Cowles Foundation Paper 854

Journal of Econometrics 59 (1993) 229–255. North-Holland

The spurious effect of unit roots on vector autoregressions

An analytical study

Hiro Y. Toda and Peter C.B. Phillips*

Yale University, New Haven, CT 06520, USA

Received May 1991, final version received May 1992

This paper analyzes whether inclusion of a statistically independent random walk in a vector autoregression can result in spurious inference. The problem was raised originally by Ohanian (1988). In a Monte Carlo simulation based on the VAR's estimated by Sims (1980, 1982), Ohanian found that 'block exogeneity' of the genuine variables with respect to an artificially generated random walk variable was rejected too often. In the present paper we attempt a full analytical study of this problem. It can be shown that if the genuine variables are nonstationary, the Wald statistic for testing the 'block exogeneity' hypothesis does not have the usual asymptotic chi-square distribution. The derived asymptotic distribution is free of nuisance parameters so that we can unambiguously determine the effect of including the random walk. Some simulated critical values for the asymptotic distribution are reported. Interestingly, it can also be shown that if the genuine variables of the model are stationary, the asymptotic distribution is still chi-square in spite of the inclusion of the random walk.

1. Introduction

Vector autoregressions (VAR's) have been used in a wide variety of econometric applications. Although most economic time series are believed to be nonstationary and difficulties in dealing with levels of such time series are well known [e.g., Granger and Newbold (1974) and Phillips (1986)], several recent studies in this field have analyzed potentially nonstationary data without detrending or differencing. Some prominent examples are Lawrence and Siow (1985), Litterman and Weiss (1985), and Sims (1980, 1982).

Correspondence to: Hiro Y. Toda, Institute of Socio-Economic Planning, University of Tsukuba, Tsukuba-shi, Ibaraki 305, Japan.

*The authors are grateful to D.W.K. Andrews, C.A. Sims, two anonymous referees, and an associate editor for their helpful comments on an earlier version. The authors also thank Glena Ames for typing the completed version of the manuscript of this paper.

0304-4076/93/$06.00

Ohanian (1988) questioned whether the use of nonstationary data in VAR's can result in spurious inferences, and he conducted a simulation study based on the empirical VAR's estimated in Sims (1980, 1982). Ohanian added an artificially generated random walk (RW) variable to the Sims' VAR model of money, real output, aggregate prices, and interest rate, and estimated the resulting five-variable VAR. Ohanian's simulations showed that 'block exogeneity' of the genuine variables with respect to the independent RW was rejected too often. This study uses actual data in conjunction with the generated RW and therefore suffers from two potential drawbacks: (i) The model is not necessarily the true data-generating mechanism; and (ii) the observed effects are conditioned on the particular realization of the empirical time series.

A general asymptotic theory for inference in multiple linear regressions with integrated processes (i.e., processes generated by ARIMA type models) has recently been developed by Park and Phillips (1988, 1989), Sims, Stock, and Watson (1990), and Tsay and Tiao (1990) among others. [See Phillips (1988) for a review of methods and results on this topic.] Sims, Stock, and Watson (1990) concentrated on VAR's and derived, as an example, a nonstandard asymptotic distribution of the Wald test statistic for a Granger noncausality hypothesis in a trivariate VAR model. Though it is closely related to Ohanian's problem, their expression for the distribution is complex and involves nuisance parameters in general so that we cannot see either the direction or size of the bias caused by nonstationarity from their results.

This paper provides a full analytical study of the problem raised by Ohanian (1988) using the methodology developed by Park and Phillips (1988, 1989). In fact, it can be shown that the Wald statistic that is of central interest in testing 'block exogeneity' has an asymptotic distribution which is free of nuisance parameters. This distribution can be computed numerically and the effect of the generated RW on inference can be determined unambiguously. For illustration, we report some percentage points of the distribution based on computer simulation.

The plan of this paper is as follows. In section 2 the model for analyzing the Ohanian's 'exogeneity test' is presented. We consider two cases: (i) the case in which the genuine variables are I(1), i.e., integrated of order one, and (ii) the case in which they are I(0), i.e., stationary. The derivation of the asymptotic distribution of the Wald statistic is given in section 3. The required treatment and the results will differ in each of the above cases. In section 4 we discuss the results obtained and make some concluding remarks. Proofs of the lemmas we need in the body of the paper are given in appendix A.

2. The model

Following Ohanian (1988), consider the n-vector time series $\{y_t\}$ generated by the pth-order VAR model

$$y_t = \alpha + A(L)y_{t-1} + u_t, \tag{1}$$

where $A(L) = \sum_{j=1}^{p} A_j L^{j-1}$, $\{u_t = (u_{1t}, \ldots, u_{nt})'\}$ is an i.i.d. sequence of n-dimensional random vectors with mean zero and covariance matrix Σ_u such that $\mathrm{E}|u_{it}|^{2+d} < \infty$ for some $d > 0$. Σ_u is assumed to be a positive definite matrix. y_t may be I(0) or I(1), and if I(1), it may be cointegrated. Let $\{\xi_t\}$ be a RW[1] generated by

$$\xi_t = \xi_{t-1} + \varepsilon_t, \tag{2}$$

where $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean zero and variance σ_ε^2 such that $\mathrm{E}|\varepsilon_t|^{2+d} < \infty$ for some $d > 0$.[2] $\{\varepsilon_t\}$ is assumed to be independent of $\{u_t\}$.

Suppose that an econometrician estimates the regression equation

$$y_t = \hat{\alpha} + \hat{A}(L)y_{t-1} + \hat{\beta}(L)\xi_{t-1} + \hat{u}_t, \tag{3}$$

where $t = 1, \ldots, T$, $\beta(L) = \sum_{j=1}^{p} \beta_j L^{j-1}$, and symbol '^' signifies 'estimated'. The lag length (p) is assumed to be specified correctly. Suppose further that the econometrician wants to know if y_t is 'block exogenous' in the $n + 1$ variable system $(y_t', \xi_t)'$ and tests the hypothesis

$$\beta_1 = \cdots = \beta_p = 0. \tag{4}$$

That is, the lagged ξ's do not 'cause' y_t in the Granger sense.[3]

Our question is whether this econometrician can correctly infer 'block exogeneity' of y_t with respect to ξ_t by appealing to conventional asymptotics for Wald tests. In order to answer this question, Ohanian generated ξ_t by a Monte Carlo simulation and used for y_t the post-war U.S. data on money, real output, price level, and interest rate. He found that the null hypothesis (4) was rejected too often.[4] Here we shall provide an analytic study and derive an asymptotic distribution of the Wald statistic[5] for testing the hypothesis (4).

[1] It does not change the results in this paper if we take $\{\xi_t\}$ to be a general vector I(1) process with innovations that are stochastically independent of u_t. However, we shall assume ξ_t to be a scalar, independent RW following the Ohanian model.

[2] Assumptions on the innovations u_t and ε_t could be weakned by allowing for martingale differences. Subsequent analysis would differ only in terms of the central limit theory we utilize in our asymptotics.

[3] Exogeneity is, of course, not equivalent to Granger noncausality. So, the term 'exogeneity test' is not very appropriate. However, we use this term throughout the paper, following Ohanian and others before him.

[4] He also found a moderate effect on the system's relative variance decomposition. But since the least squares estimators of the coefficients and the covariance matrix in nonstationary regressions are consistent, this observation is best interpreted as a small-sample or data-conditioning effect.

[5] Ohanian used the likelihood ratio test statistic, which is asymptotically equivalent to the Wald statistic that we consider in our setting.

Define

$$\bar{x}_t' = (y_{t-1}', \dots, y_{t-p}', \xi_{t-1}, \dots, \xi_{t-p}),$$

which is an $(n+1)p$-vector, and write (3) as

$$y_t = \hat{\alpha} + \hat{\Pi}\bar{x}_t + \hat{u}_t,$$

where $\Pi = [A_1, \dots, A_p, \beta_1, \dots, \beta_p]$. Then, the hypothesis (4) becomes

$$\Pi R = 0 \quad \text{or} \quad (I_n \otimes R')\text{vec}(\Pi) = 0, \tag{5}$$

where I_g is a $g \times g$ identity matrix for any integer g,

$$R = \begin{bmatrix} 0 \\ I_p \end{bmatrix} \quad (n+1)p \times p,$$

and vec(·) is the vectorization operator that stacks the *rows* of the argument matrix. Since inclusion of constant terms in the regressions is equivalent to demeaning the data prior to estimation, the Wald statistic of interest with respect to testing (4) can be written as

$$\begin{aligned}\mathscr{W} &= \text{vec}(\hat{\Pi})'(I_n \otimes R)[(I_n \otimes R')[\hat{\Sigma}_u \otimes (\bar{X}'Q_1\bar{X})^{-1}](I_n \otimes R)]^{-1} \\ &\quad \times (I_n \otimes R')\text{vec}(\hat{\Pi}) \\ &= \text{tr}[\hat{\Pi}R[R'(\bar{X}'Q_1\bar{X})^{-1}R]^{-1}R'\hat{\Pi}'\hat{\Sigma}_u^{-1}],\end{aligned}$$

where $Q_1 = I_T - T^{-1}\iota_T\iota_T'$ (ι_g is a g-vector of ones for any integer g), $\hat{\Sigma}_u$ is the least squares estimator of Σ_u, and $\bar{X}' = (\bar{x}_1, \dots, \bar{x}_T)$.

