COWLES FOUNDATION FOR RESEARCH IN ECONOMICS AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

Cowles Foundation Discussion Paper No. 978

THE SPURIOUS EFFECT OF UNIT ROOTS ON EXOGENEITY TESTS IN VECTOR AUTOREGRESSIONS: AN ANALYTICAL STUDY

by

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

May 1991

THE SPURIOUS EFFECT OF UNIT ROOTS ON EXOGENEITY TESTS IN VECTOR AUTOREGRESSIONS: AN ANALYTICAL STUDY*

Hiro Y. Toda and Peter C. B. Phillips

Cowles Foundation for Research in Economics
Yale University

ABSTRACT

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 (1980b, 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. This result is consistent with Ohanian's finding. Furthermore, the derived asymptotic distribution is free of nuisance parameters so that we can unambiguously determine the effect of including the random walk. 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.

Original draft: September,1990

Completed version: March, 1991

JEL Classification: 211

---

*The authors thank Glena Ames for typing the completed version of the manuscript of this paper.

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 analysed potentially nonstationary data without detrending or differencing. Some prominent examples are Lawrence and Siow (1985), Litterman and Weiss (1985), and Sims (1980, 1982b).

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 (1980b, 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. Indeed, some simulation results of Nankervis and Savin (1987) that are already published shed light on the problem.

The plan of this paper is as follows. In Section 2 the model for analysing 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. Some concluding remarks are made in Section 4 and the proofs of the lemmas we need in the body of the paper are given in the Appendix.

2. THE MODEL

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

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

where $A(L) = \Sigma_{j=1}^{p} A_j L^{j-1}$, $\{u_t = (u_{1t}, ..., u_{nt})'\}$ is an i.i.d. sequence of n dimensional random vectors with mean zero and covariance matrix Σ_u such that $E|u_{it}|^{2+\delta} < \infty$ for some $\delta > 0$. 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

(2) $$\xi_t = \xi_{t-1} + \epsilon_t$$

[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.

where $\{\epsilon_t\}$ is a sequence of i.i.d. random variables with mean sero and variance σ_ϵ^2 such that $E|\epsilon_t|^{2+\delta} < \infty$ for some $\delta > 0$[2].

Suppose that an econometrician estimates the regression equation

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

where $t = 1, ..., T$, $\beta(L) = \Sigma_{j=1}^{p}\beta_j L^{j-1}$, and the 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

$$(4) \qquad \beta_1 = \cdots = \beta_p = 0 .$$

This is equivalent to the lagged ξ's not "causing" y_t in the Granger sense.

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.[3] Here we shall provide an analytic study and derive an asymptotic distribution of the Wald statistic[4] for testing the hypothesis (4).

Define

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

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

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

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

[3]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.

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

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

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

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 $\text{vec}(\cdot)$ 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

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

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

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[5] separately:

1. $|I_n - A(L)L| = 0$ has at least one unit root and the rest of the roots are greater than unity;
2. All of the roots of $|I_n - A(L)L| = 0$ are greater than unity.

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 case 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 $n \geq 2$.

[5]We exclude by assumption the possibility that $|I_n - A(L)L| = 0$ has a root smaller than one.

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. We may assume that $C'C = I_k$ without loss of generality. We can write (1) in an error correction model format as

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

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.

We assume that $\alpha \notin \mathcal{R}(\Gamma)$ where $\mathcal{R}(\cdot)$ denotes the range space of the argument matrix. This implies that some or all of the variables in y_t have a time trend component[6] as we see shortly. By assumption Δy_t is I(0) with some mean μ and there exists a Wold representation such that

$$(7) \qquad \Delta y_t = \mu + \Theta(L)u_t$$

where $\Theta(L) = \Sigma_{j=0}^{\infty}\Theta_j L^j$. Define $\Delta\bar{\eta}_t = \Theta(L)u_t$ with $\bar{\eta}_0 = y_0$. Then we have for each t

$$(8) \qquad y_t = \mu t + \bar{\eta}_t .$$

Note that $C'\mu$ must be equal to zero for $C'y_t$ to be stationary (unless there is only stochastic common trending). Since $\mu \in \mathcal{R}(C)^{\perp}$ (i.e., $C'\mu = 0$) where $\mathcal{R}(\cdot)^{\perp}$ is the orthogonal complement of $\mathcal{R}(\cdot)$, there is an $n \times (n-k-1)$ matrix G such that $\mathcal{R}(\tilde{G}) = \mathcal{R}(C)^{\perp}$ and $\tilde{G}'\tilde{G} = I_{n-k}$ where $\tilde{G} = [G, \mu(\mu'\mu)^{-1/2}]$. Next, substituting (7) and (8) into (6) gives

$$(6') \qquad \mu + \Delta\bar{\eta}_t = \alpha + A^*(1)\mu + A^*(L)\Delta\bar{\eta}_{t-1} + \Gamma C'\bar{\eta}_{t-1} + u_t ,$$

---

[6] If $\alpha \in \mathcal{R}(\Gamma)$, then y_t does not possess a time trend, and the asymptotics for this case will differ slightly as we discuss later.

