

VECTOR AUTOREGRESSIONS AND CAUSALITY

BY HIRO Y. TODA AND PETER C. B. PHILLIPS¹

This paper develops a limit theory for Wald tests of Granger causality in levels vector autoregressions (VAR's) and Johansen-type error correction models (ECM's), allowing for the presence of stochastic trends and cointegration. Earlier work by Sims, Stock, and Watson (1990) on trivariate VAR systems is extended to the general case, thereby formally characterizing the circumstances when these Wald tests are asymptotically valid as χ^2 criteria. Our results for inference from unrestricted levels VAR are not encouraging. We show that without explicit information on the number of unit roots in the system and the rank of certain submatrices in the cointegration space it is impossible to determine the appropriate limit theory in advance; and, even when such information is available, the limit theory often involves both nuisance parameters and nonstandard distributions, a situation where there is no satisfactory statistical basis for mounting these tests.

The situation with regard to the use of causality tests in ECM's is also complex but more encouraging. Granger causality tests in ECM's also suffer from nuisance parameter dependencies asymptotically and, in some cases that we make explicit, nonstandard limit theory. Both these results are somewhat surprising in the light of earlier research on the validity of asymptotic χ^2 criteria in such systems. In spite of these difficulties, Johansen-type ECM's do offer a sound basis for empirical testing of the rank of the cointegration space and the rank of key submatrices that influence the asymptotics.

KEYWORDS: Error correction model, exogeneity, Granger causality, maximum likelihood, nonstandard limit theory, nuisance parameters, vector autoregression, Wald test.

1. INTRODUCTION

ONE OF THE MAJOR POTENTIAL APPLICATIONS of unrestricted estimation in systems of stochastic difference equations is to test for causality between subsets of the variables. Such tests have become common in the empirical literature following their use in Sims (1980) to test the block exogeneity of the real sector in vector autoregressions (VAR's) fitted with real and monetary variables for both Germany and the U.S.A. Such tests are routinely performed using Wald criteria that are thought to be asymptotically chi-squared, as indeed they are in stationary or trend stationary systems. In recent work, Phillips and Durlauf (1986), Park and Phillips (1988, 1989), and Sims, Stock, and Watson (1990) (hereafter SSW) have all shown that the asymptotic theory of Wald tests is typically much more complex in systems that involve variables with stochastic trends. In general, one can expect the limit theory to involve nuisance parameters and nonstandard distributions both of which substantially complicate inference procedures, as originally pointed out in the Phillips-Durlauf paper.

In their analysis of causality tests, SSW look specifically at trivariate systems and conclude that the Wald test has a limiting chi-squared distribution if the

¹ We thank D. W. K. Andrews and a co-editor for helpful comments on earlier drafts. Our thanks also go to Glena Ames for her skill in keyboarding the manuscript and to the NSF for research support under SES 8821180.

time series are cointegrated and if the long run relationship involves the variable that is excluded under the null hypothesis (SSW, p. 135, paragraph 3 and footnote 3). Given the important empirical role of causality tests in levels VAR's, it seems reasonable to us to ask the following questions: to what extent are the conclusions of SSW generally valid; what form do the qualifications on the nature of cointegration that apply in trivariate systems take in the general case; are there other special cases of interest that are worthy of mention?

One object of the present paper is to explicitly address these questions. We extend the treatment in SSW of causality tests to the case of general VAR systems with an arbitrary number of cointegrating vectors. In particular, we are able to characterize those special cases where the limit theory is indeed χ^2 . We also provide a breakdown of the general case where the limit theory has χ^2 and nonstandard components and may depend on nuisance parameters. We point to other special cases where the limit theory has nonstandard components but is free of nuisance parameters. We show that without explicit information about the number of unit roots in the system and the rank of certain submatrices in the cointegration space it is impossible to determine the appropriate limit theory in advance. Such information is typically unavailable *a priori* in much empirical work, more especially in empirical work conducted with VAR's where the emphasis is on unrestricted estimation unfettered by prior identifying information (Sims (1980)). But, even if this information is available, the limit theory in VAR estimation will still often involve nuisance parameters and then no satisfactory basis for mounting a statistical test of causality applies.

A second object of the present paper is to develop an asymptotic theory for causality tests in error correction models (ECM's) estimated by maximum likelihood. In keeping with our earlier theme of VAR estimation, we propose an asymptotic theory for Wald tests of causality in ECM's formulated as VAR's in differences with levels as additional regressor variables. Our framework is the same as Johansen (1988, 1991) and therefore has the advantage that pretests can be performed relating to key elements that drive the asymptotics, such as the dimension of the cointegration space and the rank of certain submatrices of the cointegrating matrix. In general, tests for causality in ECM's also suffer from nuisance parameter dependencies asymptotically. Moreover, in certain important cases, the limit theory of Wald tests for causality is also nonstandard and can be characterized in terms of nonlinear functions of χ^2 variates. Both these results may seem surprising given the assumed general validity of χ^2 asymptotics for Wald tests in such models. However, the situation is not as severe as it is in levels VAR estimation. In important subcases (where either the loading coefficient submatrices or the submatrices of cointegrating coefficients that are relevant under the null are of full rank) it is shown that the limit theory of Wald tests for causality is χ^2 .

The plan of the paper is as follows. Section 2 details the models we shall consider, our notation, and some theoretical preliminaries. Section 3 studies Wald tests for causality in levels VAR estimation and Section 4 extends this analysis to Johansen-type ECM's estimated by maximum likelihood under

Gaussian assumptions. Section 5 concludes the paper and an Appendix contains many of the mathematical derivations.

A summary word on notation. We use $\text{vec}(A)$ to stack the rows of a matrix A into a column vector, $[x]$ to denote the largest integer $\leq x$, and $\mathbb{R}(A)$ and $\mathbb{R}(A)^\perp$ to signify the range space and its orthogonal complement, respectively, of a matrix A . We use the symbols " \xrightarrow{d} ," " \xrightarrow{p} ," and " \equiv " to signify convergence in distribution, convergence in probability, and equality in distribution, respectively. The inequality " > 0 " denotes positive definite (p.d.) when applied to matrices. We use $I(d)$ to signify a time series that is integrated of order d , $BM(\Omega)$ to denote a vector Brownian motion with covariance matrix Ω . We write integrals with respect to Lebesgue measure such as $\int_0^1 B(s) ds$ more simply as $\int B$ to achieve notational economy. (All integrals in this paper are from 0 to 1.) Finally all limits given in this paper are taken as the sample size T tends to ∞ .

2. THE MODEL, NOTATION, AND OTHER PRELIMINARIES

Consider the n -vector time series $\{y_t\}$ generated by the k th order VAR model

$$(1) \quad y_t = J(L)y_{t-1} + u_t$$

where $J(L) = \sum_{i=1}^k J_i L^{i-1}$ and $\{u_t\}$ is an iid sequence of n dimensional random vectors with mean zero and covariance matrix $\Sigma_u > 0$, such that $E|u_{it}|^{2+\delta} < \infty$ for some $\delta > 0$ ($i = 1, \dots, n$). We shall initialize (1) at $t = -k + 1, \dots, 0$. Since the initial values $\{y_0, y_{-1}, \dots, y_{-k+1}\}$ do not affect asymptotics, we can let them be any random vectors including constants. But we will give them a certain distribution as specified below to facilitate our analysis. In setting up a likelihood function for data generated from (1) it is, of course, most convenient to take the initial values to be constant, as in Johansen (1988, 1991).

We assume that model (1) does not have a constant term, μ , say, since this simplifies considerably the presentation and derivation of our results. Of course, the basic idea in the derivation of the asymptotics is the same whether $\mu = 0$ or $\mu \neq 0$, and since the effects of deterministic trends have been investigated rather fully in the recent literature on regression with integrated processes, it is easy to see how the asymptotic distributions derived in the next two sections should be modified if $\mu \neq 0$.

We can write (1) in the equivalent ECM format

$$(2) \quad \Delta y_t = J^*(L) \Delta y_{t-1} + \underline{J}^* y_{t-1} + u_t$$

where $\underline{J}^* = J(1) - I_n$, and

$$J^*(L) = \sum_{i=1}^{k-1} J_i^* L^{i-1} \quad \text{with} \quad J_i^* = - \sum_{l=i+1}^k J_l \quad (i = 1, \dots, k-1).$$

We assume that:

$$(3a) \quad |I_n - J(L)L| = 0 \quad \text{implies} \quad |L| > 1 \quad \text{or} \quad L = 1.$$

$$(3b) \quad \underline{J}^* = \Gamma A'$$

where Γ and A are $n \times r$ matrices of full column rank r , $0 \leq r \leq n-1$. (If $\underline{J}^* = 0$, then $r = 0$ and there is no Γ or A .)

$$(3c) \quad \Gamma'_{\perp} (J^*(1) - I_n) A_{\perp} \quad \text{is nonsingular,}$$

where Γ_{\perp} and A_{\perp} are $n \times (n-r)$ matrices of full column rank such that $\Gamma'_{\perp} \Gamma = 0 = A'_{\perp} A$. (If $r = 0$, then we take $\Gamma_{\perp} = A_{\perp} = I_n$.)

Under the above conditions $\{y_t\}$ is $I(1)$, and is cointegrated with r cointegrating vectors if $r \geq 1$. Condition (3a) precludes explosive processes but allows for the model (1) to have some unit roots. Condition (3b) defines the cointegration space to be of rank r and A is a matrix whose columns span this space. Condition (3c) ensures that the Granger representation theorem applies, so that Δy_t is stationary with a Wold representation, $A'y_t$ is stationary, and y_t is an $I(1)$ process.

Suppose that we are interested in whether the first n_1 elements of y_t are "caused by" the last n_3 elements of this vector. Write

$$y_t = \begin{pmatrix} y'_{1t} & y'_{2t} & y'_{3t} \\ n_1 & n_2 & n_3 \end{pmatrix}'$$

and partition $J(L)$ conformably with y_t . Then the null hypothesis of noncausality can be formulated based on the model (1) as

$$(4) \quad \mathcal{N}_0: J_{1,13} = \cdots = J_{k,13} = 0$$

where $J_{13}(L) = \sum_{i=1}^k J_{i,13} L^{i-1}$ is the $n_1 \times n_3$ upper-right submatrix of $J(L)$. Equivalently, the noncausality null can be formulated as

$$(5) \quad \mathcal{N}_0^*: J_{1,13}^* = \cdots = J_{k-1,13}^* = 0 \quad \text{and} \quad \underline{J}_{13}^* = 0$$

based on the ECM format (2), where $J_{13}^*(L) = \sum_{i=1}^{k-1} J_{i,13}^* L^{i-1}$ and \underline{J}_{13}^* are the $n_1 \times n_3$ upper-right submatrix of $J^*(L)$ and \underline{J}^* , respectively.

