding choices of X_{1t} and X_{2t} satisfy the properties given in the previous remark. Note that the parameters A_i may not directly correspond to the parameters of interest but rather to "reduced form" functions thereof; e.g., we may be interested in a cointegrating vector involving all components of X_{0t} simultaneously, which would be defined as a suitable linear combination of β. We shall return to this issue in the next section.

Remark 4.3. One extension of the model (4.2) that will be relevant in practice is to allow for deterministic components such as a constant and trend. However, if, e.g., a constant term is included unrestrictedly in (4.2), then this will lead to a linear trend in the $I(1)$ linear combinations and a quadratic trend in the $I(2)$ directions, which may complicate the interpretability of the model and testing the cointegrating ranks p_0 and p_1 (see Paruolo, 1996). Jørgensen, Kongsted, and Rahbek (1999) propose an alternative model formulation in which all linear combinations ($I(0)$, $I(1)$ or $I(2)$) have at most a linear trend. In our parametrization, this would correspond to replacing (4.2) by

$$\begin{aligned}\Delta^2 X_t = {} & \alpha(X_{0,t-1} - A_1 X_{1,t-1} - A_2 X_{2,t-1} - A_3 \Delta X_{2,t-1} - a_1 - a_2 t) \\ & + \gamma(\Delta X_{1,t-1} - A_4 \Delta X_{2,t-1} - a_3) \\ & + \lambda(\Delta X_{0,t-1} - A_1 \Delta X_{1,t-1} - A_2 \Delta X_{2,t-1} - a_2) \\ & + \sum_{i=1}^{l-2} \Psi_i \Delta^2 X_{t-i} + \varepsilon_t, \end{aligned} \tag{4.3}$$

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

where a_1 and a_2 are p_0 vectors, and a_3 is a p_1 vector. Although we shall not consider this explicitly, the asymptotic properties for the model (4.2) derived subsequently can easily be generalized to (4.3).

LEMMA 4.1. *As* $n \to \infty$, *and under* $P_{\theta,n}$,

$$n^{-1/2}\begin{pmatrix} \sum_{i=1}^{[un]} \varepsilon_i \\ X_{1,[un]} - A_4 X_{2,[un]} \\ \Delta X_{2,[un]} \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} W(u) \\ W_1(u) \\ W_2(u) \end{pmatrix}$$

$$\equiv \begin{pmatrix} I_p \\ C_{11}\alpha_1' + C_{12}\alpha_2' \\ C_{22}\alpha_2' \end{pmatrix} W(u), \qquad u \in [0,1],$$

where $W(u) \equiv \mathrm{BM}(\Omega)$ *and where* C_{11} *and* C_{22} *are nonsingular matrices of order* $p_1 \times p_1$ *and* $p_2 \times p_2$, *respectively*, *whereas* C_{12} *is a* $p_1 \times p_2$ *matrix.*

We find that the Brownian motion corresponding to the $I(2)$ common trend (W_2) depends on $\alpha_2' W$ only, whereas $W_b = (W_1', W_2')'$, which corresponds to the full vector of common trends, depends on $\alpha_1' W$ and $\alpha_2' W$. Define $B_{\gamma\alpha} := \gamma'\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}$ and

$$W_\alpha := -\Omega_{\alpha\alpha}\alpha'\Omega^{-1}W \equiv \mathrm{BM}(\Omega_{\alpha\alpha}), \qquad \Omega_{\alpha\alpha} := (\alpha'\Omega^{-1}\alpha)^{-1},$$

$$W_{\gamma\cdot\alpha} := -\Omega_{\gamma\gamma\cdot\alpha}(\gamma' - B_{\gamma\alpha}\alpha')\Omega^{-1}W \equiv \mathrm{BM}(\Omega_{\gamma\gamma\cdot\alpha}),$$

$$\Omega_{\gamma\gamma\cdot\alpha} := (\gamma'\Omega^{-1}\gamma - B_{\gamma\alpha}\alpha'\Omega^{-1}\alpha B_{\gamma\alpha}')^{-1}.$$

Note that W_α and $W_{\gamma\cdot\alpha}$ are independent; furthermore, it can be derived that W_α is independent of W_b and that $W_{\gamma\cdot\alpha}$ is independent of W_2. Finally, define the normalizing matrix

$$D_n = \begin{pmatrix} n^{-1}I_{p_0p_1} & 0 & 0 & 0 & 0 \\ -n^{-1}(A_4' \otimes I_{p_0}) & n^{-2}I_{p_0p_2} & 0 & 0 & 0 \\ 0 & 0 & n^{-1}I_{p_0p_2} & -n^{-1}(I_{p_2} \otimes B_{\gamma\alpha}') & 0 \\ 0 & 0 & 0 & n^{-1}I_{p_1p_2} & 0 \\ 0 & 0 & 0 & 0 & n^{-1/2}I_{k_2} \end{pmatrix},$$

where $k_2 := k - p_0(p_1 + 2p_2) - p_1p_2$.

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

THEOREM 4.2. *In the model* (4.2) *under Assumptions* 4.1 *and* 4.2, *as* $n \to \infty$ *and under* $P_{\theta,n}$,

$$\Lambda_n(\theta + D_n\tau, \theta) \xrightarrow{\mathcal{L}} \Lambda(\tau) = \tau' S - \frac{1}{2}\tau' J \tau$$

$$= \sum_{i=0}^{2} \Lambda_i(\tau_i) = \sum_{i=0}^{2} \left(\tau_i' S_i - \frac{1}{2} z_i' J_i \tau_i \right), \tag{4.4}$$

where $\tau = (\tau_0', \tau_1', \tau_2')'$,

$$J = \begin{pmatrix} J_0 & 0 & 0 \\ 0 & J_1 & 0 \\ 0 & 0 & J_2 \end{pmatrix} \equiv \begin{pmatrix} \int FF' \otimes \Omega_{\alpha\alpha}^{-1} & 0 & 0 \\ 0 & \int W_2 W_2' \otimes \Omega_{\gamma\gamma\cdot\alpha}^{-1} & 0 \\ 0 & 0 & J_2 \end{pmatrix}, \tag{4.5}$$

with J_2 *some nonrandom positive definite matrix*, *and*

$$S = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \end{pmatrix} \equiv \begin{pmatrix} \operatorname{vec} \int (d\Omega_{\alpha\alpha}^{-1} W_\alpha) F' \\ \operatorname{vec} \int (d\Omega_{\gamma\gamma\cdot\alpha}^{-1} W_{\gamma\cdot\alpha}) W_2' \\ N(0, J_2) \end{pmatrix} \equiv \begin{pmatrix} J_0^{1/2} Z_0 \\ J_1^{1/2} Z_1 \\ J_2^{1/2} Z_2 \end{pmatrix}, \tag{4.6}$$

where $\mathcal{L}(F|Q_0) = \mathcal{L}((W_1', \overline{W}_2', W_2')')$ *and where* Z_0, Z_1, *and* Z_2 *are independent standard normal vectors*, *with* Z_0 *and* Z_2 *independent of* (J_0, J_1) *and* Z_1 *independent of* J_1.