from which we have

$$\mu = \alpha + A^*(1)\mu + \Gamma\tau$$

since $\Delta\bar{\eta}_t$ has mean zero by definition, where τ is the mean of the vector $C'\bar{\eta}_t = C'y_t$ (for any t) which is stationary by assumption. Define an $n \times (n-k)$ matrix S of full column rank such that $\mathcal{R}(S) = \mathcal{R}(\Gamma)^{\perp}$. Then we have from the last equation

$$S'(I_n - A^*(1))\mu = S'\alpha .$$

Furthermore, since $CC' + \tilde{G}\tilde{G}' = I_n$ and $C'\mu = 0$,

$$S'(I_n - A^*(1))\tilde{G}\tilde{G}'\mu = S'\alpha ,$$

i.e.,

$$\tilde{G}'\mu = [S'(I_n - A^*(1))\tilde{G}]^{-1}S'\alpha .$$

Premultiplying $\tilde{G}$ on both sides and noting that $\mu = (CC' + \tilde{G}\tilde{G}')\mu = \tilde{G}\tilde{G}'\mu$, we have $\mu = \tilde{G}[S'(I_n - A^*(1))\tilde{G}]^{-1}S'\alpha$. Note that if $\alpha \in \mathcal{R}(\Gamma)$, $\mu = 0$ and y_t does not have a drift or a deterministic trend. We also deduce that

$$\tau = (\Gamma'\Gamma)^{-1}\Gamma'(I_n - A^*(1))\tilde{G}[S'(I_n - A^*(1))\tilde{G}]^{-1}S'\alpha - (\Gamma'\Gamma)^{-1}\Gamma'\alpha .$$

Now we have from (6′)

$$\Delta\bar{\eta}_t = A^*(L)\Delta\bar{\eta}_{t-1} + \Gamma(C'\bar{\eta}_{t-1} - \tau) + u_t .$$

Next, define the vector $\eta_t = \bar{\eta}_t - C\tau$. Then $C'\eta_t$ has mean zero and $\Delta\bar{\eta}_t = \Delta\eta_t$. Thus, the I(1) process η_t satisfies

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

and

(10) $$\eta_t = A(L)\eta_{t-1} + u_t ,$$

Note that C is a matrix of the cointegrating vectors of η_t, also.

Unlike regressions with stationary regressors, $T^{-1}\Sigma_1^T \bar{x}_t \bar{x}_t'$ 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.

We define the matrices:

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

$$H_1 = \begin{bmatrix} H_{1a} & 0 \\ 0 & D \end{bmatrix}, \quad m \times m_1 \quad \text{with} \quad H_{1a} = [D \otimes I_n, i_p \otimes C]$$

$$H_2 = \begin{bmatrix} H_{2a} & 0 \\ 0 & i_p \end{bmatrix}, \quad m \times m_2 \quad \text{with} \quad H_{2a} = i_p \otimes G$$

$$h_3 = \begin{bmatrix} i_p \otimes \mu(\mu'\mu)^{-1} \\ 0 \end{bmatrix}, \quad m \times 1$$

where $m = (n+1)p$, $m_1 = (n+1)(p-1) + k$, $m_2 = n-k$, 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.,

$$\bar{z}_{1t} = H_1'\bar{x}_t = \begin{bmatrix} \mu + \Delta\eta_{t-1} \\ \vdots \\ \mu + \Delta\eta_{t-p+1} \\ p\tau + \zeta_t \\ \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-p+1} \end{bmatrix}$$

where $\zeta_t = C'(\eta_{t-1} + \cdots + \eta_{t-p})$,

$$z_{2t} = H_2'\bar{x}_t = \begin{bmatrix} z_{2at} \\ z_{2bt} \end{bmatrix}$$

where $z_{2at} = G'(\eta_{t-1} + \cdots + \eta_{t-p})$ and $z_{2bt} = \xi_{t-1} + \cdots + \xi_{t-p}$, and

$$z_{3t} = h_3'\bar{x}_t = pt - \frac{p(p+1)}{2} + (\mu'\mu)^{-1}\mu'(\eta_{t-1} + \cdots + \eta_{t-p}) \ . \tag{11}$$

Here we have used the fact that $\Delta\eta_t = \Delta\bar{\eta}_t$, $G'C = 0$, $C'\mu = 0$ and $G'\mu = 0$. We also define

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

and $z_t = (z_{1t}', z_{2t}', z_{3t})'$. Note that $Q_1 Z_1 = Q_1 \bar{Z}_1$ 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. z_{1t}, z_{2t}, and z_{3t} are the basic components that appear in the calculation of the asymptotic distribution of $\mathcal{W}$.

Define

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

and set

$$\Sigma = Ew_t w_t' ,$$

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

and

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

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 " $\longrightarrow_d$ " 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

$$\text{(i)} \quad \begin{bmatrix} \frac{1}{\sqrt{T}}\Sigma_{t=1}^{[Tr]} w_t \\ \frac{1}{\sqrt{T}}\Sigma_{t=1}^{T}(z_{1t} \otimes u_t) \end{bmatrix} \longrightarrow_d \begin{bmatrix} B(r) \\ \nu \end{bmatrix} = \begin{bmatrix} B_0(r) \\ B_1(r) \\ B_2(r) \\ \nu \end{bmatrix} \begin{matrix} n \\ m_1 \\ m_2 \\ nm_1 \end{matrix}$$

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 *,* ν *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* ν *.*

(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 *is positive definite.* □

The assumptions made so far do not ensure the positive definiteness of Σ_1 , which will be required below. Hence, we here assume that Σ_1 is positive definite. Since Σ_1 corresponds to the covariance matrix of the stationary components in y_t , this is a standard assumption. The next lemma follows from Lemma 1 above and Lemma 2.1 of Park and Phillips (1989). Let " $\longrightarrow_p$ " and " $\equiv$ " denote "convergence in probability" and "equivalence in distribution," respectively.

LEMMA 2

(i)(a) $\frac{1}{T}\sum_{j=1}^{T} z_{1t}z_{1t}' \longrightarrow_p \Sigma_1$.