To prepare for the analysis in the following sections we introduce some further notation and a couple of lemmas. Define $x_t = (y'_{t-1}, \dots, y'_{t-k})'$ and we can write (1) as

$$(6) \quad y_t = \Phi x_t + u_t$$

where $\Phi = [J_1, \dots, J_k]$. Define an $nk \times nk$ nonsingular matrix $H = [H_1, H_2]$ with

$$H_1 = [D \otimes I_n, e_k \otimes A],$$

and

$$H_2 = e_k \otimes A_{\perp}$$

where I_n is an $n \times n$ identity matrix, e_k is a k -dimensional unit vector, i.e.,

$(1, 0, \dots, 0)'$, and D is a $k \times (k-1)$ matrix such that

$$D = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{bmatrix}.$$

Then define $z'_t = (z'_{1t}, z'_{2t}) = (H'x_t)'$ where

$$z_{1t} = H'_1 x_t = (\Delta y'_{t-1}, \dots, \Delta y'_{t-k+1}, (A' y_{t-1})')'$$

which is an $m_1 = n(k-1) + r$ dimensional vector, and

$$z_{2t} = H'_2 x_t = A'_\perp y_{t-1}$$

which is an $m_2 = n - r$ dimensional vector. These variates z_{1t} and z_{2t} are the basic components that will appear in the large sample asymptotics developed in the next two sections. With this notation we now specify the distribution of the initial values $x_1 = (y'_0, \dots, y'_{-k+1})'$ of system (1) in a convenient way for our analysis. Note that if $z_1 = (z'_{11}, z'_{21})'$ is specified, so is x_1 through $x_1 = H'^{-1} z_1$. Our initialization assigns to z_{11} the stationary distribution of the process defined by (A1) in the Appendix, while the component z_{21} can be any random vector.

We can write (2) as

$$(7) \quad \Delta y_t = \Psi z_{1t} + u_t$$

where $\Psi = [J_1^*, \dots, J_{k-1}^*, I]$. Furthermore, since Δy_t is $I(0)$, we have the Wold representation

$$(8) \quad \Delta y_t = C(L)u_t \quad \text{where} \quad C(L) = \sum_{i=0}^{\infty} C_i L^i \quad \text{with} \quad C_0 = I_n.$$

(See (A6) in the Appendix for the explicit form of $C(L)$.)

Now write $w'_t = (u'_t, z'_{1t}, \Delta z'_{2t})$ and define for any t

$$\Sigma = E w_t w'_t,$$

$$A = \sum_{j=1}^{\infty} E w_t w'_{t+j},$$

and

$$\Omega = \Sigma + A + A'.$$

We partition Ω , Σ , and A conformably with w_t . For instance, we write

$$\Omega = \begin{bmatrix} \Omega_0 & \Omega_{01} & \Omega_{02} \\ \Omega_{10} & \Omega_1 & \Omega_{12} \\ \Omega_{20} & \Omega_{21} & \Omega_2 \end{bmatrix}$$

with indices "0", "1", and "2" corresponding to the components of w_t . Then,

we have the following lemma.

LEMMA 1:

$$(i) \quad \begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} w_t \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T (z_{1t} \otimes u_t) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} B(s) \\ \xi \end{bmatrix} = \begin{bmatrix} B_0(s) \\ B_1(s) \\ B_2(s) \\ \xi \end{bmatrix}$$

where $B(s) = (B_0(s)', B_1(s)', B_2(s)')'$ is an $(n + m_1 + m_2)$ -vector Brownian motion with covariance matrix Ω , ξ is an nm_1 -dimensional normal random vector with mean zero and covariance matrix $\Sigma_1 \otimes \Sigma_0$, and $B(s)$ and ξ are independent.

$$(ii) \quad B_2(s) = A'_{\perp} C(1) B_0(s).$$

$$(iii) \quad \Omega_0 = \Sigma_0 = \Sigma_u, \quad \Sigma_1 \text{ and } \Omega_2 \text{ are positive definite, and } A_{20} = \Sigma_{20} = 0.$$

The next lemma follows from Lemma 1 above and Lemma 2.1 of Park and Phillips (1989).

LEMMA 2:

$$(i)(a) \quad \frac{1}{T} \sum_{t=1}^T z_{1t} z'_{1t} \xrightarrow{p} \Sigma_1;$$

$$(i)(b) \quad \frac{1}{\sqrt{T}} \sum_{t=1}^T z_{1t} u'_t \xrightarrow{d} N_0 \quad \text{where } \text{vec}(N_0) = \xi;$$

$$(ii)(a) \quad \frac{1}{T} \sum_{t=1}^T z_{2t} u'_t \xrightarrow{d} \int B_2 dB'_0;$$

$$(ii)(b) \quad \frac{1}{T} \sum_{t=1}^T z_{2t} z'_{1t} \xrightarrow{d} \int B_2 dB'_1 + \Sigma_{21} + A_{21};$$

$$(iii) \quad \frac{1}{T^2} \sum_{t=1}^T z_{2t} z'_{2t} \xrightarrow{d} \int B_2 B'_2.$$

Joint convergence of all the above also applies.

Now we are ready to analyze the asymptotic distribution of the Wald statistic for testing the hypothesis (4) (or (5)).

3. CAUSALITY TESTS BASED ON LEVELS VAR ESTIMATION

Suppose we estimate the levels VAR model (1) by OLS. The estimated equation is

$$(9) \quad y_t = \hat{\Phi} x_t + \hat{u}_t \quad (t = 1, \dots, T),$$

where in this section “ $\hat{\cdot}$ ” signifies estimation by OLS. The noncausality hypothesis (4) can be written as

$$(10) \quad \mathcal{N}_0: S_1' \Phi S = 0 \quad \text{or} \quad (S_1' \times S') \text{vec}(\Phi) = 0$$

where

$$S_1 = \begin{bmatrix} I_{n_1} \\ 0 \end{bmatrix} \begin{matrix} n_1 \\ n_2 + n_3 \end{matrix} \quad \text{and} \quad S = I_k \otimes S_3 \quad \text{with} \quad S_3 = \begin{bmatrix} 0 \\ I_{n_3} \end{bmatrix} \begin{matrix} n_1 + n_2 \\ n_3 \end{matrix}.$$

Then the Wald statistic for testing (10) can be written

$$\begin{aligned} F_{LS} &= \text{vec}(\hat{\Phi})' (S_1 \otimes S) \left[(S_1' \otimes S') \left[\hat{\Sigma}_u \otimes (X'X)^{-1} \right] (S_1 \otimes S) \right]^{-1} \\ &\quad \times (S_1' \otimes S') \text{vec}(\hat{\Phi}) \\ &= \text{tr} \left[S_1' \hat{\Phi} S \left[S' (X'X)^{-1} S \right]^{-1} S' \hat{\Phi} S_1 (S_1' \hat{\Sigma}_u S_1)^{-1} \right] \end{aligned}$$

where $\hat{\Sigma}_u$ is the OLS estimator of Σ_u and $X' = (x_1, \dots, x_T)$. Under the null hypothesis $S_1' \hat{\Phi} S = S_1' U' X (X'X)^{-1} S$ where $U' = (u_1, \dots, u_T)$, and we therefore have

$$\begin{aligned} F_{LS} &= \text{tr} \left[S_1' U' X (X'X)^{-1} S \left[S' (X'X)^{-1} S \right]^{-1} \right. \\ &\quad \left. \times S' (X'X)^{-1} X' U S_1 (S_1' \hat{\Sigma}_u S_1)^{-1} \right] \\ &= \text{tr} \left[(S_1' \hat{\Sigma}_u S_1)^{-\frac{1}{2}} S_1' U' Z (Z'Z)^{-1} R' \left[R (Z'Z)^{-1} R' \right]^{-1} \right. \\ &\quad \left. \times R (Z'Z)^{-1} Z' U S_1 (S_1' \hat{\Sigma}_u S_1)^{-\frac{1}{2}} \right] \end{aligned}$$

where $R = S'H$ and $Z' = (z_1, \dots, z_T)$. We write $R = [R_1, R_2]$ with

$$R_1 = S'H_1 = [D \otimes S_3', e_k \otimes A_3]$$

and

$$R_2 = S'H_2 = e_k \otimes A_{\perp 3}$$

where A_3 and $A_{\perp 3}$ are the last n_3 rows of A and A_{\perp} , respectively.

LEMMA 3: Let $\text{rank}(A_3) = g \leq n_3$. Then, there exists a nonsingular $n_3 k \times n_3 k$ matrix \tilde{K}_T such that

$$\tilde{K}_T' R = \tilde{R}_T \mathcal{T}_T$$

with

$$\lim_{T \rightarrow \infty} \tilde{R}_T = \begin{pmatrix} \tilde{R}_{11} & 0 \\ 0 & \tilde{R}_{22} \end{pmatrix} = \tilde{R}$$

where $\mathcal{T}_T = \text{diag}(\sqrt{T} I_{m_1}, T I_{m_2})$, \tilde{R}_{11} and \tilde{R}_{22} are $(n_3(k-1) + g) \times m_1$ and $(n_3 - g) \times m_2$ matrices, both of which are of full row rank.

Using Lemma 3 we have

$$\begin{aligned}
 F_{LS} &= \text{tr} \left[\left(S_1' \hat{\Sigma}_u S_1 \right)^{-\frac{1}{2}} S_1' U' Z (Z' Z)^{-1} R' \tilde{K}_T \left[\tilde{K}_T' R (Z' Z)^{-1} R' \tilde{K}_T \right]^{-1} \right. \\
 &\quad \left. \times \tilde{K}_T' R (Z' Z)^{-1} Z' U S_1 \left(S_1' \hat{\Sigma}_u S_1 \right)^{-\frac{1}{2}} \right] \\
 &= \text{tr} \left[\left(S_1' \hat{\Sigma}_u S_1 \right)^{-\frac{1}{2}} S_1' U' Z \Upsilon_T^{-1} (\Upsilon_T^{-1} Z' Z \Upsilon_T^{-1})^{-1} \right. \\
 &\quad \times \tilde{R}_T' \left[\tilde{R}_T (\Upsilon_T^{-1} Z' Z \Upsilon_T^{-1})^{-1} \tilde{R}_T \right]^{-1} \\
 &\quad \left. \cdot \tilde{R}_T (\Upsilon_T^{-1} Z' Z \Upsilon_T^{-1})^{-1} \Upsilon_T^{-1} Z' U S_1 \left(S_1' \hat{\Sigma}_u S_1 \right)^{-\frac{1}{2}} \right].
 \end{aligned}$$

Using Lemma 2 and taking into account the consistency of $\hat{\Sigma}_u$ (see Park and Phillips (1989) for the consistency of the OLS estimator), the continuous mapping theorem gives

$$F_{LS} \xrightarrow{d} F_{LS(1)} + F_{LS(2)}$$

where

$$\begin{aligned}
 F_{LS(1)} &= \text{tr} \left[S_1' N_0' \Sigma_1^{-1} \tilde{R}_{11}' \left(\tilde{R}_{11} \Sigma_1^{-1} \tilde{R}_{11}' \right)^{-1} \tilde{R}_{11} \Sigma_1^{-1} N_0 S_1 \left(S_1' \Sigma_u S_1 \right)^{-1} \right], \\
 F_{LS(2)} &= \text{tr} \left[\left(S_1' \Sigma_u S_1 \right)^{-\frac{1}{2}} S_1' \int dB_0 B_2' \left(\int B_2 B_2' \right)^{-1} \right. \\
 &\quad \times \tilde{R}_{22}' \left[\tilde{R}_{22} \left(\int B_2 B_2' \right)^{-1} \tilde{R}_{22}' \right]^{-1} \\
 &\quad \left. \cdot \tilde{R}_{22} \left(\int B_2 B_2' \right)^{-1} \int B_2 dB_0' S_1 \left(S_1' \Sigma_u S_1 \right)^{-\frac{1}{2}} \right].
 \end{aligned}$$

Note that $F_{LS(1)}$ and $F_{LS(2)}$ are independent because N_0 is independent of $(B_0(s), B_2(s))'$ by Lemma 1(i).