The asymptotic distribution of the Wald statistic and its derivation will differ depending on whether y_t is I(1) or I(0). Thus, we need to consider the two cases[6] separately:

1. $|I_n - A(L)L| = 0$ has at least one unit root and the rest of the roots lie outside the unit circle.
2. All of the roots of $|I_n - A(L)L| = 0$ lie outside the unit circle.

Once the asymptotics for case 1 are derived, however, it is a straightforward task to obtain the corresponding results for case 2. Hence, we shall discuss the former in detail first and later give only a brief explanation for the latter case. Note that case 1 allows for cointegration among the variables in the vector y_t if $n \geq 2$.

[6] We exclude by assumption the possibility that $|I_n - A(L)L| = 0$ has a root inside the unit circle.

3. Large-sample asymptotics

3.1. The nonstationary case

In this subsection we assume that the sequence $\{y_t\}$ is I(1) and may be cointegrated with k linearly independent cointegrating vectors where $0 \leq k \leq n-1$. Let C be an $n \times k$ matrix of the cointegrating vectors. Then, we can write (1) in an error correction model format as

$$\Delta y_t = \alpha + A^*(L)\Delta y_{t-1} + \Gamma C' y_{t-1} + u_t, \tag{6}$$

where

$$A^*(L) = \sum_{j=1}^{p-1} A_j^* L^{j-1} \quad \text{with} \quad A_j^* = -\sum_{i=j+1}^{p} A_i,$$

and Γ is an $n \times k$ matrix of full column rank such that $\Gamma C' = A(1) - I_n$. If $k = 0$, there is no C and $\{y_t\}$ has a VAR representation in first-order differences.

Let G and S be $n \times (n-k)$ matrices of full column rank such that $\mathbb{R}(G) = \mathbb{R}(C)^{\perp}$ and $\mathbb{R}(S) = \mathbb{R}(\Gamma)^{\perp}$, where $\mathbb{R}(\cdot)$ denotes the range space of the argument matrix and $\mathbb{R}(\cdot)^{\perp}$ is its orthogonal complement. Define the n-vector process $\{\eta_t\}$ as $\eta_t = y_t - \mu - \delta t$, where

$$\delta = G[S'\{I_n - A^*(1)\}G]^{-1}S'\alpha, \tag{7}$$

$$\mu = \boldsymbol{C}\boldsymbol{\Gamma}'[\{I_n - A^*(1)\}\delta - \alpha], \tag{8}$$

with $\boldsymbol{C} = C(C'C)^{-1}$ and $\boldsymbol{\Gamma} = \Gamma(\Gamma'\Gamma)^{-1}$. [See Theorem 4.1 of Johansen (1991).] Substituting $\Delta y_t = \delta + \Delta\eta_t$ and $C'y_t = C'\mu + C'\eta_t$ into (6), we have

$$\Delta\eta_t = \alpha - \{I_n - A^*(1)\}\delta + \Gamma C'\mu + A^*(L)\Delta\eta_{t-1} + \Gamma C'\eta_{t-1} + u_t,$$

where

$$\begin{aligned}\alpha - \{I_n - A^*(1)\}\delta + \Gamma C'\mu &= (I_n - \Gamma\boldsymbol{\Gamma}')[\alpha - \{I_n - A^*(1)\}\delta] \\ &= S\boldsymbol{S}'[\alpha - \{I_n - A^*(1)\}\delta] = 0,\end{aligned}$$

with $\boldsymbol{S} = S(S'S)^{-1}$. Therefore, y_t may be written as

$$y_t = \mu + \delta t + \eta_t, \tag{9}$$

where η_t satisfies

$$\Delta\eta_t = A^*(L)\Delta\eta_{t-1} + \Gamma C'\eta_{t-1} + u_t. \tag{10}$$

Note from (9) and (10) that y_t contains a linear time trend as well as stochastic trend, unless $\alpha \in \mathbb{R}(\Gamma)$. If $\alpha \in \mathbb{R}(\Gamma)$, then $\delta = 0$ [see (7)] and y_t has only a stochastic trend as pointed out by Johansen (1991). In the following derivation of the asymptotic distribution, we assume $\alpha \notin \mathbb{R}(\Gamma)$, but the case $\alpha \in \mathbb{R}(\Gamma)$ will be discussed later in this subsection.

Unlike regressions with stationary regressors, $T^{-1}\sum_1^T \bar{x}_t \bar{x}_t'$ does not converge to a positive definite matrix. Hence, we need the following transformation to separate each component in $\bar{x}_t$ of different stochastic order of magnitude, so that the sample moment matrix converges properly when it is standardized appropriately.

Since $\delta \in \mathbb{R}(G)$ [see (7)], there is an $n \times (n-k-1)$ matrix G_0 such that $\mathbb{R}([G_0, \delta]) = \mathbb{R}(G)$ and $\delta' G_0 = 0$. We define the matrices:

$$D = \begin{bmatrix} 1 & & & & 0 \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ 0 & & & & -1 \end{bmatrix}, \qquad p \times (p-1),$$

$$H_1 = \left[\begin{array}{cc|c} D \otimes I_n & e_p \otimes C & 0 \\ \hline 0 & 0 & D \end{array}\right], \qquad m \times m_1,$$

$$H_2 = \left[\begin{array}{c|c} e_p \otimes G_0 & 0 \\ \hline 0 & e_p \end{array}\right], \qquad m \times m_2,$$

$$h_3 = \left[\begin{array}{c} e_p \otimes \delta(\delta'\delta)^{-1} \\ \hline 0 \end{array}\right], \qquad m \times 1,$$

where $m = (n+1)p$, $m_1 = (n+1)(p-1) + k$, $m_2 = n - k$, $e_p = (1, 0, \ldots, 0)$ which is a p-vector, and $H = [H_1, H_2, h_3]$ is nonsingular.

Next define $\bar{z}_t = H'\bar{x}_t = (\bar{z}_{1t}', z_{2t}', z_{3t})'$, i.e.,

$$\bar{z}_{1t} = H_1'\bar{x}_t = \begin{bmatrix} \delta + \Delta\eta_{t-1} \\ \vdots \\ \delta + \Delta\eta_{t-p+1} \\ C'\mu + C'\eta_{t-1} \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-p+1} \end{bmatrix}, \qquad z_{2t} = H_2'\bar{x}_t = \begin{bmatrix} z_{2at} \\ z_{2bt} \end{bmatrix},$$

where $z_{2at} = G_0'\eta_{t-1}$ and $z_{2bt} = \xi_{t-1}$, and

$$z_{3t} = h_3'\bar{x}_t = t - 1 + (\delta'\delta)^{-1}\delta'\eta_{t-1}. \tag{11}$$

Here we have used the fact that $G_0'\mu = 0$, $C'\delta = 0$, $\delta'\mu = 0$, and $G_0'\delta = 0$. [See (7) and (8).] We also define

$$z_{1t}' = (\Delta\eta_{t-1}', \ldots, \Delta\eta_{t-p+1}', C'\eta_{t-1}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-p+1}),$$

and $z_t = (z_{1t}', z_{2t}', z_{3t})'$. Note that $Q_1Z_1 = Q_1\bar{Z}_1$ where $Z_1' = (z_{11}, \ldots, z_{1T})$ and $\bar{Z}_1' = (\bar{z}_{11}, \ldots, \bar{z}_{1T})$. Note also that in (11) the first term dominates asymptotically. The variates z_{1t}, z_{2t}, and z_{3t} are the basic components that appear in the calculation of the asymptotic distribution of $\mathscr{W}$. We next analyze the asymptotic behavior of these variates and their sample moment matrices.

Define

$$w_t = \begin{bmatrix} u_t \\ z_{1t} \\ \Delta z_{2t} \end{bmatrix},$$

and set

$$\Sigma = \mathrm{E}w_tw_t',$$

$$\Lambda = \sum_{j=1}^{\infty} \mathrm{E}w_tw_{t+j}',$$

$$\Omega = \Sigma + \Lambda + \Lambda'. \tag{12}$$

We partition Ω, Σ, and Λ conformably with w_t. For instance,

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

Let '$\to_d$' signify 'convergence in distribution' and let $[s]$ denote the integer part of the real number s. Here and throughout the paper all limits are taken as T tends to ∞. We start our asymptotic analysis with the following preliminary lemma:

Lemma 1

(i) $$\begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} w_t \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (z_{1t} \otimes u_t) \end{bmatrix} \to_{\mathrm{d}} \begin{bmatrix} B(r) \\ v \end{bmatrix} = \begin{bmatrix} B_0(r) \\ B_1(r) \\ B_2(r) \\ v \end{bmatrix} \begin{matrix} n \\ m_1 \\ m_2 \\ nm_1 \end{matrix},$$

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

(ii) *Write*

$$B_2(r) = \begin{bmatrix} B_{2a}(r) \\ B_{2b}(r) \end{bmatrix} \begin{matrix} n-k-1 \\ 1 \end{matrix}.$$

Then $B_{2b}(r)$ is independent of $(B_0(r)', B_{2a}(r)')'$, and $B_{2a}(r) = K'B_0(r)$ where K is some $n \times (n-k-1)$ matrix of full column rank.