(b) $\frac{1}{\sqrt{T}}\sum_{j=1}^{T} z_{1t}u_t' \longrightarrow_d N$, *where* $\text{vec}(N) = \nu \equiv N(0, \Sigma_1 \otimes \Sigma_u)$

(ii)(a) $\frac{1}{\sqrt{T}}\sum_{j=1}^{T} u_t \longrightarrow_d \int_0^1 dB_0(r) = B_0(1)$

(b) $\frac{1}{\sqrt{T}}\sum_{j=1}^{T} z_{1t} \longrightarrow_d \int_0^1 dB_1(r) = B_1(1)$

(c) $\frac{1}{T^{3/2}}\sum_{j=1}^{T} z_{2t} \longrightarrow_d \int_0^1 B_2(r)dr$

(iii)(a) $\frac{1}{T}\sum_{j=1}^{T} z_{2t}u_t' \longrightarrow_d \int_0^1 B_2(r)dB_0(r)'$

(b) $\frac{1}{T}\sum_{j=1}^{T} z_{2t}z_{1t}' \longrightarrow_d \int_0^1 B_2(r)dB_1(r)' + \Sigma_{21} + \Lambda_{21}$

(iv)(a) $\frac{1}{T^2}\sum_{j=1}^{T} z_{2t}z_{2t}' \longrightarrow_d \int_0^1 B_2(r)B_2(r)'dr$

(v)(a) $\frac{1}{T^{3/2}}\sum_{j=1}^{T} z_{3t}u_t' \longrightarrow_d p\int_0^1 rdB_0(r)'$

(b) $\frac{1}{T^{3/2}}\sum_{j=1}^{T} z_{3t}z_{1t}' \longrightarrow_d p\int_0^1 rdB_1(r)'$

(c) $\frac{1}{T^{5/2}}\sum_{j=1}^{T} z_{3t}z_{2t}' \longrightarrow_d p\int_0^1 rB_2(r)'dr$.

Joint convergence of all the above also applies. □

Now we are ready to analyze the asymptotics of $\mathcal{W}$. 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

$$\mathcal{W} = \text{tr}\left[U'Q_1\bar{X}(\bar{X}'Q_1\bar{X})^{-1}R[R'(\bar{X}'Q_1\bar{X})^{-1}R]^{-1}R'(\bar{X}'Q_1\bar{X})^{-1}\bar{X}'Q_1U\hat{\Sigma}_u^{-1}\right]$$

$$= \text{tr}\left[U'Q_1Z(Z'Q_1Z)^{-1}P[P'(Z'Q_1Z)^{-1}P]^{-1}P'(Z'Q_1Z)^{-1}Z'Q_1U\hat{\Sigma}_u^{-1}\right] ,$$

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

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

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

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

Define further the matrices

$$\tilde{P}' = \begin{bmatrix} D' \\ p^{-1} i_p' \end{bmatrix} P' = \left[\begin{array}{c|c|c} \tilde{P}_{11}' & 0 & 0 \\ \hline 0 & \tilde{p}_{22}' & 0 \end{array}\right] = \left[\begin{array}{c|c} \tilde{P}_{11}' & 0 \\ \hline 0 & \tilde{p}_2' \end{array}\right]$$

where

$$\tilde{P}_{11}' = D'P_1' = [0, D'D]\ ,\ (p-1) \times m_1\ ,$$

$$\tilde{p}_{22}' = p^{-1} i_p' P_2' = (0, \ldots, 0, 1)\ ,\ 1 \times m_2\ ,$$

and

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

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

$$\Upsilon_T = \begin{bmatrix} T^{1/2} I_{m_1} & 0 & 0 \\ 0 & T I_{m_2} & 0 \\ 0 & 0 & T^{3/2} \end{bmatrix},$$

and

$$\Upsilon_T^* = \begin{bmatrix} T^{1/2} I_{p-1} & 0 \\ 0 & T \end{bmatrix}.$$

Then we have

$$\mathcal{W} = \mathrm{tr}\left[U'Q_1 Z(Z'Q_1 Z)^{-1} \tilde{P} \Upsilon_T^* [\Upsilon_T^* \tilde{P}'(Z'Q_1 Z)^{-1} \tilde{P} \Upsilon_T^*]^{-1} \Upsilon_T^* \tilde{P}'(Z'Q_1 Z)^{-1} Z'Q_1 U \hat{\Sigma}_u^{-1}\right]. \tag{13}$$

We need the following lemma:

LEMMA 3

(i) $$\Upsilon_T^{-1} Z' Q_1 Z \Upsilon_T^{-1} \to_d \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \int_0^1 \tilde{B}_{2*}(r)\tilde{B}_{2*}(r)' dr \end{bmatrix},$$