Since $\text{vec}(\tilde{R}_{11} \Sigma_1^{-1} N_0 S_1) = (\tilde{R}_{11} \Sigma_1^{-1} \otimes S_1') \text{vec}(N_0) \equiv N(0, \tilde{R}_{11} \Sigma_1^{-1} \tilde{R}_{11}' \otimes S_1' \Sigma_u S_1)$, by Lemma 2(i)(b), we see that

$$\begin{aligned}
 F_{LS(1)} &= \text{vec}(\tilde{R}_{11} \Sigma_1^{-1} N_0 S_1)' \left[\tilde{R}_{11} \Sigma_1^{-1} \tilde{R}_{11}' \otimes S_1' \Sigma_u S_1 \right]^{-1} \text{vec}(\tilde{R}_{11} \Sigma_1^{-1} N_0 S_1) \\
 &\equiv \chi_{n_1[n_3(k-1)+g]}^2.
 \end{aligned}$$

On the other hand, $F_{LS(2)}$ component has a nonstandard distribution and depends on nuisance parameters in general.

There are, however, two special cases that are noteworthy. First, note that if $\text{rank}(A_3) = n_3$, we may take $K_1 = I_{n_3}$ and there is no K_2 in the proof of Lemma

3. Hence $\tilde{R} = (\tilde{R}_{11}, 0)$ with \tilde{R}_{11} of full row rank n_3 . Therefore, there is no $F_{LS(2)}$ component in this case. That is, we have the usual chi-square asymptotics.

THEOREM 1: *If y_t is cointegrated and $\text{rank}(A_3) = n_3$, then under the null hypotheses (4), $F_{LS} \xrightarrow{d} \chi^2_{n_1 n_3 k}$.*

This theorem generalizes the SSW's (1990) result from their analysis of trivariate VAR(p) systems with one cointegrating vector—"if there is a linear combination involving X_{2t} which is stationary," then "the F -test will have an asymptotic χ^2_p/p distribution" (SSW, 1990, p. 135, paragraph 3 and footnote 3).

Next, we investigate the other extreme case where there is no cointegration. In this case, it will be shown that the limit distribution of the F_{LS} statistic is nonstandard but free of nuisance parameters. If y_t is not cointegrated, there is no A and we may take $A_\perp = I_n$. Hence we have $R_1 = D \otimes S'_3$ and $R_2 = e_k \otimes S'_3$. Obviously we can set $\tilde{K}_1 = D \otimes I_{n_3}$ and $\tilde{K}_2 = i_k \otimes I_{n_3}$ in the proof of Lemma 3 so that we have $\tilde{R}_{11} = D'D \otimes S'_3$ and $\tilde{R}_{22} = S'_3$. We shall rotate coordinates and cast $F_{LS(2)}$ into a canonical form that eliminates nuisance parameters.

Define $S'_{1+2} = (I_{n_1+n_2}, 0)$ and let

$$B_a(s) = \tilde{R}_{22} B_2(s) = S'_3 B_2(s),$$

$$B_b(s) = S'_{1+2} B_2(s),$$

and

$$W_1(s) = (S'_1 \Sigma_u S_1)^{-\frac{1}{2}} S'_1 B_0(s)$$

which is a standard Brownian motion since $\Omega_0 = \Sigma_u$ by Lemma 1(iii). Then we can write $F_{LS(2)}$ as

$$(11) \quad F_{LS(2)} = \text{tr} \left[\int dW_1 \underline{B}'_a \left(\int \underline{B}_a \underline{B}'_a \right)^{-1} \int \underline{B}_a dW'_1 \right]$$

where

$$(12) \quad \underline{B}_a(s) = B_a(s) - \int B_a B'_b \left(\int B_b B'_b \right)^{-1} B_b(s).$$

The covariance matrix of the m_2 -vector Brownian motion $(B_b(s), B_a(s))'$ is Ω_2 , which is p.d. by Lemma 1(iii). We partition this as

$$\begin{bmatrix} \Omega_b & \Omega_{ba} \\ \Omega_{ab} & \Omega_a \end{bmatrix}$$

conformably with $(B_b(s), B_a(s))'$. Then define

$$B_{a \cdot b}(s) = B_a(s) - \Omega_{ab} \Omega_b^{-1} B_b(s)$$

which is a Brownian motion independent of $B_b(s)$. Substituting $B_a(s) = B_{a \cdot b}(s) + \Omega_{ab}\Omega_b^{-1}B_b(s)$ into (12) gives

$$\underline{B}_a(s) = B_{a \cdot b}(s) - \int B_{a \cdot b} B'_b \left(\int B_b B'_b \right)^{-1} B_b(s).$$

Furthermore, define

$$\begin{aligned} W_a(s) &= (\Omega_a - \Omega_{ab}\Omega_b^{-1}\Omega_{ba})^{-\frac{1}{2}} B_{a \cdot b}(s) = \Omega_{a \cdot b}^{-\frac{1}{2}} B_{a \cdot b}(s), \\ W_b(s) &= \Omega_b^{-\frac{1}{2}} B_b(s), \end{aligned}$$

where $W_a(s)$ and $W_b(s)$ are standard Brownian motions independent of each other. Then we can reduce (11) to the canonical form

$$F_{LS(2)} = \text{tr} \left[\int dW_1 \underline{W}'_a \left(\int \underline{W}_a \underline{W}'_a \right)^{-1} \int \underline{W}_a dW'_1 \right]$$

where

$$(13) \quad \underline{W}_a(s) = W_a(s) - \int W_a W'_b \left(\int W_b W'_b \right)^{-1} W_b(s).$$

Moreover, we have Lemma 4.

LEMMA 4: $W_1(s)$ may be taken as the first n_1 elements of $W_b(s)$.

Therefore we have obtained the following theorem.

THEOREM 2: If y_t is not cointegrated, then under the null hypothesis (4)

$$F_{LS} \xrightarrow{d} \chi^2_{n_1 n_3(k-1)} + \text{tr} \left[\int dW_1 \underline{W}'_a \left(\int \underline{W}_a \underline{W}'_a \right)^{-1} \int \underline{W}_a dW'_1 \right]$$

where the first and the second terms are independent,

$$\underline{W}_a(s) = W_a(s) - \int W_a W'_b \left(\int W_b W'_b \right)^{-1} W_b(s),$$

$$\begin{bmatrix} W_a(s) \\ W_b(s) \end{bmatrix}_{n_1 + n_2} \equiv BM(I_n),$$

and $W_1(s)$ is the vector of the first n_1 elements of $W_b(s)$.

Theorem 2 shows that if processes are not cointegrated, causality tests in levels VAR's are asymptotically similar, and therefore the critical values for the tests can be tabulated conveniently. Once the critical values have been computed, the tests can easily be implemented in practice. Of course, if it is known

that the system is $I(1)$ with no cointegration, causality tests based on differences VAR's are also valid, and in these tests the usual chi-square critical values are employed. Hence, an interesting question is the following: which causality test should be implemented in empirical work if it is known that the process is $I(1)$ with no cointegration. To provide some insight into this matter, we consider the situation where the null hypothesis (4) is false.

For this purpose it is convenient to consider the equivalent formulation (5) of the null hypothesis (4). The alternative hypothesis corresponding to (5) is:

$$\mathcal{N}_1^*: J_{i,13}^* \neq 0 \quad \text{for some } i \quad (i = 1, \dots, k-1).$$

Note that J_{13}^* must be zero both under the null and alternative hypotheses since $J^* = 0$ if the process, y_t , is $I(1)$ with no cointegration. Hence, if it is known that there is no cointegration, the constraints $J_{13}^* = 0$ in (5) are redundant. That is, tests of the noncausality hypothesis (4) in levels models contain redundant parameter restrictions. On the other hand, causality tests in difference models take account of the unit root constraint, $J^* = 0$, and test the null hypothesis that $J_{1,13}^* = \dots = J_{k-1,13}^* = 0$. Therefore, causality tests in difference VAR's are likely to have higher power in finite samples.

The above two theorems show that in two extreme cases the asymptotic distribution of the Wald test is free of nuisance parameters: (i) If there is "sufficient cointegration with respect to y_{3t} " in the sense that $\text{rank}(A_3) = n_3$, then the asymptotic distribution is $\chi_{n_1 n_3 k}^2$; and (ii) if there is no cointegration, the Wald test statistic converges to a nonstandard but nuisance parameter free distribution. In the intermediate cases, however, the asymptotic distribution is not only nonstandard but also dependent on nuisance parameters, i.e., if there is cointegration but it is "insufficient with respect to y_{3t} " in the sense that $\text{rank}(A_3) < n_3$, then the asymptotic distribution depends on nuisance parameters in a rather complicated manner. We illustrate this by an example. (The reader is referred to Toda and Phillips (1991a) for the full development of the asymptotics in the general case.)

EXAMPLE 1: Let the true model be the following trivariate system with one cointegrating vector (given by the first equation) and error covariance matrix $\Sigma_u = (\sigma_{u_i u_j})$:

$$y_{1t} = y_{2t-1} + u_{1t},$$

$$y_{2t} = y_{2t-1} + u_{2t},$$

$$y_{3t} = y_{3t-1} + u_{3t}.$$

Suppose we set the lag length $k = 1$ and $J(L) = (b_{ij})$ in (1), estimate an unrestricted VAR(1), and test \mathcal{N} : $b_{13} = 0$ (i.e., y_3 has no causal effect on y_1) using the statistic F_{LS} . In this case we have $A_3 = 0$ and $g = 0$, so there is insufficient cointegration with respect to y_3 .

Let us now define the correlation matrix (ρ_{ij}) where $\rho_{ij} = \sigma_{u_i u_j} / (\sigma_{u_i u_i} \sigma_{u_j u_j})^{1/2}$. Then, after a little manipulation we find the following limiting expression for

the F_{LS} in terms of three scalar Brownian motions:

$$F_{LS} \xrightarrow{d} \left(\int dW_1 \underline{W}_a \right)^2 / \int \underline{W}_a^2$$

where $\underline{W}_a(s) = W_a(s) - \int W_a W_b (f W_b^2)^{-1} W_b(s)$,

$$\begin{pmatrix} W_1(s) \\ W_a(s) \\ W_b(s) \end{pmatrix} \equiv BM(\Omega_w) \quad \text{with} \quad \Omega_w = \begin{pmatrix} 1 & \omega_{a1} & \omega_{b1} \\ \omega_{a1} & 1 & 0 \\ \omega_{b1} & 0 & 1 \end{pmatrix},$$

$\omega_{a1} = (1 - \rho_{23}^2)^{-\frac{1}{2}}(\rho_{31} - \rho_{32}\rho_{31}) = \rho_{13 \cdot 2}(1 - \rho_{12}^2)^{\frac{1}{2}}$, and $\omega_{b1} = \rho_{12}$.

We deduce that the limit distribution of F_{LS} is dependent on the nuisance parameters ρ_{12} (the correlation between u_{1t} and u_{2t}) and $\rho_{13 \cdot 2}$ (the partial correlation between u_{1t} and u_{3t} given u_{2t}).

In sum, to detect noncausality in a nonstationary VAR model like (1) we need first to know if y_t is cointegrated or not. If there is no cointegration, we may formulate the model in terms of the first differences, or we can apply Theorem 2. If there is cointegration, we need further to know if the cointegration is sufficient with respect to the variables whose causal effects are being tested. If it is sufficient, we can apply Theorem 1. If the cointegration is insufficient, however, it is necessary to know $\text{rank}(A_3)$, to estimate nuisance parameters, and to simulate the asymptotic distribution that is relevant for the particular model we have using estimated nuisance parameters. This procedure seems too complicated and computationally demanding in practice besides having no sound statistical basis. We conclude that causality tests based on OLS estimation in levels VAR's are not to be recommended in general. We shall therefore propose an alternative and more promising procedure based on ML estimation of the model in ECM format in Section 4.