(iii) $\Lambda_{20} = \Sigma_{20} = 0$, $\Omega_0 = \Sigma_0 = \Sigma_u$ *which is positive definite, and Ω_2 and Σ_1 are positive definite.* □

The next lemma follows from Lemma 1 above and Lemma 2.1 of Park and Phillips (1989). Let '$\to_{\mathrm{p}}$' and '$\equiv$' denote 'convergence in probability' and 'equivalence in distribution', respectively.

Lemma 2

(i) (*a*) $$\frac{1}{T} \sum_{t=1}^{T} z_{1t} z_{1t}' \to_{\mathrm{p}} \Sigma_1,$$

(*b*) $$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{1t} u_t' \to_{\mathrm{d}} N \quad \textit{where} \quad \mathrm{vec}(N) = v \equiv \mathrm{N}(0, \Sigma_1 \otimes \Sigma_u);$$

(ii) (*a*) $$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \to_{\mathrm{d}} \int_0^1 \mathrm{d}B_0(r) = B_0(1),$$

(*b*) $$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{1t} \to_{\mathrm{d}} \int_0^1 \mathrm{d}B_1(r) = B_1(1),$$

(c) $\frac{1}{T^{3/2}} \sum_{t=1}^{T} z_{2t} \to_d \int_0^1 B_2(r)\mathrm{d}r;$

(iii) (a) $\frac{1}{T} \sum_{t=1}^{T} z_{2t}u_t' \to_d \int_0^1 B_2(r)\mathrm{d}B_0(r)',$

(b) $\frac{1}{T} \sum_{t=1}^{T} z_{2t}z_{1t}' \to_d \int_0^1 B_2(r)\mathrm{d}B_1(r)' + \Sigma_{21} + \Lambda_{21};$

(iv) (a) $\frac{1}{T^2} \sum_{t=1}^{T} z_{2t}z_{2t}' \to_d \int_0^1 B_2(r)B_2(r)'\mathrm{d}r;$

(v) (a) $\frac{1}{T^{3/2}} \sum_{t=1}^{T} z_{3t}u_t' \to_d \int_0^1 r\mathrm{d}B_0(r)',$

(b) $\frac{1}{T^{3/2}} \sum_{t=1}^{T} z_{3t}z_{1t}' \to_d \int_0^1 r\mathrm{d}B_1(r)',$

(c) $\frac{1}{T^{5/2}} \sum_{t=1}^{T} z_{3t}z_{2t}' \to_d \int_0^1 rB_2(r)'\mathrm{d}r.$

Joint convergence of all the above also applies. □

Now we are ready to analyze the asymptotics of $\mathscr{W}$. Since $\hat{\Pi} - \Pi = U'Q_1\bar{X}(\bar{X}'Q_1\bar{X})^{-1}$ where $U' = (u_1, \ldots, u_T)$ and $Q_1\bar{Z}_1 = Q_1Z_1$, we have

$$\begin{aligned}\mathscr{W} &= \mathrm{tr}[U'Q_1\bar{X}(\bar{X}'Q_1\bar{X})^{-1}R[R'(\bar{X}'Q_1\bar{X})^{-1}R]^{-1} \\ &\quad \times R'(\bar{X}'Q_1\bar{X})^{-1}\bar{X}'Q_1U\hat{\Sigma}_u^{-1}] \\ &= \mathrm{tr}[U'Q_1Z(Z'Q_1Z)^{-1}P[P'(Z'Q_1Z)^{-1}P]^{-1}P'(Z'Q_1Z)^{-1}Z'Q_1U\hat{\Sigma}_u^{-1}],\end{aligned}$$

where $P' = R'H$ and $Z' = (z_1, \ldots, z_T)$. Note that $P' = [P_1', P_2', p_3]$, where

$$P_1' = R'H_1 = [0, D], \qquad p \times m_1,$$

$$P_2' = R'H_2 = [0, e_p], \qquad p \times m_2,$$

$$p_3 = R'h_3 = 0, \qquad p \times 1.$$

Define further the matrices

$$\tilde{P}' = \begin{bmatrix} D' \\ i_p' \end{bmatrix} P' = \left[\begin{array}{c|c|c} \tilde{P}'_{11} & \tilde{P}'_{12} & 0 \\ \hline 0 & \tilde{p}'_{22} & 0 \end{array}\right],$$

where

$$\tilde{P}'_{11} = D'P'_1 = [0, D'D], \qquad (p-1) \times m_1,$$

$$\tilde{P}'_{12} = D'P'_2 = [0, e_{p-1}], \qquad (p-1) \times m_2,$$

$$\tilde{p}'_{22} = i'_p P'_2 = (0, \ldots, 0, 1), \qquad 1 \times m_2,$$

and write

$$\tilde{p}'_2 = (\tilde{p}'_{22}, 0), \qquad 1 \times (m_2 + 1).$$

Note that each of $\tilde{P}'_{11}$, $\tilde{p}'_{22}$, and $\tilde{p}'_2$ is of full row rank. Define also

$$\Upsilon_T = \begin{bmatrix} T^{1/2} I_{m_1} & 0 & 0 \\ 0 & T I_{m_2} & 0 \\ 0 & 0 & T^{3/2} \end{bmatrix} \quad \text{and} \quad \Upsilon_T^* = \begin{bmatrix} T^{1/2} I_{p-1} & 0 \\ 0 & T \end{bmatrix}.$$

Then we have

$$\begin{aligned} \mathscr{W} = \operatorname{tr}[&U'Q_1Z(Z'Q_1Z)^{-1}\tilde{P}\Upsilon_T^*[\Upsilon_T^*\tilde{P}'(Z'Q_1Z)^{-1}\tilde{P}\Upsilon_T^*]^{-1} \\ &\times \Upsilon_T^*\tilde{P}'(Z'Q_1Z)^{-1}Z'Q_1U\hat{\Sigma}_u^{-1}]. \end{aligned} \tag{13}$$

We need the following lemma:

Lemma 3

(i) $$\Upsilon_T^{-1}Z'Q_1Z\Upsilon_T^{-1} \to_d \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \int_0^1 \tilde{B}_{2*}(r)\tilde{B}_{2*}(r)'\,dr \end{bmatrix},$$

(ii) $$\Upsilon_T^{-1}Z'Q_1U \to_d \begin{bmatrix} N \\ \int_0^1 \tilde{B}_{2*}(r)\,dB_0(r)' \end{bmatrix},$$

where

$$\tilde{B}_{2*}(r) = B_{2*}(r) - \int_0^1 B_{2*}(s)\mathrm{d}s, \qquad B_{2*}(r) = \begin{bmatrix} B_2(r) \\ r \end{bmatrix}. \quad \square$$

Now we have in (13):

$$\Upsilon_T^* \tilde{P}'(Z'Q_1Z)^{-1}\tilde{P}\Upsilon_T^* = \tilde{P}'_T[\Upsilon_T^{-1}Z'Q_1Z\Upsilon_T^{-1}]^{-1}\tilde{P}_T$$
$$\rightarrow_{\mathrm{d}} \begin{bmatrix} \tilde{P}'_{11}\Sigma_1^{-1}\tilde{P}_{11} & 0 \\ 0 & \tilde{p}'_2[\int_0^1 \tilde{B}_{2*}(r)\tilde{B}_{2*}(r)'\mathrm{d}r]^{-1}\tilde{p}_2 \end{bmatrix},$$

by Lemma 3, where

$$\tilde{P}'_T = \begin{bmatrix} \tilde{P}'_{11} & \tilde{P}'_{12}/\sqrt{T} & 0 \\ 0 & \tilde{p}'_{22} & 0 \end{bmatrix}.$$

Similarly,

$$\Upsilon_T^* \tilde{P}'(Z'Q_1Z)^{-1}Z'Q_1U = \tilde{P}'_T[\Upsilon_T^{-1}Z'Q_1Z\Upsilon_T^{-1}]^{-1}\Upsilon_T^{-1}Z'Q_1U$$
$$\rightarrow_{\mathrm{d}} \begin{bmatrix} \tilde{P}'_{11}\Sigma_1^{-1}N \\ \tilde{p}'_2[\int_0^1 \tilde{B}_{2*}(r)\tilde{B}_{2*}(r)'\mathrm{d}r]^{-1}\int_0^1 \tilde{B}_{2*}(r)\mathrm{d}B_0(r)' \end{bmatrix}.$$

Thus, taking account of the consistency of $\hat{\Sigma}_u$,[7] the continuous mapping theorem gives

$$\mathscr{W} \rightarrow_{\mathrm{d}} \mathscr{W}_1 + \mathscr{W}_2,$$

where

$$\mathscr{W}_1 = \mathrm{tr}[N'\Sigma_1^{-1}\tilde{P}_{11}[\tilde{P}'_{11}\Sigma_1^{-1}\tilde{P}_{11}]^{-1}\tilde{P}'_{11}\Sigma_1^{-1}N\Sigma_u^{-1}],$$

and

$$\mathscr{W}_2 = \mathrm{tr}\left[\int_0^1 \mathrm{d}B_0(r)\tilde{B}_{2*}(r)'\left[\int_0^1 \tilde{B}_{2*}(r)\tilde{B}_{2*}(r)'\mathrm{d}r\right]^{-1}\right.$$
$$\times \tilde{p}_2\left[\tilde{p}'_2\left[\int_0^1 \tilde{B}_{2*}(r)\tilde{B}_{2*}(r)'\mathrm{d}r\right]^{-1}\tilde{p}_2\right]^{-1}$$
$$\left.\times \tilde{p}'_2\left[\int_0^1 \tilde{B}_{2*}(r)\tilde{B}_{2*}(r)'\mathrm{d}r\right]^{-1}\int_0^1 \tilde{B}_{2*}(r)\mathrm{d}B_0(r)'\Sigma_u^{-1}\right]. \tag{14}$$

[7] See Park and Phillips (1989) for the consistency of least squares estimators in this context.