Under $P_{\theta + D_n\tau, n}$, (4.4) *and* (4.5) *still hold but with* $\mathcal{L}(F|Q_\tau)$ *as defined in* (3.9), *so that*

$$\mathcal{L}(J|Q_\tau) = \mathcal{L}(H_\tau^{*\prime} J H_\tau^* | Q_0), \tag{4.7}$$

where $H_\tau^* := \operatorname{diag}(H_\tau, I_{k_2})$, *with* H_τ *as defined in* (3.11).

Remark 4.4. Theorem 4.2 shows that likelihood inference on $I(2)$ cointegration in VAR models has the same characteristics as in triangular systems. In fact, it is easily checked that the model (3.1)–(3.3) corresponds to (4.2) with $\alpha = (I_{p_0}, 0, 0)'$, $\gamma = (0, I_{p_1}, 0)'$, $\lambda = 0$, and $\Psi_i = 0, i = 1, \ldots, l-2$. The additional LAN part of the log-likelihood ratio $\Lambda_2(\tau_2)$ corresponds to the parameters that are assumed known in the triangular systems. Most interestingly, (4.7) implies the same lack of asymptotic ancillarity as before, so that the full log-likelihood ratio is not LAMN.

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

Remark 4.5. As in the previous section, the i.i.d. $N(0,\Omega)$ assumption can be relaxed to allow for i.i.d. $(0,\Omega)$ innovations with finite and positive definite information matrix $\mathcal{I}$. The analysis is quite analogous to Remark 3.8, and we shall not consider it explicitly here. In such cases the maximum likelihood algorithm suggested in Remark 4.2 will of course no longer be applicable.

5. INFERENCE ON PARAMETER RESTRICTIONS

In the previous sections we have seen that in general, inference on the cointegration parameters in $I(2)$ systems is not mixed normal, whether one considers triangular systems or vector autoregressions. However, it has also been shown that certain partial mixed normality results can be obtained. It is therefore of interest to establish for which functions of the parameters inference is mixed normal. Furthermore, it is possible that by imposing particular restrictions on the parameter space, a LAMN model results. Both issues are addressed in this section. Here, $\theta := (\mathrm{vec}(A_1,A_2,A_3)',\mathrm{vec}(A_4)')'$, the parameter vector in the triangular system. However, the results will also apply to the analysis of Section 4, as long as no restrictions are imposed on the remaining (short-run) parameters.

The original (unrestricted) parameter space is $\Theta = \mathbb{R}^k$, with $k = p_0(p_1 + 2p_2) + p_1p_2$. We consider restricting the parameter space to

$$\Theta_0 = \{\theta \in \mathbb{R}^k : g(\theta) = 0\} = \{\theta = h(\phi) : \phi \in \mathbb{R}^q\}, \tag{5.1}$$

with $g : \mathbb{R}^k \to \mathbb{R}^r$ and $h : \mathbb{R}^q \to \mathbb{R}^k$ vector functions, $1 \le r = k - q \le k$, and ϕ a q vector of free parameters. Note that we formulate the restrictions on θ both explicitly (as $g(\theta) = 0$) and implicitly (as $\theta = h(\phi)$). If $q = 0$, then θ is fully specified, so that ϕ is void. Throughout this section the true value is denoted by θ_0, which is assumed to lie in Θ_0, so that $g(\theta_0) = 0$ and $\theta_0 = h(\phi_0)$ for some true value ϕ_0.

In ergodic models, the asymptotic analysis proceeds via a Taylor series expansion, assuming that g and h are continuously differentiable with derivative matrices of rank r and q, respectively, in a neighborhood around the true value. In the present context, however, this is not sufficient because of different convergence rates in different directions, as embodied in D_n. For example, suppose that $\theta = (\theta_1,\theta_2)'$ and consider the function $g(\theta) = \theta_1^2 + \theta_2$, so that $G(\theta) = \partial g(\theta)/\partial\theta' = (2\theta_1,1)$. Note that G has rank one for all parameter values, including $\theta_1 = 0$, which we take to be the true value. If both $\hat\theta_{1n}$ and $\hat\theta_{2n}$ are root-n consistent, then a Taylor series expansion will correctly indicate $\sqrt{n}(g(\hat\theta_n) - g(\theta_0)) = \sqrt{n}(\hat\theta_{2n} - \theta_{2,0}) + o_p(1)$. However, when $\hat\theta_{1n} = O_p(n^{-1})$ and $(\hat\theta_{2n} - \theta_{2,0}) = O_p(n^{-2})$, then $n^2(g(\hat\theta_n) - g(\theta_0)) = (n\hat\theta_{1n})^2 + n^2(\hat\theta_{2n} - \theta_{2,0})$, whereas a (first-order) Taylor series expansion would incorrectly suggest $n^2(g(\hat\theta_n) - g(\theta_0)) = n^2(\hat\theta_{2n} - \theta_{2,0}) + o_p(1)$. Therefore, a full row rank of G (or a full

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

column rank of $\partial h(\phi)/\partial \phi'$) is not sufficient for the validity of first-order Taylor series approximation. Instead, we require the following assumption.

Assumption 5.1. The function h is continuously differentiable, with derivative $H(\phi) = \partial h(\phi)/\partial \phi'$. Furthermore, there exist a $q \times q$ normalizing matrix C_n and a $k \times q$ matrix K of full column rank, such that for all $\delta > 0$,

$$\sup_{\|\psi\| \leq \delta} \|D_n^{-1} H(\phi_0 + C_n \psi) C_n - K\| \to 0.$$

The assumption allows the function h to be locally linearized: for a sequence $\phi_n = \phi_0 + C_n \psi$, it implies via the mean value theorem

$$D_n^{-1}(h(\phi_n) - h(\phi_0)) = K C_n^{-1}(\phi_n - \phi_0) + o(1) = K\psi + o(1).$$

Let $\tilde{\Lambda}_n(\phi_1, \phi_2) := \Lambda(h(\phi_1), h(\phi_2))$, likelihood ratio of the restricted model. Let $\tilde{\phi}_n$ denote the MLE for this model and define the likelihood ratio test statistic for $\mathcal{H}_0 : \theta \in \Theta_0$,

$$LR_n := -2(\log L_n(h(\tilde{\phi}_n)) - \log L_n(\hat{\theta}_n)) = -2\Lambda_n(h(\tilde{\phi}_n), \hat{\theta}_n).$$

Define also $k_0 := p_0(p_1 + 2p_2)$ and $k_1 := p_1 p_2$, the dimensions of τ_0 and τ_1, respectively.