But before proceeding to the next section we consider the extension of the above theorems to the VAR model with a nonzero constant term, μ , viz.

$$(14) \quad y_t = \mu + \Phi x_t + u_t.$$

If (14) has some unit roots, the process y_t may contain linear trends as well as stochastic trends.² Effects on the asymptotics of the presence of deterministic trends in $I(1)$ regressors are discussed in Park and Phillips (1988) in a general framework. Therefore, we report only the results very briefly.

Suppose the process y_t is cointegrated. If (14) is estimated by OLS, the asymptotic distribution of the Wald statistic for noncausality, in general, differs depending on whether y_t contains linear trends or not. But if the rank condition of Theorem 1 is satisfied, the Wald statistic still has an asymptotic chi-square

² If $\Gamma'_1 \mu \neq 0$, then linear trends are present, otherwise y_t possesses only stochastic trends. See Johansen (1991).

distribution with the same degrees of freedom. However, if y_t is not cointegrated and (14) is estimated, then the nuisance parameter independence of the causality test no longer applies. This is because, if there is no cointegration, the nonzero μ in (14) produces linear trends in y_t , which break the collinearity of the Brownian motions in Lemma 4.

To obtain asymptotically similar tests for causality in the model with nonzero μ , we need to include a time trend term, t , as well as a constant term in the estimated system. Then, the linear trend components in y_t are eliminated since including time as a regressor is equivalent to detrending the data prior to estimation. Accordingly, the asymptotic distribution of the causality test becomes free of nuisance parameters if the process is not cointegrated. But the Brownian motions in Theorem 2 must be replaced with "detrended Brownian motions." Of course, Theorem 1 continues to hold even if a time trend term (in addition to a constant term) is included in the estimated equation.

4. CAUSALITY TESTS BASED ON ML ESTIMATION OF THE MODEL IN ECM FORMAT

As we saw in the last section, causality tests based on OLS estimators of unrestricted levels VAR's are not very useful in general because of uncertainties regarding the relevant asymptotic theory and potential nuisance parameters in the limit. In this section we consider an alternative way to test noncausality hypotheses in cointegrated VAR's. Our testing procedure is based on Johansen's (1988, 1991) ML method. This method has two advantages over the levels VAR approach considered in the last section. First, the ML procedure gives estimators of the system's cointegrating vectors, A , and their weights Γ . Hence if the asymptotic distribution of the tests depends on the structure of A (or Γ), as in the case of OLS based tests, then we may use these estimators to test relevant hypotheses about the structure of A (or Γ). Moreover, the ML estimators of the cointegrating vectors are asymptotically median unbiased and have mixed normal limit distributions, unlike those that would be obtained from levels VAR estimation, and they are therefore much better suited to perform inference. Second, since ML methods take into account information on the presence of unit roots in the system, we can avoid unit root asymptotics altogether, i.e., the asymptotic distribution of the ML estimator of A will be mixed normal and conventional normal asymptotics will apply to the estimators of the other parameters. (See Phillips (1991).)

We deal with the ECM representation of the system given in (2), and estimate the parameters J_1^*, \dots, J_{k-1}^* , Γ , and A . We continue to assume that the model has no constant term. The asymptotic theory does differ if the model has a constant term. However, our results on causality tests obtained below are not affected by that difference.

Although Johansen assumes normality of the innovation sequence $\{u_t\}$ in addition to the assumptions we made in Section 2, it is obvious in view of our Lemma 1 and Lemma 2 that all the asymptotic results in Johansen (1988)

continue to hold without the extra assumption.³ Thus, suppose that by Johansen's likelihood ratio test about the number of cointegrating vectors we have decided that there are $r \geq 1$ cointegrating vectors. (If there is no cointegration, we can formulate the model in terms of first order differences $\{\Delta y_t\}$, or we may apply Theorem 2 in the last section.) Then the ML estimator \hat{A} of A is given by the eigenvectors corresponding to the r largest eigenvalues that solve equation (9) of Johansen (1988). Also, let \hat{A}_\perp be the $n-r$ eigenvectors⁴ corresponding to the $n-r$ smallest eigenvalues and assume that all the eigenvectors are normalized in the manner prescribed by Johansen (1988, p. 235). The estimator of $\Psi = [J_1^*, \dots, J_{k-1}^*, \Gamma]$ (see (7)) is given by

$$\hat{\Psi} = \Delta Y' \hat{Z}_1 (\hat{Z}_1' \hat{Z}_1)^{-1}$$

where $\Delta Y' = [\Delta y_1, \dots, \Delta y_T]$, and $\hat{Z}_1' = [\hat{z}_{11}, \dots, \hat{z}_{1T}]$ with $\hat{z}_{1t} = [\Delta y_{t-1}', \dots, \Delta y_{t-k+1}', (\hat{A}' y_{t-1})']'$. In this section the symbol " $\hat{\cdot}$ " on top of a parameter signifies that the parameter is estimated by ML.

We shall construct a Wald test statistic based on these ML estimators to test the noncausality hypothesis (5). But before proceeding further, we summarize the limit behavior of the ML estimators and some sample moment matrices. Define

$$\tilde{A} = \hat{A} \hat{\Pi}^{-1},$$

$$\tilde{\Gamma} = \hat{\Gamma} \hat{\Pi}',$$

and

$$\tilde{\Psi} = [\hat{J}_1^*, \dots, \hat{J}_{k-1}^*, \tilde{\Gamma}],$$

where $\hat{\Pi} = \underline{A}' \hat{A}$ with $\underline{A}' = (A' A)^{-1} A'$, and define

$$\tilde{A}_\perp = \hat{A}_\perp \hat{\Pi}_\perp^{-1},$$

where $\hat{\Pi}_\perp = \underline{A}'_\perp \hat{A}_\perp$ with $\underline{A}'_\perp = (A'_\perp A_\perp)^{-1} A'_\perp$. The limit theory we need is given in the following lemma, which is a consequence of results in Johansen (1988).

LEMMA 5:

$$(i) \quad T(\tilde{A} - A) \xrightarrow{d} A_\perp \left(\int B_2 B_2' \right)^{-1} \int B_2 dB_c',$$

where $B_2(s)$ is the $m_2 = n - r$ dimensional Brownian motion defined in Lemma 1, $B_c(s)$ is an r dimensional Brownian motion with covariance matrix $\Omega_c =$

³ If u_t is not normally distributed, the estimators considered below are no longer ML estimators. Nevertheless, we shall continue to refer to them as "ML estimators."

⁴ These eigenvectors do not provide a consistent estimator of the space spanned by A_\perp . But we call them \hat{A}_\perp since their role in the derivation of the asymptotic distribution is the same as that of A_\perp . (See the Proof of Lemma 5.)

$(\Gamma' \Sigma_u^{-1} \Gamma)^{-1}$, and $B_2(s)$ and $B_c(s)$ are independent.

$$(ii) \quad \sqrt{T}(\tilde{\Psi} - \Psi) \xrightarrow{d} N = \begin{bmatrix} N_1 & N_2 \end{bmatrix},$$

$n(k-1) \quad r$

where $\text{vec}(N') \equiv N(0, \Sigma_1^{-1} \otimes \Sigma_u)$ which is independent of $B_2(s)$ and $B_c(s)$.

$$(iii) \quad \hat{\Sigma}_u \xrightarrow{p} \Sigma_u$$

where $\hat{\Sigma}_u = T^{-1}(\Delta Y' \Delta Y - \Delta Y' \hat{Z}_1 (\hat{Z}_1' \hat{Z}_1)^{-1} \hat{Z}_1' \Delta Y)$ is the ML estimator of Σ_u .

$$(iv) \quad \tilde{\Omega}_c = (\tilde{\Gamma}' \hat{\Sigma}_u^{-1} \tilde{\Gamma})^{-1} \xrightarrow{p} \Omega_c.$$

$$(v) \quad T(\tilde{Z}_1' \tilde{Z}_1)^{-1} \xrightarrow{p} \Sigma_1^{-1}$$

where $\tilde{Z}_1' = [\tilde{z}_{11}, \dots, \tilde{z}_{1T}]$ with $\tilde{z}_{1t} = [\Delta y'_{t-1}, \dots, \Delta y'_{t-k+1}, (\tilde{A}' y_{t-1})']$.

(vi)(a) For a matrix M such that $M'A = 0$, $M'A_{\perp} = M'A_{\perp}$,

$$(b) \quad T^2(\tilde{Z}_2' \tilde{Z}_2)^{-1} \xrightarrow{d} \left(\int B_2 B_2' \right)^{-1}$$

where $\tilde{Z}_2' = [\tilde{z}_{21}, \dots, \tilde{z}_{2T}]$ with $\tilde{z}_{2t} = \tilde{A}'_{\perp} y_{t-1}$.

Now the null hypothesis of noncausality is given by (5). This can be written alternatively as

$$(15) \quad \Phi_{13}^* = S_1' \Phi^* S = 0 \quad \text{or} \quad \text{vec}(\Phi_{13}^*) = (S' \otimes S_1') \text{vec}(\Phi^*) = 0$$

where $\Phi^* = [J_1^*, \dots, J_{k-1}^*, \underline{J}^*]$ and $\Phi_{13}^* = [J_{1,13}^*, \dots, J_{k-1,13}^*, \underline{J}_{13}^*]$. Note that (15) involves some nonlinear restrictions, viz., $\underline{J}_{13}^* = \Gamma_1 A_3' = 0$ where Γ_1 denotes the first n_1 rows of Γ . Since

$$\hat{\Gamma} \hat{A}' - \Gamma A' = \hat{\Gamma} \hat{\Pi}' (\hat{\Pi}^{-1})' \hat{A}' - \Gamma A' = (\tilde{\Gamma} - \Gamma) \tilde{A}' + \Gamma (\tilde{A}' - A'),$$

we have under the null hypothesis (15)

$$(16) \quad \text{vec}(\hat{\Phi}_{13}^*) = (S' \otimes S_1') \text{vec}(\hat{\Phi}^* - \Phi^*) = \tilde{P} \tilde{\eta}$$

where $\hat{\Phi}^* = [\hat{J}_1^*, \dots, \hat{J}_{k-1}^*, \hat{\underline{J}}^*]$ with $\hat{\underline{J}}^* = \hat{\Gamma} \hat{A}'$, $\tilde{\eta} = (\text{vec}(\tilde{\Psi}' - \Psi'), \text{vec}(\tilde{A} - A)')'$, $\tilde{P} = (\tilde{P}_1, P_2)$ with

$$\tilde{P}_1 = \begin{bmatrix} I_{k-1} \otimes S_3' \otimes S_1' & 0 \\ 0 & \tilde{A}_3 \otimes S_1' \end{bmatrix} \begin{matrix} n_1 n_3 (k-1) \\ n_1 n_3 \end{matrix} \quad \text{and}$$

$n^2(k-1) \quad nr$

$$P_2 = \begin{bmatrix} 0 \\ S_3' \otimes \tilde{\Gamma}_1 \end{bmatrix} \begin{matrix} n_1 n_3 (k-1) \\ n_1 n_3 \end{matrix},$$

nr

and \tilde{A}_3 denotes the last n_3 rows of \tilde{A} .