In the above, $\mathscr{W}_1$ and $\mathscr{W}_2$ are independent because N is independent of $(B_0(r)', B_2(r)')'$ by Lemma 1(i). Note that, since $\operatorname{vec}(\tilde{P}_{11}'\Sigma_1^{-1}N) = (\tilde{P}_{11}'\Sigma_1^{-1} \otimes I_n)\operatorname{vec}(N) \equiv \mathrm{N}(0, \tilde{P}_{11}'\Sigma_1^{-1}\tilde{P}_{11} \otimes \Sigma_u)$ by Lemma 2(i)(b), we have

$$\mathscr{W}_1 = \operatorname{vec}(\tilde{P}_{11}'\Sigma_1^{-1}N)'[\tilde{P}_{11}'\Sigma_1^{-1}\tilde{P}_{11} \otimes \Sigma_u]^{-1}\operatorname{vec}(\tilde{P}_{11}'\Sigma_1^{-1}N) \equiv \chi^2_{n(p-1)}.$$

Furthermore, since $\tilde{p}_2' = (0, \ldots, 0, 1, 0)$ and $\Omega_0 = \Sigma_u$ by Lemma 1(iii), we can write (14) as

$$\mathscr{W}_2 = \operatorname{tr}\left[\int_0^1 \Omega_0^{-1/2}\mathrm{d}B_0(r)\underline{\tilde{B}}_{2b}(r)\left[\int_0^1 \underline{\tilde{B}}_{2b}(r)^2\mathrm{d}r\right]^{-1} \times \int_0^1 \underline{\tilde{B}}_{2b}(r)\mathrm{d}B_0(r)'\Omega_0^{-1/2}\right], \tag{15}$$

where

$$\underline{\tilde{B}}_{2b}(r) = \tilde{B}_{2b}(r) - \int_0^1 \tilde{B}_{2b}(s)\tilde{B}_{2a^*}(s)'\,\mathrm{d}s\left[\int_0^1 \tilde{B}_{2a^*}(s)\tilde{B}_{2a^*}(s)'\,\mathrm{d}s\right]^{-1}\tilde{B}_{2a^*}(r),$$

$$\tilde{B}_{2a^*}(r) = B_{2a^*}(r) - \int_0^1 B_{2a^*}(s)\mathrm{d}s,$$

$$B_{2a^*}(r) = \begin{bmatrix} B_{2a}(r) \\ r \end{bmatrix}.$$

Note from Lemma 1(ii),

$$B_{2a}(r) = J_1'\Omega_0^{-1/2}B_0(r),$$

where $J_1' = K'\Omega_0^{1/2}$. Multiplying $(J_1'J_1)^{-1/2}$ on both sides,

$$W_1(r) = (J_1'J_1)^{-1/2}J_1'\Omega_0^{-1/2}B_0(r),$$

where $W_1(r) = (J_1'J_1)^{-1/2}B_{2a}(r)$. Let J_2 be an $n \times (k+1)$ matrix such that $J_1'J_2 = 0$, and define

$$J' = \begin{bmatrix} (J_1'J_1)^{-1/2}J_1' \\ (J_2'J_2)^{-1/2}J_2' \end{bmatrix},$$

where $J'J = I_n$ and $JJ' = I_n$ by uniqueness of the inverse matrix. Then write

$$W(r) = J'\Omega_0^{-1/2} B_0(r) = \begin{bmatrix} W_1(r) \\ W_2(r) \end{bmatrix},$$

where $W_2(r) = (J_2'J_2)^{-1/2} J_2'\Omega_0^{-1/2} B_0(r)$. Note that $W(r)$ is an n-vector standard Brownian motion and hence $W_1(r)$ and $W_2(r)$ are independent.

We also write

$$V(r) = \omega_{2b}^{-1/2} B_{2b}(r),$$

where $V(r)$ is a scalar standard Brownian motion independent of $W(r)$ since $B_{2b}(r)$ is independent of $B_0(r)$ by Lemma 1(ii), and ω_{2b} is the variance of $B_{2b}(r)$. Hence we have

$$\tilde{J}_1^{-1} \tilde{B}_{2a*}(r) = \tilde{W}_{1*}(r),$$

$$\omega_{2b}^{-1/2} \tilde{B}_{2b}(r) = \tilde{V}(r),$$

where

$$\tilde{J}_1 = \begin{bmatrix} (J_1'J_1)^{1/2} & 0 \\ 0 & 1 \end{bmatrix},$$

$$\tilde{W}_{1*}(r) = W_{1*}(r) - \int_0^1 W_{1*}(s)\,\mathrm{d}s, \tag{16}$$

$$W_{1*}(r) = \begin{bmatrix} W_1(r) \\ r \end{bmatrix},$$

$$\tilde{V}(r) = V(r) - \int_0^1 V(s)\,\mathrm{d}s. \tag{17}$$

Combining the above results (15) can be written as

$$\begin{aligned} \mathscr{W}_2 &= \mathrm{tr}\left[\int_0^1 J'\Omega_0^{-1/2}\,\mathrm{d}B_0(r) \underline{\tilde{B}}_{2b}(r) \left[\int_0^1 \underline{\tilde{B}}_{2b}(r)^2 \,\mathrm{d}r \right]^{-1} \right. \\ &\qquad \left. \times \int_0^1 \underline{\tilde{B}}_{2b}(r) \mathrm{d}B_0(r)'\Omega_0^{-1/2} J \right] \\ &= \mathrm{tr}\left[\int_0^1 \mathrm{d}W(r) \underline{\tilde{V}}_*(r) \left[\int_0^1 \underline{\tilde{V}}_*(r)^2 \mathrm{d}r \right]^{-1} \int_0^1 \underline{\tilde{V}}_*(r) \mathrm{d}W(r)' \right] \\ &= \mathscr{W}_{21} + \mathscr{W}_{22}, \end{aligned}$$

where

$$\mathscr{W}_{21} = \mathrm{tr}\left[\int_0^1 \mathrm{d}W_1(r)\tilde{\underline{V}}_*(r)\left[\int_0^1 \tilde{\underline{V}}_*(r)^2 \mathrm{d}r\right]^{-1}\int_0^1 \tilde{\underline{V}}_*(r)\mathrm{d}W_1(r)'\right],$$

$$\mathscr{W}_{22} = \mathrm{tr}\left[\int_0^1 \mathrm{d}W_2(r)\tilde{\underline{V}}_*(r)\left[\int_0^1 \tilde{\underline{V}}_*(r)^2 \mathrm{d}r\right]^{-1}\int_0^1 \tilde{\underline{V}}_*(r)\mathrm{d}W_2(r)'\right], \tag{18}$$

and

$$\tilde{\underline{V}}_*(r) = \tilde{V}(r) - \int_0^1 \tilde{V}(s)\tilde{W}_{1\bullet}(s)'\mathrm{d}s\left[\int_0^1 \tilde{W}_{1\bullet}(s)\tilde{W}_{1\bullet}(s)'\,\mathrm{d}s\right]^{-1}\tilde{W}_{1\bullet}(r). \tag{19}$$

Now we write

$$\mathscr{W}_{22} = \mathrm{vec}\left[\int_0^1 \mathrm{d}W_2(r)\tilde{\underline{V}}_*(r)\right]'\left[I_{k+1}\otimes\int_0^1 \tilde{\underline{V}}_*(r)^2\mathrm{d}r\right]^{-1}$$
$$\times\,\mathrm{vec}\left[\int_0^1 \mathrm{d}W_2(r)\,\tilde{\underline{V}}_*(r)\right].$$

Let the symbol '$|_{W_1,V}$' signify the conditional distribution given realization of W_1 and V. By the same argument as that of Lemma 5.1 of Park and Phillips (1989), we have

$$\left[I_{k+1}\otimes\int_0^1 \tilde{\underline{V}}_*(r)^2\mathrm{d}r\right]^{-1/2}\mathrm{vec}\left[\int_0^1 \mathrm{d}W_2(r)\tilde{\underline{V}}_*(r)\right]\Bigg|_{W_1,V} \equiv \mathrm{N}(0, I_{k+1}).$$

Since this conditional distribution does not depend on W_1 and V, it is also the unconditional distribution. Thus, we deduce that $\mathscr{W}_{22} \equiv \chi^2_{k+1}$. Furthermore, $\mathscr{W}_{22}$ is independent of $W_1(r)$ and $V(r)$ and, hence, $\mathscr{W}_{21}$.