THEOREM 5.1. *Consider the model* (3.1)–(3.3) *under the restrictions* (5.1) *satisfying Assumption* 5.1. *As* $n \to \infty$ *and under* $P_{\theta_0, n}$, $\theta_0 \in \Theta_0$,

$$\tilde{\Lambda}_n(\phi_0 + C_n \psi, \phi_0) \xrightarrow{\mathcal{L}} \tilde{\Lambda}(\psi) = \psi' \tilde{S} - \tfrac{1}{2} \psi' \tilde{J} \psi, \tag{5.2}$$

where $\tilde{S} = K'S$ *and* $\tilde{J} = K'JK$, *with* S *and* J *as given in Theorem* 3.1. *Furthermore*,

$$LR_n \xrightarrow{\mathcal{L}} S' J^{-1} K_\perp (K_\perp' J^{-1} K_\perp)^{-1} K_\perp' J^{-1} S. \tag{5.3}$$

A sufficient condition for $LR_n \xrightarrow{\mathcal{L}} \chi^2(r)$ *is that* K *may be expressed as*

$$K = \begin{pmatrix} K_{00} & 0 \\ 0 & K_{11} \end{pmatrix}, \tag{5.4}$$

where K_{ii} *are* $k_i \times q_i$ *matrices*, $i = 0,1$, *with* $0 \leq q_0 = q - q_1 \leq q$.

Remark 5.1. Note that C_n is not unique, because it may always be replaced by $C_n Q$ for some arbitrary nonsingular matrix Q. Because this will change K to KQ, the question is whether (5.4) is satisfied for at least one choice of C_n. Observe also that this is a *sufficient* condition; its *necessity* has not been proved.

Remark 5.2. As an application of Theorem 5.1, consider the problem of testing the null hypothesis $\mathcal{H}_0 : A_3 = 0$, corresponding to the case where the poly-

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

nomial cointegrating relation reduces to a $CI(2,2)$ linear combination. Letting $\phi = (\mathrm{vec}(A_1, A_2)', \mathrm{vec}(A_4)')'$, the function $h(\phi)$ is linear with derivative

$$H(\phi) = \begin{pmatrix} I_{p_0(p_1+p_2)} & 0 \\ 0 & 0 \\ 0 & I_{p_1 p_2} \end{pmatrix}.$$

The block-diagonality of H may suggest that (5.4) is already satisfied. However, because D_n is not block diagonal, it can be shown that for some suitably chosen C_n,

$$D_n^{-1} H(\phi) C_n = K = \begin{pmatrix} I_{p_0(p_1+p_2)} & 0 \\ 0 & (I_{p_2} \otimes \Omega^{10}(\Omega^{00})^{-1}) \\ 0 & I_{p_1 p_2} \end{pmatrix}.$$

This satisfies the conditions of Theorem 5.1 either if $p_1 = 0$ or if $p_1 > 0$ and $\Omega^{10} = 0$, which in turn corresponds to the restriction $\Omega_{01} - \Omega_{02}\Omega_{22}^{-1}\Omega_{21} = 0$; i.e., u_{0t} and u_{1t} are uncorrelated, conditional upon u_{2t}. The corresponding conditions for mixed normality in the VAR model were obtained by Paruolo (1995).

Remark 5.3. To see whether the restricted likelihood ratio $\tilde{\Lambda}$ is LAMN, we simply would have to check the asymptotic local ancillarity condition, i.e., to show that the distribution of $\tilde{J}$ does not vary with ψ. An obvious example of this is when in the restricted model A_4 is fully specified. The lack of ancillarity noted in Theorem 3.1 will then completely vanish, because it was only caused by variation in $\mathcal{T}_1$, which corresponds to local variation in A_4. Another example would be where A_1 is fully specified, because this corresponds to removing the first block $W_1 + \mathcal{T}_1 \overline{W}_2$ from F in (3.9).

Remark 5.4. The three cases analyzed by Kitamura (1995) are also examples of restricted $I(2)$ models such that LAMN obtains. His Case A corresponds, in our notation, to a restricted version of the model with $A_3 = 0, A_4 = 0$, which given the previous remark leads to a LAMN model. Case B is obtained, after some rewriting (see the Appendix), as a restricted version of a model in which X_{1t} is partitioned as $(X_{11,t}', X_{12,t}')'$ and correspondingly

$$A_1 = (0, A_{12}), \quad A_4 = \begin{pmatrix} A_{41} \\ 0 \end{pmatrix}.$$

Thus, the distribution of the part of X_{1t} that appears as a regressor in the first p_0 equations does not vary with A_4, which again eliminates the source of the lack of ancillarity. Finally, Case C of Kitamura (1995) is an extension of Case B allowing for polynomial cointegration. However, exactly the same restrictions are imposed on A_1 and A_4 as in Case B, which again yields ancillarity of

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

the information matrix and hence mixed normality of the likelihood ratio; for details see the Appendix.

Remark 5.5. Theorem 5.1 implies that inference on a hypothesis of the form $\mathcal{H}_0: A_4 = A_{4,0}$ is mixed normal. Furthermore, Remark 5.3 implies that the model under this restriction is LAMN. This suggests that in practice, one may test hypotheses on, e.g., A_3 in two steps: first restrict A_4 to a value $A_{4,0}$ that is not rejected by a likelihood ratio test (using χ^2 critical values) and next test the hypothesis of interest on A_3 in the restricted model, again using χ^2 critical values. Of course, this procedure would suffer from the usual drawbacks of multistep testing procedures. Note that most applications of $I(2)$ cointegration that may be found in the literature involve prices and nominal variables being $I(2)$, whereas the corresponding real variables or price differentials are $I(1)$. In such instances one has rather strong prior information about the value of A_4, and the present analysis suggests that inference is considerably simplified by testing and (if not rejected) imposing these restrictions.

6. DISCUSSION

In this paper we have shown that, in contrast with $I(1)$ models, imposing the unit roots in $I(2)$ systems is not sufficient to yield ancillarity of the observed information and hence mixed normality of the log-likelihood ratio and of the MLE. We have noted that in this particular case, the fact that the likelihood may be approximated locally by a quadratic function does not imply mixed normality almost everywhere in the parameter space. Finally, we have given a characterization theorem to determine whether inference on a particular parameter of interest is mixed normal, and we have shown that in general, the so-called polynomial cointegration parameter does not belong to this class. However, a two-step procedure to circumvent this problem was also suggested, where in the first step the parameter A_4 is restricted to an economically plausible value, after testing for this restriction using a χ^2 likelihood ratio test, and in the second step hypotheses on the remaining parameters are tested in the restricted model (with A_4 known), again using χ^2 likelihood ratio tests.