(ii) $$\Upsilon_T^{-1} Z' Q_1 U \to_d \begin{bmatrix} N \\ \int_0^1 \tilde{B}_{2*}(r)\, dB_0(r)' \end{bmatrix},$$

where

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

and

$$B_{2*}(r) = \begin{bmatrix} B_2(r) \\ pr \end{bmatrix} . \ \square$$

From Lemma 3

$$\Upsilon_T^* \tilde{P}' (Z' Q_1 Z)^{-1} \tilde{P} \Upsilon_T^* = \tilde{P}' \left[\Upsilon_T^{-1} Z' Q_1 Z \Upsilon_T^{-1} \right]^{-1} \tilde{P}$$

$$\to_d \begin{bmatrix} \tilde{P}_{11}' \Sigma_1^{-1} \tilde{P}_{11} & 0 \\ 0 & \tilde{p}_2' [\int_0^1 \tilde{B}_{2*}(r)\tilde{B}_{2*}(r)' dr]^{-1} \tilde{p}_2 \end{bmatrix},$$

and

$$\Upsilon_T^* \tilde{P}' (Z' Q_1 Z)^{-1} Z' Q_1 U = \tilde{P} \left[\Upsilon_T^{-1} Z' Q_1 Z \Upsilon_T^{-1} \right]^{-1} \Upsilon_T^{-1} Z' Q_1 U$$

$$\to_d \begin{bmatrix} \tilde{P}_{11}' \Sigma_1^{-1} N \\ \tilde{p}_2' [\int_0^1 \tilde{B}_{2*}(r)\tilde{B}_{2*}(r)' dr]^{-1} \int_0^1 \tilde{B}_{2*}(r) dB_0(r)' \end{bmatrix}.$$

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

$$\mathcal{W} \to_d \mathcal{W}_1 + \mathcal{W}_2$$

where

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

---

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

and

$$(14)\qquad \mathcal{W}_2 = \mathrm{tr}\left[\int_0^1 dB_0(r)\tilde{B}_{2*}(r)'\left[\int_0^1 \tilde{B}_{2*}(r)\tilde{B}_{2*}(r)'dr\right]^{-1}\tilde{p}_2\left[\tilde{p}_2'\left[\int_0^1 \tilde{B}_{2*}(r)\tilde{B}_{2*}(r)'dr\right]^{-1}\tilde{p}_2\right]^{-1}\right.$$

$$\left.\tilde{p}_2'\left[\int_0^1 \tilde{B}_{2*}(r)\tilde{B}_{2*}(r)'dr\right]^{-1}\int_0^1 \tilde{B}_{2*}(r)dB_0(r)'\Sigma_u^{-1}\right],$$

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

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

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

$$(15)\qquad \mathcal{W}_2 = \mathrm{tr}\left[\int_0^1 \Omega_0^{-1/2}dB_0(r)\underline{\tilde{B}}_{2b}(r)\left[\int_0^1 \underline{\tilde{B}}_{2b}(r)^2dr\right]^{-1}\int_0^1 \underline{\tilde{B}}_{2b}(r)dB_0(r)'\Omega_0^{-1/2}\right]$$

where

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

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

and

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

Now 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

$$\tilde{J}_1 = \left[\begin{array}{c|c} (J_1'J_1)^{1/2} & 0 \\ \hline 0 & p \end{array}\right] ,$$

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

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

and

(17) $$\tilde{V}(r) = V(r) - \int_0^1 V(s)ds .$$

Combining the above results (15) can be written as

$$\mathscr{W}_2 = \mathrm{tr}\left[\int_0^1 J'\Omega_0^{-1/2} dB_0(r)\underline{\tilde{B}}_{2b}(r)\left[\int_0^1 \underline{\tilde{B}}_{2b}(r)^2 dr\right]^{-1}\int_0^1 \underline{\tilde{B}}_{2b}(r) dB_0(r)'\Omega_0^{-1/2} J\right]$$

$$= \mathrm{tr}\left[\int_0^1 dW(r)\underline{\tilde{V}}_*(r)\left[\int_0^1 \underline{\tilde{V}}_*(r)^2 dr\right]^{-1}\int_0^1 \underline{\tilde{V}}_*(r) dW(r)'\right]$$

$$= \mathscr{W}_{21} + \mathscr{W}_{22}$$

where

(18)
$$\mathscr{W}_{21} = \mathrm{tr}\left[\int_0^1 dW_1(r)\underline{\tilde{V}}_*(r)\left[\int_0^1 \underline{\tilde{V}}_*(r)^2 dr\right]^{-1}\int_0^1 \underline{\tilde{V}}_*(r) dW_1(r)'\right],$$
$$\mathscr{W}_{22} = \mathrm{tr}\left[\int_0^1 dW_2(r)\underline{\tilde{V}}_*(r)\left[\int_0^1 \underline{\tilde{V}}_*(r)^2 dr\right]^{-1}\int_0^1 \underline{\tilde{V}}_*(r) dW_2(r)'\right],$$

and

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

Now we write

$$\mathscr{W}_{22} = \mathrm{vec}\left[\int_0^1 dW_2(r)\underline{\tilde{V}}_*(r)\right]'\left[I_{k+1}\otimes\int_0^1 \underline{\tilde{V}}_*(r)^2 dr\right]^{-1}\mathrm{vec}\left[\int_0^1 dW_2(r)\underline{\tilde{V}}_*(r)\right].$$

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

$$\left[I_{k+1}\otimes\int_0^1 \underline{\tilde{V}}_*(r)^2 dr\right]^{-1/2}\mathrm{vec}\left[\int_0^1 dW_2(r)\underline{\tilde{V}}_*(r)\right]\Bigg|_{W_1,V} \equiv N(0, I_{k+1}).$$

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 $(0 \leq k \leq n-1)$ *unit roots and the rest of the roots are greater than unity and if* $\alpha \notin \mathcal{R}(\Gamma)$, *then*

$$\mathcal{W} \to_d \chi^2_{n(p-1)+k+1} + \mathrm{tr}\left[\int_0^1 dW_1(r)\underline{\tilde{V}}_*(r)\left[\int_0^1 \underline{\tilde{V}}_*(r)^2 dr\right]^{-1}\int_0^1 \underline{\tilde{V}}_*(r)dW_1(r)'\right] .$$

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 $\mathcal{W}$ converges in distribution to a sum of the usual chi–square distribution and a unit root type distribution. If $k = n-1$, $\mathcal{W}$ converges in distribution to χ^2_{np} , because then there is no $\mathcal{W}_{21}$ term. This is because y_t has only one stochastic trend in that case and it is dominated by a deterministic trend. In fact, one can show that if y_t is trend stationary, $\mathcal{W}$ converges to χ^2_{np} , though we do not report the derivation since it is somewhat obvious given the methodology and the results proved in this paper. If $k \leq n-2$, however, the $\mathcal{W}_{21}$ term comes into play and causes a bias in the block exogeneity test (4). The bias of the test thus depends on the $\mathcal{W}_{21}$ component of the limit distribution. Since $\mathcal{W}_{21}$ 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.