Therefore, we have obtained the following theorem.

Theorem 1. If $|I_n - A(L)L| = 0$ has $n-k$ $(0 \le k \le n-1)$ unit roots and the rest of the roots lie outside the unit circle and if $\alpha \notin \mathbb{R}(\Gamma)$, then

$$\mathscr{W} \to_{\mathrm{d}} \chi^2_{n(p-1)+k+1} + \mathrm{tr}\left[\int_0^1 \mathrm{d}W_1(r)\tilde{\underline{V}}_*(r)\left[\int_0^1 \tilde{\underline{V}}_*(r)^2\mathrm{d}r\right]^{-1}\right.$$
$$\left.\times\int_0^1 \tilde{\underline{V}}_*(r)\mathrm{d}W_1(r)'\right].$$

Here, the first and the second terms on the right-hand side are independent, $W_1(r)$ is an $(n-k-1)$-dimensional standard Brownian motion, and $\underline{\tilde{V}}_*(r)$ is defined in (16), (17), and (19), where the scalar standard Brownian motion $V(r)$ is independent of $W_1(r)$. □

Observe that $\mathscr{W}$ converges in distribution to a sum of the usual chi-square distribution and a unit root type distribution. If $k = n-1$, $\mathscr{W}$ converges in distribution to χ^2_{np}, because then there is no $\mathscr{W}_{21}$ term. This is because y_t has only one stochastic trend in that case and it is dominated by a deterministic trend. If $k \leq n-2$, however, the $\mathscr{W}_{21}$ term comes into play and causes a bias in the 'block exogeneity' test (4). The bias of the test thus depends on the $\mathscr{W}_{21}$ component of the limit distribution. Since $\mathscr{W}_{21}$ depends only on the number of the variables, n, and the dimension of the cointegration space, k, we can determine the size and direction of the bias unambiguously by computing the distribution numerically in any specific case. Some simulated critical values for the nonstandard distributions are reported in the next section.

Before proceeding to the stationary case, we note that if $\alpha \in \mathbb{R}(\Gamma)$ (including the case of α being equal to zero), then we have a different asymptotic distribution since in that case y_t does not contain a time trend. It should be apparent from the above derivation that r, the component corresponding to a time trend, in $W_{1*}(r)$ in (16) will be replaced with a Brownian motion. Thus, we have:

Theorem 1′. If $|I_n - A(L)L| = 0$ has $n-k$ $(0 \leq k \leq n-1)$ unit roots and the rest of the roots lie outside the unit circle and if $\alpha \in \mathbb{R}(\Gamma)$, then

$$\mathscr{W} \to_{\mathrm{d}} \chi^2_{n(p-1)+k} + \operatorname{tr}\left[\int_0^1 \mathrm{d}W_1(r)\underline{\tilde{V}}(r)\left[\int_0^1 \underline{\tilde{V}}(r)^2\,\mathrm{d}r\right]^{-1}\int_0^1 \underline{\tilde{V}}(r)\mathrm{d}W_1(r)'\right].$$

Here, the first and the second terms on the right-hand side are independent, $W_1(r)$ is an $(n-k)$-dimensional Brownian motion,

$$\underline{\tilde{V}}(r) = \tilde{V}(r) - \int_0^1 \tilde{V}(s)\tilde{W}_1(s)'\,\mathrm{d}s\left[\int_0^1 \tilde{W}_1(s)\tilde{W}_1(s)'\,\mathrm{d}s\right]^{-1}\tilde{W}_1(r),$$

$$\tilde{W}_1(r) = W_1(r) - \int_0^1 W_1(s)\,\mathrm{d}s,$$

and $\tilde{V}(r)$ is defined in (17) where the scalar standard Brownian motion $V(r)$ is independent of $W_1(r)$. □

Note that Theorem 1′ implies that if y_t is I(1) and does not have a deterministic trend, the Wald statistic $\mathscr{W}$ always converges to a nonstandard distribution.

Unlike Theorem 1, the second term in the asymptotic distribution does not disappear even when $k = n - 1$.

3.2. The stationary case

We now consider the case in which the sequence $\{y_t\}$ is stationary. Since the derivation of the asymptotic distribution is similar to that in the nonstationary case discussed above, we shall give only a brief explanation in the present case.

We can write, for each t,

$$y_t = \mu + \eta_t, \tag{20}$$

where $\mu = (I_n - A(1))^{-1}\alpha$ and $\eta_t = (I_n - A(L)L)^{-1}u_t$. The H matrix is now defined as

$$H_1 = \begin{bmatrix} I_{np} & 0 \\ \hline 0 & D \end{bmatrix}, \qquad h_2 = \begin{bmatrix} 0 \\ e_p \end{bmatrix},$$

and $H = [H_1, h_2]$, which is clearly nonsingular. Note that, since y_t is I(0) with the fixed mean μ, we no longer need h_3 to isolate the time trend component. Accordingly, we define the new $\bar{z}_{1t}$, z_{2t}, and z_{1t} as follows:

$$\bar{z}_{1t} = H_1'\bar{x}_t = (y_{t-1}', \ldots, y_{t-p}', \varepsilon_{t-1}, \ldots, \varepsilon_{t-p+1})',$$

$$z_{2t} = h_2'\bar{x}_t = \xi_{t-1},$$

$$z_{1t} = (\eta_{t-1}', \ldots, \eta_{t-p}', \varepsilon_{t-1}, \ldots, \varepsilon_{t-p+1})'.$$

Note that $Q_1\bar{Z}_1 = Q_1Z_1$, as before.

Now it should be apparent that, for the redefined z_{1t} and z_{2t} above, Lemma 1(i) and (iii) still hold with obvious changes in the dimension of the Brownian motions and the covariance matrices. In the stationary case, we have no z_{2at}, and z_{2t} corresponds to z_{2bt} of the nonstationary case. Hence Lemma 1(ii) now becomes:

Lemma 1(ii)′. *$B_2(r)$ is a scalar Brownian motion independent of $B_0(r)$.* □

Thus Lemmas 2(i)–(iv) also hold (with obvious modifications) for the redefined z_{1t} and z_{2t}. With the present definition of H above we now have

$$P_1' = [0, D], \qquad p \times [(n+1)p - 1],$$

$$p_2' = e_p,$$

and there is no p_3. Hence,

$$\tilde{P}'_{11} = [0, D'D], \qquad (p-1)\times[(n+1)p-1],$$

$$\tilde{P}'_{12} = e_{p-1},$$

$$\tilde{p}_{22} = 1.$$

Redefine the normalization matrices Υ_T and Υ_T^* accordingly, and Lemma 3 becomes:

Lemma 3′

(i) $$\Upsilon_T^{-1} Z'Q_1 Z \Upsilon_T^{-1} \rightarrow_{\mathrm{d}} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \int_0^1 \tilde{B}_2(r)^2 \mathrm{d}r \end{bmatrix},$$

(ii) $$\Upsilon_T^{-1} Z'Q_1 U \rightarrow_{\mathrm{d}} \begin{bmatrix} N \\ \int_0^1 \tilde{B}_2(r)\mathrm{d}B_0(r)' \end{bmatrix},$$

where

$$\tilde{B}_2(r) = B_2(r) - \int_0^1 B_2(s)\mathrm{d}s. \quad \square$$

Since $\tilde{p}_{22} = 1$, it follows from Lemma 3′ that, with suitably redefined $\tilde{P}_T$,

$$\Upsilon_T^* \tilde{P}'(Z'Q_1 Z)^{-1} \tilde{P} \Upsilon_T^* = \tilde{P}'_T(\Upsilon_T^{-1} Z'Q_1 Z \Upsilon_T^{-1})^{-1} \tilde{P}_T$$
$$\rightarrow_{\mathrm{d}} \begin{bmatrix} \tilde{P}'_{11}\Sigma_1^{-1}\tilde{P}_{11} & 0 \\ 0 & [\int_0^1 \tilde{B}_2(r)^2 \mathrm{d}r]^{-1} \end{bmatrix} \tag{21}$$

and

$$\Upsilon_T^* \tilde{P}'(Z'Q_1 Z)^{-1} Z'Q_1 U = \tilde{P}'_T[\Upsilon_T^{-1} Z'Q_1 Z \Upsilon_T^{-1}]^{-1} \Upsilon_T^{-1} Z'Q_1 U$$
$$\rightarrow_{\mathrm{d}} \begin{bmatrix} \tilde{P}'_{11}\Sigma_1^{-1} N \\ [\int_0^1 \tilde{B}_2(r)^2 \mathrm{d}r]^{-1} \int_0^1 \tilde{B}_2(r)\mathrm{d}B_0(r)' \end{bmatrix}. \tag{22}$$

