Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers | Cowles Foundation

5-1-1991

Vector Autoregression and Causality

Hiro Y. Toda

Peter C.B. Phillips

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

Part of the Economics Commons

Recommended Citation

Toda, Hiro Y. and Phillips, Peter C.B., "Vector Autoregression and Causality" (1991). *Cowles Foundation Discussion Papers*. 1220.

https://elischolar.library.yale.edu/cowles-discussion-paper-series/1220

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

Cowles Foundation Discussion Paper No. 977

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

VECTOR AUTOREGRESSION and CAUSALITY

by

H. Y. Toda and P. C. B. Phillips

May 1991

VECTOR AUTOREGRESSIONS AND CAUSALITY*

by

H. Y. Toda and P. C. B. Phillips

Cowles Foundation for Research in Economics
Yale University

February 1991

*Our thanks to Glena Ames for her skill in keyboarding the manuscript and to the NSF for research support under SES 8821180.

0. ABSTRACT

This paper develops a complete 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. In consequence, we conclude that there is no sound asymptotic justification for the empirical use of Granger causality tests in unrestricted levels VAR's.

The situation with regard to the use of causality tests in ECM's is also complex but somewhat more encouraging. We demonstrate that 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. In consequence, we recommend an operational procedure for conducting Granger causality tests in the important practical case of testing the causal effects of one variable on another group of variables and vice versa.

JEL Classification: 211

Key Words: Bessel function; 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 tests of 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 USA. 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 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). This conclusion was expressed in SSW by the following general statement:

> In the causality test, the statistic has a χ^2 distribution if the process is cointegrated; otherwise its asymptotic distribution is nonstandard and must be computed numerically (SSW, p. 115).

Since macroeconomic time series for most countries display nonstationarity and at least some degree of cointegration, many investigators are likely to see this broad conclusion of SSW as encouraging with regard to the continued use of Wald tests for causality as asymptotic χ^2 criteria. Is this optimism warranted? Given the important empirical role of causality tests in levels VAR's, it seems reasonable to us to ask the following questions in addressing this problem: 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; finally, is the optimism just referred to really warranted, i.e. are causality tests in levels VAR's to be recommended at all, given the inferential complications of a multiplicity of different asymptotics and potential nuisance parameter dependencies?

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 and we provide a complete analysis of the asymptotics of Wald tests for causality. 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. In consequence, our analysis of the general case leads us away from an optimistic conclusion. We submit that causality tests in unrestricted levels VAR's are not to be recommended in practice.

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 a general 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, 1989) 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. We demonstrate that, in general, tests for causality in ECM's also suffer from nuisance parameters 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. We therefore suggest an operational procedure for causality tests in ECM's which is applicable in the most important practical case, viz. where we wish to test causal effects of one variable on another group of variables and vice versa.

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 vec(A) to stack the *rows* of a matrix A into a column vector, [x] to denote the smallest 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(Ω) 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

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

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 $\delta > 0$ $(i = 1, ..., n)$. We shall initialize (1) at $t = -k+1, ..., 0$ and allow the initial values $\{y_0, y_{-1}, ..., y_{-k+1}\}$ to be any random vectors including constants. In setting up a likelihood function for data generated from (1) it is, of course, most convenient to require the initial values to be constant, as in Johansen (1988, 1989).

We can write (1) in the equivalent ECM format

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

where $J_* = J(1) - I_n$, and

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

We assume that:

(3a) $|I_n - J(L)L| = 0$ implies $|L| \geq 1$.

(3b) $J_* = \Gamma A'$ where Γ and A are $n \times r$ matrices of full column rank r, $0 \leq r \leq n-1$. (If $r = 0$, then we take $J_* = 0$.)

(3c) $\Gamma_\perp'(J^*(1) - I_n)A_\perp$ is nonsingular, where $\Gamma_\perp$ 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 $r \geq 1$. (See Johansen (1989).)

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 Wold representation in terms of the process $\{u_t - \mu^*\}$ where μ^* is some constant vector, $A'y_t$ is stationary and y_t is in general an I(1) process with a deterministic trend or drift (see Theorem 3.1 of Johansen (1989)). In fact, if $\Gamma_{\perp}'\mu \neq 0$, the process y_t has a deterministic trend as well as a stochastic trend, while if $\Gamma_{\perp}'\mu = 0$ (including the case where $\mu = 0$), there is no deterministic trend in y_t. As is well known, the presence or absence of a deterministic trend affects the asymptotic distribution of parameter estimators and test statistics based on them.

We shall assume that $\mu = 0$ in the following sections 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 $\mu \neq 0$. Thus, we assume that the true model is

$$(1)' \qquad y_t = J(L)y_{t-1} + u_t$$

or in its ECM representation:

$$(2)' \qquad \Delta y_t = J^*(L)\Delta y_{t-1} + \Gamma A' y_{t-1} + u_t \, .$$

We shall report the results for the case where $\mu \neq 0$ only after the models (1)′ and (2)′ have been analyzed in detail.

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

$$y_t = \begin{array}{c} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} \end{array} \begin{array}{l} n_1 \\ n_2 \\ n_3 \end{array}$$

and partition $J(L)$ conformably with y_t. Then we have under the null hypothesis of noncausality the following levels VAR formulation:

$$(1)'' \qquad y_t = \begin{bmatrix} J_{11}(L) & J_{12}(L) & 0 \\ J_{21}(L) & J_{22}(L) & J_{23}(L) \\ J_{31}(L) & J_{32}(L) & J_{33}(L) \end{bmatrix} y_{t-1} + u_t .$$

That is, the null hypothesis can be formulated based on the model (1)′ as

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

where $J_{13}(L) = \sum_{i=1}^{k} J_{i,13} L^{i-1}$. Equivalently, the noncausality null can be formulated as

$$(5) \qquad \mathcal{H}_0^* : J_{1,13}^* = \cdots = J_{k-1,13}^* = 0 \;\; \textit{and} \;\; J_{*13} = 0$$

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

To prepare for the analysis in the following sections we introduce some further notation and a couple of lemmas. Define

$$x_t = \begin{bmatrix} y_{t-1} \\ \vdots \\ y_{t-k} \end{bmatrix}$$

and we can write (1)′ as

$$(7) \qquad y_t = \Phi x_t + u_t$$

where $\Phi = [J_1, \ldots, J_k]$. Define an $nk \times nk$ 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., $(1, 0, \ldots, 0)'$, 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

$$z_{1t} = H_1'x_t = \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-k+1} \\ A'y_{t-1} \end{bmatrix}$$

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. With this notation we can write (2)′ as

$$\Delta y_t = \Psi z_{1t} + u_t \tag{8}$$

where $\Psi = [J_1^*, \ldots, J_{k-1}^*, \Gamma]$. 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.

Furthermore, since Δy_t is $I(0)$, we have the Wold representation

$$\Delta y_t = C(L)u_t \quad \text{where} \quad C(L) = \sum_{i=0}^{\infty} C_i L^i . \tag{9}$$

(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

$$\Sigma = Ew_t w_t' ,$$

$$\Lambda = \sum_{j=1}^{\infty} Ew_t w_{t+j}' ,$$

and

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

We partition Ω , Σ , and Λ 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 .

Now we have the following lemma:

LEMMA 1

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

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) $B_2(s) = A_\perp' C(1) B_0(s)$

(iii) $\Omega_0 = \Sigma_0 = \Sigma_u$ *and* Ω_2 *are p.d.* $\Lambda_{20} = \Sigma_{20} = 0$. □

In the development of the theory below we will also require Σ_1 to be p.d. Since the assumptions which have been made so far do not ensure this, we now assume that Σ_1 is p.d. Note that Σ_1 is the covariance matrix of the stationary component in y_t , so this is a standard assumption.

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

LEMMA 2.

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

(b) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{1t}u_t' \xrightarrow{d} N_0$ *where* $\mathrm{vec}(N_0) = \xi$

(ii)(a) $\frac{1}{T}\sum_{t=1}^{T} z_{2t}u_t' \xrightarrow{d} \int B_2 dB_0'$

(b) $\frac{1}{T}\sum_{t=1}^{T} z_{2t}z_{1t}' \xrightarrow{d} \int B_2 dB_1' + \Sigma_{21} + \Lambda_{21}$

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

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

(10) $$y_t = \hat{\Phi}x_t + \hat{u}_t\ ,\ \ t = 1, \ldots, T$$

where in this section "^" signifies estimation by OLS. The noncausality hypothesis (4) can be written as

(11) $$\mathcal{H}_0 : S_1'\Phi S = 0 \ \text{ or } \ (S_1' \otimes S')\mathrm{vec}(\Phi) = 0$$

where

$$S_1 = \begin{array}{c} \begin{bmatrix} I_{n_1} \\ 0 \end{bmatrix} \end{array} \begin{array}{l} n_1 \\ n - n_1 \end{array}$$

$$S = I_k \otimes S_3 \ \text{ with } \ S_3 = \begin{array}{c} \begin{bmatrix} 0 \\ I_{n_3} \end{bmatrix} \end{array} \begin{array}{l} n - n_3 \\ n_3 \end{array} .$$

Then the Wald statistic for testing (11) can be written

$$F_{LS} = \text{vec}(\hat{\Phi})'(S_1 \otimes S)\Big[(S_1' \otimes S')[\hat{\Sigma}_u \otimes (X'X)^{-1}](S_1 \otimes S)\Big]^{-1}(S_1' \otimes S')\text{vec}(\hat{\Phi})$$
$$= \text{tr}\Big[S_1'\hat{\Phi}S\Big[S'(X'X)^{-1}S\Big]^{-1}S'\hat{\Phi}'S_1(S_1'\hat{\Sigma}_uS_1)^{-1}\Big]$$

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

$$F_{LS} = \text{tr}\Big[S_1'U'X(X'X)^{-1}S\Big[S'(X'X)^{-1}S\Big]^{-1}S'(X'X)^{-1}X'US_1(S_1'\hat{\Sigma}_uS_1)^{-1}\Big]$$
$$= \text{tr}\Big[(S_1'\hat{\Sigma}_uS_1)^{-\frac{1}{2}}S_1'U'Z(Z'Z)^{-1}R\Big[R'(Z'Z)^{-1}R\Big]^{-1}R'(Z'Z)^{-1}Z'US_1(S_1'\hat{\Sigma}_uS_1)^{-\frac{1}{2}}\Big]$$

where $R' = S'H$ and $Z' = (z_1, ..., z_T)$. We write

$$R' = [R_1', R_2']$$

with

$$R_1' = S'H_1 = [D \otimes S_3', e_k \otimes A_3]$$
$$R_2' = S'H_2 = e_k \otimes A_{\perp 3}$$

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

Let $g = \text{rank}(A_3)$. Since A_3 is an $n_3 \times r$ matrix, it must be that $0 \le g \le \min(n_3, r)$. Define a nonsingular matrix $\tilde{K} = [\tilde{K}_a, \tilde{K}_b]$ with

$$\tilde{K}_a = [D \otimes I_{n_3}, i_k \otimes K_a]$$

and

$$\tilde{K}_b = i_k \otimes K_b .$$

Here i_k is a k—vector of ones, and K_a and K_b are $n_3 \times g$ and $n_3 \times (n_3 - g)$ matrices, respectively, such that

$$\mathbb{R}(K_a) = \mathbb{R}(A_3)$$

and

$$\mathbb{R}(K_b) = \mathbb{R}(A_3)^\perp \text{ , i.e., } K_b'A_3 = 0 .$$

Then define

$$\tilde{R}' = \tilde{K}'R' = \begin{bmatrix} \tilde{K}_a'R_1' & \tilde{K}_a'R_2' \\ \tilde{K}_b'R_1' & \tilde{K}_b'R_2' \end{bmatrix} = \begin{bmatrix} \tilde{R}_{11}' & \tilde{R}_{12}' \\ 0 & \tilde{R}_{22}' \end{bmatrix}$$

where

$$\tilde{R}_{11}' = \begin{bmatrix} D'D \otimes S_3' & e_{k-1} \otimes A_3 \\ 0 & K_a'A_3 \end{bmatrix}$$

$$\tilde{R}_{12}' = \begin{bmatrix} e_{k-1} \otimes A_{\perp 3} \\ K_a'A_{\perp 3} \end{bmatrix}$$

and

$$\tilde{R}_{22}' = K_b'A_{\perp 3} \, .$$

Note that $\tilde{R}_{11}'$ is of full row rank since $D'D \otimes S_3'$ and $K_a'A_3$ are of full row rank by construction, and that $\tilde{R}_{22}'$ is of full row rank since otherwise $\tilde{R}'$ could not be of full row rank. ($\tilde{R}'$ must be of full row rank because $\tilde{K}'$ is nonsingular and R' is of full row rank.)