Before proceeding to the stationary case, we note that if $\alpha \in \mathcal{R}(\Gamma)$ (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 are greater than unity and if* $\alpha \in \mathcal{R}(\Gamma)$, *then*

$$\mathcal{W} \to_d \chi^2_{n(p-1)+k} + \mathrm{tr}\left[\int_0^1 dW_1(r)\underline{\tilde{V}}(r)\left[\int_0^1 \underline{\tilde{V}}(r)^2 dr\right]^{-1}\int_0^1 \underline{\tilde{V}}(r)dW_1(r)'\right]$$

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

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

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

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 $\mathcal{W}$ 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

$$H_1 = \left[\begin{array}{c|c} I_{np} & 0 \\ \hline 0 & D \end{array}\right]$$

$$h_2 = \begin{bmatrix} 0 \\ i_p \end{bmatrix}$$

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:

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

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

$$z_{1t} = (\eta_{t-1}', \dots, \eta_{t-1}', \epsilon_{t-p}, \dots, \epsilon_{t-p})' .$$

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

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

$$p_2' = i_p ,$$

and there is no p_3 . Hence

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

$$\tilde{p}_2 = 1 .$$

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

LEMMA 3′

(i) $$\Upsilon_T^{-1}Z'Q_1Z\Upsilon_T^{-1} \rightarrow_d \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \int_0^1 \tilde{B}_2(r)^2 dr \end{bmatrix}$$

(ii) $$\Upsilon_T^{-1}Z'Q_1U \rightarrow_d \begin{bmatrix} N \\ \int_0^1 \tilde{B}_2(r) dB_0(r)' \end{bmatrix}$$

where $\tilde{B}_2(r) = B_2(r) - \int_0^1 B_2(s)ds$. □

Since $\tilde{p}_2 = 1$, it follows from Lemma 3′ that

$$\Upsilon_T^* \tilde{P}'(Z'Q_1Z)^{-1}\tilde{P}\Upsilon_T^* = \tilde{P}'(\Upsilon_T^{-1}Z'Q_1Z\Upsilon_T^{-1})^{-1}\tilde{P} \to_d \begin{bmatrix} \tilde{P}_{11}'\Sigma_1^{-1}\tilde{P}_{11} & 0 \\ 0 & \left[\int_0^1 \tilde{B}_2(r)^2 dr\right]^{-1} \end{bmatrix} \tag{21}$$

and

$$\Upsilon_T^* \tilde{P}'(Z'Q_1Z)^{-1}Z'Q_1U = \tilde{P}'\left[\Upsilon_T^{-1}Z'Q_1Z\Upsilon_T^{-1}\right]^{-1}\Upsilon_T^{-1}Z'Q_1U \tag{22}$$

$$\to_d \begin{bmatrix} \tilde{P}_{11}'\Sigma_1^{-1}N \\ [\int_0^1 \tilde{B}_2(r)^2 dr]^{-1}\int_0^1 \tilde{B}_2(r)dB_0(r)' \end{bmatrix} .$$

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

$$\mathcal{W} \to_d \mathcal{W}_1 + \mathcal{W}_2$$

where

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

$$\mathcal{W}_2 = \mathrm{tr}\left[\int_0^1 dB_0(r)\tilde{B}_2(r)\left[\int_0^1 \tilde{B}_2(r)^2 dr\right]^{-1}\int_0^1 \tilde{B}_2(r)dB_0(r)'\Sigma_u^{-1}\right] , \tag{23}$$

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

$$\mathcal{W}_2 = \mathrm{vec}\left[\int_0^1 \Omega_0^{-1/2}dB_0(r)\tilde{B}_2(r)\right]'\left[I_n \otimes \int_0^1 \tilde{B}_2(r)dr\right]^{-1}\mathrm{vec}\left[\int_0^1 \Omega_0^{-1/2}dB_0(r)\tilde{B}_2(r)\right] ,$$

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

$$\left[I_n \otimes \int_0^1 \tilde{B}_2(r)dr\right]^{-1/2}\mathrm{vec}\left[\int_0^1 \Omega_0^{-1/2}dB_0(r)\tilde{B}_2(r)\right] \equiv N(0, I_n)$$

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

Thus, we have obtained:

<u>THEOREM 2</u>. *If* $|I_n - A(L)L| \neq 0$ *for* $|L| \leq 1$ *, then* $\mathcal{W} \rightarrow_d \chi^2_{np}$. □

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

Since macroeconomic time series such as those used in Ohanian's experiment are generally regarded as nonstationary, Theorem 1 (or 1′) would be more relevant than Theorem 2 in practice. The former are consistent with Ohanian's findings that: (i) inclusion of an artificially generated RW in Sims' empirical VAR's resulted in overrejecting the exogeneity hypothesis; whereas (ii) bias did not arise when a white noise process was included in the model. Conversely, Ohanian's simulation results together with our Theorem 1 and Theorem 2 imply that those macroeconomic variables have stochastic trends (if Sims' VAR model is the true data generating mechanism) for otherwise there would be no bias in the exogeneity tests, at least in large samples. Moreover, as was noted earlier, it is easy to see that if these variables had only deterministic trends, the $\mathcal{W}$ statistic would converge to the usual chi–square distribution.