Substituting (21) and (22) into (13) gives

$$\mathscr{W} \rightarrow_{\mathrm{d}} \mathscr{W}_1 + \mathscr{W}_2,$$

where

$$\mathscr{W}_1 = \mathrm{tr}[N'\Sigma_1^{-1}\tilde{P}_{11}[\tilde{P}'_{11}\Sigma_1^{-1}\tilde{P}_{11}]^{-1}\tilde{P}'_{11}\Sigma_1^{-1}N\Sigma_u^{-1}],$$

$$\mathscr{W}_2 = \mathrm{tr}\left[\int_0^1 \mathrm{d}B_0(r)\tilde{B}_2(r)\left[\int_0^1 \tilde{B}_2(r)^2\mathrm{d}r\right]^{-1}\int_0^1 \tilde{B}_2(r)\mathrm{d}B_0(r)'\Sigma_u^{-1}\right], \quad (23)$$

and $\mathscr{W}_1$ and $\mathscr{W}_2$ are indepenent by Lemma 1(i), as before. We can easily show that $\mathscr{W}_1 \equiv \chi^2_{n(p-1)}$. As for $\mathscr{W}_2$, write

$$\mathscr{W}_2 = \mathrm{vec}\left[\int_0^1 \Omega_0^{-1/2}\mathrm{d}B_0(r)\tilde{B}_2(r)\right]'\left[I_n \otimes \int_0^1 \tilde{B}_2(r)^2\mathrm{d}r\right]^{-1}$$

$$\times\, \mathrm{vec}\left[\int_0^1 \Omega_0^{-1/2}\mathrm{d}B_0(r)\tilde{B}_2(r)\right],$$

since $\Sigma_u = \Omega_0$. By Lemma 5.1 of Park and Phillips (1989),

$$\left[I_n \otimes \int_0^1 \tilde{B}_2(r)^2\mathrm{d}r\right]^{-1}\mathrm{vec}\left[\int_0^1 \Omega_0^{-1/2}\mathrm{d}B_0(r)\tilde{B}_2(r)\right] \equiv \mathrm{N}(0, I_n),$$

since $\tilde{B}_2(r)$ is independent of $B_0(r)$ by Lemma 1(ii)'. Therefore $\mathscr{W}_2 \equiv \chi^2_n$.

Thus, we have obtained:

Theorem 2. If $|I_n - A(L)L| \neq 0$ *for* $|L| \leq 1$, *then* $\mathscr{W} \to_d \chi^2_{np}$. □

Interestingly, the inclusion of an independent RW variable in a stationary VAR estimation does not cause any bias in the 'exogeneity test' at least asymptotically.

4. Conclusions

Findings in Ohanian (1988, table 9) that are relevant for our asymptotic analysis may be summarized as follows:

1. 'Block exogeneity' of genuine variables was rejected, on average, 21% of the time for a 5% level test when an artificially generated RW was included in Sims' empirical VAR's, while

2. inclusion of a white noise process did not cause any bias in the test.

Table 1
Percentiles of the nonstandard distributions.

	10%	5%	1%	p-values at 5% χ^2_{np} critical values
		$n = 4$, $p = 4$, $k = 0$		
Theorem 1	27.85	31.22	38.08	13.24%
Thoerem 1′	29.36	32.76	39.65	17.52%
χ^2_{16}	23.54	26.30	32.00	5.00%
		$n = 4$, $p = 1$, $k = 0$		
Thoerem 1	13.10	15.53	21.39	23.82%
Theorem 1′	14.71	17.30	22.87	32.97%
χ^2_4	7.78	9.49	13.28	5.00%

Finding 2 is consistent with the asymptotic theory. Even in the case where the genuine variables are nonstationary, the test is a valid asymptotically chi-square test since the independent white noise process is trivially cointegrated with the other nonstationary variables, i.e., the white noise itself is a stationary linear combination of those variables. [See Sims, Stock, and Watson (1990, example 2) and Toda and Phillips (1991, corollary 1.1).]

If the macroeconomic time series used in Ohanian's study are nonstationary, then our Theorems 1 and 1′ are relevant, and they are consistent with Ohanian's finding 1. We simulated percentile points for the asymptotic distributions given in Theorems 1 and 1′, with different n, p, and k. Table 1 shows some of those simulated percentile points. (Details of the computation are given in appendix B.) Ohanian's VAR consists of four macroeconomic variables plus a RW with a lag length of one year (four quarters). Hence $n = 4$ and $p = 4$. As for k, we report here only the case where $k = 0$ since cointegration among the macroeconomic variables used in Ohanian (1988) is not likely. [See, for instance, Stock and Watson (1989).] Of course, the deviation from the χ^2_{np} distribution becomes smaller as k increases.[8] Thus, the upper half of table 1 shows the limit distributions with $n = 4$, $p = 4$, and $k = 0$. It is interesting to observe that those simulated distributions are not only qualitatively but also quantitatively comparable with the Ohanian results. For comparison purposes, the cases where $p = 1$ are reported in the lower half of table 1. Notice that if $p = 1$, the limit distribution in Theorem 1′ does not have a chi-square part at all.

[8] For example, percentile points of the limit distribution in Theorem 1′ with $k = 1$ are very similar to those of the distribution in Theorem 1 with $k = 0$. [Recall that the degrees of freedom for the chi-square part are $n(p - 1) + k + 1$ in Theorem 1, but $n(p - 1) + k$ in Theorem 1′.]

Conversely, results in Ohanian's experiment together with our Theorems 1, 1′, and 2 imply that those macroeconomic variables used in his study are likely to possess stochastic trends (possibly with deterministic trends), provided that Sim's VAR model is the true data-generating process. For otherwise there would have been no such notable bias in the 'exogeneity test'; if the genuine variables are stationary, then, by Theorem 2, the $\mathscr{W}$ statistic converges to the usual chi-square variate.

In this paper we have concentrated specifically on the spurious inference problem for 'exogeneity tests' that was raised by Ohanian. This is, however, only a special case of the problems that arise from using nonstationary data. In general, as Park and Phillips (1988, 1989) show, commonly used test statistics such as the Wald statistic not only converge to nonstandard distributions but also the asymptotic distributions typically involve nuisance parameters. These problems make inference under nonstationarity difficult although, as the Park–Phillips analysis shows, it is still possible to transform the test statistic so that it has a nuisance parameter free distribution. In this sense, the fact that our $\mathscr{W}$ statistic has a limit distribution that is free of nuisance parameters is itself noteworthy.

One might hope that this property would carry over to a more general case. Unfortunately, this is not the case. Indeed, the possibility that the variables may be cointegrated is a substantial complication, as suggested by the analysis of the trivariate system in Sims, Stock, and Watson (1990). A related paper by the authors [Toda and Phillips (1991)] studies the general case and shows that the Wald statistic for the Granger noncausality hypothesis test in a general VAR framework has a limit distribution which, in general, has a nonstandard component that is commonly dependent on nuisance parameters. However, the limit distribution is the same as the usual asymptotic chi-square distribution if the system has sufficiently many cointegrating vectors.

Appendix A

Proof of Lemma 1

(i) We have from (10),

$$z_{1at+1} = \Phi_{1a} z_{1at} + F_{1a} u_t, \tag{A.1}$$

where $z'_{1at} = (\Delta\eta'_{t-1}, \ldots, \Delta\eta'_{t-p+1}\,(C'\eta_{t-1})')$, and

$$\Phi_{1a} = \left[\begin{array}{c|c|c} A_1^* \cdots A_{p-2}^* & A_{p-1}^* & \Gamma \\ \hline I_{n(p-2)} & 0 & 0 \\ \hline C'A_1^* \cdots C'A_{p-2}^* & C'A_{p-1}^* & I_k + C'\Gamma \end{array}\right], \quad F_{1a} = \begin{bmatrix} e_{p-1} \otimes I_n \\ C' \end{bmatrix},$$

where $e_{p-1} = (1,0, \ldots, 0)'$ which is a $(p-1)$-vector. Similarly we have from (2),

$$z_{1bt+1} = \Phi_{1b} z_{1bt} + e_{p-1}\varepsilon_t, \tag{A.2}$$

where $z_{1bt} = (\varepsilon_{t-1}, \ldots, \varepsilon_{t-p+1})'$ and Φ_{1b} is suitably defined. Thus we can write (A.1) and (A.2) together as

$$z_{1t+1} = \Phi_1 z_{1t} + F_1 v_t,$$

where $v_t = (u_t', \varepsilon_t)'$, and

$$\Phi_1 = \begin{bmatrix} \Phi_{1a} & 0 \\ 0 & \Phi_{1b} \end{bmatrix}, \qquad F_1 = \begin{bmatrix} F_{1a} & 0 \\ 0 & e_{p-1} \end{bmatrix}.$$

Since z_{1t} is I(0) by assumption, the eigenvalues of Φ_1 must be all less than unity in absolute values.