Given Lemma 5(i) and (ii), the expansion (16) motivates us to employ the Wald statistic

$$(17) \quad F_{ML} = \text{vec}(\hat{\Phi}_{13}^{*'})' (\hat{P}\hat{V}\hat{P}')^{-1} \text{vec}(\hat{\Phi}_{13}^{*'})$$

where \hat{P} is defined as \tilde{P} with \tilde{A}_3 and Γ_1 replaced by \hat{A}_3 and $\hat{\Gamma}_1$, respectively, and

$$\hat{V} = \begin{pmatrix} (\hat{Z}'_1 \hat{Z}_1)^{-1} \otimes \hat{\Sigma}_u & 0 \\ 0 & \hat{A}_\perp (\hat{Z}'_2 \hat{Z}_2)^{-1} \hat{A}'_\perp \otimes \hat{\Omega}_c \end{pmatrix}.$$

\hat{A}_3 and $\hat{\Gamma}_1$ are the last n_3 rows of \hat{A} and the first n_1 rows of $\hat{\Gamma}_1$, and $\hat{\Omega}_c = (\hat{\Gamma}'_1 \hat{\Sigma}_u^{-1} \hat{\Gamma}_1)^{-1}$. Note that \hat{P} and \hat{V} can be constructed using the estimates \hat{A} , $\hat{\Gamma}$, \hat{A}_\perp , and $\hat{\Sigma}_u$.

It turns out that $\hat{P}\hat{V}\hat{P}'$ might be singular (even in the limit). However, we will be interested in the case where $\hat{P}\hat{V}\hat{P}'$ is nonsingular in the limit. In order to guarantee this we need to assume that either Γ_1 or A_3 is of full row rank, and consequently in the above formula of the Wald statistic it is assumed that $(\hat{P}\hat{V}\hat{P}')^{-1}$ exists. This problem arises due to the nonlinear restrictions $\Gamma_1 A'_3 = 0$ involved in the noncausality hypothesis in ECM's. In the case of a Wald test of a nonlinear hypothesis $f(\theta) = 0$, say, in stationary models it is well known that the Jacobian matrix, $\partial f(\theta)/\partial \theta'$, evaluated at the true parameter value must have full row rank for the Wald statistic to converge to a chi-square distribution with the usual degrees of freedom. The same principle applies to the restriction $\Gamma_1 A'_3 = 0$ in our nonstationary model, and the rank conditions on Γ_1 and A_3 correspond to the conditions on the Jacobian matrix, $\partial f(\theta)/\partial \theta'$.

Let us first assume that $\text{rank}(A_3) = g < n_3$ and $\text{rank}(\Gamma_1) = n_1$ in the following development of the asymptotics. The treatment of the case in which $\text{rank}(A_3) = n_3$ and $\text{rank}(\Gamma_1) \leq n_1$ is easier and will be briefly discussed later.

Now it is convenient to rewrite the matrix form $\hat{P}\hat{V}\hat{P}'$ as

$$(18) \quad \hat{P}\hat{V}\hat{P}' = \tilde{P}^\dagger \tilde{V}^\dagger \tilde{P}^\dagger',$$

Here

$$\tilde{V}^\dagger = \begin{pmatrix} (\tilde{Z}'_1 \tilde{Z}_1)^{-1} \otimes \hat{\Sigma}_u & 0 \\ 0 & (\tilde{Z}'_2 \tilde{Z}_2)^{-1} \otimes \tilde{\Omega}_c \end{pmatrix},$$

and $\tilde{P}^\dagger = (\tilde{P}_1, \tilde{P}_2^\dagger)$ with

$$\tilde{P}_2^\dagger = \begin{pmatrix} 0 \\ \tilde{A}_{\perp 3} \otimes \tilde{\Gamma}_1 \end{pmatrix}$$

where $\tilde{A}_{\perp 3}$ is the last n_3 rows of \tilde{A}_\perp . It is straightforward to show the above

equality (18) using the relations $\tilde{A} = \hat{A}\hat{\Pi}^{-1}$, $\tilde{I} = \hat{I}\hat{\Pi}'$, and $\tilde{A}_\perp = \hat{A}_\perp\hat{\Pi}_\perp^{-1}$. For instance, $\tilde{A}_\perp(\tilde{Z}_2'\tilde{Z}_2)^{-1}\tilde{A}'_\perp = \hat{A}_\perp(\tilde{Z}_2'\tilde{Z}_2)^{-1}\tilde{A}'_\perp$.

To obtain the limit distribution of the F_{ML} , we need the next lemma.

LEMMA 6: *If $\text{rank}(A_3) = g < n_3$ and $\text{rank}(\Gamma_1) = n_1$, there exists an $n_1n_3k \times n_1n_3k$ nonsingular matrix K_T^* such that*

$$K_T^{*'}\tilde{P} = \tilde{P}_T\Upsilon_T^*,$$

$$K_T^{*'}\tilde{P}^\dagger = \tilde{P}_T^\dagger\Upsilon_T^\dagger,$$

where \tilde{P}_T is an $n_1n_3k \times (nm_1 + rm)$ matrix, \tilde{P}_T^\dagger is an $n_1n_3k \times (nm_1 + rm_2)$ matrix, $\Upsilon_T^* = \text{diag}(\sqrt{T}I_{nm_1}, TI_{nr})$, and $\Upsilon_T^\dagger = \text{diag}(\sqrt{T}I_{nm_1}, TI_{rm_2})$. Moreover,

$$\tilde{P}_T \xrightarrow{P} \begin{pmatrix} P_{11} & 0 \\ 0 & P_{22} \end{pmatrix}$$

where P_{11} is an $n_1(n_3(k-1) + g) \times nm_1$ matrix of full row rank and P_{22} is an $n_1(n_3 - g) \times rm$ matrix, and

$$\tilde{P}_T^\dagger \xrightarrow{P} \begin{pmatrix} P_{11} & 0 \\ 0 & P_{22}^\dagger \end{pmatrix}$$

where $P_{22}^\dagger = P_{22}(A_\perp \otimes I_r)$ is an $n_1(n_3 - g) \times rm_2$ matrix of full row rank.

Using (18) and Lemma 6 we have from (17)

$$\begin{aligned} F_{ML} &= \tilde{\eta}'\tilde{P}'(\tilde{P}^\dagger\tilde{V}^\dagger\tilde{P}^\dagger)^{-1}\tilde{P}\tilde{\eta} \\ &= \tilde{\eta}'\tilde{P}'K_T^*(K_T^{*'}\tilde{P}^\dagger\tilde{V}^\dagger\tilde{P}^\dagger K_T^*)^{-1}K_T^{*'}\tilde{P}\tilde{\eta} \\ &= \tilde{\eta}'\Upsilon_T^*\tilde{P}_T'(\tilde{P}_T^\dagger\Upsilon_T^\dagger\tilde{V}^\dagger\Upsilon_T^\dagger\tilde{P}_T')^{-1}\tilde{P}_T\Upsilon_T^*\tilde{\eta}. \end{aligned}$$

From the last expression, we get by Lemma 5

$$F_{ML} \xrightarrow{d} F_{ML(1)} + F_{ML(2)}$$

where

$$\begin{aligned} F_{ML(1)} &= \text{vec}(N')'P_{11}'[P_{11}(\Sigma_1^{-1} \otimes \Sigma_u)P_{11}']^{-1}P_{11}\text{vec}(N'), \\ F_{ML(2)} &= \text{vec}\left(\left(\int B_2B_2'\right)^{-1}\int B_2dB_c'\right)'P_{22}^{\dagger'}\left[P_{22}^\dagger\left(\left(\int B_2B_2'\right)^{-1} \otimes \Omega_c\right)P_{22}^{\dagger'}\right]^{-1} \\ &\quad \cdot P_{22}^\dagger \text{vec}\left(\left(\int B_2B_2'\right)^{-1}\int B_2dB_c'\right), \end{aligned}$$

and $F_{ML(1)}$ and $F_{ML(2)}$ are independent since N is independent of $B_2(s)$ and $B_c(s)$ by Lemma 5(ii).

Because $\text{vec}(N') \equiv N(0, \Sigma_1^{-1} \otimes \Sigma_u)$, we easily see that $F_{ML(1)} \equiv \chi_{n_1 n_3(k-1)+n_1 g}^2$. As for $F_{ML(2)}$, note that by the same argument as that of Lemma 5.1 of Park and Phillips (1988)

$$\begin{aligned} & \left[P_{22}^\dagger \left(\left(\int B_2 B_2' \right)^{-1} \otimes \Omega_c \right) P_{22}^\dagger \right]^{-\frac{1}{2}} P_{22}^\dagger \text{vec} \left(\left(\int B_2 B_2' \right)^{-1} \int B_2 dB_c' \right) \\ & \equiv N(0, I_{n_1(n_3-g)}) \end{aligned}$$

since $B_2(s)$ and $B_c(s)$ are independent by Lemma 5(i). Hence, $F_{ML(2)} \equiv \chi_{n_1(n_3-g)}^2$. Since $F_{ML(1)}$ and $F_{ML(2)}$ are independent, we deduce that

$$F_{ML} \xrightarrow{d} \chi_{n_1 n_3 k}^2.$$

To interpret the decomposition of the limiting distribution into the $F_{ML(1)}$ and $F_{ML(2)}$ components, consider testing the following hypotheses in ECM's:

$$\mathcal{N}_{(1)}^*: J_{1,13}^* = \cdots = J_{k-1,13}^* = 0 \quad \text{and} \quad \Gamma_1 = 0$$

and

$$\mathcal{N}_{(2)}^*: A_3 = 0.$$

These are sufficient (but obviously not necessary) conditions for noncausality. It follows easily from Lemma 5(ii) that under the null hypothesis the Wald statistic for testing $\mathcal{N}_{(1)}^*$ converges to a chi-square distribution with $n_1 n_3(k-1) + n_1 r$ degrees of freedom. It is also easy to see that under $\mathcal{N}_{(2)}^*$ the Wald test for $\mathcal{N}_{(2)}^*$ has an asymptotic chi-square distribution with $n_3 r$ degrees of freedom by Lemma 5(i). Moreover, these two Wald statistics are asymptotically independent by Lemma 5(i) and (ii). The $F_{ML(1)}$ and $F_{ML(2)}$ terms, to some extent, come from testing $\mathcal{N}_{(1)}^*$ and $\mathcal{N}_{(2)}^*$, respectively. However, the derivation of the asymptotic distributions in the case of testing \mathcal{N}_0^* in (5) is much more complicated and the ranks of the matrices Γ_1 and A_3 play an important role in the final result.

Next suppose that $\text{rank}(A_3) = n_3$ and $\text{rank}(\Gamma_1) \leq n_1$. In this case it is easy to show that Lemma 6 holds with no P_{22} , viz., $\tilde{P}_T \xrightarrow{p} (P_{11}, 0)$ and $\tilde{P}_T^\dagger \xrightarrow{p} (P_{11}, 0)$ where P_{11} is an $n_1 n_3 k \times nm_1$ matrix of full row rank. Therefore

$$F_{ML} \xrightarrow{d} \text{vec}(N') P_{11}' [P_{11}(\Sigma_1^{-1} \otimes \Sigma_u) P_{11}']^{-1} P_{11} \text{vec}(N')$$

which also has a $\chi_{n_1 n_3 k}^2$ distribution.

We summarize these results in our next theorem.