Although we do not actually compute the asymptotic distribution derived above, closely related distributions are tabulated in Nankervis and Savin (1987) and Johansen and Juselius (1990). The latter authors' Table A1 report simulated distributions of the trace of a stochastic matrix representing a limit unit root distribution that is similar to our $\mathcal{W}_{21}$ in (18). Their statistic is not exactly the same as ours, but the table shows clearly a significant deviation from the usual chi–square distribution.

In addition, Nankervis and Savin (1987) happened to compute one special case of (the square root of) the $\mathcal{W}_{21}$ distribution in Theorem 1′. They call the statistic $t(\delta)$ and it is tabulated in the first row of their Table 3. In our notation, $t(\delta)$ corresponds to the t–statistic for the coefficient β_1 in a simple bivariate model:

$$y_t = \alpha + a_1 y_{t-1} + \beta_1 \xi_{t-1} + u_t$$

$$\xi_t = \xi_{t-1} + \epsilon_t$$

where the true values of the coefficients are such that $\alpha = 0$, $a_1 = 1$ and $\beta_1 = 0$, and u_t and ϵ_t are independent normal random variables. In fact, the asymptotic distribution of the t–statistic is by Theorem 1′

$$\left[\int_0^1 \underline{\tilde{V}}(r)^2 dr\right]^{-1/2} \int_0^1 \underline{\tilde{V}}(r) dW_1(r)$$

where $W_1(r)$ is a scalar standard Brownian motion independent of $V(r)$. The Nankervis–Savin Table 3 shows that the asymptotic distribution (their sample size is 500) has much more dispersion than the $N(0,1)$, which is of course the asymptotic distribution of the t–statistic in regressions with stationary regressors. This implies that the null hypothesis that $\beta_1 = 0$ would be rejected too often using a nominal asymptotic $N(0,1)$ critical value, confirming our theory.

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 $\mathcal{W}$ 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

PROOF OF LEMMA 1

(i) Define

$$v_t = (u_t', \epsilon_t)'$$

$$x_t = (\eta_{t-1}', \dots, \eta_{t-p}', \xi_{t-1}, \dots, \xi_{t-p})' .$$

Then (10) and (2) together can be written in the VAR(1) representation

$$x_{t+1} = \Phi x_t + F v_t \tag{24}$$

where

$$\Phi = \begin{bmatrix} \Phi_a & 0 \\ 0 & \Phi_b \end{bmatrix},$$

$$\Phi_a = \left[\begin{array}{ccc|c} A_1 & \cdots & A_{p-1} & A_p \\ \hline & I_{n(p-1)} & & 0 \end{array}\right],$$

$$\Phi_b = \left[\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & 0 \\ \hline & & I_{p-1} & & 0 \end{array}\right],$$

$$F = \begin{bmatrix} F_a & 0 \\ 0 & e_p \end{bmatrix},$$

$e_p = (1, 0, \dots, 0)'$ which is a p—vector, and $F_a = e_p \otimes I_n$. Define $\tilde{H}_{2a}$ and hence $\tilde{H}_2$ by replacing G with $\tilde{G}$ in the definition of H_{2a} and H_2, respectively. Let $\tilde{H} = [H_1, \tilde{H}_2]$ accordingly and define

$$\tilde{z}_t = \begin{bmatrix} z_{1t} \\ \tilde{z}_{2t} \end{bmatrix} = \tilde{H}' x_t = \begin{bmatrix} H_1' x_t \\ \tilde{H}_2' x_t \end{bmatrix}.$$

Since $\tilde{H}$ is clearly nonsingular, we can write (24) as

$$\tilde{z}_{t+1} = \tilde{H}' \Phi \tilde{H} (\tilde{H}' \tilde{H})^{-1} \tilde{z}_t + \tilde{H}' F v_t . \tag{25}$$

Using the fact that $C'\tilde{G} = 0$ and hence $(I_n - A(1))\tilde{G} = 0$, we have

$$\tilde{H}'\Phi\tilde{H}(\tilde{H}'\tilde{H})^{-1} = \begin{bmatrix} \Phi_1 & 0 \\ \Phi_2 & I_{n-k+1} \end{bmatrix}$$

where

$$\Phi_1 = \begin{bmatrix} \Phi_{1a} & 0 \\ 0 & \Phi_{1b} \end{bmatrix},$$

$$\Phi_2 = \begin{bmatrix} \Phi_{2a} & 0 \\ 0 & \phi'_{2b} \end{bmatrix},$$

$$\Phi_{1a} = H'_{1a}\Phi_a H_{1a}(H'_{1a}H_{1a})^{-1},$$

$$\Phi_{1b} = D'\Phi_b D(D'D)^{-1},$$

$$\Phi_{2a} = \tilde{H}'_{2a}\Phi_a H_{1a}(H'_{1a}H_{1a})^{-1},$$

$$\phi'_{2b} = i'_p\Phi_b D(D'D)^{-1}.$$

Furthermore

$$\tilde{H}'F = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix},$$

where

$$F_1 = H'_1F = \begin{bmatrix} F_{1a} & 0 \\ 0 & e_{p-1} \end{bmatrix} \text{ with } F_{1a} = \begin{bmatrix} e_{p-1} \otimes I_n \\ C' \end{bmatrix},$$