Now by the same argument as that of Theorem 2.2 in Chan and Wei (1988),

$$\frac{1}{T}\sum_{t=1}^{T} z_{1t}z_{1t}' \to_{\mathrm{p}} \Sigma_1 \tag{A.3}$$

and

$$\begin{bmatrix} \dfrac{1}{\sqrt{T}} \displaystyle\sum_{t=1}^{[Tr]} v_t \\ \dfrac{1}{\sqrt{T}} \displaystyle\sum_{t=1}^{T} (z_{1t} \otimes v_t) \end{bmatrix} \to_{\mathrm{d}} \begin{bmatrix} B_v(r) \\ \tilde{v} \end{bmatrix}, \tag{A.4}$$

where $B_v(r) = (B_u(r)', B_\varepsilon(r))'$ is an $(n+1)$-vector Brownian motion with covariance matrix $\Sigma_v = \mathrm{E}v_t v_t'$, $\tilde{v}$ is an $(m_1(n+1))$-dimensional normal random vector with mean zero and covariance matrix $\Sigma_1 \otimes \Sigma_v$, and $B_v(r)$ and $\tilde{v}$ are independent. Note also that $B_u(r)$ and $B_\varepsilon(r)$ are obviously independent.

From (A.1) we have

$$z_{1at} = \Psi_{1a}(L) F_{1a} u_{t-1}, \tag{A.5}$$

where $\Psi_{1a}(L) = (I - \Phi_{1a}L)^{-1} = \sum_{j=0}^{\infty} \Psi_{1a,j} L^j$ with $\Psi_{1a,0} = I$. Since $\Psi_{1a}(L)$ is the inverse of $(I - \Phi_{1a}L)$ and $|I - \Phi_{1a}L| = 0$ has only stable roots, we know by

Brillinger (1981, p. 77) that, for all $g \geq 0$,

$$\sum_{j=1}^{\infty} j^g \| \Psi_{1a,j} \|_a < \infty,$$

where $\| \Psi_{1a,j} \|_a$ denotes the sum of the absolute values of the entries of $\Psi_{1a,j}$. This in turn implies

$$\sum_{j=1}^{\infty} j^2 \| \Psi_{1a,j} \|^2 < \infty,$$

where $\| \Psi_{1a,j} \|^2 = \operatorname{tr}(\Psi_{1a,j} \Psi'_{1a,j})$. Then, (A.4) and a multivariate extension of Theorem 3.4 of Phillips and Solo (1992) implies

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} z_{1at} \to_{\mathrm{d}} \Psi_{1a}(1) F_{1a} B_u(r). \tag{A.6}$$

Furthermore, we obviously have from (A.4) and the definition of z_{1bt},

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} z_{1bt} \to_{\mathrm{d}} i_{p-1} \otimes B_\varepsilon(r), \tag{A.7}$$

where i_{p-1} is a $(p-1)$-vector of ones.

Next, define $\tilde{z}_{2at} = G'\eta_{t-1} = (z'_{2at}, \delta'\eta'_{t-1})'$, and we have from (10)

$$\Delta \tilde{z}_{2at+1} = G'\Phi_{2a} z_{1at} + G'u_t,$$

where $\Phi_{2a} = [A_1^*, \ldots, A_{p-1}^*, \Gamma]$. Hence

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \Delta \tilde{z}_{2at} = G'\Phi_{2a} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} z_{1at} + G' \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t + \mathrm{o}_p(1)$$

$$\to_{\mathrm{d}} G'\Phi_{2a} \Psi_{1a}(1) F_{1a} B_u(r) + G'B_u(r) = \tilde{K}' B_u(r), \tag{A.8}$$

by virtue of (A.4) and (A.6), where $\tilde{K}' = G'[I_n + \Phi_{2a}\Psi_{1a}(1) F_{1a}]$. Also, we apparently have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \Delta z_{2bt} \to_{\mathrm{d}} B_\varepsilon(r). \tag{A.9}$$

To obtain result (i) of Lemma 1, we set

$$B_0(r) = B_u(r), \tag{A.10}$$

$$B_1(r) = \begin{bmatrix} \Psi_{1a}(1) F_{1a} B_u(r) \\ i_{p-1} \otimes B_\varepsilon(r) \end{bmatrix}, \tag{A.11}$$

$$B_{2a}(r) = K' B_u(r) \quad \text{where} \quad K' = [I_{n-k-1}, 0] \tilde{K}', \tag{A.12}$$

$$B_{2b}(r) = B_\varepsilon(r), \tag{A.13}$$

$$v = (I_{m_1} \otimes \tilde{I}_n) \tilde{v} \quad \text{where} \quad \tilde{I}_n = [I_n, 0]. \tag{A.14}$$

See (A.4), (A.6)–(A.9). $B(r) = (B_0(r)', B_1(r)', B_2(r)')'$ is independent of v because $B_v(r)$ is independent of $\tilde{v}$. The covariance matrix of $B(r)$ is given by (12).

(ii) $B_{2b}(r)$ is independent of $(B_0(r)', B_{2a}(r)')'$ from (A.10), (A.12), and (A.13) because $B_\varepsilon(r)$ is independent of $B_u(r)$.

We have already shown that we may write $B_{2a}(r) = K' B_u(r)$. [See (A.12).] So we need here to show that K' is of full row rank. From (10) and (A.5), $\Delta\eta_t$ has a Wold representation

$$\Delta\eta_t = \Phi_{2a} z_{1at} + u_t = \Phi_{2a} \Psi_{1a}(L) F_{1a} u_{t-1} + u_t = \Theta(L) u_t,$$

where $\Theta(L) = I_n + \Phi_{2a} \Psi_{1a}(L) F_{1a} L$. By Granger Representation Theorem, $\mathbb{R}(G) = \mathbb{R}(C)^\perp = \mathbb{R}(\Theta(1))$, and hence $G'\Theta(1)$ is of full row rank. But $\tilde{K}' = G'[I_n + \Phi_{2a} \Psi_{1a}(1) F_{1a}] = G'\Theta(1)$. [See (A.8).] Therefore K' in (A.12) is of full row rank as required.

(iii) $\Omega_0 = \Sigma_0 = \Sigma_u$ is obvious. The positive definiteness of Σ_1 is proved from (A.1) in the same way as Lemma 5.5.5 of Anderson (1971). Ω_2 is given by

$$\Omega_2 = \begin{bmatrix} K'\Sigma_u K & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix},$$

which is positive definite because K is of full column rank. Since Δz_{2t} is a function of only the past history of the innovations, i.e., $\{v_{t-1}, v_{t-2}, \ldots\}$, we have

$$\mathrm{E}\Delta z_{2t} u_{t+j} = 0 \quad \text{for all} \quad j \geq 0.$$

Hence $\Sigma_{20} = \Lambda_{20} = 0$. □

Proof of Lemma 2

(i) was proved in the proof of Lemma 1(i). [See (A.3).] (ii)–(iv) follow immediately from Lemma 2.1 of Park and Phillips (1989) noting that $\Sigma_{20} = \Lambda_{20} = 0$ and $z_{3t} = t + o_p(t)$. □

Proof of Lemma 3

(i)
$$\Upsilon_T^{-1} Z' Q_1 Z \Upsilon_T^{-1} = \Upsilon_T^{-1} Z'Z \Upsilon_T^{-1} - T^{-1} \Upsilon_T^{-1} Z' i_T i_T' Z \Upsilon_T^{-1}$$

$$= \begin{bmatrix} T^{-1} Z_1'Z_1 & T^{-3/2} Z_1'Z_2 & T^{-2} Z_1'Z_3 \\ T^{-3/2} Z_2'Z_1 & T^{-2} Z_2'Z_2 & T^{-5/2} Z_2'Z_3 \\ T^{-2} Z_3'Z_1 & T^{-5/2} Z_3'Z_2 & T^{-3} Z_3'Z_3 \end{bmatrix}$$

$$- \begin{bmatrix} T^{-1} Z_1' i_T \\ T^{-3/2} Z_2' i_T \\ T^{-2} Z_3' i_T \end{bmatrix} (T^{-1} i_T' Z_1, T^{-3/2} i_T' Z_2, T^{-2} i_T' Z_3)$$

$$\to_d \left[\begin{array}{c|cc} \Sigma_1 & 0 & \\ \hline 0 & \int_0^1 B_2(r) B_2(r)' \mathrm{d}r & \int_0^1 B_2(r) r \mathrm{d}r \\ 0 & \int_0^1 r B_2(r)' \mathrm{d}r & \int_0^1 r^2 \mathrm{d}r \end{array}\right]$$

$$- \left[\begin{array}{c|cc} 0 & 0 & \\ \hline 0 & \int_0^1 B_2(r) \mathrm{d}r \int_0^1 B_2(r)' \mathrm{d}r & \int_0^1 B_2(r) \mathrm{d}r \int_0^1 r \mathrm{d}r \\ 0 & \int_0^1 r \mathrm{d}r \int_0^1 B_2(r)' \mathrm{d}r & \int_0^1 r \mathrm{d}r \int_0^1 r \mathrm{d}r \end{array}\right]$$

$$= \left[\begin{array}{c|c} \Sigma_1 & 0 \\ \hline 0 & \int_0^1 \tilde{B}_{2\cdot}(r) \tilde{B}_{2\cdot}(r)' \mathrm{d}r \end{array}\right]$$

by Lemma 2.