To compare results previously obtained in triangular systems and vector autoregressions, we have proposed a new parametrization of the $I(2)$ VAR model. The parametrization requires some a priori information, in particular information on which variables may represent the common trends (X_{1t} and X_{2t}) and, of these, which variables are $I(2)$ with no $I(1)$ linear combinations (X_{2t}). However, such information is always required to obtain an interpretable, identified parametrization. Although the emphasis in the present paper is on asymptotic inference rather than practical statistical analysis, we note that the proposed parametrization suggests an easy switching algorithm to obtain the MLE's and various likelihood ratio statistics. For example, the analysis of weak exogeneity in $I(2)$ systems (see Paruolo and Rahbek, 1999) will lead to simple and easily testable restrictions on the adjustment parameters of the model.

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

REFERENCES

Boswijk, H.P. (1995) Identifiability of Cointegrated Systems. Discussion paper TI 95–78, Tinbergen Institute.
Granger, C.W.J. & T.-H. Lee (1990) Multicointegration. *Advances in Econometrics* 8, 71–84.
Haldrup, N. (1998) An econometric analysis of $I(2)$ variables. *Journal of Economic Surveys* 12, 595–650.
Hansen, B.E. (1992) Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.
Jeganathan, P. (1995) Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–887.
Jeganathan, P. (1997) On asymptotic inference in linear cointegrated time series systems. *Econometric Theory* 13, 692–745.
Johansen, S. (1992) A representation of vector autoregressive processes integrated of order 2. *Econometric Theory* 8, 188–201.
Johansen, S. (1995a) A statistical analysis of cointegration for $I(2)$ variables. *Econometric Theory* 11, 25–59.
Johansen, S. (1995b) The role of ancillarity in inference for non-stationary variables. *Economic Journal* 105, 302–320.
Johansen, S. (1995c) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford : Oxford University Press.
Johansen, S. (1997) Likelihood analysis of the $I(2)$ model. *Scandinavian Journal of Statistics* 24, 433–462.
Jørgensen, C., H.C. Kongsted, & A.C. Rahbek (1999) Trend-stationarity in the $I(2)$ cointegration model. *Journal of Econometrics* 90, 265–289.
Kitamura, Y. (1995) Estimation of cointegrated systems with $I(2)$ processes. *Econometric Theory* 11, 1–24.
Le Cam, L. & G.L. Yang (1990) *Asymptotics in Statistics. Some Basic Concepts*. New York: Springer-Verlag.
Magnus, J.R. & H. Neudecker (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester: Wiley.
Paruolo, P. (1995) *Topics in Statistical Inference for Vector Autoregressive Processes Integrated of Order 1 and 2*. Ph.D. Thesis, University of Copenhagen.
Paruolo, P. (1996) On the determination of integration indices in $I(2)$ systems. *Journal of Econometrics* 72, 313–356.
Paruolo, P. & A.C. Rahbek (1999) Weak exogeneity in $I(2)$ VAR systems. *Journal of Econometrics* 93, 281–308.
Phillips, P.C.B. (1989) Partially identified econometric models. *Econometric Theory* 5, 181–240.
Phillips, P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
Sims, C.A., J.H. Stock, & M.W. Watson (1990) Inference in linear time series models with some unit roots. *Econometrica* 58, 113–144.
Stock, J.H. & M.W. Watson (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.

APPENDIX

Proof of Theorem 3.1. The log-likelihood function is given by

$$\log L_n(\theta) = -\frac{np}{2}\log 2\pi - \frac{n}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{n}\varepsilon_t(\theta)'\Omega^{-1}\varepsilon_t(\theta),$$

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

with

$$\varepsilon_t(\theta) = \begin{pmatrix} \varepsilon_{0t}(\theta) \\ \varepsilon_{1t}(\theta) \\ \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} X_{0t} - A_1 X_{1t} - A_2 X_{2t} - A_3 \Delta X_{2t} \\ \Delta X_{1t} - A_4 \Delta X_{2t} \\ \Delta^2 X_{2t} \end{pmatrix}.$$

Note that $\varepsilon_t(\theta) = \varepsilon_t$ under $P_{\theta,n}$. Because $\varepsilon_t(\theta)$ is linear in θ, the log-likelihood function is exactly quadratic in θ, so that the $o_p(1)$ term in (2.5) vanishes, and it suffices to determine the limiting behavior of the first and second derivatives of $\log L_n(\theta)$. Define

$$Z_t(\theta) = \begin{pmatrix} X_{1t} - A_4 X_{2t} \\ n^{-1} X_{2t} \\ \Delta X_{2t} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{t} \varepsilon_{1i}(\theta) \\ n^{-1} \sum_{i=1}^{t} \sum_{j=1}^{i} \varepsilon_{2j} \\ \sum_{i=1}^{t} \varepsilon_{2i} \end{pmatrix}.$$

Using well-known properties of the partitioned inverse of Ω, it is easily checked that

$$D_n(\theta)' \frac{\partial L_n(\theta)}{\partial \theta} = \begin{pmatrix} n^{-1} \sum_{t=1}^{n} (Z_t(\theta) \otimes \Omega_{00\cdot b}^{-1}) \varepsilon_{0\cdot b,t}(\theta) \\ n^{-1} \sum_{t=1}^{n} (\Delta X_{2t} \otimes \Omega_{11\cdot 2}^{-1}) \varepsilon_{1\cdot 2,t}(\theta) \end{pmatrix},$$

where $\varepsilon_{0\cdot b,t}(\theta) := \varepsilon_{0t}(\theta) - \Omega_{0b}\Omega_{bb}^{-1}\varepsilon_{bt}(\theta)$ and $\varepsilon_{1\cdot 2,t}(\theta) := \varepsilon_{1t}(\theta) - \Omega_{12}\Omega_{22}^{-1}\varepsilon_{2t}$. Similarly, we find

$$D_n(\theta)' \frac{\partial^2 L_n(\theta)}{\partial \theta \partial \theta'} D_n(\theta) = \begin{pmatrix} n^{-2} \sum_{t=1}^{n} (Z_t(\theta) Z_t(\theta)' \otimes \Omega_{00\cdot b}^{-1}) & 0 \\ 0 & n^{-2} \sum_{t=1}^{n} (\Delta X_{2t} \Delta X_{2t}' \otimes \Omega_{11\cdot 2}^{-1}) \end{pmatrix}. \tag{A.1}$$

Under $P_{\theta + D_n(\theta)\tau, n}$, it is the case that $\Delta X_{1t} = (A_4 + n^{-1}\mathcal{T}_1)\Delta X_{2t} = \varepsilon_{1t}$, so that

$$\varepsilon_{1t}(\theta) = \Delta X_{1t} - A_4 \Delta X_{2t} = \varepsilon_{1t} + \mathcal{T}_1 n^{-1} \sum_{i=1}^{t} \varepsilon_{2i}.$$