THEOREM 3: Suppose that in the model (2) (or equivalently (1)) y_t is cointegrated. If $\text{rank}(A_3) = n_3$ or $\text{rank}(\Gamma_1) = n_1$, then under the null hypothesis (5) (or equivalently (4)), $F_{ML} \xrightarrow{d} \chi_{n_1 n_3 k}^2$.

Unfortunately, even the ML method does not always guarantee the usual chi-square asymptotics because the rank condition in Theorem 3 is not always satisfied under the null. To illustrate the problem that arises when there is rank

deficiency both in A_3 and Γ_1 we provide the following example. (See Toda and Phillips (1991a) for more details.)

EXAMPLE 2: Consider the trivariate cointegrated system

$$(19) \quad \Delta y_t = \gamma \alpha' y_{t-1} + u_t$$

where $y'_t = (y_{1t}, y_{2t}, y_{3t})$, $u'_t = (u_{1t}, u_{2t}, u_{3t})$, $\gamma' = (\gamma_1, \gamma_2, \gamma_3) = (0, 1, 1)$, and $\alpha' = (\alpha_1, \alpha_2, \alpha_3) = (1, -\frac{1}{2}, 0)$. In this example we use lower case letters to signify vectors and scalars. (For example, α corresponds to A .)

Suppose that we want to test whether y_{3t-1} causes y_{1t} . Then the null hypothesis is

$$(20) \quad \gamma_1 \alpha_3 = 0,$$

and the Wald statistic given by (17) becomes

$$F_{ML} = \frac{(\hat{\gamma}_1 \hat{\alpha}_3)^2}{\hat{\alpha}_3^2 \hat{\sigma}_{u_1} (\hat{Z}'_1 \hat{Z}_1)^{-1} + \hat{\gamma}_1^2 \hat{\omega}_c \hat{\alpha}_{\perp 3} (\hat{Z}'_2 \hat{Z}_2)^{-1} \hat{\alpha}'_{\perp 3}}$$

where $\sigma_{u_1} = \text{var}(u_{1t})$ and $\hat{Z}'_1 = (\hat{\alpha}' y_0, \dots, \hat{\alpha}' y_{T-1})$. This can be rewritten as

$$F_{ML} = \frac{1}{\frac{T(\hat{Z}'_1 \hat{Z}_1)^{-1} \hat{\sigma}_{u_1}}{(\sqrt{T} \tilde{\gamma}_1)^2} + \frac{T^2 \alpha_{\perp 3} (\tilde{Z}'_2 \tilde{Z}_2)^{-1} \alpha'_{\perp 3} \tilde{\omega}_c}{(T \tilde{\alpha}_3)^2}}$$

because $S'_3 \alpha = \alpha_3 = 0$ and hence $\tilde{\alpha}_{\perp 3} = S'_3 \tilde{A}_{\perp} = S'_3 A_{\perp} = \alpha_{\perp 3}$ by Lemma 5(vi)(a). Since by Lemma 5 $\hat{\sigma}_{u_1} \xrightarrow{p} \sigma_{u_1}$, $\hat{\omega}_c \xrightarrow{p} \omega_c$, $T(\hat{Z}'_1 \hat{Z}_1)^{-1} \xrightarrow{p} \sigma_1^{-1}$, $T^2(\tilde{Z}'_2 \tilde{Z}_2)^{-1} \xrightarrow{d} (JB_2 B'_2)^{-1}$, $\sqrt{T} \tilde{\gamma}_1 = \sqrt{T}(\tilde{\gamma}_1 - \gamma_1) \xrightarrow{d} N(0, \sigma_{u_1} \sigma_1^{-1})$, and $T \tilde{\alpha}_3 = T(\tilde{\alpha}_3 - \alpha_3) \xrightarrow{d} \alpha_{\perp 3} (JB_2 B'_2)^{-1} JB_2 dB_c$, we have

$$(21) \quad F_{ML} \xrightarrow{d} \frac{1}{\chi_a^{-1} + \chi_b^{-1}} = \frac{\chi_a \chi_b}{\chi_a + \chi_b}$$

where each of χ_a and χ_b is distributed as chi-square with one degree of freedom, and χ_a and χ_b are independent.

Thus, F_{ML} does not converge to a χ^2_1 distribution but to a nonlinear function of independent chi-square variates. This occurs because both of γ_1 and α_3 are zero. Under the null hypothesis (20), we can expand $\tilde{\gamma}_1 \tilde{\alpha}_3$ as

$$\tilde{\gamma}_1 \tilde{\alpha}_3 = \alpha_3 (\tilde{\gamma}_1 - \gamma_1) + \gamma_1 (\tilde{\alpha}_3 - \alpha_3) + (\tilde{\gamma}_1 - \gamma_1) (\tilde{\alpha}_3 - \alpha_3)$$

where $\tilde{\gamma}_1 - \gamma_1 = O_p(T^{-\frac{1}{2}})$, $\tilde{\alpha}_3 - \alpha_3 = O_p(T^{-1})$, and $(\tilde{\gamma}_1 - \gamma_1)(\tilde{\alpha}_3 - \alpha_3) = O_p(T^{-\frac{3}{2}})$. If $\alpha_3 \neq 0$, $\tilde{\gamma}_1 \tilde{\alpha}_3$ is asymptotically dominated by the first term. If $\alpha_3 = 0$ but $\gamma_1 \neq 0$, then $\tilde{\gamma}_1 \tilde{\alpha}_3$ is asymptotically dominated by the second term. In either case the Wald statistic F_{ML} will have an asymptotic chi-square distribution with one degree of freedom because $\sqrt{T}(\tilde{\gamma}_1 - \gamma_1)$ is asymptotically normal and $T(\tilde{\alpha}_3 - \alpha_3)$ is asymptotically mixed normal. If $\alpha_3 = \gamma_1 = 0$, however, $\tilde{\gamma}_1 \tilde{\alpha}_3$ is equal to the third term and therefore the usual chi-square asymptotics do not hold.

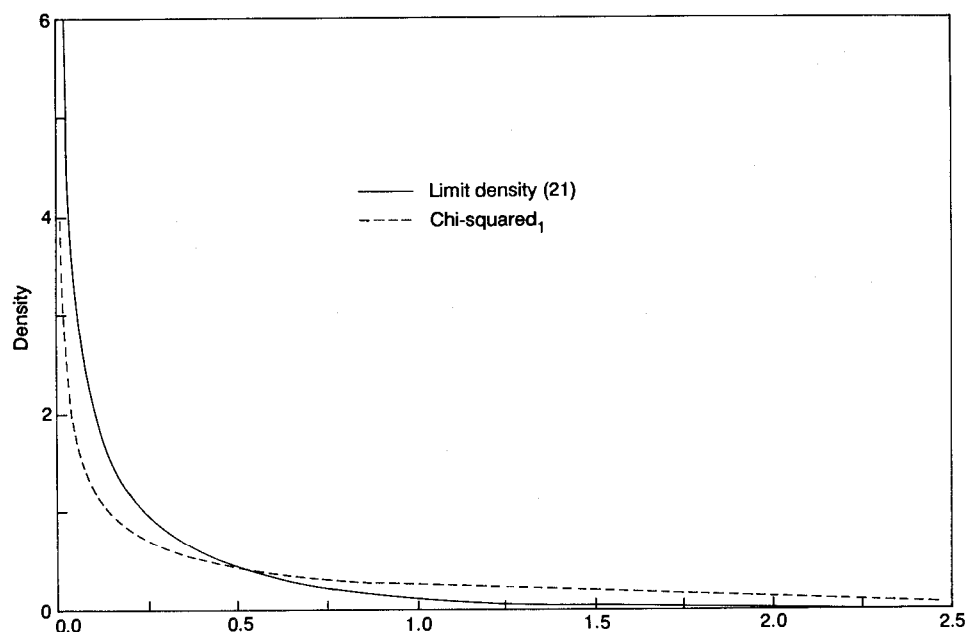


FIGURE 1

The density of the distribution in (21) is graphed in Figure 1 against that of a χ_1^2 variate. (The analytic form of this density is given in Toda and Phillips (1991a).) As is apparent from the figure, the density of (21) is much more concentrated near the origin and has a much thinner tail than the χ_1^2 distribution. If we were to test the null hypothesis (20) using a critical value obtained from (21) for the Wald statistic in this case, the test would have much greater power than a test that employed a nominal χ_1^2 critical value. In practice, of course, we do not know that both γ_1 and α_3 are zero, and an investigator who routinely employed a χ_1^2 critical value in this case would suffer from both the size distortion and the resulting power loss.

In sum, if the system is subject to cointegration, causality tests based on ML estimation may well collapse and not satisfy the usual chi-square asymptotics, not because of failure to use information on unit roots (as in the levels VAR estimation), but because of the nonlinear constraints $\Gamma_1 A_3' = 0$ that are necessarily involved in the null hypothesis. Thus, we need to know whether the condition that (i) $\text{rank}(\Gamma_1) = n_1$ or that (ii) $\text{rank}(A_3) = n_3$ holds. Unless we have a reason to believe a priori that either condition (i) or (ii) holds, we have to test the conditions empirically. This can, of course, be done using the ML estimates of Γ_1 and A_3 . In particular, condition (i) or (ii) can easily be tested if $n_1 = 1$ or $n_3 = 1$, respectively. A companion paper, Toda and Phillips (1991b) proposes some sequential test procedures for causality that are applicable when $n_1 = 1$ and/or $n_3 = 1$.

In the above development of the asymptotics of causality tests, we ignored the issue of the estimation of r . A natural question is how severe its impact on the causality test can be in small or moderately sized samples. Furthermore, even if r were known, some size distortion and loss of power are inevitable in the causality test in ECM's because of its sequential nature (i.e., verify the rank condition and then test causality). To address these issues, Toda and Phillips (1991b) conducted a simulation study which investigated the performance of the causality tests in ECM's when $n_1 = 1$ or $n_3 = 1$, in comparison with conventional causality tests in levels VAR's and in difference VAR's. These simulations found some favorable evidence in support of the sequential causality test based on the ECM formulation at least in relatively large samples (more than 100 observations) and showed that the sequential procedure generally outperforms the conventional tests.

The results for the Wald test derived in this section are not affected by the presence of a constant term in the true model and the inclusion of a constant term (and a time trend term) in the estimated equation. The sequential test suggested in Toda and Phillips (1991b) can also be implemented in such models.

5. CONCLUSION

This paper has studied the asymptotics of Granger causality tests in unrestricted levels VAR's and Johansen-type ECM's. The results of our analysis are not encouraging for these tests in levels VAR's. Our main conclusions regarding the use of Wald tests in levels VAR's are as follows.

(i) Causality tests are valid asymptotically as χ^2 criteria only when there is *sufficient* cointegration with respect to the variables whose causal effects are being tested. The precise condition for sufficiency involves a rank condition on a submatrix of the cointegrating matrix. Since the estimates of such matrices in levels VAR's suffer from simultaneous equations bias (as shown in Phillips (1991)) there is no valid statistical basis for determining whether the required sufficient condition applies.

(ii) When the rank condition for sufficiency fails, the limit distribution is more complex and involves a mixture of a χ^2 and a nonstandard distribution, which generally involves nuisance parameters. The precise form of the distribution depends on the actual rank of a submatrix of the cointegrating matrix and again no valid statistical basis for mounting a Wald test of causality applies.

In view of these results we conclude the empirical use of Granger causality tests in levels VAR's is not to be encouraged in general when there are stochastic trends and the possibility of cointegration.