Define

$$\Upsilon_T = \mathrm{diag}(\sqrt{T}\, I_{m_1}, TI_{m_2}) \, ,$$

$$\Upsilon_T^* = \mathrm{diag}(\sqrt{T}\, I_{n_3(k-1)+g}, TI_{n_3-g})$$

and

$$\tilde{R}_T' = \begin{bmatrix} \tilde{R}_{11}' & T^{-1/2}\tilde{R}_{12}' \\ 0 & \tilde{R}_{22}' \end{bmatrix} .$$

Note that $\Upsilon_T^*\tilde{R}' = \tilde{R}_T'\Upsilon_T$. Thus we have

$$F_{LS} = \mathrm{tr}\left[(S_1'\hat{\Sigma}_u S_1)^{-\frac{1}{2}} S_1'U'Z(Z'Z)^{-1}\tilde{R}\Upsilon_T^*\left[\Upsilon_T^*\tilde{R}'(Z'Z)^{-1}\tilde{R}\Upsilon_T^*\right]^{-1}\Upsilon_T^*\tilde{R}'(Z'Z)^{-1}Z'US_1(S_1'\hat{\Sigma}_u S_1)^{-\frac{1}{2}}\right]$$

$$(12) \qquad = \mathrm{tr}\left[(S_1'\hat{\Sigma}_u S_1)^{-\frac{1}{2}} S_1'U'Z\Upsilon_T^{-1}\left[\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1}\right]^{-1}\tilde{R}_T\left[\tilde{R}_T'\left[\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1}\right]^{-1}\tilde{R}_T\right]^{-1}\right.$$

$$\left. \cdot\tilde{R}_T'\left[\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1}\right]^{-1}\Upsilon_T^{-1}Z'US_1(S_1'\hat{\Sigma}_u S_1)^{-\frac{1}{2}}\right] .$$

$$F_{LS(1)} = \mathrm{vec}(\tilde{R}_{11}'\Sigma_1^{-1}N_0S_1)'\left[\tilde{R}_{11}'\Sigma_1^{-1}\tilde{R}_{11}' \otimes S_1'\Sigma_u S_1\right]^{-1}\mathrm{vec}(\tilde{R}_{11}'\Sigma_1^{-1}N_0S_1)$$
$$\equiv \chi^2_{n_1[n_3(k-1)+g]} .$$

On the other hand, more calculation is necessary in order to reduce $F_{LS(2)}$ to a simpler expression. The idea is to rotate coordinates and cast $F_{LS(2)}$ into a canonical form that eliminates as many nuisance parameters as possible.

Define a nonsingular matrix

$$R_*' = \begin{bmatrix} R_a' \\ R_b' \end{bmatrix}$$

where $R_a' = \tilde{R}_{22}'$ and R_b' is an $[(n-r)-(n_3-g)] \times (n-r)$ matrix of full row rank such that $R_b'\tilde{R}_{22} = 0$. Let

$$B_a(s) = R_a'B_2(s) = \tilde{R}_{22}'B_2(s) ,$$
$$B_b(s) = R_b'B_2(s) ,$$

and

$$W_1(s) = \left[S_1'\Sigma_u S_1\right]^{-\frac{1}{2}} S_1'B_0(s) ,$$

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

$$F_{LS(2)} = \mathrm{tr}\Bigg[\int dW_1B_2'R_*\left[\int R_*'B_2B_2'R_*\right]^{-1}R_*'R_a(R_a'R_a)^{-1}$$
$$\cdot\left\{(R_a'R_a)^{-1}R_a'R_*\left[\int R_*'B_2B_2'R_*\right]^{-1}R_*'R_a(R_a'R_a)^{-1}\right\}^{-1}$$
$$\cdot(R_a'R_a)^{-1}R_a'R_*\left[\int R_*'B_2B_2'R_*\right]^{-1}\int R_*'B_2dW_1'\Bigg] .$$

Noting that $(R_a'R_a)^{-1}R_a'R_* = [I_{n_3-g},\ 0]$, this expression reduces to

$$F_{LS(2)} = \mathrm{tr}\left[\int dW_1\underline{B}_a'\left[\int \underline{B}_a\underline{B}_a'\right]^{-1}\int \underline{B}_a dW_1'\right] \tag{15}$$

where

Now we need the following lemma.

LEMMA 3

(i) $\Upsilon_T^{-1} Z'Z \Upsilon_T^{-1} \xrightarrow{d} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \int B_2 B_2' \end{bmatrix}$

(ii) $\Upsilon_T^{-1} Z'U \xrightarrow{d} \begin{bmatrix} N_0 \\ \int B_2 B_2' \end{bmatrix}$. □

Since $\tilde{R}_T' \longrightarrow \mathrm{diag}(\tilde{R}_{11}', \tilde{R}_{22}')$, we have from Lemma 3

$$\tilde{R}_T' \left[\Upsilon_T^{-1} Z'Z \Upsilon_T^{-1} \right]^{-1} \tilde{R}_T \xrightarrow{d} \begin{bmatrix} \tilde{R}_{11}' \Sigma_1^{-1} \tilde{R}_{11} & 0 \\ 0 & \tilde{R}_{22}' (\int B_2 B_2')^{-1} \tilde{R}_{22} \end{bmatrix} \tag{13}$$

and

$$\tilde{R}_T' \left[\Upsilon_T^{-1} Z'Z \Upsilon_T^{-1} \right]^{-1} \Upsilon_T^{-1} Z'U \xrightarrow{d} \begin{bmatrix} \tilde{R}_{11}' \Sigma_1^{-1} N_0 \\ \tilde{R}_{22}' (\int B_2 B_2')^{-1} \int B_2 dB_0' \end{bmatrix} . \tag{14}$$

Thus combining (12), (13), and (14), and taking into account the consistency of $\hat{\Sigma}_u$ (see Park and Phillips (1989) as for the consistency of the OLS estimator), the continuous mapping theorem gives

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

where

$$F_{LS(1)} = \mathrm{tr}\left[S_1' N_0' \Sigma_1^{-1} \tilde{R}_{11} \left[\tilde{R}_{11}' \Sigma_1^{-1} \tilde{R}_{11} \right]^{-1} \tilde{R}_{11}' \Sigma_1^{-1} N_0 S_1 \left[S_1' \Sigma_u S_1 \right]^{-1} \right] ,$$

$$F_{LS(2)} = \mathrm{tr}\left[\left[S_1' \Sigma_u S_1 \right]^{-\frac{1}{2}} S_1' \int dB_0 B_2' \left[\int B_2 B_2' \right]^{-1} \tilde{R}_{22} \left[\tilde{R}_{22}' \left[\int B_2 B_2' \right]^{-1} \tilde{R}_{22} \right]^{-1} \right.$$
$$\left. \cdot \tilde{R}_{22}' \left[\int B_2 B_2' \right]^{-1} \int B_2 dB_0' S_1 \left[S_1' \Sigma_u S_1 \right]^{-\frac{1}{2}} \right] .$$

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

(16) $$\underline{B}_a(s) = B_a(s) - \int B_a B_b' \left[\int B_b B_b'\right]^{-1} B_b(s) .$$

The covariance matrix of the (n−r)−vector Brownian motion $(B_a(s)', B_b(s)')'$ is $R_*'\Omega_2 R_*$, which is p.d. since Ω_2 is p.d. by Lemma 1(iii). We partition this as

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

conformably with $(B_a(s)', B_b(s)')'$. 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 $B_b(s)$. Note that the positive definiteness of $R_*'\Omega_2 R_*$ implies that of the covariance matrix $\Omega_{a\cdot b} = \Omega_a - \Omega_{ab}\Omega_b^{-1}\Omega_{ba}$ of $B_{a\cdot b}(s)$. Substituting $B_a(s) = B_{a\cdot b}(s) + \Omega_{ab}\Omega_b^{-1}B_b(s)$ into (16) 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

$$W_a(s) = \left[\Omega_a - \Omega_{ab}\Omega_b^{-1}\Omega_{ba}\right]^{-\frac{1}{2}} B_{a\cdot b}(s) = \Omega_{a\cdot b}^{-1/2} B_{a\cdot b}(s)$$
$$W_b(s) = \Omega_b^{-1/2} B_b(s)$$

where $W_a(s)$ and $W_b(s)$ are standard Brownian motions independent of each other. Then we can reduce (15) to the canonical form

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

where

(17) $$\underline{W}_a(s) = W_a(s) - \int W_a W_b' \left[\int W_b W_b'\right]^{-1} W_b(s) .$$

Notice that $W_a(s)$ and $W_b(s)$ are independent of each other but not of $W_1(s)$ in general.

Thus, we have the following formal statement of the asymptotic distribution of a general purpose causality test in a levels VAR.

THEOREM 1 *In the model* (1)′ *if* $|I_n - J(L)L| = 0$ *has* n−r *unit roots* $(0 \le r \le n-1)$ *and the remaining roots lie outside the unit circle, and if* $\mathrm{rank}(A_3) = g\ (\le n_3)$, *then under the null hypothesis* (4)

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

where the first and the second terms are independent,

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

and

$$\begin{bmatrix} W_1(s) \\ W_a(s) \\ W_b(s) \end{bmatrix} \equiv BM(\Omega_W) \ \text{ with } \ \Omega_W = \begin{bmatrix} I_{n_1} & \Omega_{1a} & \Omega_{1b} \\ \Omega_{a1} & I_{n_3-g} & 0 \\ \Omega_{b1} & 0 & I_{(n-r)-(n_3-g)} \end{bmatrix}$$

and where

$$\Omega_{a1} = \Omega_{a\cdot b}^{-1/2}[R_a' - \Omega_{ab}\Omega_b^{-1}R_b']\Omega_{20}S_1\left[S_1'\Sigma_u S_1\right]^{-\frac{1}{2}} ,$$

$$\Omega_{b1} = \Omega_b^{-1/2}R_b'\Omega_{20}S_1\left[S_1'\Sigma_u S_1\right]^{-\frac{1}{2}} . \ \square$$

The condition that $\mathrm{rank}(A_3) = g$ has the following interpretation. Suppose that the subvector $\tilde{y}_{1t} = (y_{1t}', y_{2t}')'$ of y_t is cointegrated with $\tilde{r}_1$ linearly independent cointegrating vectors $(0 \le \tilde{r}_1 \le n - n_3 - 1)$. Then there exists some $n \times r$ matrix

$$\tilde{A} = \begin{array}{c} \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{13} \\ 0 & \tilde{A}_{33} \end{bmatrix} \begin{array}{l} n - n_3 \\ n_3 \end{array} \\ \begin{array}{cc} \tilde{r}_1 & r - \tilde{r}_1 \end{array} \end{array}$$

such that $\mathbb{R}(A) = \mathbb{R}(\tilde{A})$. It can be shown that $\mathrm{rank}(\tilde{A}_{33}) = r - \tilde{r}_1 \le n_3$. Since $\mathbb{R}(A_3) = \mathbb{R}(\tilde{A}_{33})$, we deduce that $\mathrm{rank}(A_3) = r - \tilde{r}_1$. Thus, we may say that $g = \mathrm{rank}(A_3)$ is the number of remaining cointegrating vectors that involve some elements of y_{3t} after the cointegrating vectors for $\tilde{y}_{1t}$ are exhausted.