This implies, still under $P_{\theta + D_n(\theta)\tau, n}$, that

$$n^{-1/2} Z_{[un]}(\theta) \xrightarrow{\mathcal{L}} F(u),$$

with $\mathcal{L}(F|Q_\tau)$ given in (3.9). Furthermore, $n^{-1/2}\Delta X_{2,[un]} \xrightarrow{\mathcal{L}} W_2(u)$ from (3.4). Combining this with (A.1), we obtain (3.7), (3.10), and (3.11) by the continuous mapping theorem. Furthermore, the stochastic integral expression in (3.8) follows from well-

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

known results on weak convergence to matrix stochastic integrals (see Hansen, 1992). Independence of $W_{0\cdot b}$ and W_b implies that $Z_0 = J_0^{-1/2} S_0$ is standard normal, independently of W_b (and hence (J_0, J_1)); similarly, $Z_1 = J_1^{-1/2} S_1$ is standard normal, independently of J_1, because $W_{1\cdot 2}$ is independent of W_2. ■

Proof of Theorem 4.1. The starting point is that we may impose, under Assumption 4.2, the restrictions

$$c'\beta = I_{p_0}, \quad d'\eta = I_{p_1}, \quad \bar{c}'_\perp \beta_\perp = I_{p_1+p_2}, \quad \bar{d}'_\perp \eta_\perp = I_{p_2}.$$

These restrictions are allowed because β, η, $\beta_\perp$, and $\eta_\perp$ are only defined up to the space spanned by their columns; imposing these restrictions leads to a unique choice of a basis for these column spaces.

We shall first show that the model (4.1) under Assumption 4.1(ii) and (iii) may be rewritten as (4.2). Using $I_p = c\bar{c}' + c_1\bar{c}'_1 + c_2\bar{c}'_2$, we have

$$\alpha\beta' X_{t-1} = \alpha(\beta' c, \beta' c_1, \beta' c_2)(\bar{c}, \bar{c}_1, \bar{c}_2)' X_{t-1} = \alpha(X_{0,t-1} - A_1 X_{1,t-1} - A_2 X_{2,t-1}),$$

with $A_1 = -\beta' c_1$ and $A_2 = -\beta' c_2$.

Next, because $I_p = c(\beta' c)^{-1}\beta' + \beta_\perp(\bar{c}'_\perp \beta_\perp)^{-1}\bar{c}'_\perp = c\beta' + \beta_\perp \bar{c}'_\perp$, we have

$$\Gamma = \Gamma c\beta' + \Gamma\beta_\perp \bar{c}'_\perp = \lambda\beta' + \Gamma\beta_\perp \bar{c}'_\perp,$$

where $\lambda = \Gamma c$. Using $I_p = \alpha\bar{\alpha}' + \bar{\alpha}_\perp \alpha'_\perp$, rewrite the second term as

$$\begin{aligned}\Gamma\beta_\perp \bar{c}'_\perp &= \alpha\bar{\alpha}'\Gamma\beta_\perp \bar{c}'_\perp + \bar{\alpha}_\perp(\alpha'_\perp \Gamma\beta_\perp)\bar{c}'_\perp \\ &= \alpha\bar{\alpha}'\Gamma\beta_\perp \bar{c}'_\perp + \bar{\alpha}_\perp \xi\eta'\bar{c}'_\perp \\ &= \alpha\bar{\alpha}'\Gamma\beta_\perp \bar{c}'_\perp + \alpha_1\zeta',\end{aligned}$$

where $\zeta = \bar{c}_\perp \eta$. Because $I_{p_1+p_2} = d(\eta' d)^{-1}\eta' + \eta_\perp(\bar{d}'_\perp \eta_\perp)^{-1}\bar{d}'_\perp = d\eta' + \eta_\perp \bar{d}'_\perp$, this reduces to

$$\begin{aligned}\alpha\bar{\alpha}'\Gamma\beta_\perp \bar{c}'_\perp + \alpha_1\zeta' &= \alpha\bar{\alpha}'\Gamma\beta_\perp \eta_\perp \bar{d}'_\perp \bar{c}'_\perp + \alpha\bar{\alpha}'\Gamma\beta_\perp d\eta'\bar{c}'_\perp + \alpha_1\zeta' \\ &= \alpha(\bar{\alpha}'\Gamma\beta_2)\bar{c}'_2 + \alpha\bar{\alpha}'\Gamma\beta_\perp d\zeta' + \alpha_1\zeta' \\ &= -\alpha A_3 \bar{c}'_2 + \gamma\zeta',\end{aligned}$$

where $A_3 = -\bar{\alpha}'\Gamma\beta_2$ and $\gamma = \alpha\bar{\alpha}'\Gamma\beta_\perp d + \alpha_1$. Thus we find

$$\Gamma\Delta X_{t-1} = \lambda(\Delta X_{0,t-1} - A_1\Delta X_{1,t-1} - A_2\Delta X_{2,t-1}) - \alpha A_3 \Delta X_{2,t-1} + \gamma\zeta'\Delta X_{t-1}.$$

Finally, using $I_{p_1+p_2} = d\bar{d}' + d_\perp \bar{d}'_\perp$, we have

$$\zeta'\Delta X_{t-1} = \eta' d\bar{d}'\bar{c}'_\perp \Delta X_{1,t-1} + \eta' d_\perp \bar{d}'_\perp \bar{c}'_\perp \Delta X_{t-1} = \Delta X_{1,t-1} - A_4 \Delta X_{2,t-1},$$

where $A_4 = -\eta' d_\perp$.

In summary, we have $A_1 = -\beta' c_1$, $A_2 = -\beta' c_2$, $A_3 = -\bar{\alpha}'\Gamma\beta_2$, $A_4 = -\eta' d_\perp$, $\lambda = \Gamma c$, and $\gamma = \alpha\bar{\alpha}'\Gamma\beta_\perp d + \alpha_1$. What remains to be shown is that for all $(A_1, A_2, A_3, A_4, \alpha, \gamma, \lambda)$ such that the matrix (α, γ) has full column rank $p_0 + p_1$, the corresponding matrices α

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

and β have full column rank and $\alpha_\perp'\Gamma\beta_\perp$ has rank p_1, as required by Assumption 4.1. For α and $\beta = (\bar{c}, \bar{c}_1, \bar{c}_2)(I_{p_0}, -A_1, -A_2)'$ this is obvious. Furthermore, we have

$$\alpha_\perp'\Gamma\beta_\perp = \alpha_\perp'\gamma\zeta'\beta_\perp = \alpha_\perp'\gamma(I_{p_1}, -A_4)(\bar{d}, \bar{d}_\perp)'.$$

This matrix has rank p_1 if $\alpha_\perp'\gamma$ has rank p_1, and this is implied by the condition that (α, γ) has full column rank. ■