(iii) If there is no cointegration the Wald test statistic for causality has a nonstandard but nuisance parameter free limit distribution. This distribution could be used for tests when it is known that there are stochastic trends but no cointegration in the system.

Testing for causality in ECM's is more promising than in levels VAR's but it still involves some difficulties. Our main results are as follows.

(iv) Wald tests for causality in ECM's are not always valid asymptotic χ^2 criteria.

(v) Problems of nuisance parameter dependencies and nonstandard distributions enter the limit theory in the general case.

(vi) Sufficient rank conditions for causality tests to be asymptotically valid χ^2 tests are given. These rank conditions relate to submatrices of both the cointegrating matrix and the loading coefficient matrix. They can, in principle, be tested empirically using the ML estimates of these submatrices.

Granger causality tests in systems of stochastic difference equations are fraught with many complications when there are stochastic trends and cointegration in the system. The results for causality tests in ECM's are deserving of some emphasis in view of the fact that other types of Wald tests in ECM's are known to be asymptotically valid χ^2 tests. Since ML estimation of ECM's delivers optimal estimates of the cointegrating space, ECM's provide a more promising basis than VAR's for the sequential inference procedures that are needed to adequately test causality hypothesis in these models. Indeed, simulation exercises reported in Toda and Phillips (1991b) indicate that the sequential test in ECM's works reasonably well in small systems (3 or 4 variables) and moderately large sample sizes (more than 100 observations).

Institute of Socio-Economic Planning, University of Tsukuba, Tsukuba-shi, Ibaraki 305, Japan

and

Cowles Foundation for Research in Economics, Yale University, P.O. Box 2125 Yale Station, New Haven, CT 06520-2125, U.S.A.

Manuscript received July, 1991; final revision received February, 1993.

APPENDIX

PROOF OF LEMMA 1: (i) From (2) we have

$$(A1) \quad z_{1t+1} = Gz_{1t} + Fu_t$$

where

$$G = \begin{bmatrix} J_1^* & \dots & J_{k-1}^* & \Gamma \\ I_{n(k-2)} & 0 & 0 \\ AJ_1^* & \dots & AJ_{k-1}^* & AT + I_r \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} e_{k-1} \otimes I_n \\ A' \end{bmatrix}.$$

Since z_{1t} is $I(0)$ by assumption, the eigenvalues of G must be all less than unity. Hence we can write (A1) as

$$(A2) \quad z_{1t} = \Theta(L)Fu_{t-1}$$

where $\Theta(L) = \sum_{j=0}^{\infty} \Theta_j L^j = \sum_{j=0}^{\infty} G^j L^j$.

Now by the same argument as that of Theorem 2.2 of Chan and Wei (1988), $T^{-1} \sum_{t=1}^T z_{1t} z'_{1t} \xrightarrow{p} \Sigma_1$ and

$$(A3) \quad \begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} u_t \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T (z_{1t} \otimes u_t) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} B_0(s) \\ \xi \end{bmatrix}$$

where $B_0(s)$ is an n -vector Brownian motion with covariance matrix Σ_u , and ξ is an nm_1 -dimensional normal random vector with mean zero and covariance matrix $\Sigma_1 \otimes \Sigma_u$ with $\Sigma_1 = E z_{1t} z'_{1t} = \sum_{j=0}^{\infty} G^j F \Sigma_u F' G^j$. $B_0(s)$ and ξ are independent.

Since $\Theta(L)$ is the inverse of $I - GL$ and $|I - GL| = 0$ has only stable roots, we see from Brillinger (1981, p. 77) that for all $p \geq 0$

$$\sum_{j=1}^{\infty} j^p \|\Theta_j\|_a < \infty$$

where $\|\Theta_j\|_a$ denotes the sum of the absolute value of the entries of Θ_j . This in turn implies that

$$\sum_{j=1}^{\infty} j^2 \|\Theta_j\|^2 < \infty$$

where $\|\Theta_j\| = \text{tr}(\Theta_j \Theta_j')^{\frac{1}{2}}$. Thus, by the multivariate extension of Theorem 3.4 of Phillips and Solo (1992),

$$(A4) \quad \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} z_{1t} \xrightarrow{d} \Theta(1) F B_0(s).$$

Since $\Delta z_{2t} = A'_{\perp} \Delta y_{t-1}$, we also have from (7) and (A4)

$$(A5) \quad \begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} \Delta z_{2t} &= A'_{\perp} \Psi \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} z_{1t} + A'_{\perp} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} u_t + o_p(1) \\ &\xrightarrow{d} A'_{\perp} \Psi \Theta(1) F B_0(s) + A'_{\perp} B_0(s) = A'_{\perp} [I_n + \Psi \Theta(1) F] B_0(s). \end{aligned}$$

Next, we set $B_1(s) = \Theta(1) F B_0(s)$ and $B_2(s) = A'_{\perp} [I_n + \Psi \Theta(1) F] B_0(s)$. Combining (A3), (A4), and (A5) we have (i) as stated and the covariance matrix of $B(s) = (B_0(s), B_1(s), B_2(s))'$ is given by Ω .

(ii) Inserting (A2) into (7) gives

$$(A6) \quad \Delta y_t = \Psi \Theta(L) F u_{t-1} + u_t = [I_n + \Psi \Theta(L) F L] u_t.$$

Hence, $C(L) = I_n + \Psi \Theta(L) F L$ (cf. (8)), and $C(1) = I_n + \Psi \Theta(1) F$. Therefore, from (A5) we have

$$(A7) \quad B_2(s) = A'_{\perp} C(1) B_0(s).$$

(iii) It is obvious that $\Omega_0 = \Sigma_0 = \Sigma_u$, which is p.d. by assumption. From (A7) we have

$$(A8) \quad \Omega_2 = A'_{\perp} C(1) \Sigma_u C(1)' A_{\perp},$$

which is also p.d. since $R(C(1)) = R(A)^{\perp} = R(A_{\perp})$. The positive definiteness of Σ_1 is proved from (A1) in the same way as Lemma 5.5.5 of Anderson (1971).

Since $\Delta z_{2t} = A'_{\perp} \Delta y_{t-1}$ is a function of the past history of the innovations $\{u_{t-1}, u_{t-2}, \dots\}$ we have $E \Delta z_{2t} u'_{t+j} = 0$ for all $j \geq 0$. Hence $\Sigma_{20} = \Lambda_{20} = 0$.

PROOF OF LEMMA 2: (i) was proved in Lemma 1(i). The rest of Lemma 2 immediately follows from Lemma 1 and from Lemma 2.1 of Park and Phillips (1989) noting that $\Sigma_{20} = \Lambda_{20} = 0$.

PROOF OF LEMMA 3: Define an $n_3 \times g$ matrix K_1 with $\text{rank}(K_1) = g$ and an $n_3 \times (n_3 - g)$ matrix K_2 with $\text{rank}(K_2) = n_3 - g$ such that $R(K_1) = R(A_3)$ and $R(K_2) = R(A_3)^{\perp}$, i.e., $K_2' A_3 = 0$. Let

$\tilde{K}_T = [\sqrt{T}\tilde{K}_1, T\tilde{K}_2]$ with $\tilde{K}_1 = [D \otimes I_{n_3}, i_k \otimes K_1]$ and $\tilde{K}_2 = i_k \otimes K_2$ where i_k is a k -vector of ones. Then, the required results hold with

$$\tilde{R}_{11} = \begin{pmatrix} D'D \otimes S'_3 & e_{k-1} \otimes A_3 \\ 0 & K'_1 A_3 \end{pmatrix}$$

and $\tilde{R}_{22} = K'_2 A_{\perp 3}$.

PROOF OF LEMMA 4: If y_t is not cointegrated, we have the VAR($k-1$) representation in first order differences such that

$$\Delta y_t = \begin{bmatrix} J_{11}^*(L) & J_{12}^*(L) & 0 \\ J_{21}^*(L) & J_{22}^*(L) & J_{23}^*(L) \\ J_{31}^*(L) & J_{32}^*(L) & J_{33}^*(L) \end{bmatrix} \Delta y_{t-1} + u_t.$$

Since $I_n - J^*(L)L$ is invertible, (8) becomes

$$\Delta y_t = [I_n - J^*(L)L]^{-1} u_t, \quad \text{i.e.,} \quad C(L) = [I_n - J^*(L)L]^{-1}.$$

Hence from Lemma 1(ii) we have $B_2(s) = [I_n - J^*(1)]^{-1} B_0(s)$ since $A_{\perp} = I_n$. Thus

$$B_0(s) = [I_n - J^*(1)] \begin{bmatrix} B_b(s) \\ B_a(s) \end{bmatrix}$$

since $B_a(s) = S'_3 B_2(s)$ and $B_b(s) = S'_{1+2} B_2(s)$. Multiplying $(S'_1 \Sigma_a S_1)^{-\frac{1}{2}} S'_1$ on both sides of this last equation, we have

$$W_1(s) = L'_1 W_b(s)$$

where $L'_1 = (S'_1 \Sigma_a S_1)^{-\frac{1}{2}} [I_{n_1} - J_{11}^*(1), -J_{12}^*(1)] \Omega_b^{\frac{1}{2}}$. Note that L'_1 is of full row rank, and that $L'_1 L_1 = I_{n_1}$ since $W_1(s)$ is a standard Brownian motion. Define a nonsingular matrix $\tilde{L} = [L_1, L_2]$ where L_2 is an $(n_1 + n_2) \times n_2$ matrix such that $L'_2 L_1 = 0$ and $L'_2 L_2 = I_{n_2}$. Then we can write

$$W_1(s) = [I_{n_1}, 0] \tilde{L}' W_b(s).$$

Notice that in (13) we may replace $W_b(s)$ with $\tilde{L}' W_b(s)$ which is also a standard Brownian motion independent of $W_a(s)$. Therefore, redefining $W_b(s)$ as $\tilde{L}' W_b(s)$ gives the required result.

PROOF OF LEMMA 5: (i) This is proved in Lemma 8 of Johansen (1988).

(ii) Recall that $z_{1t} = H'_1 x_t$ where x_t and H_1 were defined in Section 2. Define $\tilde{H}_1 = [D \otimes I_n, e_k \otimes A]$. Then $\tilde{z}_{1t} = \tilde{H}'_1 x_t$ and

$$\begin{aligned} \text{(A9)} \quad \tilde{Z}'_1 \tilde{Z}_1 &= \tilde{H}'_1 X' X \tilde{H}_1 \\ &= [(\tilde{H}_1 - H_1)' X' + H'_1 X'] [X(\tilde{H}_1 - H_1) + XH_1] \\ &= Z'_1 Z_1 + (\tilde{H}_1 - H_1)' X' Z_1 + Z'_1 X(\tilde{H}_1 - H_1) + (\tilde{H}_1 - H_1)' X' X(\tilde{H}_1 - H_1) \end{aligned}$$

where $\tilde{H}_1 - H_1 = [0, e_k \otimes (\tilde{A} - A)] = O_p(T^{-1})$ by virtue of (i). By the same argument as (A5) and (A7)

$$T^{-1/2} \sum_{t=1}^{[Ts]} \Delta y_t \xrightarrow{d} C(1) B_0(s).$$

Hence, by Lemma 2.1 of Park and Phillips (1989), $X' Z_1 = O_p(T)$ and $X' X = O_p(T^2)$. Therefore, from (A9)

$$\text{(A10)} \quad T^{-1} \tilde{Z}'_1 \tilde{Z}_1 = T^{-1} Z'_1 Z_1 + o_p(1) \xrightarrow{p} \Sigma_1$$

by Lemma 2(i)(a). Also

$$(A11) \quad (Z_1' - \tilde{Z}_1') \tilde{Z}_1 = (H_1 - \tilde{H}_1)' X' Z_1 + (H_1 - \tilde{H}_1)' X' X (\tilde{H}_1 - H_1) = O_p(1).$$

Furthermore $U' \tilde{Z}_1 = U' Z_1 + U' X (\tilde{H}_1 - H_1)$ where $U' X = O_p(T)$. Hence, by Lemma 2(i)(b)

$$(A12) \quad \frac{1}{\sqrt{T}} U' \tilde{Z}_1 = \frac{1}{\sqrt{T}} U' Z_1 + o_p(1) \\ \xrightarrow{d} N'_0$$

where $\text{vec}(N'_0) \equiv N(0, \Sigma_1^{-1} \otimes \Sigma_u)$.

