Econometrica, Vol. 58, No. 1 (January, 1990), 113–144

INFERENCE IN LINEAR TIME SERIES MODELS WITH SOME UNIT ROOTS

BY CHRISTOPHER A. SIMS, JAMES H. STOCK, AND MARK W. WATSON[1]

This paper considers estimation and hypothesis testing in linear time series models when some or all of the variables have unit roots. Our motivating example is a vector autoregression with some unit roots in the companion matrix, which might include polynomials in time as regressors. In the general formulation, the variable might be integrated or cointegrated of arbitrary orders, and might have drifts as well. We show that parameters that can be written as coefficients on mean zero, nonintegrated regressors have jointly normal asymptotic distributions, converging at the rate $T^{1/2}$. In general, the other coefficients (including the coefficients on polynomials in time) will have nonnormal asymptotic distributions. The results provide a formal characterization of which t or F tests—such as Granger causality tests—will be asymptotically valid, and which will have nonstandard limiting distributions.

KEYWORDS: Cointegration, error correction models, vector autoregressions.

1. INTRODUCTION

VECTOR AUTOREGRESSIONS have been used in an increasingly wide variety of econometric applications. In this paper, we investigate the distributions of least squares parameter estimators and Wald test statistics in linear time series models that might have unit roots. The general model includes several important special cases. For example, all the variables could be integrated of order zero (be "stationary"), possibly around a polynomial time trend. Alternatively, all the variables could be integrated of order one, with the number of unit roots in the multivariate representation equaling the number of variables, so that the variables have a VAR representation in first differences. Another special case is that all the variables are integrated of the same order, but there are linear combinations of these variables that exhibit reduced orders of integration, so that the system is cointegrated in the sense of Engle and Granger (1987). In addition to VAR's, this model contains as special cases linear univariate time series models with unit roots as studied by White (1958), Fuller (1976), Dickey and Fuller (1979), Solo (1984), Phillips (1987), and others.

The model and notation are presented in Section 2. Section 3 provides an asymptotic representation of the ordinary least squares (OLS) estimator of the coefficients in a regression model with "canonical" regressors that are a linear transformation of Y_t, the original regressors. An implication of this result is that the OLS estimator is consistent whether or not the VAR contains integrated components, as long as the innovations in the VAR have enough moments and a

[1] The authors thank Lars Peter Hansen, Hal White, and an anonymous referee for helpful comments on an earlier draft. This research was supported in part by the National Science Foundation through Grants SES-83-09329, SES-84-08797, SES-85-10289, and SES-86-18984. This is a revised version of two earlier papers, "Asymptotic Normality of Coefficients in a Vector Autoregression with Unit Roots," March, 1986, by Sims, and "Wald Tests of Linear Restrictions in a Vector Autoregression with Unit Roots," June, 1986, by Stock and Watson.

zero mean, conditional on past values of Y_t. When an intercept is included in a regression based on the canonical variables, the distribution of coefficients on the stationary canonical variates with mean zero is asymptotically normal with the usual covariance matrix, converging to its limit at the rate $T^{1/2}$. In contrast, the estimated coefficients on nonstationary stochastic canonical variates are nonnormally distributed, converging at a faster rate. These results imply that estimators of coefficients in the original untransformed model have a joint nondegenerate asymptotic normal distribution if the model can be rewritten so that these original coefficients correspond in the transformed model to coefficients on mean zero stationary canonical regressors.

The limiting distribution of the Wald F statistic is obtained in Section 4. In general, the distribution of this statistic does not have a simple form. When all the restrictions being tested in the untransformed model correspond to restrictions on the coefficients of mean zero stationary canonical regressors in the transformed model, then the test statistic has the usual limiting χ^2 distribution. In contrast, when the restrictions cannot be written solely in terms of coefficients on mean zero stationary canonical regressors and at least one of the canonical variates is dominated by a stochastic trend, then the test statistic has a limiting representation involving functionals of a multivariate Wiener process and in general has a nonstandard asymptotic distribution.

As a special case, the results apply to a VAR with some roots equal to one but with fewer unit roots than variables, a case that has recently come to the fore as the class of cointegrated VAR models. Engle and Granger (1987) have pointed out that such models can be handled with a two-step procedure, in which the cointegrating vector is estimated first and used to form a reduced, stationary model. The asymptotic distribution theory for the reduced model is as if the cointegrating vector were known exactly. One implication of our results is that such two-step procedures are unnecessary, at least asymptotically: if the VAR is estimated on the original data, the asymptotic distribution for the coefficients normalized by $T^{1/2}$ is a singular normal and is identical to that for a model in which the cointegrating vector is known exactly *a priori*. This result is important because the two-step procedures have so far been justified only by assuming that the number of cointegrating vectors is known. This paper shows that, at a minimum, as long as one is not interested in drawing inferences about intercepts or about linear combinations of coefficients that have degenerate limiting distributions when normalized by $T^{1/2}$, it is possible to avoid such two-step procedures in large samples. However, when there are unit roots in the VAR, the coefficients on any intercepts or polynomials in time included in the regression and their associated t statistics will typically have nonstandard limiting distributions.