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

$$
\begin{aligned}
y_{1t} &= y_{2t-1} + u_{1t} \\
y_{2t} &= y_{2t-1} + u_{2t} \\
y_{3t} &= y_{3t-1} + u_{3t} \,.
\end{aligned}
$$

Suppose we set the lag length $k = 1$ and $J(L) = B = (b_{ij})$ in (1)′, estimate an unrestricted VAR(1) and test $\mathcal{H}_0 : b_{13} = 0$ (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 $g = 0$ in Theorem 1, which together with $k = 1$ imply the limit

$$F_{LS} \xrightarrow{d} \int dW_1 \underline{W}_a \left[\int \underline{W}_a^2\right]^{-1} \int \underline{W}_a dW_1 \,.$$

Let us now define the correlation matrix (ρ_{ij}) where

$$\rho_{ij} = \sigma_{u_i u_j} / (\sigma_{u_i u_i} \sigma_{u_j u_j})^{1/2} \,.$$

After a little manipulation we find the following formulae for the covariances Ω_{a1} and Ω_{b1} in the matrix Ω_W of the theorem:

$$
\begin{aligned}
\Omega_{a1} &= \left[1 - \rho_{23}^2\right]^{-1/2} (\rho_{31} - \rho_{32}\rho_{21}) = \rho_{13\cdot 2}\left[1 - \rho_{12}^2\right]^{1/2} \,, \\
\Omega_{b1} &= \rho_{12} \,.
\end{aligned}
$$

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}). Sufficient conditions for this limit distribution to be nuisance parameter free are $\rho_{12} = \rho_{13\cdot 2} = 0$ or equivalently $\rho_{12} = \rho_{13} = 0$ i.e. no correlation between u_{1t} , u_{2t} and u_{3t} . Interestingly, these conditions are not necessary. An alternative set of sufficient conditions are $\rho_{12} = \rho_{23} = 0$ and $\rho_{31} = 1$. Then $\Omega_{b1} = 0$ and $\Omega_{a1} = 1$ so that the limit process $W_a(s) = W_1(s)$ a.s. The limit distribution can then be written entirely in terms of the independent Wiener processes $W_1(s)$ and $W_b(s)$. □

If $\text{rank}(A_3) = n_3$, we may take $K_a = I_{n_3}$ and there is no K_b. Hence $\tilde{R}' = [\tilde{R}'_{11}, \tilde{R}'_{12}]$ with $\tilde{R}'_{11}$ of full row rank n_3. Thus, $\tilde{R}'_T \to [\tilde{R}'_{11}, 0]$ and in the calculation following Lemma 3, the nonstandard blocks that involve $\tilde{R}'_{22}$ disappear. Therefore, in this case there is no component $F_{LS(2)}$. That is, we have the usual chi–square asymptotics.

COROLLARY 1.1 *If* $|I_n - J(L)L| = 0$ *has* $n-r$ *unit roots and the remaining roots lie outside the unit circle, and if* $\text{rank}(A_3) = n_3$, *then under the null hypothesis* (4)

$$F_{LS} \xrightarrow{d} \chi^2_{n_1 n_3 k} . \ \Box$$

By Corollary 1.1 together with the remark following Theorem 1, one can say that if there are as many cointegrating vectors involving some elements of y_{3t} as the dimension of the subvector y_{3t} after the cointegrating vectors for $\tilde{y}_{1t}$ are exhausted, then the asymptotic distribution is chi–square. This generalizes the statement made in Sims, Stock and Watson (1990) that 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" (Sims, Stock and Watson, 1990, p. 135, paragraph 3 and footnote 3).

EXAMPLE 2. Let the true model be the following trivariate system with one cointegrating vector that involves all variables (given by the second equation):

$$y_{1t} = y_{1t-1} + u_{1t}$$
$$y_{2t} = y_{1t-1} + y_{3t-1} + u_{2t}$$
$$y_{3t} = y_{3t-1} + u_{3t} .$$

Suppose we proceed as in Example 1 setting $k = 1$ and testing

$$\mathcal{H}_0 : b_{13} = 0 \ \text{(no causal effect of } y_3 \text{ on } y_1\text{)}$$

using the statistic F_{LS} from a fitted VAR(1). In this case it is easy to see that $A_3 = 1$, $g = n_3 = 1$ and Corollary 1.1 applies, so that $F_{LS} \xrightarrow{d} \chi^2_1$.

Next suppose we wish to test the joint hypothesis

$$\mathcal{H}_0 : b_{12} = b_{13} = 0 \ \text{(no causal effect of } y_2 \text{ and } y_3 \text{ on } y_1\text{)}.$$

The cointegrating matrix in this case is

$$A' = (1, -1, 1)$$

and

$$A_3 = \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

which does not have full row rank, so that

$$g = 1 < 2 = n_3 .$$

The limit distribution of the Wald test of $\mathcal{X}_0$ is the mixture

$$F_{LS} \xrightarrow{d} \chi_1^2 + \int dW_1 \underline{W}_a \left[\int \underline{W}_a^2\right]^{-1} \int \underline{W}_a dW_1 .$$

After a little calculation we find

$$\underline{W}_a(s) = W_a(s) - \int W_a W_1 \left[\int W_1^2\right]^{-1} W_1(s)$$

where $W_1(s)$ and $W_a(s)$ are independent Wiener processes. The limit distribution of F_{LS} in this case is therefore nonstandard but nuisance parameter free.

Observe that in this final example the cointegrating relation involves both y_2 and y_3. Thus, there is a linear combination involving both these variables that is stationary. Yet, nevertheless, the limit distribution of the Wald test of causality is nonstandard. This example serves to illustrate that the limit theory is more complex than the discussion in SSW might suggest. In fact, as Corollary 1.1 makes clear, the appropriate sufficient condition for the Wald test to be asymptotically χ^2 is that there are as many cointegrating vectors involving the variables y_3 in (1)'' (that are concerned in the null hypothesis) in a linearly independent way as the dimension of y_3. □

Now suppose that y_t is not cointegrated. Then there is no A and we may take $A_\perp = I_n$ in the above calculation. Hence we have $R_1' = D \otimes S_3'$ and $R_2' = e_k \otimes S_3'$. Obviously we can set $\tilde{K}_a = D \otimes I_{n_3}$ and $\tilde{K}_b = i_k \otimes I_{n_3}$ so that we have $\tilde{R}_{11}' = D'D \otimes S_3'$ and $\tilde{R}_{22}' = S_3'$. Thus we may take $R_b' = [I_{n-n_3}, 0]$.

Furthermore if y_t is not cointegrated, we have the VAR(k—1) representation in first order differences such that

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

corresponding to (1)". Since $I_n - J^*(L)L$ is invertible, (9) becomes

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

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 $B_b(s) = R_b' B_2(s)$ with $R_b' = [I_{n-n_3}, 0]$. Multiplying $(S_1' \Sigma_u S_1)^{-1/2} S_1'$ on both sides of this last equation, we have

$$W_1(s) = K_1' W_b(s)$$

where $K_1' = (S_1' \Sigma_u S_1)^{-1/2} [I_{n_1} - J^*_{11}(1), -J^*_{12}(1)] \Omega_b^{1/2}$. Note that K_1' is of full row rank, and that $K_1' K_1 = I_{n_1}$ since $W_1(s)$ is a standard Brownian motion. Define a nonsingular matrix

$$K = [K_1, K_2]$$

where K_2 is an $(n - n_3) \times n_2$ matrix such that $K_2' K_1 = 0$ and $K_2' K_2 = I_{n_2}$. Then we can write

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

Notice that in (17) we may replace $W_b(s)$ with $K' W_b(s)$ which is also a standard Brownian motion. Therefore, redefining $W_b(s)$ as $K' W_b(s)$ we obtain the following result.

COROLLARY 1.2. *If* y_t *is not cointegrated, i.e.,* $|I_n - J(L)L| = 0$ *has* n *unit roots, then under the null hypothesis* (4)

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

where the first and the second terms are independent,

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

$$\begin{array}{cl} \begin{bmatrix} W_a(s) \\ W_b(s) \end{bmatrix} & \begin{array}{l} n_3 \\ n-n_3 \end{array} \end{array} \equiv BM(I_n) ,$$

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

The above two corollaries show that in two extreme cases the asymptotic distribution of the Wald test is free of nuisance parameters. Thus: (i) If there is "sufficient cointegration with respect to y_{3t}" in the sense that $\mathrm{rank}(A_3) = n_3$, then the asymptotic distribution is $\chi^2_{n_1 n_3 k}$; and (ii) If there is no cointegration, the Wald test statistic converges to a nonstandard but nuisance parameter free distribution, and hence critical values for the test in this case can be conveniently tabulated.

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 $\mathrm{rank}(A_3) < n_3$, then the asymptotic distribution depends on nuisance parameters in a rather complicated manner. Hence, in order to test causality using this approach, estimation of the nuisance parameters would be required and critical values for the particular distribution would need to be calculated by simulation based on those estimated parameters.

In sum, to detect noncausality in a VAR model like (1)′, we need first to know whether y_t is cointegrated or not, and second to see if $\mathrm{rank}(A_3) = n_3$. If there happens to be no cointegration or if there is sufficient cointegration with respect to y_{3t}, then we can apply Corollary 1.2 and Corollary 1.1, respectively. (Of course in the case of no cointegration we can also formulate a $(k-1)^{th}$ order VAR in Δy_t.) But otherwise it is necessary to know $\mathrm{rank}(A_3)$, 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 cannot but conclude that causality tests based on OLS estimation in levels VAR's are far from recommendable. We shall therefore propose an alternative and somewhat more promising procedure based on ML estimation of the model in ECM format in Section 4.

But before proceeding to the next section we extend the above theorem and its corollaries to the model (1) where $\mu \neq 0$. Write the true model (1) as

$$y_t = \mu + \Phi x_t + u_t \tag{18}$$

and consider two types of estimated equations:

$$y_t = \hat{\mu} + \hat{\Phi} x_t + \hat{u}_t \tag{19}$$

and

$$y_t = \hat{\mu} + \hat{\theta} t + \hat{\Phi}_t x_t + \hat{u}_t \tag{20}$$

where a time trend term is included in (20) since some or all elements of y_t may have deterministic trends as well as stochastic trends if $\mu \neq 0$.

Effects on the asymptotics of the presence of deterministic trends in I(1) regressors, and those of the inclusion of a time trend term in estimated equations are discussed in Park and Phillips (1988) in a general framework. Therefore, we briefly report here only the results and the intuition behind them.

We first consider the estimated system of equations (20). Including both a constant and a time trend term as regressors is equivalent to detrending the process $\{y_t\}$ prior to the estimation of (7). Hence the basic components that appear in the asymptotic distribution in Theorem 1 become "detrended Brownian motions." Furthermore this distribution is unchanged whether or not some elements of y_t actually have deterministic time trends. Thus whether $\Gamma_\perp' \mu \neq 0$ or $\Gamma_\perp' \mu = 0$ (including the case that $\mu = 0$), if we estimate the equations (20), then Theorem 1 and Corollary 1.2 still hold with the Brownian motions $W_a(s)$ and $W_b(s)$ in the asymptotic distribution replaced by the "detrended Brownian motions" $\tilde{W}_a(s)$ and $\tilde{W}_b(s)$, respectively, where

$$\tilde{W}_a(s) = W_a(s) - \int W_a \tau' \left[\int \tau\tau'\right]^{-1} \tau(s)$$

with $\tau(s)' = (1,s)$ and $\tilde{W}_b(s)$ is similarly defined. Clearly Corollary 1.1 also holds.