Now since $\tilde{\Psi} = \Delta Y' \tilde{Z}_1 (\tilde{Z}_1' \tilde{Z}_1)^{-1}$, we have from (A10), (A11), and (A12)

$$\sqrt{T}(\tilde{\Psi} - \Psi) = \frac{1}{\sqrt{T}} U' \tilde{Z}_1 \left(\frac{\tilde{Z}_1' \tilde{Z}_1}{T} \right)^{-1} + \frac{1}{\sqrt{T}} \Psi (Z_1 - \tilde{Z}_1)' \tilde{Z}_1 \left(\frac{\tilde{Z}_1' \tilde{Z}_1}{T} \right)^{-1} \xrightarrow{d} N'_0 \Sigma_1^{-1} = N$$

where $\text{vec}(N') = (\Sigma_1^{-1} \otimes I_n) \text{vec}(N'_0) \equiv N(0, \Sigma_1^{-1} \otimes \Sigma_u)$. Furthermore, N is independent of $(B_2(s)', B_c(s)')$ because $B_2(s)$ and $B_c(s)$ are linear combinations of elements of $B_0(s)$ (see Lemma 1 in Section 2 and Lemma 8 of Johansen (1988)) and N_0 is independent of $B_0(s)$ by Lemma 1(i).

(iii) This is proved in Theorem 3 of Johansen (1988).

(iv) This follows immediately from (ii) and (iii).

(v) This was proved in (ii) above.

(vi) Write

$$\hat{A}_\perp = A \hat{\Xi} + A_\perp \hat{\Pi}_\perp$$

where $\hat{\Xi} = A' \hat{A}_\perp$ and $\hat{\Pi}_\perp = A'_\perp \hat{A}_\perp$. Since each column of \hat{A}_\perp is an eigenvector of the equation (9) of Johansen (1988),

$$\hat{A}_\perp S_{10} S_{00}^{-1} S_{01} \hat{A}_\perp = \hat{A}_\perp S_{11} \hat{A}_\perp \cdot \text{diag}(\hat{\lambda}_{r+1}, \dots, \hat{\lambda}_n) \\ = \text{diag}(\hat{\lambda}_{r+1}, \dots, \hat{\lambda}_n)$$

where S_{ij} ($i, j = 0, 1$) are the product moment matrices defined by (8) of Johansen (1988) with k -lagged level variables replaced by one-lagged levels (i.e., S_{11} , S_{01} , and S_{10} correspond to Johansen's S_{kk} , S_{0k} , and S_{k0} , respectively),⁵ $\hat{\lambda}_j$ ($j = r+1, \dots, n$) are the eigenvalues corresponding to \hat{A}_\perp , and the last equality follows from the normalization condition: $\hat{A}'_\perp S_{11} \hat{A}_\perp = I_{n-r}$. Since $\hat{\lambda}_j$ ($j = r+1, \dots, n$) are $O_p(T^{-1})$ by Lemma 6 of Johansen (1988), \hat{A}_\perp and hence $\hat{\Xi}$ and $\hat{\Pi}_\perp$ are $O_p(T^{-\frac{1}{2}})$.

Also from the normalization condition

$$I_{n-r} = \hat{A}'_\perp S_{11} \hat{A}_\perp \\ = (\hat{\Xi}' A' + \hat{\Pi}'_\perp A'_\perp) S_{11} (A \hat{\Xi} + A_\perp \hat{\Pi}_\perp) \\ = \hat{\Xi}' A' S_{11} A \hat{\Xi} + \hat{\Pi}'_\perp A'_\perp S_{11} A \hat{\Xi} + \hat{\Xi}' A' S_{11} A_\perp \hat{\Pi}_\perp + \hat{\Pi}'_\perp A'_\perp S_{11} A_\perp \hat{\Pi}_\perp.$$

Hence

$$\hat{\Pi}'_\perp A'_\perp S_{11} A_\perp \hat{\Pi}_\perp \xrightarrow{p} I_{n-r}$$

since $A' S_{11} A = O_p(1)$ by Lemma 3 of Johansen (1988) and $A' S_{11} A_\perp = O_p(1)$ similarly. Thus

$$\det(\hat{\Pi}_\perp)^2 \det(A'_\perp S_{11} A_\perp) \xrightarrow{p} 1.$$

⁵ Notice that we have the level variables y_{t-1} in (2), while Johansen formulates the model so as to have y_{t-k} as the level variables. This difference, of course, does not affect the asymptotics.

Since $T^{-1}A'_{\perp}S_{11}A_{\perp} \xrightarrow{d} B_2B_2'$ (in our notation) by Lemma 3 of Johansen (1988), it follows that $\hat{\Pi}_{\perp}^{-1} = O_p(T^{\frac{1}{2}})$ and therefore

$$(A13) \quad \hat{\Xi}\hat{\Pi}_{\perp}^{-1} = O_p(1).$$

$$(a) \quad M'\tilde{A}_{\perp} = M'(A\hat{\Xi}\hat{\Pi}_{\perp}^{-1} + A_{\perp}) = M'A_{\perp}.$$

$$(b) \quad \text{Substituting } \tilde{A}_{\perp} = A\hat{\Xi}\hat{\Pi}_{\perp}^{-1} + A_{\perp},$$

$$\frac{1}{T^2}\tilde{Z}_2'\tilde{Z}_2 = \frac{1}{T^2}(A\hat{\Xi}\hat{\Pi}_{\perp}^{-1} + A_{\perp})'Y'_{-1}Y_{-1}(A\hat{\Xi}\hat{\Pi}_{\perp}^{-1} + A_{\perp})$$

where $Y'_{-1} = [y_0, \dots, y_{T-1}]$. Hence

$$\begin{aligned} \frac{1}{T^2}\tilde{Z}_2'\tilde{Z}_2 &= \frac{1}{T^2}Z_2'Z_2 + \frac{1}{T^2}(\hat{\Xi}\hat{\Pi}_{\perp}^{-1})'A'Y'_{-1}Y_{-1}A_{\perp} + \frac{1}{T^2}A'_{\perp}Y'_{-1}Y_{-1}A\hat{\Xi}\hat{\Pi}_{\perp}^{-1} \\ &\quad + \frac{1}{T^2}(\hat{\Xi}\hat{\Pi}_{\perp}^{-1})'A'Y'_{-1}Y_{-1}A\hat{\Xi}\hat{\Pi}_{\perp}^{-1} \\ &\xrightarrow{d} \int B_2B_2' \end{aligned}$$

by (A13) and Lemma 2.

PROOF OF LEMMA 6: Define an $n_1n_3k \times n_1n_3k$ matrix $K_T^* = [\sqrt{T}K_1^*, TK_2^*]$ with

$$K_1^* = \begin{pmatrix} I_{n_3(k-1)} \otimes I_{n_1} & 0 \\ 0 & K_1 \otimes I_{n_1} \end{pmatrix} \quad \text{and} \quad K_2^* = \begin{pmatrix} 0 \\ K_2 \otimes I_{n_1} \end{pmatrix}$$

where K_1 and K_2 are defined in the proof of Lemma 3.

We give a proof of the result for \tilde{P}^\dagger . The result for \tilde{P} can be proved in an entirely analogous manner. Now

$$K_T^{*'}\tilde{P}^\dagger = \begin{pmatrix} K_1^{*'}\tilde{P}_1 & T^{-\frac{1}{2}}K_1^{*'}\tilde{P}_2^\dagger \\ \sqrt{T}K_2^{*'}\tilde{P}_1 & K_2^{*'}\tilde{P}_2^\dagger \end{pmatrix} T_T^\dagger.$$

Since $\tilde{A}_3 \xrightarrow{p} A_3$, $\sqrt{T}K_1'\tilde{A}_3 = \sqrt{T}K_1'(A_3 - A_3) \xrightarrow{p} 0$, $\tilde{I}_1 \xrightarrow{p} \Gamma_1$ and $K_2'\tilde{A}_{3\perp} = K_2'S_3'\tilde{A}_{\perp} = K_2'S_3'A_{\perp} = K_2'A_{3\perp}$ by Lemma 5(i), (ii), and (vi)(a), we have

$$K_1^{*'}\tilde{P}_1 = \begin{pmatrix} I_{n_3(k-1)} \otimes I_{n_1} & 0 \\ 0 & K_1'\tilde{A}_3 \otimes S_1' \end{pmatrix} \xrightarrow{p} \begin{pmatrix} I_{n_3(k-1)} \otimes I_{n_1} & 0 \\ 0 & K_1'A_3 \otimes S_1' \end{pmatrix} = P_{11},$$

$$T^{-\frac{1}{2}}K_1^{*'}\tilde{P}_2^\dagger \xrightarrow{p} 0,$$

$$\sqrt{T}K_2^{*'}\tilde{P}_1 = (0, \sqrt{T}K_2'\tilde{A}_3 \otimes S_1') \xrightarrow{p} 0,$$

and

$$K_2^{*'}\tilde{P}_2^\dagger = K_2'\tilde{A}_{\perp 3} \otimes \tilde{I}_1 \xrightarrow{p} K_2'A_{3\perp} \otimes \Gamma_1 = P_{22}^\dagger$$

where P_{11} and P_{22}^\dagger are full row rank matrices as required.

REFERENCES

- ANDERSON, T. W. (1971): *The Statistical Analysis of Time Series*. New York: Wiley.
 BRILLINGER, D. R. (1981): *Time Series: Data Analysis and Theory*. San Francisco: Holden-Day.
 CHAN, N. H., AND C. Z. WEI (1988): "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics*, 16, 367-401.

- JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.
- (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.
- PARK, J., AND P. C. B. PHILLIPS (1988): "Statistical Inference in Regressions with Integrated Processes: Part I," *Econometric Theory*, 4, 468–497.
- (1989): "Statistical Inference in Regressions with Integrated Processes: Part II," *Econometric Theory*, 5, 95–131.
- PHILLIPS, P. C. B. (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.
- PHILLIPS, P. C. B., AND S. N. DURLAUF (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 573–496.
- PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.
- SIMS, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.
- SIMS, C. A., J. H. STOCK, AND M. W. WATSON (1990): "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113–144.
- TODA, H. Y., AND P. C. B. PHILLIPS (1991a): "Vector Autoregressions and Causality," Cowles Foundation Discussion Paper No. 977.
- (1991b): "Vector Autoregression and Causality: A Theoretical Overview and Simulation Study," University of Western Australia Working Paper 91-07, to appear in *Econometric Reviews*, 1993.