In Sections 5 and 6, these general results are applied to several examples. Section 5 considers a univariate AR(2) with a unit root with and without a drift; the Dickey-Fuller (1979) tests for a unit root in these models follow directly from the more general results. Section 6 examines two common tests of linear restrictions performed in VAR's: a test for the number of lags that enter the true VAR and a "causality" or predictability test that lagged values of one variable do not enter the equation for a second variable. These examples are developed for a trivariate system of integrated variables with drift. In the test for lag length, the F test has a chi-squared asymptotic distribution with the usual degrees of freedom. In the causality test, the statistic has a χ^2 asymptotic distribution if the process is cointegrated; otherwise, its asymptotic distribution is nonstandard and must be computed numerically. Some conclusions are summarized in Section 7.

2. THE MODEL

We consider linear time series models that can be written in first order form,

$$(2.1) \qquad Y_t = AY_{t-1} + G\Omega^{1/2}\eta_t \qquad (t=1,\ldots,T),$$

where Y_t is a k-dimensional time series variable and A is a $k\times k$ matrix of coefficients. The $N\times 1$ vector of disturbances $\{\eta_t\}$ is assumed to be a sequence of martingale differences with $E[\eta_t|\eta_1,\ldots,\eta_{t-1}]=0$ and $E[\eta_t\eta_t'|\eta_1,\ldots,\eta_{t-1}]=I_N$ for $t=1,\ldots,T$. The $N\times N$ matrix $\Omega^{1/2}$ is thought of as the square root of the covariance matrix of some "structural" errors $\Omega^{1/2}\eta_t$. The $k\times N$ constant matrix G is thought of as known *a priori*, and typically contains ones and zeros indicating which errors enter which equations. Note that because N might be less than k, some of the elements of Y_t (or more generally, some linear combinations of Y_t) might be nonrandom. It is assumed that A has k_1 eigenvalues with modulus less than one and that the remaining $k-k_1$ eigenvalues exactly equal one. As is shown below, this formulation is sufficiently general to include a VAR of arbitrary finite order with arbitrary orders of integration, constants and finite order polynomials in t. The assumptions do not, however, allow complex unit roots so, for example, seasonal nonstationarity is not treated.

The regressors Y_t will in general consist of random variables with various orders of integration, of constants, and of polynomials in time. These components in general are of different orders in t. Often there will be linear combinations of Y_t having a lower order in probability than the individual elements of Y_t itself. Extending Engle and Granger's (1987) terminology, we refer to the vectors that form these linear combinations as *generalized cointegrating vectors*. As long as the system has some generalized cointegrating vectors, the calculations below demonstrate that $T^{-p}\Sigma Y_tY_t'$ will converge to a singular (possibly random) limit, where p is a suitably chosen constant; that is, some elements of Y_t will exhibit perfect multicolinearity, at least asymptotically. Thus we work with a transformation of Y_t, say Z_t, that uses the generalized cointegrating vectors of Y_t to isolate those components having different orders in probability. Specifically, let

$$(2.2) \qquad Z_t = DY_t.$$

(Note that in the dating convention of (2.1) the actual regressors are Y_{t-1} or, after transforming by D, Z_{t-1}.) The nonsingular $k\times k$ matrix D is chosen in such a way that Z_t has a simple representation in terms of the fundamental stochastic and nonstochastic components. Let $\xi_t^1=\Sigma_{s=1}^t\eta_s$, and let ξ_t^j be defined recursively by $\xi_t^j = \Sigma_{s=1}^t \xi_s^{j-1}$, so that ξ_t^1 is the N-dimensional driftless random walk with innovations η_t and ξ_t^j is the j-fold summation of η_t. The transformation D is chosen so that

(2.3)

$$Z_t = \begin{bmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \\ Z_t^4 \\ \vdots \\ Z_t^{2g} \\ Z_t^{2g+1} \end{bmatrix}$$

$$= \begin{bmatrix} F_{11}(L) & 0 & 0 & 0 & \cdots & 0 & 0 \\ F_{21}(L) & F_{22} & 0 & 0 & \cdots & 0 & 0 \\ F_{31}(L) & F_{32} & F_{33} & 0 & \cdots & 0 & 0 \\ F_{41}(L) & F_{42} & F_{43} & F_{44} & \cdots & 0 & 0 \\ \vdots & \vdots & & & & \vdots & \vdots \\ F_{2g,1}(L) & F_{2g,2} & \cdots & \cdots & \cdots & F_{2g,2g} & 0 \\ F_{2g+1,1}(L) & F_{2g+1,2} & \cdots & \cdots & \cdots & F_{2g+1,2g} & F_{2g+1,2g+1} \end{bmatrix} \begin{bmatrix} \eta_t \\ 1 \\ \xi_t^1 \\ t \\ \vdots \\ t^{g-1} \\ \xi_t^g \end{bmatrix}$$