Proof of Lemma 4.1. Define $U_{0t} := X_{0t} - A_1X_{1t} - A_2X_{2t} - A_3\Delta X_{2t}$, $U_{1t} := \Delta X_{1t} - A_4\Delta X_{2t}$, and $U_{2t} := \Delta^2 X_{2t}$. We shall express the model (4.2) as a vector autoregression in $U_t := (U_{0t}', U_{1t}', U_{2t}')'$. Let $Q = (c, c_1, c_2)$ and hence $Q^{-1} = (\bar{c}, \bar{c}_1, \bar{c}_2)'$, so that $X_t = Q(X_{0t}', X_{1t}', X_{2t}')'$. Note that

$$\Delta^2 X_{0t} = A_1\Delta U_{1t} + (A_2 + A_1A_4)U_{2t} + A_3\Delta U_{2t} + \Delta^2 U_{0t}$$

and $\Delta^2 X_{1t} = A_4U_{2t} + \Delta U_{1t}$. Therefore, using the lag operator L, (4.2) can be expressed as

$$\Psi(L)Q\begin{pmatrix} I_{p_0}\Delta^2 & A_1\Delta & A_3\Delta + (A_2 + A_1A_4) \\ 0 & I_{p_1}\Delta & A_4 \\ 0 & 0 & I_{p_2} \end{pmatrix} U_t = (\alpha + \lambda\Delta, \gamma, \lambda A_3)LU_t + \varepsilon_t$$

or $\Phi(L)U_t = \varepsilon_t$. By Assumption 4.1, $|\Phi(z)| = 0$ has all its roots equal to one or outside the unit circle; hence if rank $\Phi(1) = p$, $\Phi(L)$ is invertible and U_t is stationary. Define the nonsingular matrix $P = (\alpha, \alpha_1, \alpha_2)$. We find

$$P'\Phi(1) = \begin{pmatrix} -\alpha'\alpha & -\alpha'\gamma & \alpha'(\Psi\beta_2 - \lambda A_3) \\ 0 & -\alpha_1'\alpha_1 & \alpha_1'(\Psi\beta_2 - \lambda A_3) \\ 0 & 0 & \alpha_2'(\Psi\beta_2 - \lambda A_3) \end{pmatrix},$$

where we have used $\alpha_1'\gamma = \alpha_1'\alpha_1$ and $\beta_2' = ([A_2 + A_1A_4]', A_4', I_{p_2})Q'$. It is easily checked that

$$\alpha_2'(\Gamma\bar{\beta}\bar{\alpha}'\Gamma + \Psi)\beta_2 = \alpha_2'(\Psi\beta_2 - \lambda A_3),$$

which explains the role of condition (iv) of Assumption 4.1: U_t is stationary if and only if this condition is satisfied. In that case

$$U_t = [P'\Phi(L)]^{-1}P'\varepsilon_t = C(L)P'\varepsilon_t = CP'\varepsilon_t + C^*(L)\Delta P'\varepsilon_t,$$

where $C(L) = [P'\Phi(L)]^{-1}$, $C = C(1)$, and $C^*(L)$ is a matrix power series such that $C(L) = C + C^*(L)\Delta$. Because the matrix $P'\Phi(1)$ is upper block triangular, so is C:

$$C = [P'\Phi(1)]^{-1} = \begin{pmatrix} C_{00} & C_{01} & C_{02} \\ 0 & C_{11} & C_{12} \\ 0 & 0 & C_{22} \end{pmatrix}.$$

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

This implies

$$n^{-1/2}\sum_{i=1}^{[un]} U_{1i} = (C_{11}\alpha_1' + C_{12}\alpha_2')\sum_{i=1}^{[un]}\varepsilon_i + o_p(1),$$

$$n^{-1/2}\sum_{i=1}^{[un]} U_{2i} = C_{22}\alpha_2'\sum_{i=1}^{[un]}\varepsilon_i + o_p(1)$$

and hence Lemma 4.1. ■

Proof of Theorem 4.2. We use the fact that the log-likelihood is twice continuously differentiable, which allows us to focus on the normalized first and second derivatives (see (2.5)):

$$S_n = D_n'\frac{\partial \log L_n(\theta)}{\partial\theta}, \qquad J_n = -D_n'\frac{\partial^2 \log L_n(\theta)}{\partial\theta\partial\theta'}D_n.$$

Define $\varepsilon_t(\theta)$ as the residual vector (which equals ε_t under $P_{\theta,n}$) and the vector

$$Z_t(\theta) = \begin{pmatrix} Z_{1t}(\theta) \\ Z_{2t} \\ Z_{3t} \end{pmatrix} = \begin{pmatrix} X_{1,t-1} - A_4 X_{2,t-1} \\ n^{-1}X_{2,t-1} \\ \Delta X_{2,t-1} \end{pmatrix}.$$

Partition $S_n = (S_{0n}', S_{1n}', S_{2n}')'$. Using straightforward algebra, it can be shown that

$$S_{0n} = -\frac{1}{n}\sum_{t=1}^{n} Z_t(\theta)\otimes(\alpha'\Omega^{-1}\varepsilon_t(\theta)) - \frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix}\Delta Z_{1t}(\theta)\\ \Delta Z_{2t}\\ 0\end{pmatrix}\otimes(\lambda'\Omega^{-1}\varepsilon_t(\theta)),$$

$$S_{1n} = -\frac{1}{n}\sum_{t=1}^{n}\Delta X_{2,t-1}\otimes[\gamma'\Omega^{-1}\varepsilon_t(\theta) - B_{\gamma\alpha}\alpha'\Omega^{-1}\varepsilon_t(\theta)].$$

The second term in S_{0n} (involving the first difference of Z_t) is of lower order in probability than the first term and hence asymptotically negligible. Similarly, we find

$$J_n = \begin{pmatrix} J_{0n} & 0 & 0 \\ 0 & J_{1n} & 0 \\ 0 & 0 & J_{2n}\end{pmatrix} + o_p(1),$$

where

$$J_{0n} = \frac{1}{n^2}\sum_{t=1}^{n} Z_t(\theta)Z_t(\theta)'\otimes\Omega_{\alpha\alpha}^{-1} + o_p(1),$$

$$J_{1n} = \frac{1}{n^2}\sum_{t=1}^{n}\Delta X_{2t}\Delta X_{2t}'\otimes\Omega_{\gamma\gamma.\alpha}^{-1}.$$

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

We do not give an explicit expression for S_{2n} and J_{2n}, but from standard asymptotics for ergodic processes, we have

$$(S_{2n}, J_{2n}) \xrightarrow{\mathcal{L}} (J_2^{1/2} Z_2, J_2),$$

where J_2 is a nonrandom positive definite matrix. The block-diagonality of J, and also the asymptotic independence of Z_2 from S_{0n}, S_{1n}, J_{0n}, and J_{1n}, is a standard result in asymptotics for integrated processes (for a discussion in the present context, see Johansen, 1997). The remaining convergence results follow immediately from Lemma 4.1 and the continuous mapping theorem. Mixed normality follows because W_α is uncorrelated with and hence independent of W_1 and W_2 and $W_{\gamma.\alpha}$ is independent of W_2. Furthermore, Z_0 is independent of Z_1 and J_1, because W_α is independent of $W_{\gamma.\alpha}$.