(ii)
$$\Upsilon_T^{-1} Z' Q_1 U = \Upsilon_T^{-1} Z' U - T^{-1} \Upsilon_T^{-1} Z' i_T i_T' U$$

$$= \begin{bmatrix} T^{-1/2} Z_1' U \\ T^{-1} Z_2' U \\ T^{-3/2} Z_3' U \end{bmatrix} - \begin{bmatrix} T^{-1} Z_1' i_T \\ T^{-3/2} Z_2' i_T \\ T^{-2} Z_3' i_T \end{bmatrix} T^{-1/2} i_T' U$$

$$\to_d \begin{bmatrix} N \\ \int_0^1 B_2(r)\mathrm{d}B_0(r)' \\ \int_0^1 r\mathrm{d}B_0(r)' \end{bmatrix} - \begin{bmatrix} 0 \\ \int_0^1 B_2(r)\mathrm{d}r\int_0^1 \mathrm{d}B_0(r)' \\ \int_0^1 r\mathrm{d}r\int_0^1 \mathrm{d}B_0(r)' \end{bmatrix}$$

$$= \begin{bmatrix} N \\ \int_0^1 \tilde{B}_{2\cdot}(r)\mathrm{d}B_0(r)' \end{bmatrix} .$$

by Lemma 2. □

Appendix B

The limit distribution in Theorem 1 can be represented as

$$\zeta^*(d_0, d_1, d_2) = \zeta_0^*(d_0) + \zeta_{12}^*(d_1, d_2) , \tag{B.1}$$

where $\zeta_0^*(d_0) \equiv \chi_{d_0}^2$ that is independent of $\zeta_{12}^*(d_1, d_2)$, and

$$\zeta_{12}^*(d_1, d_2) \equiv \chi_1^2 + \operatorname{tr}\left[\int_0^1 \mathrm{d}W_1(r)\,\underline{\tilde{V}}_*(r)' \left[\int_0^1 \underline{\tilde{V}}_*(r)\,\underline{\tilde{V}}_*(r)'\,\mathrm{d}r \right]^{-1} \times \int_0^1 \underline{\tilde{V}}_*(r)\mathrm{d}W_1(r)' \right] ,$$

with

$$\begin{bmatrix} W_1(r) \\ V\ (r) \end{bmatrix} \equiv BM \begin{bmatrix} I_{d_1-1} & 0 \\ 0 & I_{d_2} \end{bmatrix} ,$$

and $\underline{\tilde{V}}_*(r)$ being defined as in (16), (17), and (19). Similarly the distribution in Theorem 1′ can be represented as

$$\zeta(d_0, d_1, d_2) = \zeta_0(d_0) + \zeta_{12}(d_1, d_2) , \tag{B.2}$$

where $\zeta_0(d_0) \equiv \chi_{d_0}^2$ that is independent of $\zeta_{12}(d_1, d_2)$, and

$$\zeta_{12}(d_1, d_2) \equiv \operatorname{tr}\left[\int_0^1 \mathrm{d}W_1(r)\,\underline{\tilde{V}}(r)' \left[\int_0^1 \underline{\tilde{V}}(r)\,\underline{\tilde{V}}(r)'\,\mathrm{d}r \right]^{-1} \times \int_0^1 \underline{\tilde{V}}(r)\mathrm{d}W_1(r)' \right] , \tag{B.3}$$

with

$$\begin{bmatrix} W_1(r) \\ V\ (r) \end{bmatrix} \equiv BM \begin{bmatrix} I_{d_1} & 0 \\ 0 & I_{d_2} \end{bmatrix},$$

and $\tilde{\underline{V}}(r)$ being defined as in Theorem 1′.

Note that d_0, d_1, and d_2 in (B.1) and (B.2) correspond to $(n-1)p+k$, $n-k$, and 1 in Theorems 1 and 1′. Although we have analyzed only the case in which $d_2 = 1$, i.e., ξ_t is a scalar, ξ_t can be a general d_2-vector process independent of y_t, as noted in footnote 1.

To simulate the distribution of $\zeta^*(d_0, d_1, d_2)$, we generated a 10,000 replications of 1,100 observations of y_t and ξ_t, according to

$$y_t = \alpha + Ay_{t-1} + u_t \quad \text{with} \quad y_0 = 0\,, \tag{B.4}$$

where $\alpha = (0, \ldots, 0, 1)$ that is a d_1-vector, $A = I_{d_1}$, and $u_t \equiv$ iid $\mathrm{N}(0, I_{d_1})$, and

$$\xi_t = \xi_{t-1} + \varepsilon_t \quad \text{with} \quad \xi_0 = 0\,,$$

where $\xi_t \equiv$ iid $\mathrm{N}(0, I_{d_2})$ independent of u_t. We discarded the first 100 observations of $\{(y_t', \xi_t')'\}$, giving the series of length 1,000.

Next, for each replication of $\{(y_t', \xi_t')'\}$, we computed the Wald statistic, $\hat{\xi}^*_{12}(d_1, d_2)$ say, for the null hypothesis that $\beta = 0$ in the estimated equation

$$y_t = \hat{\alpha} + \hat{A}y_{t-1} + \hat{\beta}\xi_{t-1} + \hat{u}_t, \qquad t = 101, \ldots, 1100\,.$$

Then, adding an independent $\chi^2_{d_0}$ variate to each $\hat{\zeta}^*_{12}(d_1, d_2)$, we obtained a 10,000 realizations, $\hat{\zeta}^*(d_0, d_1, d_2)$ say, of the (approximate) distribution of $\zeta^*(d_0, d_1, d_2)$, from which we computed the percentile points reported in section 4.

We simulated the $\zeta(d_0, d_1, d_2)$ in (B.2) in the same way as above, except that we set $\alpha = 0$ in (B.4). The justification for the approximation adopted here is, of course, provided by Theorems 1 and 1′. All computations were performed in the Gauss Matrix Programming Language.

To get an idea of how well the $\zeta^*(d_0, d_1, d_2)$ and $\zeta(d_0, d_1, d_2)$ distributions are approximated by the method adopted here, we also computed 10,000 replications of $\hat{\zeta}^*(0,4,1) = \hat{\zeta}^*_{12}(4,1)$ and $\hat{\zeta}(0,4,1) = \hat{\zeta}_{12}(4,1)$ with $A = 0.5I_4$ in (B.4). From those series, 10%, 5%, and 1% points were calculated. These values should coincide with the corresponding percentile points of the χ^2_4 distribution by Theorem 2, since y_t is stationary in that case. The deviations of the simulated values from the χ^2_4 critical values were within 0.13 for $\hat{\zeta}(0,4,1)$ and 0.43 for $\hat{\zeta}^*(0,4,1)$. Simulated p-values at the 5% χ^2_4 critical value were 5.29% and 5.46%, respectively.

References

Anderson, T.W., 1971, The statistical analysis of time series (Wiley, New York, NY).

Brillinger, D.R., 1981, Time series: Data analysis and theory (Holden Day, San Francisco, CA).

Chan, N.H. and C.Z. Wei, 1988, Limiting distributions of least squares estimates of unstable autoregressive processes, Annals of Statistics 16, 367–401.

Granger, C.W. and P. Newbold, 1974, Spurious regressions in econometrics, Journal of Econometrics 2, 111–120.

Johansen, S., 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, Econometrica 59, 1551–1580.

Lawrence, C. and A. Siow, 1985, Interest rates and investment spending: Some empirical evidence for postwar U.S. producer equipment 1947–1980, Journal of Business 58, 359–375.

Litterman, R.B. and L. Weiss, 1985, Money, real interest rates, and output: A reinterpretation of postwar of US data, Econometrica 53, 129–156.

Ohanian, L.E., 1988, The spurious effects of unit roots on vector autoregressions: A Monte Carlo study, Journal of Econometrics 39, 251–266.

Park, J.Y. and P.C.B. Phillips, 1988, Statistical inference in regressions with integrated processes: Part 1, Econometric Theory 4, 468–497.

Park, J.Y. and P.C.B. Phillips, 1989, Statistical inference in regressions with integrated processes: Part 2, Econometric Theory 5, 95–131.

Phillips, P.C.B., 1986, Understanding spurious regressions in econometrics, Journal of Econometrics 33, 311–340.

Phillips, P.C.B., 1988, Multiple regression with integrated processes, in: N.U. Prabhu, ed., Statistical inference from stochastic processes, Contemporary Mathematics 80, 79–106.

Phillips, P.C.B. and V. Solo, 1992, Asymptotics for linear processes, Annals of Statistics, forthcoming.

Sims, C.A., 1980, Comparison of interwar and postwar business cycles: Monetarism reconsidered, American Economic Review 70, 250–256.

Sims, C.A., 1982, Policy analysis with econometric models, Brookings Paper on Economic Activity, 107–164.

Sims, C.A., J.H. Stock, and M.W. Watson, 1990, Inference in linear time series models with some unit roots, Econometrica 58, 113–144.

Stock, J.H. and M.W. Watson, 1989, Interpreting the evidence on money–income causality, Journal of Econometrics 40, 161–181.

Toda, H.Y. and P.C.B. Phillips, 1991, Vector autoregressions and causality, Mimco. (Yale University, New Haven, CT).

Tsay, R.S. and G.C. Tiao, 1990, Asymptotic properties of multivariate nonstationary processes with applications to autoregressions, Annals of Statistics 18, 220–250.