$$F_2 = \tilde{H}'_2F = \begin{bmatrix} \tilde{G}' & 0 \\ 0 & 1 \end{bmatrix}.$$

Thus, from (25) we have

$$z_{1t+1} = \Phi_1 z_{1t} + F_1 v_t \tag{26}$$

and

$$(27) \qquad \Delta\tilde{z}_{2t+1} = \Phi_2 z_{1t} + F_2 v_t \, .$$

Since z_{1t} is I(0) by assumption, the eigenvalues of Φ_1 must be all less than unity. Hence we can write (26) as

$$(28) \qquad z_{1t} = \Psi_1(L)F_1 v_{t-1} = \begin{bmatrix} \Psi_{1a}(L) & 0 \\ 0 & \Psi_{1b}(L) \end{bmatrix} \begin{bmatrix} F_{1a} u_{t-1} \\ e_{p-1}\epsilon_{t-1} \end{bmatrix}$$

where $\Psi_1(L) = \Sigma_{j=0}^{\infty}\Psi_{1,j}L^j = \Sigma_{j=0}^{\infty}\Phi_1^j L^j$, $\Psi_{1a}(L) = \Sigma_{j=0}^{\infty}\Psi_{1a,j}L^j = \Sigma_{j=0}^{\infty}\Phi_{1a}^j L^j$, and $\Psi_{1b}(L) = \Sigma_{j=0}^{\infty}\Psi_{1b,j}L^j = \Sigma_{j=0}^{\infty}\Phi_{1b}^j L^j$. Then

$$\Sigma_1 = Ez_{1t}z_{1t}' = \sum_{j=0}^{\infty} \Phi_1^j F_1 \Sigma_v F_1' \Phi_1^{j\prime}$$

where $\Sigma_v = Ev_t v_t'$.

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

$$(29) \qquad \frac{1}{T}\sum_{t=1}^{T} z_{1t}z_{1t}' \longrightarrow_p \Sigma_1$$

and

$$(30) \qquad \begin{bmatrix} \frac{1}{\sqrt{T}}\Sigma_{t=1}^{[Tr]} v_t \\ \frac{1}{\sqrt{T}}\Sigma_{t=1}^{[Tr]}(z_{1t} \otimes v_t) \end{bmatrix} \longrightarrow_d \begin{bmatrix} B_v(r) \\ \tilde{\nu} \end{bmatrix}$$

where $B_v(r) = (B_u(r)', B_\epsilon(r))'$ is an (n+1)—vector Brownian motion with covariance matrix Σ_v, $\tilde{\nu}$ 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 $\tilde{\nu}$ are independent. Note also that $B_u(r)$ and $B_\epsilon(r)$ are obviously independent.

From (27) and (28) we have

$$(31) \qquad \Delta\tilde{z}_{2t+1} = \Psi_2(L)v_t$$

where $\Psi_2(L) = F_2 + \Phi_2\Psi_1(L)F_1 L$. Write (31) as

$$\begin{bmatrix} \Delta\tilde{z}_{2at+1} \\ \Delta z_{2bt+1} \end{bmatrix} = \begin{bmatrix} \Psi_{2a}(L) & 0 \\ 0 & \psi_{2b}(L) \end{bmatrix} \begin{bmatrix} u_t \\ \epsilon_t \end{bmatrix}$$

where $\Psi_{2a}(L) = \tilde{G}' + \Phi_{2a}\Psi_{1a}(L)F_{1a}L$ and $\psi_{2b}(L) = 1 + \phi'_{2b}\Psi_{1b}(L)e_{p-1}L$. But recall that we write $\Delta\eta_t = \Delta\bar{\eta}_t = \Theta(L)u_t$ in (7). Hence, $\Delta\tilde{z}_{2at+1}$ can be also written as

$$\Delta\tilde{z}_{2at+1} = \tilde{G}'\Delta\eta_t + \cdots + \tilde{G}'\Delta\eta_{t-p+1}$$
$$= \tilde{G}'(\Theta(L) + \Theta(L)L + \cdots + \Theta(L)L^{p-1})u_t \,.$$

Therefore $\Psi_{2a}(1) = p\tilde{G}'\Theta(1)$. Since $\mathcal{R}(\tilde{G}) = \mathcal{R}(C)^{\perp} = \mathcal{R}(\Theta(1))$, $\Psi_{2a}(1)$ is of full row rank. Similarly we deduce that $\psi_{2b}(1) = p$.

Next, since in (28) $\Psi_1(L)$ is the inverse of $(I_n - \Phi_1 L)$ and $|I_n - \Phi_1 L| = 0$ has only stable roots, we know by Brillinger (1981, p. 77) that for all $g \geq 0$

$$\sum_{j=1}^{\infty} j^g \|\Psi_{1,j}\|_a < \infty \tag{32}$$

where $\|\Psi_{1,j}\|_a$ denotes the sum of the absolute value of the entries of $\Psi_{1,j}$. (32) in turn implies

$$\sum_{j=1}^{\infty} j^2 \|\Psi_{1,j}\|^2 < \infty \tag{33}$$

where $\|\Psi_{1,j}\|^2 = \mathrm{tr}(\Psi_{1,j}\Psi'_{1,j})$. Then, by a multivariate extension of Theorem 3.3 of Phillips and Solo (1989),

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} z_{1t} \rightarrow_d \Psi_1(1)F_1B_v(r)$$

and from (27)