The remainder of the proof is analogous to the proof of Theorem 3.1. Note that under $P_{\theta+D_n\tau,n}$, $Z_{1t}(\theta) = X_{1,t-1} - A_4 X_{2,t-1} = \sum_{s=1}^{t-1} U_{1s} + n^{-1}\mathcal{T}_1 \sum_{s=1}^{t-1}\sum_{j=1}^{s} U_{2j}$. ■

Proof of Theorem 5.1. Defining the likelihood function $\tilde{L}_n(\phi) := L_n(h(\phi))$, we have, analogously to (2.5),

$$\tilde{\Lambda}_n(\phi_0 + C_n\psi, \phi_0) = \psi'\tilde{S}_n + \tfrac{1}{2}\psi'\tilde{J}_n\psi + o_p(1),$$

where

$$\tilde{S}_n = C_n' \frac{\partial \tilde{L}_n(\phi_0)}{\partial \phi} = C_n' H(\phi)' D_n^{-1\prime} S_n,$$

$$\tilde{J}_n = C_n' \frac{\partial^2 \tilde{L}_n(\phi_0)}{\partial \phi \partial \phi'} C_n = C_n' H(\phi)' D_n^{-1\prime} J_n D_n^{-1} H(\phi) C_n,$$

so that, by Assumption 5.1,

$$(\tilde{S}_n, \tilde{J}_n) \xrightarrow{\mathcal{L}} (K'S, K'JK).$$

This proves (5.2). Because $L_n(\theta_0) = \tilde{L}_n(\phi_0)$, we may write LR_n as

$$\begin{aligned} LR_n &= -2\Big(\max_{\phi\in\mathbb{R}^q} \tilde{\Lambda}_n(\phi, \phi_0) - \max_{\theta\in\mathbb{R}^k} \Lambda_n(\theta, \theta_0)\Big) \\ &= -2\Big(\max_{\psi\in\mathbb{R}^q} \tilde{\Lambda}_n(\phi_0 + C_n\psi, \phi_0) - \max_{\tau\in\mathbb{R}^k} \Lambda_n(\theta_0 + D_n\tau, \theta_0)\Big). \end{aligned}$$

Assuming that the MLE's are asymptotically centering, so that the maximizers of the log-likelihood ratios $\tilde{\Lambda}_n(\phi_0 + C_n\psi, \phi_0)$ and $\Lambda_n(\theta_0 + D_n\tau, \theta_0)$ converge in distribution to the maximizers of the limiting log-likelihood ratios $\tilde{\Lambda}(\psi)$ and $\Lambda(\tau)$, respectively, we find

$$LR_n \xrightarrow{\mathcal{L}} S'J^{-1}S - S'K(K'JK)^{-1}K'S = S'J^{-1}K_\perp(K_\perp' J^{-1} K_\perp)^{-1} K_\perp' J^{-1} S.$$

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

Condition (5.4) implies $K_\perp = \mathrm{diag}(K_{00,\perp}, K_{11,\perp})$, and combining this with $J = \mathrm{diag}(J_0, J_1)$ and $S_i = J_i^{1/2} Z_i, i = 0,1$, we obtain

$$(K_\perp' J^{-1} K_\perp)^{-1/2} K_\perp' J^{-1} S = \begin{pmatrix} (K_{00,\perp} J_0^{-1} K_{00,\perp}')^{-1/2} K_{00,\perp} J_0^{-1/2} Z_0 \\ (K_{11,\perp} J_1^{-1} K_{11,\perp}')^{-1/2} K_{11,\perp} J_1^{-1/2} Z_1 \end{pmatrix} = \begin{pmatrix} Z_0^* \\ Z_1^* \end{pmatrix}.$$

Using the independence between Z_i and J_j, $j \geq i$, it follows that $Z^* = (Z_0^{*\prime}, Z_1^{*\prime})'$ is a standard normal vector, and hence $LR_n \xrightarrow{\mathcal{L}} Z^{*\prime} Z^* \equiv \chi^2(r)$. ■

Proof of Remark 5.4. Case A of Kitamura (1995) is

$$y_{1t} = B_{12} y_{2t} + u_{1t},$$

$$y_{at} = B_{a3} y_{3t} + u_{at},$$

$$\Delta y_{2t} = u_{2t},$$

$$\Delta^2 y_{3t} = u_{3t}.$$

Defining $X_{0t} = (y_{1t}', y_{at}')'$, $X_{1t} = y_{2t}$, and $X_{2t} = y_{3t}$ and similarly $\varepsilon_{0t} = (u_{1t}', u_{at}')'$, $\varepsilon_{1t} = u_{2t}$, and $\varepsilon_{2t} = u_{3t}$, the model corresponds to (3.1)–(3.3), with

$$A_1 = \begin{pmatrix} B_{12} \\ 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 \\ B_{a3} \end{pmatrix}, \quad A_3 = 0, \quad A_4 = 0.$$

Kitamura's Case B is

$$y_{1t}^{(1)} = B_{12}^{(1)} y_{2t} + u_{1t}^{(1)},$$

$$\bar{y}_{bt}^{(1)} = B_{12}^{(2)} y_{2t} + u_{1t}^{(2)},$$

$$y_{at} = B_{a3} y_{3t} + u_{at},$$

$$y_{bt}^{(1)} = B_{b3}^{(1)} y_{3t} + \bar{y}_{bt}^{(1)},$$

$$y_{bt}^{(2)} = B_{b3}^{(2)} y_{3t} + \bar{y}_{bt}^{(2)},$$

$$\Delta \bar{y}_{bt}^{(2)} = u_{bt},$$

$$\Delta y_{2t} = u_{2t},$$

$$\Delta^2 y_{3t} = u_{3t}.$$

After substitution of the unobserved components $\bar{y}_{bt}^{(1)}$ and $\bar{y}_{bt}^{(2)}$, this leads to (3.1)–(3.3) with $X_{0t} = (y_{1t}^{(1)\prime}, y_{at}', y_{bt}^{(1)\prime})'$, $X_{1t} = (y_{bt}^{(2)\prime}, y_{2t}')'$, and $X_{2t} = y_{3t}$; similarly $\varepsilon_{0t} = (u_{1t}^{(1)\prime}, u_{at}', u_{bt}^{(1)\prime})'$, $\varepsilon_{1t} = (u_{bt}^{(2)\prime}, u_{2t}')'$, and $\varepsilon_{2t} = u_{3t}$; and finally