Next suppose that we estimate the system of equations (19). Then the asymptotic distribution is affected by whether or not there actually are deterministic trends. Since the inclusion of a constant term as a regressor is equivalent to demeaning the process $\{y_t\}$ prior to estimation, if $\Gamma_\perp'\mu = 0$, then Theorem 1 and Corollary 1.2 again hold with the Brownian motions $W_a(s)$ and $W_b(s)$ replaced by the "demeaned Brownian motions" $\bar{W}_a(s)$ and $\bar{W}_b(s)$, respectively, where

$$\bar{W}_a(s) = W_a(s) - \int W_a$$

and $\bar{W}_b(s)$ is defined analogously. Of course, Corollary 1.1 holds.

However if $\Gamma_\perp'\mu \neq 0$ and hence some or all elements of y_t have deterministic trends, the extension is not as straightforward as it is in the earlier cases. Since the time trend component is not eliminated by demeaning the data, one of the elements, say the last one, of $\bar{W}_a(s)$ or $\bar{W}_b(s)$ must be replaced by a component corresponding to the time trend, t. More precisely, instead of $W_a(s)$ or $W_b(s)$ define

$$W_{j*}(s) = \begin{bmatrix} W_{j1}(s) \\ s \end{bmatrix} \quad (j = a \text{ or } b)$$

where $W_{j1}(s)$ is an $n_3 - g - 1$ (if $j = a$) or $(n-r) - (n_3 - g) - 1$ (if $j = b$) dimensional standard Brownian motion. Then, for example, if $j = b$, $W_b(s)$ in the distribution of Theorem 1 is replaced by the demeaned $W_{b*}(s)$, i.e., $\bar{W}_{b*}(s)$ where

$$\bar{W}_{b*}(s) = W_{b*}(s) - \int W_{b*},$$

and $W_a(s)$ is replaced by the demeaned Brownian motion $\bar{W}_a(s)$. It depends on the structure of $A_{\perp 3}$ which of $\bar{W}_a(s)$ and $\bar{W}_b(s)$ should be modified to $\bar{W}_{j*}(s)$ $(j = a$ or $b)$.

Corollary 1.1 obviously holds, but Corollary 1.2 is not true any more if $\mu \neq 0$ and (19) is estimated. (Note that if there is no cointegration, nonzero μ always produces deterministic trends in y_t.)

The reason can be explained as follows. Recall that in the argument proceeding to Corollary 1.2, we had $[I_n - J^*(1)]B_2(s) = B_0(s)$ where $I_n - J^*(1)$ is nonsingular. This implies that each element of $B_0(s)$ can be expressed as a linear combination of the elements of $B_2(s)$. Furthermore, because of the restriction that $J^*_{13}(1) = 0$ it turns out that the first n_1 elements of $B_0(s)$, $S_1'B_0(s)$, are collinear with $R_b'B_2(s)$ and this leads to the nuisance parameter free property of the Wald test in the case of no cointegration. However, if there are deterministic trends in y_t, $B_2(s)$ is an $n-1$ dimensional Brownian motion and is equal to $G_0'[I_n - J^*(1)]^{-1}B_0(s)$ where G_0 is an $n \times (n-1)$ matrix such that $G_0'[I_n - J^*(1)]^{-1}\mu = 0$ since only the $(n-1)$-vector $G_0'y_t$ is asymptotically dominated by stochastic trends while y_t itself is dominated by deterministic trends. Therefore, all the elements of $B_0(s)$ cannot be expressed as linear combinations of $B_2(s)$, and hence in general $S_1'B_0(s)$ cannot be collinear with $B_b(s)$.[1] Thus if we estimate (19) when $\mu \neq 0$, then Corollary 1.2 no longer holds.

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 of little practical use. 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, 1989) 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 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

---

[1]We can similarly explain why the asymptotic distribution of Theorem 1 is not free of nuisance parameters in general; if there is cointegration, then $B_2(s) = A_\perp'C(1)B_0(s)$ and hence all elements of $S_1'B_0(s)$ cannot be necessarily expressed as linear combinations of the elements of $B_2(s)$.

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 (1988/1991).)

We deal with the ECM representation of the system given in (2) and (2)′, and estimate the parameters $J_1^*, ..., J_{k-1}^*$, Γ, and A. The asymptotic theory does differ depending on whether or not there are deterministic trends in y_t (i.e., $\Gamma_\perp'\mu = 0$ or not) as is shown in Johansen (1989). The difference in the asymptotic distributions, however, does not affect our results about causality tests obtained below. Therefore, as in the last section, we shall assume for simplicity that $\mu = 0$, i.e., the true model is (2)′.

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.[2] 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 Corollary 1.2 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 $n-r$ eigenvectors[3] 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 $\Psi = [J_1^*, ..., J_{k-1}^*, \Gamma]$ is given by

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

where $\Delta Y' = [\Delta y_1, ..., \Delta y_T]$, and $\hat{Z}_1' = [\hat{z}_{11}, ..., \hat{z}_{1T}]$ with $\hat{z}_{1t} = [\Delta y_{t-1}', ..., \Delta y_{t-k+1}',$

[2]If u_t is not normally distributed, the estimators considered below are not ML estimators any more. Nevertheless we shall continue to refer to them as "ML estimators."

[3]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 4.)

$(\hat{A}'y_{t-1})']'$. In this section the symbol "^" 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

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

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 4.

(i) $T(\tilde{A}-A) \xrightarrow{d} A_\perp \left[\int B_2 B_2'\right]^{-1} \int B_2 dB_c' = A_\perp M_2^{-1} M_c$

where $M_2 = \int B_2 B_2'$, $M_c = \int B_2 dB_c'$, $B_2(s)$ *is the* $n-r$ *dimensional Brownian motion defined in Lemma* 1, $B_c(s)$ *is an* r *dimensional Brownian motion with covariance matrix* $\Omega_c = (\Gamma'\Sigma_u^{-1}\Gamma)^{-1}$, *and* $B_2(s)$ *and* $B_c(s)$ *are independent.*

(ii) $\sqrt{T}(\tilde{\Psi}-\Psi) \xrightarrow{d} N = [\underset{n(k-1)}{N_1} , \underset{r}{N_2}]$

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

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

where $\hat{\Sigma}_u$ *is given by* (7) *or* (12) *of Johansen* (1988).

(iv) $\tilde{\Omega}_c = \left[\tilde{\Gamma}'\hat{\Sigma}_u^{-1}\tilde{\Gamma}\right]^{-1} \xrightarrow{p} \Omega_c$

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

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

(vi) (a) $\underline{A}'\tilde{A}_\perp = O_p(1)$

(b) $T^2(\tilde{Z}_2'\tilde{Z}_2)^{-1} \xrightarrow{d} \left[\int B_2B_2'\right]^{-1} = M_2^{-1}$

where $\tilde{Z}_2' = [\tilde{z}_{21}, ..., \tilde{z}_{2T}]$ *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

(21) $$S_1'\Phi_*S = 0 \quad \text{or} \quad (S' \otimes S_1')\text{vec}(\Phi_*') = 0$$

where $\Phi_* = [J_1^*, ..., J_{k-1}^*, J_*]$. Note that (21) involves some nonlinear restrictions, viz.

$$S_1'J_*S_3 = S_1'\Gamma A'S_3 = \Gamma_1 A_3' = 0$$

where Γ_1 denotes the first n_1 rows of Γ . Since

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

we have

(22) $$\text{vec}(\hat{\Phi}_*' - \Phi_*') = \tilde{P}\tilde{\eta}$$

where $\hat{\Phi}_* = [\hat{J}_1^*, ..., \hat{J}_{k-1}^*, \hat{J}_*]$ with $\hat{J}_* = \hat{\Gamma}\hat{A}'$,

$$\tilde{P} = \begin{array}{c} \left[\begin{array}{c|c|c} I_{n^2(k-1)} & 0 & 0 \\ \hline 0 & A \otimes I_n & I_n \otimes \tilde{\Gamma} \end{array}\right] \begin{array}{l} n^2(k-1) \\ n^2 \end{array} \\ \begin{array}{ccc} n^2(k-1) & \quad nr & \quad nr \end{array} \end{array}$$

and

$$\tilde{\eta} = \begin{bmatrix} \text{vec}(\tilde{\Psi}'-\Psi') \\ \text{vec}(\tilde{A} - A) \end{bmatrix} .$$

Furthermore define

$$\hat{P}_* = \begin{array}{c} \left[\begin{array}{c|c|c} I_{n^2(k-1)} & 0 & 0 \\ \hline 0 & \hat{A} \otimes I_n & \hat{A}_\perp \otimes \hat{\Gamma} \end{array}\right] \begin{array}{l} n^2(k-1) \\ n^2 \end{array} \\ \begin{array}{ccc} n^2(k-1) & nr & (n-r)r \end{array} \end{array}$$

and

$$\hat{\Omega}_* = \begin{array}{c} \begin{array}{cc} n^2(k-1) + nr & (n-r)r \end{array} \\ \left[\begin{array}{c|c} (\hat{Z}_1'\hat{Z}_1)^{-1} \otimes \hat{\Sigma}_u & 0 \\ \hline 0 & (\hat{Z}_2'\hat{Z}_2)^{-1} \otimes \hat{\Omega}_c \end{array}\right] \begin{array}{l} n^2(k-1) + nr \\ (n-r)r \end{array} \end{array}$$

where $\hat{\Omega}_c = (\hat{\Gamma}'\hat{\Sigma}_u^{-1}\hat{\Gamma})^{-1}$, and $\hat{Z}_2' = [\hat{z}_{21}, \ldots, \hat{z}_{2T}]$ with $\hat{z}_{2t} = \hat{A}_\perp' y_{t-1}$.

Then, in order to test the hypothesis (21) we consider the Wald statistic

$$F_{ML} = \text{vec}(\hat{\Phi}_*')' (S \otimes S_1)[(S' \otimes S_1')\hat{P}_*\hat{\Omega}_*\hat{P}_*'(S \otimes S_1)]^-(S' \otimes S_1')\text{vec}(\hat{\Phi}_*') . \tag{23}$$

Note that $(S' \otimes S_1')\hat{P}_*\hat{\Omega}_*\hat{P}_*'(S \otimes S_1)$ might be singular since $\hat{P}_*\hat{\Omega}_*\hat{P}_*'$ is singular and this possibility is accommodated through the use of the generalized inverse in (23).

To analyze the asymptotics of F_{ML} it will be convenient to transform F_{ML} as follows. Define a nonsingular matrix

$$H_0 = \begin{array}{c} \left[\begin{array}{c|c|c|c} I_{n^2(k-1)} & 0 & 0 & 0 \\ \hline 0 & A \otimes I_n & A_\perp \otimes \Gamma & A_\perp \otimes \Gamma_\perp \end{array}\right] \begin{array}{l} n^2(k-1) \\ n^2 \end{array} \\ \begin{array}{cccc} n^2(k-1) & nr & (n-r)r & (n-r)^2 \end{array} \end{array}$$

and $\underline{H}_0 = H_0(H_0'H_0)^{-1}$. Then since $H_0\underline{H}_0' = I_{n^2k}$ and $S\Phi_*'S_1 = 0$ under the null hypothesis, we can write (23) as

$$F_{ML} = \text{vec}(\hat{\Phi}_*' - \Phi_*')' \underline{H}_0 Q[Q'\underline{H}_0'\hat{P}_*\hat{\Omega}_*\hat{P}_*'\underline{H}_0 Q]^- Q'\underline{H}_0' \text{vec}(\hat{\Phi}_*' - \Phi_*') \tag{24}$$

where

$$Q' = (S' \otimes S_1')H_0$$

$$= \begin{array}{cc} \left[\begin{array}{c|c|c|c} I_{k-1} \otimes S_3' \otimes S_1' & 0 & 0 & 0 \\ \hline 0 & A_3 \otimes S_1' & A_{\perp 3} \otimes \Gamma_1 & A_{\perp 3} \otimes \Gamma_{\perp 1} \end{array}\right] & \begin{array}{c} n_1n_3(k-1) \\ n_1n_3 \end{array} \\ \begin{array}{cccc} n^2(k-1) & nr & (n-r)r & (n-r)^2 \end{array} & \end{array}$$

where $\Gamma_{\perp 1}$ is the first n_1 rows of $\Gamma_\perp$.