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} \Delta\tilde{z}_{2t} = \Phi_2 \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} z_{1t} + F_2 \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} v_t + o_p(1)$$
$$\rightarrow_d \Phi_2\Psi_1(1)F_1B_v(r) + F_2B_v(r)$$
$$= \Psi_2(1)B_v(r) = \begin{bmatrix} \Psi_{2a}(1)B_u(r) \\ \psi_{2b}(1)B_\epsilon(r) \end{bmatrix} . \tag{34}$$

To obtain result (i) of Lemma 1 we need only notice that

(35) $$z_{2t} = \left[\begin{array}{c|c|c} I_{n-k-1} & 0 & 0 \\ \hline 0 & 0 & 1 \end{array}\right] \tilde{z}_{2t}$$

and hence $z_{2at} = [I_{n-k-1}, 0]\tilde{z}_{2at}$ so that we can set

(36) $$B_0(r) = B_u(r)$$

(37) $$B_1(r) = \Psi_1(1)F_1B_v(r)$$

(38) $$B_{2a}(r) = [I_{n-k-1}, 0]\Psi_{2a}(1)B_u(r)$$

(39) $$B_{2b}(r) = \psi_{2b}(1)B_\epsilon(r)$$

(40) $$\nu = (I_{m_1} \otimes \tilde{I}_n)\tilde{\nu} \text{ where } \tilde{I}_n = [I_n, 0] .$$

$B(r) = (B_0(r)', B_1(r)', B_2(r)')'$ is independent of ν because $B_v(r)$ is independent of $\tilde{\nu}$. The covariance matrix of $B(r)$ is given by (12).

(ii) $B_{2b}(r)$ is independent of $(B_0(r)', B_{2a}(r)')'$ from (36), (38) and (39) because $B_\epsilon(r)$ is independent of $B_u(r)$. From (38) we can take $K' = [I_{n-k-1}, 0]\Psi_{2a}(1)$. Then K is of full column rank as required.

(iii) $\Omega_0 = \Sigma_0 = \Sigma_u$ is obvious. Ω_2 is given by

$$\Omega_2 = \begin{bmatrix} K'\Sigma_u K & 0 \\ 0 & \psi_{2b}(1)^2\sigma_\epsilon^2 \end{bmatrix},$$

which is positive definite because K is of full column rank and $\psi_{2b}(1) = p \neq 0$. Since from (31) and (35) Δz_{2t} is a function of only the past history of the innovations, i.e., $\{v_{t-1}, v_{t-2}, \ldots\}$, we have

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

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

PROOF OF LEMMA 2

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

PROOF OF LEMMA 3

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

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

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

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

$$- \left[\begin{array}{c|cc} 0 & \multicolumn{2}{c}{0} \\ \hline 0 & \int_0^1 B_2(r)dr \int_0^1 B_2(r)'dr & p\int_0^1 B_2(r)dr \int_0^1 r dr \\ 0 & p\int_0^1 r dr \int_0^1 B_2(r)'dr & p^2 \int_0^1 r dr \int_0^1 r dr \end{array}\right]$$

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

by Lemma 2.

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

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

$$\longrightarrow_d \begin{bmatrix} N \\ \int_0^1 B_2(r) dB_0(r)' \\ p\int_0^1 r dB_0(r)' \end{bmatrix} - \begin{bmatrix} 0 \\ \int_0^1 B_2(r) dr \int_0^1 dB_0(r)' \\ p\int_0^1 r dr \int_0^1 dB_0(r)' \end{bmatrix}$$

$$= \begin{bmatrix} N \\ \int_0^1 \tilde{B}_{2*}(r) dB_0(r)' \end{bmatrix} .$$

by Lemma 2. □

8. REFERENCES

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

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

Granger, C. W. and P. Newbold (1974). "Spurious regressions in econometrics," Journal of *Econometrics,* 2, 111—120.

Johansen, S. and Katarina Juselius (1990). "Maximum likelihood estimation and inference on cointegration — with applications to the demand for money," *Oxford Bulletin of Economics and Statistics,* 52, 169—210.

Lawrence, Colin and Alyoysius 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, Robert B. and Laurence Weiss (1985). "Money, real interest rates, and output: A reinterpretation of postwar U.S. data," *Econometrica,* 53, 129—156.

Nankervis, J. C. and N. E. Savin (1987). "Finite sample distributions of t and F statistics in an AR(1) with an exogenous variable," *Econometric Theory,* 3, 387—408.

Ohanian, Lee 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). "Statistic 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, Peter 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. Prahbu (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics,* 80, 79—106.

Phillips, Peter C. B. and Victor Solo (1989). "Asymptotics for linear processes," Cowles Foundation Discussion Paper No. 932.

Sims, Christopher A. (1980a). "Macroeconomics and reality," *Econometrics,* 48, 1—48.

Sims, Christopher A. (1980b). "Comparison of interwar and postwar business cycles: Monetarism reconsidered," *American Economic Review,* 70, 250—256.

Sims, Christopher A. (1982). "Policy analysis with econometric models," *Brookings Paper on Economic Activity,* 107—164.

Sims, Christopher A., James H. Stock and Mark W. Watson (1990). "Inference in linear time series models with some unit roots," *Econometrics,* 58, 113—144.

Toda, H. Y. and P. C. B. Phillips (1991). "Vector autoregressions and causality," Yale University, mimeographed.

Tsay, Ruey S. and George C. Tiao (1990). "Asymptotic properties of multivariate nonstationary processes with applications to autoregressions," *Annals of Statistics*, 18, 220–250.