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

$$A_1 = \begin{pmatrix} 0 & B_{12}^{(1)} \\ 0 & 0 \\ 0 & B_{12}^{(2)} \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 \\ B_{a3} \\ B_{b3}^{(1)} \end{pmatrix}, \quad A_3 = 0, \quad A_4 = \begin{pmatrix} B_{b3}^{(2)} \\ 0 \end{pmatrix}.$$

Finally, Case C is an extension of Case B, where the first two equations are replaced by

$$y_{1t}^{(1)} = B_{12}^{(1)} y_{2t} + B_{1\Delta b}^{(11)} \Delta y_{bt}^{(1)} + B_{1\Delta b}^{(12)} \Delta y_{bt}^{(2)} + u_{1t}^{(1)},$$

$$\bar{y}_{bt}^{(1)} = B_{12}^{(2)} y_{2t} + B_{1\Delta b}^{(21)} \Delta y_{bt}^{(1)} + B_{1\Delta b}^{(22)} \Delta y_{bt}^{(2)} + u_{1t}^{(2)}.$$

Using the same definition for X_{it} and ε_{it}, $i = 0,1,2$ as in Case B, we obtain the system

$$X_{0t} = A_1 X_{1t} + A_2 X_{2t} + A_{\Delta 0} \Delta X_{0t} + A_{\Delta 1} \Delta X_{1t} + \varepsilon_{0t}, \tag{A.2}$$

$$\Delta X_{1t} = A_4 \Delta X_{2t} + \varepsilon_{1t}, \tag{A.3}$$

$$\Delta^2 X_{2t} = \varepsilon_{2t}, \tag{A.4}$$

where A_1, A_2, and A_4 are the same as in Case B and

$$A_{\Delta 0} = \begin{pmatrix} 0 & 0 & B_{1\Delta b}^{(11)} \\ 0 & 0 & 0 \\ 0 & 0 & B_{1\Delta b}^{(21)} \end{pmatrix}, \quad A_{\Delta 1} = \begin{pmatrix} B_{1\Delta b}^{(12)} & 0 \\ 0 & 0 \\ B_{1\Delta b}^{(22)} & 0 \end{pmatrix}.$$

The system (A.2)–(A.4) is not yet of the form (3.1)–(3.3), because of the ΔX_{0t} and ΔX_{1t} terms in (A.2). By substitution of (A.3) in (A.2), we find

$$X_{0t} = A_1 X_{1t} + A_2 X_{2t} + A_{\Delta 0} \Delta X_{0t} + A_3^* \Delta X_{2t} + \varepsilon_{0t}^*, \tag{A.5}$$

where

$$A_3^* = \begin{pmatrix} B_{1\Delta b}^{(12)} B_{b3}^{(2)} \\ 0 \\ B_{1\Delta b}^{(22)} B_{b3}^{(2)} \end{pmatrix}$$

and $\varepsilon_{0t}^* = \varepsilon_{0t} + A_{\Delta 1} \varepsilon_{1t}$. What remains is the ΔX_{0t} regressor in (A.5). Note that this term, because it involves a lagged dependent variable, makes the model dynamic rather than static, which implies that we shall not be able to write the system in the form of (3.1)–(3.3) with i.i.d. disturbances (assuming that the u_{it}'s are jointly i.i.d.). We may however obtain this representation with stationary (linear process) disturbances, by deriving the final form of (A.5). This final form involves the inverse of the matrix lag polynomial $M(L) = [I - A_{\Delta 0}(1 - L)]$. It is assumed that this matrix lag polynomial only has stable roots, which requires that the roots of $|I - B_{1\Delta b}^{(21)}(1 - L)| = 0$ are all outside the unit circle (see also Kitamura, 1995, p. 23). Then $M(L)$ is invertible, and its inverse may be expanded as

$$M(L)^{-1} = \sum_{i=0}^{\infty} A_{\Delta 0}^i (1 - L)^i = I + A_{\Delta 0}(1 - L) + \tilde{A}(L)(1 - L)^2,$$

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press

where $\tilde{A}(L)$ is a matrix power series with exponentially decaying weights. This leads to

$$\begin{aligned} X_{0t} &= A_1 X_{1t} + A_2 X_{2t} + A_{\Delta 0} A_1 \Delta X_{1t} + (A_3^* + A_{\Delta 0} A_2) \Delta X_{2t} + \varepsilon_{0t}^* + A_{\Delta 0} \Delta \varepsilon_{0t}^* \\ &\quad + A_{\Delta 0} A_3^* \Delta^2 X_{2t} + \tilde{A}(L)(A_1 \Delta^2 X_{1t} + A_2 \Delta^2 X_{2t} + A_3^* \Delta^3 X_{2t} + \Delta^2 \varepsilon_{0t}^*) \\ &= A_1 X_{1t} + A_2 X_{2t} + A_3 \Delta X_{2t} + \varepsilon_{0t}^{**}, \end{aligned} \tag{A.6}$$

where $A_3 = A_3^* + A_{\Delta 0} A_2 + A_{\Delta 0} A_1 A_4$ and

$$\begin{aligned} \varepsilon_{0t}^{**} &= \varepsilon_{0t}^* + A_{\Delta 0} \Delta \varepsilon_{0t}^* + A_{\Delta 0} A_1 \varepsilon_{1t} + A_{\Delta 0} A_3^* \Delta^2 X_{2t} \\ &\quad + \tilde{A}(L)(A_1 \Delta^2 X_{1t} + A_2 \Delta^2 X_{2t} + A_3^* \Delta^3 X_{2t} + \Delta^2 \varepsilon_{0t}^*), \end{aligned}$$

a mean-zero stationary process. In summary, Case C may be rewritten as a system consisting of equations (A.6), (A.3), and (A.4), which is of the form (3.1)–(3.3), but with stationary rather than i.i.d. disturbances. The important thing to note is that A_1 and A_4 are exactly the same as in Case B. In both cases the part of X_{1t} that enters the levels equation (A.6), namely, $X_{12,t} = y_{2t}$, does not depend on the unknown parameter $B_{b3}^{(2)}$ in A_4. Imposing these restrictions leads us to replace, in the expression for the regressor vector $Z_t(\theta)$ used in the proof of Theorem 3.1, the first entry $(X_{1t} - A_4 X_{2t})$ by $X_{12,t} = \sum_{i=1}^{t} \varepsilon_{12,i}$, so that $Z_t(\theta)$ and hence the observed information matrix (A.1) no longer depends on θ. This yields local asymptotic ancillarity of the observed information and hence LAMN. ■

https://doi.org/10.1017/S0266466600166046 Published online by Cambridge University Press