Let $g = \text{rank}(A_3)$ as before, and $d = \text{rank}(\Gamma_1) \le \min(n_1, r)$. Next define an $n_1 \times d$ matrix L_a and an $n_1 \times (n_1 - d)$ matrix L_b such that

$$\mathbb{R}(L_a) = \mathbb{R}(\Gamma_1)$$

and

$$\mathbb{R}(L_b) = \mathbb{R}(\Gamma_1)^\perp .$$

Now introduce the nonsingular matrix

$$\tilde{L}' = \begin{array}{cc} \left[\begin{array}{c|c} I_{n_1n_3(k-1)} & 0 \\ \hline 0 & K_a' \otimes I_{n_1} \\ \hline 0 & K_b' \otimes L_a' \\ \hline 0 & K_b' \otimes L_b' \end{array}\right] & \begin{array}{l} n_1n_3(k-1) \\ n_1g \\ (n_3 - g)d \\ (n_3 - g)(n_1-d) \end{array} \\ \begin{array}{cc} n_1n_3(k-1) & n_1n_3 \end{array} & \end{array}$$

where K_a and K_b were defined in Section 3. Let $\tilde{Q}' = \tilde{L}'Q'$ where

$$\tilde{Q}' = \begin{array}{cc} \left[\begin{array}{c|c|c|c} I_{k-1} \otimes S_3' \otimes S_1' & 0 & 0 & 0 \\ \hline 0 & K_a'A_3 \otimes S_1' & K_a'A_{\perp 3} \otimes \Gamma_1 & K_a'A_{\perp 3} \otimes \Gamma_{\perp 1} \\ \hline 0 & 0 & K_b'A_{\perp 3} \otimes L_a'\Gamma_1 & K_b'A_{\perp 3} \otimes L_a'\Gamma_{\perp 1} \\ \hline 0 & 0 & 0 & K_b'A_{\perp 3} \otimes L_b'\Gamma_{\perp 1} \end{array}\right] & \begin{array}{l} n_1n_3(k-1) \\ n_1g \\ (n_3-g)d \\ (n_3-g)(n_1-d) \end{array} \end{array}$$

Also define

$$\tilde{Q}_T' = \left[\begin{array}{c|c|c|c} I_{k-1} \otimes S_3' \otimes S_1' & 0 & 0 & 0 \\ \hline 0 & K_a'A_3 \otimes S_1' & (K_a'A_{\perp 3} \otimes \Gamma_1)/\sqrt{T} & (K_a'A_{\perp 3} \otimes \Gamma_{\perp 1})/T \\ \hline 0 & 0 & K_b'A_{\perp 3} \otimes L_a'\Gamma_1 & (K_b'A_{\perp 3} \otimes L_a'\Gamma_{\perp 1})/\sqrt{T} \\ \hline 0 & 0 & 0 & K_b'A_{\perp 3} \otimes L_b'\Gamma_{\perp 1} \end{array}\right]$$

and

$$\begin{array}{cl} Q_*' = \left[\begin{array}{c|c|c} Q_{*1}' & 0 & 0 \\ \hline 0 & Q_{*2}' & 0 \\ \hline 0 & 0 & Q_{*3}' \end{array}\right] & \begin{array}{l} n_1n_3(k-1) + n_1g \\ (n_3 - g)d \\ (n_3 - g)(n_1 - d) \end{array} \\ \quad\quad n^2(k-1)+nr \quad (n-r)r \quad (n-r)^2 & \end{array}$$

where

$$Q_{*1}' = \left[\begin{array}{c|c} I_{k-1} \otimes S_3' \otimes S_1' & 0 \\ \hline 0 & K_a'A_3 \otimes S_1' \end{array}\right],$$

$$Q_{*2}' = K_b'A_{\perp 3} \otimes L_a'\Gamma_1$$

and

$$Q_{*3}' = K_b'A_{\perp 3} \otimes L_b'\Gamma_{\perp 1} .$$

Note that each of Q_{*j}' $(j = 1, 2, 3)$ is of full row rank by construction, and that $\tilde{Q}_T' \to Q_*'$.

With these transformations we may rewrite (24) as

$$F_{ML} = \mathrm{vec}(\hat{\Phi}_*' - \Phi_*')' \underline{H}_0\tilde{Q}[\tilde{Q}'\underline{H}_0'\hat{P}_*\hat{\Omega}_*\hat{P}_*'\underline{H}_0\tilde{Q}]^{-}\tilde{Q}'\underline{H}_0' \mathrm{vec}(\hat{\Phi}_*' - \Phi_*') .$$

Further, since $\mathrm{diag}(\sqrt{T}\, I_{n_1n_3(k-1)+n_1g}, TI_{(n_3-g)d}, T^{3/2}I_{(n_3-g)(n_1-d)})\tilde{Q}' = \tilde{Q}_T\Upsilon_{0T}^*$ where $\Upsilon_{0T}^* = \mathrm{diag}(\sqrt{T}\, I_{n^2(k-1)+nr}, TI_{(n-r)r}, T^{3/2}I_{(n-r)^2})$, we have

(25) $$F_{ML} = \operatorname{vec}(\hat{\Phi}'_* - \Phi'_*)' \underline{H}_0 \Upsilon^*_{0T} \tilde{Q}_T [\tilde{Q}'_T \Upsilon^*_{0T} \underline{H}'_0 \hat{P}_* \hat{\Omega}_* \hat{P}'_* \underline{H}_0 \Upsilon^*_{0T} \tilde{Q}_T]^- \tilde{Q}'_T \Upsilon^*_{0T} \underline{H}'_0 \operatorname{vec}(\hat{\Phi}'_* - \Phi'_*) .$$

Thus, we now consider the limit behavior of $\Upsilon^*_{0T}\underline{H}'_0 \operatorname{vec}(\hat{\Phi}'_* - \Phi'_*)$ and $\Upsilon^*_{0T}\underline{H}'_0\hat{P}_*\hat{\Omega}_*\hat{P}'_*\underline{H}_0\Upsilon^*_{0T}$. First let us look at the former. From (22)

$$\Upsilon^*_{0T}\underline{H}'_0 \operatorname{vec}(\hat{\Phi}'_* - \Phi'_*) = \Upsilon^*_{0T}\tilde{P}_0\tilde{\eta}$$

where

$$\tilde{P}_0 = \underline{H}'_0\tilde{P} = \begin{array}{cl} \left[\begin{array}{c|c|c} I_{n^2(k-1)} & 0 & 0 \\ \hline 0 & I_{nr} & \underline{A}' \otimes \tilde{\Gamma} \\ \hline 0 & 0 & \underline{A}'_\perp \otimes \underline{\Gamma}'\tilde{\Gamma} \\ \hline 0 & 0 & \underline{A}'_\perp \otimes \underline{\Gamma}'_\perp\tilde{\Gamma} \end{array}\right] & \begin{array}{l} n^2(k\text{-}1) \\ nr \\ (n\text{-}r)r \\ (n\text{-}r)^2 \end{array} \\ \begin{array}{ccc} \quad n^2(k-1) & \quad nr & \qquad nr \end{array} & \end{array}$$

and $\underline{\Gamma}' = (\Gamma'\Gamma)^{-1}\Gamma'$ with $\underline{\Gamma}'_\perp$ being defined analogously. Letting $\Upsilon_{0T} = \operatorname{diag}(\sqrt{T}\, I_{n^2(k-1)+nr}, TI_{nr})$, we have by Lemma 4(i) & (ii)

(26) $$\Upsilon^*_{0T}\underline{H}'_0 \operatorname{vec}(\hat{\Phi}'_* - \Phi'_*) = \left[\begin{array}{c|c|c} I_{n^2(k-1)} & 0 & 0 \\ \hline 0 & I_{nr} & (\underline{A}' \otimes \tilde{\Gamma})/\sqrt{T} \\ \hline 0 & 0 & \underline{A}'_\perp \otimes \Gamma'\tilde{\Gamma} \\ \hline 0 & 0 & \underline{A}'_\perp \otimes \Gamma'_\perp\tilde{\Gamma}\sqrt{T} \end{array}\right] \Upsilon_{0T}\tilde{\eta} \xrightarrow{d} P_0\eta$$

where

$$P_0 = \begin{array}{c} \begin{array}{cc} n^2(k-1)+nr & (n-r)r \end{array} \\ \left[\begin{array}{c|c} I_{n^2(k-1)+nr} & 0 \\ \hline 0 & I_{(n-r)r} \\ \hline 0 & I_{n-r} \otimes \underline{\Gamma}_{\perp}' N_2 \end{array}\right] \end{array} \begin{array}{l} \\ n^2(k-1)+nr \\ (n-r)r \\ (n-r)^2 \end{array}$$

and $\eta' = [\mathrm{vec}(N')', \mathrm{vec}(M_2^{-1}M_c)']$.

Next we consider $\Upsilon^*_{0T}\underline{H}_0'\hat{P}_*\hat{\Omega}_*\hat{P}_*'\underline{H}_0\Upsilon^*_{0T}$. Let

$$\hat{V} = \begin{array}{c} \left[\begin{array}{c|c|c} I_{n^2(k-1)} & 0 & 0 \\ \hline 0 & \hat{\Pi}^{-1} \otimes I_n & 0 \\ \hline 0 & 0 & \hat{\Pi}_{\perp}^{-1} \otimes \hat{\Pi}' \end{array}\right] \\ \begin{array}{ccc} n^2(k-1) & nr & (n-r)r \end{array} \end{array} \begin{array}{l} n^2(k-1) \\ nr \\ (n-r)r \\ \\ \end{array}$$

and define

$$\tilde{P}_* = \hat{P}_*\hat{V} = \left[\begin{array}{c|c|c} I_{n^2(k-1)} & 0 & 0 \\ \hline 0 & \tilde{A} \otimes I_n & \tilde{A}_{\perp} \otimes \tilde{\Gamma} \end{array}\right],$$

$$\tilde{\Omega}_* = \hat{V}^{-1}\hat{\Omega}_*(\hat{V}^{-1})' = \left[\begin{array}{c|c} (\tilde{Z}_1'\tilde{Z}_1)^{-1} \otimes \hat{\Sigma}_u & 0 \\ \hline 0 & (\tilde{Z}_2'\tilde{Z}_2)^{-1} \otimes \tilde{\Omega}_c \end{array}\right].$$

Note that $\hat{P}_*\hat{\Omega}_*\hat{P}_*' = \hat{P}_*\hat{V}\hat{V}^{-1}\hat{\Omega}_*(\hat{V}^{-1})'\hat{V}'\hat{P}_*' = \tilde{P}_*\tilde{\Omega}_*\tilde{P}_*'$, and that $\Upsilon^*_{0T}\underline{H}_0'\tilde{P}_*\tilde{\Omega}_*\tilde{P}_*'\underline{H}_0\Upsilon^*_{0T}$ $= \tilde{P}_{*T}\tilde{\Upsilon}_{0T}\tilde{\Omega}_*\tilde{\Upsilon}_{0T}\tilde{P}_{*T}'$ where $\tilde{\Upsilon}_{0T} = \mathrm{diag}(\sqrt{T}\, I_{n^2(k-1)+nr}, TI_{(n-r)r})$ and

$$\tilde{P}_{*T} = \left[\begin{array}{c|c|c} I_{n^2(k-1)} & 0 & 0 \\ \hline 0 & I_{nr} & \underline{A}'\tilde{A}_\perp/\sqrt{T} \otimes \tilde{\Gamma} \\ \hline 0 & \underline{A}_\perp'\tilde{A}\sqrt{T} \otimes \underline{\Gamma}' & I_{n-r} \otimes \underline{\Gamma}'\tilde{\Gamma} \\ \hline 0 & \underline{A}_\perp'\tilde{A}T \otimes \underline{\Gamma}_\perp' & I_{n-r} \otimes \underline{\Gamma}_\perp'\tilde{\Gamma}\sqrt{T} \end{array}\right]$$

since $\underline{A}'\tilde{A} = \underline{A}'\hat{A}(A'\hat{A})^{-1}A'A = I_r$ and $\underline{A}_\perp'\tilde{A}_\perp = \underline{A}_\perp\hat{A}_\perp(A_\perp'\hat{A}_\perp)^{-1}A_\perp'A_\perp = I_{n-r}$. Hence, from Lemma 4(i), (ii) & (vi),

$$\tilde{P}_{*T} \xrightarrow{d} P_* = \begin{array}{c} \left[\begin{array}{c|c|c} I_{n^2(k-1)} & 0 & 0 \\ \hline 0 & I_{nr} & 0 \\ \hline 0 & 0 & I_{(n-r)r} \\ \hline 0 & M_2^{-1}M_c \otimes \underline{\Gamma}_\perp' & I_{n-r} \otimes \underline{\Gamma}_\perp'N_2 \end{array}\right] \begin{array}{l} n^2(k\text{-}1) \\ nr \\ (n\text{-}r)r \\ (n\text{-}r)^2 \end{array} \\ \begin{array}{ccc} n^2(k-1) & \quad nr \quad & (n-r)r \end{array} \end{array}$$

and

$$\tilde{\Upsilon}_{0T}\tilde{\Omega}_*\tilde{\Upsilon}_{0T} \xrightarrow{d} \Omega_* = \left[\begin{array}{c|c} \Sigma_1^{-1} \otimes \Sigma_u & 0 \\ \hline 0 & M_2^{-1} \otimes \Omega_c \end{array}\right].$$

Therefore, we have

(27) $$\Upsilon_{0T}^*\underline{H}_0'\hat{P}_*\hat{\Omega}_*\hat{P}_*'\underline{H}_0\Upsilon_{0T}^* \xrightarrow{d} P_*\Omega_*P_*'.$$

Thus, from (26) and (27)

(28) $$\tilde{Q}_T'\Upsilon_{0T}^*\underline{H}_0' \operatorname{vec}(\hat{\Phi}_*' - \Phi_*') \xrightarrow{d} Q_*'P_0\eta$$

and

(29) $$\tilde{Q}_T'\Upsilon_{0T}^*\underline{H}_0'\hat{P}_*\hat{\Omega}_*\hat{P}_*'\underline{H}_0\Upsilon_{0T}^*\tilde{Q}_T \xrightarrow{d} Q_*'P_*\Omega_*P_*'Q_*.$$

Note that $Q_*'P_*\Omega_*P_*'Q_*$ is not necessarily nonsingular. For example, suppose that $k = 1$ and $r = 1$, and that n is even and $n_1 = n_3 = n/2$. Then $Q_*'P_*$ is an $n^2/4 \times (2n-1)$ matrix. If $n \geq 8$, $Q_*'P_*$ has more rows than columns, and hence $Q_*'P_*\Omega_*P_*'Q_*$ cannot be nonsingular. If $Q_*'P_*\Omega_*P_*'Q_*$ is not nonsingular, we cannot apply the continuous mapping theorem directly to obtain the asymptotic distribution of (25) from (28) and (29). Furthermore, even if $Q_*'P_*\Omega_*P_*'Q_*$ is nonsingular, the asymptotic distribution of (25) can be shown to depend on nuisance parameters Σ_1, Σ_u, and Ω_c in general. See Example 4 below for an illustration of this.

Thus we assume that $g = \text{rank}(A_3) = n_3$ or $d = \text{rank}(\Gamma_1) = n_1$. (In the operational procedure we shall suggest below, one or the other of these assumptions will be made and tested when $n_1 = 1$ or $n_3 = 1$.) First suppose that $g < n_3$ and $d = n_1$. Then from the definition of Q_*, we have

$$
Q_*' = \begin{array}{c} \left[\begin{array}{c|c|c} Q_{*1}' & 0 & 0 \\ \hline 0 & Q_{*2}' & 0 \end{array}\right] \begin{array}{l} n_1n_3(k-1) + n_1g \\ (n_3 - g)n_1 \end{array} \\ \begin{array}{ccc} n^2(k-1)+nr & (n-r)r & (n-r)^2 \end{array} \end{array}
$$

where Q_{*1}' and Q_{*2}' are of full row rank. Hence

$$
(30) \qquad Q_*'P_*\Omega_*P_*'Q_* = \left[\begin{array}{c|c} Q_{1*}'(\Sigma_1^{-1} \otimes \Sigma_u)Q_{1*} & 0 \\ \hline 0 & Q_{2*}'(M_2^{-1} \otimes \Omega_c)Q_{2*} \end{array}\right]
$$

which is nonsingular. Also

$$
(31) \qquad Q_*'P_0\eta = \left[\begin{array}{c} Q_{1*}' \ \text{vec}(N') \\ \hline Q_{2*}' \ \text{vec}(M_2^{-1}M_c) \end{array}\right].
$$

Thus, combining (25), (28), (29), (30), and (31) we obtain by the continuous mapping theorem

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

where

$$F_{ML(1)} = \mathrm{vec}(N')' Q_{1*}\left[Q'_{1*}(\Sigma_1^{-1} \otimes \Sigma_u)Q_{1*}\right]^{-1} Q'_{1*}\ \mathrm{vec}(N') ,$$

$$F_{ML(2)} = \mathrm{vec}(M_2^{-1}M_c)' Q_{2*}\left[Q'_{2*}(M_2^{-1} \otimes \Omega_c)Q_{2*}\right]^{-1} Q'_{2*}\ \mathrm{vec}(M_2^{-1}M_c) ,$$

and $F_{ML(1)}$ and $F_{ML(2)}$ are independent since N is independent of M_2 and M_c by Lemma 4(ii).

Because $\mathrm{vec}(N') \equiv N(0, \Sigma_1^{-1} \otimes \Sigma_u)$, we easily see that $F_{ML(1)} \equiv \chi^2_{n_1n_3(k-1)+n_1g}$. As for $F_{ML(2)}$, note that by the same argument as that of Lemma 5.1 of Park and Phillips (1988)

$$\left[Q'_{2*}(M_2^{-1} \otimes \Omega_c)Q_{2*}\right]^{-\frac{1}{2}} Q'_{2*}\ \mathrm{vec}(M_2^{-1}M_c) \equiv N(0, I_{n_1(n_3-g)})$$

since $B_2(s)$ and $B_c(s)$ are independent by Lemma 4(i). Hence, $F_{ML(2)} \equiv \chi^2_{n_1(n_3-g)}$. Since $F_{ML(1)}$ and $F_{ML(2)}$ are independent, we deduce that

$$F_{ML} \xrightarrow{d} \chi^2_{n_1n_3k} .$$

Next suppose that $g = n_3$ and $d \leq n_1$. Then there will be neither Q_{2*} nor Q_{3*} , and hence a similar calculation to the above gives

$$F_{ML} \xrightarrow{d} \mathrm{vec}(N')Q_{1*}\left[Q'_{1*}(\Sigma_1^{-1} \otimes \Sigma_u)Q_{1*}\right]^{-1} Q'_{1*}\ \mathrm{vec}(N')$$

which also has a $\chi^2_{n_1n_3k}$ distribution.

We summarize these results in our next theorem.

THEOREM 2. *Suppose that in the model* (1)′ *(or equivalently* (2)′*)* $|I_n - J(L)L| = 0$ *has* $n-r$ *unit roots* $(1 \leq r \leq n-1)$ *and the remaining roots lie outside the unit circle. If* $\mathrm{rank}(A_3) = n_3$ *or* $\mathrm{rank}(\Gamma_1) = n_1$ *, then under the null hypothesis* (4) *(or equivalently* (5)*)*

$$F_{ML} \xrightarrow{d} \chi^2_{n_1n_3k} . \ \square$$

Unfortunately, even the ML method does not always guarantee the usual chi-square asymptotics because the rank condition in Theorem 2 is not always satisfied. To illustrate the problem that arises when there is rank deficiency both in A_3 and Γ_1 we provide the following example.

EXAMPLE 3. Consider the trivariate cointegrated system

$$\Delta y_t = \gamma\alpha' y_{t-1} + u_t \tag{32}$$

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 $\alpha' = (\alpha_1, \alpha_2, \alpha_3) = (1, -1/2, 0)$. 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

$$\gamma_1\alpha_3 = 0, \tag{33}$$

and the Wald statistic given by (23) becomes

$$F_{ML} = \frac{(\hat{\gamma}_1\hat{\alpha}_3)^2}{\hat{\alpha}_3^2\hat{\sigma}_{u1}(\hat{Z}_1'\hat{Z}_1)^{-1} + \hat{\gamma}_1^2\hat{\omega}_c\hat{\alpha}_{\perp 3}(\hat{Z}_2'\hat{Z}_2)^{-1}\hat{\alpha}_{\perp 3}'}$$

where $\sigma_{u1} = \mathrm{var}(u_{1t})$ and $\hat{Z}_1' = (\hat{\alpha}' y_0, \ldots, \hat{\alpha}' y_{T-1})$. Though we could use the formulae given earlier, it is easier to proceed as follows, taking advantage of the simplicity of the model. We write

$$F_{ML} = \frac{1}{\dfrac{T(\tilde{Z}_1'\tilde{Z}_1)^{-1}\hat{\sigma}_{u1}}{(\sqrt{T}\,\tilde{\gamma}_1)^2} + \dfrac{T^2\alpha_{\perp 3}(\tilde{Z}_2'\tilde{Z}_2)^{-1}\alpha_{\perp 3}'\tilde{\omega}_c}{(T\tilde{\alpha}_3)^2}}$$

because $\tilde{\alpha}_{\perp 3} = S_3'\tilde{A}_\perp = S_3'(A_\perp \underline{A}_\perp'\tilde{A}_\perp + \alpha\underline{\alpha}'\tilde{A}_\perp) = \alpha_{\perp 3} + \alpha_3\underline{\alpha}'\tilde{A}_\perp = \alpha_{3\perp}$. (For $\underline{A}_\perp'\tilde{A}_\perp = I_2$ and $\alpha_3 = 0$.) Since by Lemma 4 $\hat{\sigma}_{u1} \xrightarrow{p} \sigma_{u1}$, $\tilde{\omega}_c \xrightarrow{p} \omega_c$, $T(\tilde{Z}_1'\tilde{Z}_1)^{-1} \xrightarrow{p} \sigma_1^{-1}$, $T^2(\tilde{Z}_2'\tilde{Z}_2)^{-1} \xrightarrow{d} M_2^{-1}$, $\sqrt{T}\,\tilde{\gamma}_1 = \sqrt{T}(\tilde{\gamma}_1 - \gamma_1) \xrightarrow{d} N(0, \sigma_{u1}\sigma_1^{-1})$, and $T\tilde{\alpha}_3 = T(\tilde{\alpha}_3 - \alpha_3) \xrightarrow{d} \alpha_{\perp 3}M_2^{-1}M_c$, we have

$$F_{ML} \xrightarrow{d} \frac{1}{\chi_a^{-1} + \chi_b^{-1}} = \frac{\chi_a\chi_b}{\chi_a + \chi_b} \tag{34}$$

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_1^2 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 that

$\gamma_1 \alpha_3 = 0$, 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 $\tilde{\gamma}_1 - \gamma_1 = O_p(T^{-1/2})$, $\tilde{\alpha}_3 - \alpha_3 = O_p(T^{-1})$, and $(\tilde{\gamma}_1 - \gamma_1)(\tilde{\alpha}_3 - \alpha_3) = O_p(T^{-3/2})$. 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 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.

The following theorem gives the analytic form of the density of the limit variate in (34).

THEOREM 3. *If* $Y = \chi_a \chi_b/(\chi_a + \chi_b)$ *where* χ_a *and* χ_b *are independent chi–square variates, both with degrees of freedom* n , *then the density of* Y *is given by:*

(35) $$\mathrm{pdf}(y) = 2^{1-n}\Gamma(n/2)^{-2} e^{-y} y^{n-1} \Sigma_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix} K_{-k+n/2}(y) .$$

When $n = 1$ *as in* (34) *we have:*

(36) $$\mathrm{pdf}(y) = (2/\pi) e^{-y} K_{1/2}(y) .$$

In (35) *and* (36) $K_\nu(y)$ *denotes a Macdonald function or modified Bessel function of the third kind with parameter* ν *(c.f. Erdelyi* (1953), *p.* 5*).*

The density (36) is graphed in Figure 1 against that of a χ_1^2 variate. As is apparent from the figure, the density of (36) 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 (33) using a critical value obtained from (36) 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 employed a conventional Wald test in this case would be unaware of both the size distortion in the use of nominal χ^2 critical values and the resulting power loss. □

EXAMPLE 4. Consider the trivariate system with two linearly independent cointegrating vectors

$$\Delta y_t = \Gamma A' y_{t-1} + u_t$$

where $\Gamma' = (0, \gamma_2', \gamma_3')$ and $A' = (\alpha_1', \alpha_2', 0)$ are 2×3 matrices. Again in this example lower case letters denote scalars and vectors.

Suppose that as in the last example we want to test the hypothesis

$$\gamma_1 \alpha_3' = 0 \, .$$

This time we shall apply the formulae given earlier. Since $\alpha_3 = \gamma_1 = 0$, we have from (28) and (29)

$$Q_*' P_0 \eta = \alpha_{3\perp} \gamma_{1\perp} \underline{\gamma}_\perp' N_2 m_c' m_2^{-1}$$

and

$$Q_*' P_* \Omega_* P_*' Q_* = (\alpha_{3\perp} \gamma_{1\perp})^2 m_2^{-1} m_c \Sigma_1^{-1} m_c' m_2^{-1} \underline{\gamma}_\perp' \Sigma_u \underline{\gamma}_\perp + (\alpha_{3\perp} \gamma_{1\perp})^2 m_2^{-1} \underline{\gamma}_\perp' N_2 \Omega_c N_2' \underline{\gamma}_\perp \, .$$

Note that since there are no lagged differences in the system, the covariance matrix of N_2 is $\Sigma_u \otimes \Sigma_1^{-1}$. Now define

$$\zeta_a = \Omega_c^{-1/2} m_c' m_2^{-1/2} = \Omega_c^{-1/2} \int dB_c B_2 \left[\int B_2^2 \right]^{-1/2} \equiv N(0, I_2)$$

and

$$\zeta_b = \Sigma_1^{1/2} N_2' \underline{\gamma}_\perp (\underline{\gamma}_\perp' \Sigma_u \underline{\gamma}_\perp)^{-1/2} \equiv N(0, I_2)$$

where ζ_a and ζ_b are independent. Then

$$Q_*' P_0 \eta = \alpha_{3\perp} \gamma_{1\perp} (\underline{\gamma}_\perp' \Sigma_u \underline{\gamma}_\perp)^{1/2} \zeta_b' \Sigma_1^{-1/2} \Omega_c^{1/2} \zeta_a m_2^{-1/2}$$

and

$$\begin{aligned} Q_*' P_* \Omega_* P_*' Q_* = {} & (\alpha_{3\perp} \gamma_{1\perp})^2 m_2^{-1/2} \zeta_a' \Omega_c^{1/2} \Sigma_1^{-1} \Omega_c^{1/2} \zeta_a m_2^{-1/2} (\underline{\gamma}_\perp' \Sigma_u \underline{\gamma}_\perp) \\ & + (\alpha_{3\perp} \gamma_{1\perp})^2 m_2^{-1} (\underline{\gamma}_\perp' \Sigma_u \underline{\gamma}_\perp)^{1/2} \zeta_b' \Sigma_1^{-1/2} \Omega_c^{1/2} \Sigma_1^{-1/2} \zeta_b (\underline{\gamma}_\perp' \Sigma_u \underline{\gamma}_\perp)^{1/2} \, . \end{aligned}$$

Therefore

$$F_{ML} \xrightarrow{d} \frac{\left[\zeta_b' \Sigma_1^{-1/2} \Omega_c^{1/2} \zeta_a \right]^2}{\zeta_a' \Omega_c^{1/2} \Sigma_1^{-1} \Omega_c^{1/2} \zeta_a + \zeta_b' \Sigma_1^{-1/2} \Omega_c \Sigma_1^{-1/2} \zeta_b} \, ,$$

which shows that the asymptotic distribution is nuisance parameter dependent when $r = 2$ and $\alpha_3 = \gamma_1 = 0$ in the above trivariate system. Notice that in this last expression if r is equal to one, then ζ_a, ζ_b, Σ_1^{-1}, and Ω_c are scalar, and the limit distribution reduces to the one given in Example 3. □

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, but will necessarily complicate the inference procedure, as we now discuss.

In particular, condition (i) or (ii) can be easily tested if $n_1 = 1$ or $n_3 = 1$, respectively. For example, suppose that $n_1 = 1$ and $n_3 \geq 1$. Then Γ_1 is an r–vector, and hence condition (i) is equivalent to $\Gamma_1 \neq 0$. Furthermore, notice that Γ_1 being a zero vector implies that $J_{*13} = \Gamma_1 A_3' = 0$ in (5). Therefore if $n_1 = 1$, we propose the following (sequential) testing procedure.

(I) Test the hypothesis that $\Gamma_1 = 0$ by

$$F_{ML}^{I} = \hat{\Gamma}_1 \hat{\Sigma}_\gamma^{-1} \hat{\Gamma}_1' / \hat{\sigma}_{u1}$$

where $\hat{\Sigma}_\gamma$ is the $r \times r$ lower–right block of $(\hat{Z}_1'\hat{Z}_1)^{-1}$. F_{ML}^{I} converges in distribution to χ_r^2 under the null hypothesis that $\Gamma_1 = 0$.

(II) If the above hypothesis is rejected, then test the hypothesis (5) using the statistic (23). If not rejected, then test the hypothesis

$$J_{1,13}^* = \cdots = J_{k-1,13}^* = 0 \qquad (37)$$

by the Wald statistic

$$F_{ML}^{II} = \hat{\phi}_+'[S_+'\hat{\Sigma}_+S_+]^{-1}\hat{\phi}_+/\hat{\sigma}_{u1}$$

where $\phi_+' = [J_{1,13}^*, \ldots, J_{k-1,13}^*]$, $S_+ = I_{k-1} \otimes S_3$, and $\hat{\Sigma}_+$ is the $n(k-1) \times n(k-1)$ upper-left block of $(\hat{Z}_1'\hat{Z}_1)^{-1}$. F_{ML}^{II} has an asymptotic $\chi^2_{n_3(k-1)}$ distribution under the null hypothesis (37).

Note that the convergence results of F_{ML}^{I} and F_{ML}^{II} stated above follow immediately from Lemma 4. In the case of $n_3 = 1$, we can proceed with a similar testing procedure.

In most applications of causality tests we can expect that $n_1 = 1$ (and probably $n_3 = 1$, also). The above testing procedure therefore should be useful, although some loss of power in relation to an exact test such as that based on the limit (34) will be inevitable when there is composite rank deficiency as in Example 3.

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:

(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 (1988/1991)) 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 recommend against the empirical use of Granger causality tests in levels VAR's when there are stochastic trends and the possibility of cointegration. Next

(iii) If there is no cointegration the Wald test statistic for causality has a nonstandard but nuisance parameter free limit distribution. This distribution could conceivably 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 is still unsatisfactory in general. Our main results are as follows.

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

(v) Problems of nuisance parameter dependencies and nonstandard distributions enter the limit theory in the general case. Both these problems compromise the validity of conventional theory. These results may be considered surprising and deserving of some emphasis in view of the fact that other types of Wald test in ECM's are known to be asymptotically valid χ^2 tests.

(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.

(vii) For the special but important case of testing when there is either one causal variable or one dependent variable of interest, a sequential operational procedure for inference is suggested.

We conclude that Granger causality tests in systems of stochastic difference equations are fraught with many complications when there are stochastic trends and cointegration in the system. Neither levels VAR models nor systems formulated in ECM format lead to a satisfactory basis of inference. But since ML estimation of ECM's delivers optimal estimates of the cointegration space it would seem that ECM's provide a more promising basis than VAR's for the sequential inference procedures that are needed to adequately test causality hypotheses in these models.

APPENDIX

PROOF OF LEMMA 1

(i) From (2)′ we have

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

where

$$G = \left[\begin{array}{c|c|c} J_1^*, \cdots & J_{k-1}^* & \Gamma \\ \hline I_{n(k-2)} & 0 & 0 \\ \hline A'J_1^*, \cdots & A'J_{k-1}^* & A'\Gamma + I_r \end{array}\right]$$

and

$$F = \begin{bmatrix} e_{k-1} \otimes I_n \\ A' \end{bmatrix}.$$

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

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

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

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 $\Sigma_1 = Ez_{1t}z_{1t}'$ $= \Sigma_{j=0}^{\infty} G^j F \Sigma_u F' G^{j'}$. $B_0(s)$ and ξ are independent.

Since $\Theta(L)$ is the inverse of $I_n - GL$ and $|I_n - GL| = 0$ 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 $\|\Theta_j\| = \mathrm{tr}(\Theta_j \Theta_j')^{1/2}$. Thus by the multivariate extension of Theorem 3.3 of Phillips and Solo (1989)

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

Since $\Delta z_{2t} = A_\perp' \Delta y_{t-1}$, we also have from (8) and (A4)

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

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 (8) gives

$$\text{(A6)} \qquad \begin{aligned} \Delta y_t &= \Psi \Theta(L) F u_{t-1} + u_t \\ &= [I_n + \Psi \Theta(L) F L] u_t . \end{aligned}$$

Hence, $C(L) = I_n + \Psi \Theta(L) F L$ (cf. (9)), and $C(1) = I_n + \Psi \Theta(1) F$. Therefore, from (A5) we have

(A7) $$B_2(s) = A_\perp' C(1) B_0(s) .$$

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

(A8) $$\Omega_2 = A_\perp' C(1) \Sigma_u C(1)' A_\perp ,$$

which is also p.d. since $\mathbb{R}(C(1)) = \mathbb{R}(A)^\perp = \mathbb{R}(A_\perp)$.

Since $\Delta z_{2t} = A_\perp' \Delta y_{t-1}$ is a function of the past history of the innovations $\{u_{t-1}, u_{t-2}, \ldots\}$ we have

$$E\Delta z_{2t} u_{t+j}' = 0 \quad \text{for all} \quad j \geq 0 .$$

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

(i) $$\Upsilon_T^{-1} Z' Z \Upsilon_T^{-1} = \begin{bmatrix} T^{-1} \sum_{t=1}^{T} z_{1t} z_{1t}' & T^{-3/2} \sum_{t=1}^{T} z_{1t} z_{2t}' \\ T^{-3/2} \sum_{t=1}^{T} z_{2t} z_{1t}' & T^{-2} \sum_{t=1}^{T} z_{2t} z_{2t}' \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \int B_2 B_2' \end{bmatrix} ,$$

and

$$\text{(ii)} \qquad \Upsilon_T^{-1} Z'U = \begin{bmatrix} T^{-1/2} \sum_{t=1}^{T} z_{1t} u_t' \\ T^{-1} \sum_{t=1}^{T} z_{2t} u_t' \end{bmatrix}$$

$$\text{(A9)} \qquad \xrightarrow{d} \begin{bmatrix} N_0 \\ \int B_2 dB_0' \end{bmatrix}$$

by Lemma 2.

PROOF OF LEMMA 4

(i) This is proved in Lemma 8 of Johansen (1988).

(ii) Recall that $z_{1t} = H_1' x_t$ where x_t is given by (6). Define

$$\tilde{H}_1 = [D \otimes I_n, e_k \otimes \tilde{A}] .$$

Then $\tilde{z}_{1t} = \tilde{H}_1' x_t$ and

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

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

$$\text{(A10)} \qquad \begin{aligned} T^{-1} \tilde{Z}_1' \tilde{Z}_1 &= T^{-1} Z_1' Z_1 + o_p(1) \\ &\xrightarrow{p} \Sigma_1 \end{aligned}$$

by Lemma 2(i)(a). Also

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

Furthermore

$$U'\tilde{Z}_1 = U'Z_1 + U'X(\tilde{H}_1 - H_1)$$

where $U'X = O_p(T)$ similarly. Hence, by (A9) or Lemma 2(i)(b)

$$\text{(A12)} \qquad \frac{1}{\sqrt{T}}U'\tilde{Z}_1 = \frac{1}{\sqrt{T}}U'Z_1 + o_p(1) \xrightarrow{d} N_0'$$

where $\mathrm{vec}(N_0) \equiv N(0, \Sigma_1 \otimes \Sigma_u)$.

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 $\mathrm{vec}(N') = (\Sigma_1^{-1} \otimes I_n)\mathrm{vec}(N_0) \equiv N(0, \Sigma_1^{-1} \otimes \Sigma_u)$. 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 $\hat{\Xi} = \underline{A}'\hat{A}_\perp$ and $\hat{\Pi}_\perp = \underline{A}_\perp'\hat{A}_\perp$.

(a) Since each column of $\hat{A}_\perp$ is an eigenvector of the equation (9) of Johansen (1988),

$$\begin{aligned}\hat{A}_\perp' S_{10} S_{00}^{-1} S_{01} \hat{A}_\perp &= \hat{A}_\perp S_{11} \hat{A}_\perp \cdot \mathrm{diag}(\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_n) \\ &= \mathrm{diag}(\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_n)\end{aligned}$$

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),[4] $\hat{\lambda}_j$ $(j = r+1, \ldots, n)$ 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 $\hat{\lambda}_j$ $(j = r+1, \ldots, n)$ are $O_p(T^{-1})$ by Lemma 6 of Johansen (1988), $\hat{A}_\perp$ and hence $\hat{\Xi}$ and $\hat{\Pi}_\perp$ are $O_p(T^{-1/2})$.

Also from the normalization condition

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

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

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

Since $T^{-1} A_\perp' S_{11} A_\perp \xrightarrow{d} \int B_2 B_2'$ (in our notation) by Lemma 3 of Johansen (1988), it follows that $\hat{\Pi}_\perp^{-1} = O_p(T^{1/2})$.

Therefore $\underline{A}' \tilde{A}_\perp = \hat{\Xi} \hat{\Pi}_\perp^{-1} = O_p(1)$.

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

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

[4]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.

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

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

by our Lemma 2.

PROOF OF THEOREM 3. Write $y = z_a z_b/(z_a + z_b)$ where the joint density of (z_a, z_b) is

$$\mathrm{pdf}(z_a, z_b) = c\ \exp\{-(1/2)(z_a + z_b)\}z_a^{n/2-1}z_b^{n/2-1}$$

and $c = \left[2^n\Gamma(n/2)^2\right]^{-1}$. We transform variables as $(z_a, z_b) \to (s,y)$ where $z_a = s+y$ and $z_b = y + y^2/s$ with $y \geq 0$, $s > 0$. The jacobian is $((s+y)/s)^2$ and, hence,

$$\begin{aligned}\mathrm{pdf}(y,s) &= c\ \exp\{-(1/2)(s + 2y + y^2/s)\}(s+y)^{n/2-1}y^{n/2-1}((s+y)/s)^{n/2-1}((s+y)/s)^2\\ &= ce^{-y}y^{n/2-1}\exp\{-(1/2)(s + y^2/s)\}(s+y)^n s^{-n/2-1}\\ &= ce^{-y}y^{n/2-1}\Sigma_{k=0}^n\begin{bmatrix}n\\k\end{bmatrix}\exp\{-(1/2)(s + y^2/s)\}s^{k-n/2-1}y^{n-k}\ .\end{aligned}\tag{A13}$$

Next observe the following integral representation for Macdonald's function $K_\nu(y)$, i.e. the modified Bessel function of the third kind, see Erdelyi (1953, pp. 5 and 82, formula (23)):

$$K_\nu(y) = (1/2)y^\nu\int_0^\infty \exp\{-(1/2)(s + y^2/s)\}s^{-\nu-1}ds\ .\tag{A14}$$

Integrating (A13) with respect to s and using (A14) we deduce that:

$$\begin{aligned}\mathrm{pdf}(y) &= ce^{-y}y^{n/2-1}\Sigma_{k=0}^n\begin{bmatrix}n\\k\end{bmatrix}[2K_{-k+n/2}(y)y^{k-n/2}]y^{n-k}\\ &= 2ce^{-y}y^{n/2-1}\Sigma_{k=0}^n\begin{bmatrix}n\\k\end{bmatrix}K_{-k+n/2}(y)y^{n/2}\\ &= 2^{1-n}\Gamma(n/2)^{-2}e^{-y}y^{n-1}\Sigma_{k=0}^n\begin{bmatrix}n\\k\end{bmatrix}K_{-k+n/2}(y)\end{aligned}$$

giving (35). When $n = 1$ we have

$$\begin{aligned}\mathrm{pdf}(y) &= \Gamma(1/2)^{-2} e^{-y}\{K_{1/2}(y) + K_{-1/2}(y)\} \\ &= (2/\pi) e^{-y} K_{1/2}(y)\end{aligned}$$

since $K_\nu(y) = K_{-\nu}(y)$ (e.g. Erdelyi (1953), p. 5, formula (14)), leading to (36).

REFERENCES

[1] Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory.* San Francisco: Holden-Day.

[2] Chan, N. H. and C. Z. Wei (1988). "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics* 16, pp. 367—401.

[3] Erdelyi, A. (1953). *Higher Transcendental Functions*, Vol II. New York: MacGraw—Hill.

[4] Johansen, S. (1988). "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control* 12, pp. 231—254.

[5] Johansen, S. (1989). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," University of Copenhagen, Institute of Mathematical Statistics Preprint No. 3.

[6] Johansen, S. and K. Juselius (1989). "Maximum Likelihood Estimation and Inference on Cointegration — with Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics* 52, pp. 169—210.

[7] Park, J. (1988). "Canonical Cointegrating Regressions," Cornell University, CAE Working Paper No. 88—29.

[8] Park, J. and P. C. B. Phillips (1988). "Statistical Inference in Regressions with Integrated Processes: Part I," *Econometric Theory* 4, pp. 468—497.

[9] Park, J. and P. C. B. Phillips (1989). "Statistical Inference in Regressions with Integrated Processes: Part II," *Econometric Theory* 5, pp. 95—131.

[10] Phillips, P. C. B. (1988/1991). "Optimal Inference in Cointegrated Systems," *Econometrica*, forthcoming.

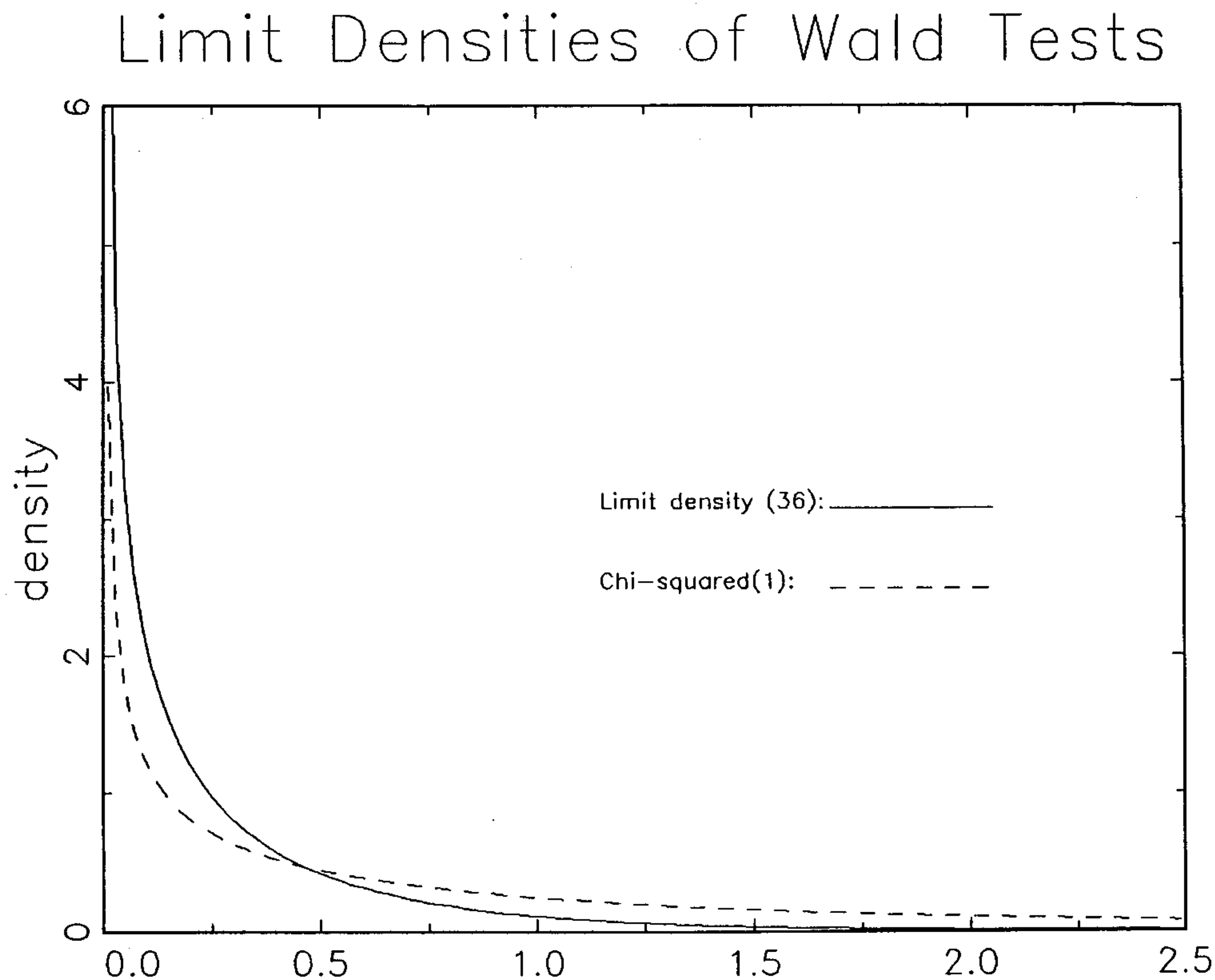
[11] Phillips, P. C. B. and S. N. Durlauf (1986). "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies* 53, pp. 473—496.

[12] Phillips, P. C. B. and B. E. Hansen (1990). "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies* 57, pp. 99—125.

[13] Phillips, P. C. B. and V. Solo (1989). "Asymptotics for Linear Processes," Cowles Foundation Discussion Paper No. 932.

[14] Sims, C. A. (1980). "Macroeconomics and Reality," *Econometrica* 48, pp. 1—48.

[15] Sims, C. A., J. H. Stock and M. W. Watson (1990). "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica* 58, pp. 113—144.



Limit Densities of Wald Tests
density
Limit density (36):
Chi-squared(1):
0
2
4
6
0.0
0.5
1.0
1.5
2.0
2.5

Figure 1