$$\equiv F(L)\nu_t$$

where L is the lag operator and $\nu_t = (\eta_t' \cdots \xi_t^{g\prime})'$. Note that the stochastic and deterministic elements in ν_t alternate and that ν_t has dimension $(g+1)N+g$. The variates ν_t will be referred to as the canonical regressors associated with Y_t. In general, $F(L)$ need not be square even though D will be. In addition, for specific models fitting in the general framework (2.1), some of the rows given in (2.3) will be absent altogether.

The lag polynomial $F_{11}(L)$ has dimension $k_1 \times N$, and it is assumed that $\Sigma_{j=0}^{\infty} F_{11j} F_{11j}'$ is nonsingular. Without loss of generality, F_{jj} is assumed to have full row rank k_j (possibly equal to zero) for $j = 2, \ldots, 2g+1$, so that $k = \Sigma_{j=1}^{2g+1} k_j$. These assumptions ensure that, after appropriate rescaling, the moment matrix $\Sigma Z_t Z_t'$ is (almost surely) invertible—i.e., no elements of Z_t are perfectly multicollinear asymptotically—so that the OLS estimator of AD^{-1} is unique.

From (2.2) and (2.3), it is clear that D must be chosen so that its rows select linear combinations of Y_t that are different orders in probability. Thus some of the rows of D can be thought of as generalizations of cointegrating vectors: partitioning $D = [D_1' \cdots D_{2g+1}']'$, so that $Z_t^j = D_j Y_t$, D_1 forms a linear combination of Y_t such that Z_t^1 has mean zero and is $O_p(1)$; D_2 forms a linear combination with mean F_{22} that is also $O_p(1)$. The linear combinations formed with D_3 are $O_p(t^{1/2})$, those formed with D_4 are $O_p(t)$, and so on. In this framework these linear combinations include first differences of the data, in addition to including cointegrating vectors in the sense of Engle and Granger

(1987). The row space of $D_1, \ldots, D_{2g}$ is the subspace of $\Re^k$ spanned by the generalized cointegrating vectors of Y_t.

Derivation of (2.2) *and* (2.3) *from the Jordan Form of A*

Specific examples of (2.2) and (2.3) are given at the end of this section and in Sections 5 and 6. As these examples demonstrate, D and $F(L)$ in general are not unique, although the row spaces of $D_1, [D_1' \; D_2']'$, etc. are. This poses no difficulty for the asymptotic analysis; indeed, it will be seen that only the blocks along the diagonal in $F(L)$ enter into the asymptotic representation for the estimator and F statistic. This nonuniqueness means that in many cases a set of generalized cointegrating vectors can be deduced by inspection of the system, and that $F(L)$ is then readily calculated. For completeness, however, we now sketch how (2.2) and (2.3) can be derived formally from the Jordan canonical form of A.

Let $A = B^{-1}JB$ be the Jordan decomposition of A, so that the matrix J is block diagonal with the eigenvalues of A on its diagonal. Suppose that the Jordan blocks are ordered so that the final block contains all the unit eigenvalues and no eigenvalues less than one in modulus. Let J_1 denote the $k_1 \times k_1$ block with eigenvalues less than one in modulus, let J_2 denote the $(k-k_1) \times (k-k_1)$ block with unit eigenvalues, and partition B conformably with J so that $B = (B_1' \; B_2')'$. The representation (2.2) and (2.3) can be constructed by considering the linear combinations of Y_t formed using B. Let $Z_t^1 = B_1 Y_t$. These definitions and (2.1) imply that

$$Z_t^1 = J_1 Z_{t-1}^1 + B_1 G \Omega^{1/2} \eta_t. \tag{2.4}$$

Because the eigenvalues of J_1 are less than one in modulus by construction, Z_t^1 is integrated of order zero and the autoregressive representation (2.4) can be inverted to yield

$$Z_t^1 = F_{11}(L)\eta_t, \tag{2.5}$$

where $F_{11}(L) = (I - J_1 L)^{-1} B_1 G \Omega^{1/2}$. Thus (2.5) provides the canonical representation for the mean zero stationary elements of Z_t.

The representation for the integrated and deterministic terms comes from considering the final Jordan block, J_2. This block in general has ones on the diagonal, zeros or ones on the first superdiagonal, and zeros elsewhere; the location of the ones above the diagonal determines the number and orders of the polynomials in time and integrated stochastic processes in the representation (2.3). The structure of this Jordan block makes it possible to solve recursively for each of the elements of $B_2 Y_t$. Because J_2 consists of only ones and zeros, each element of $B_2 Y_t$ will be a linear combination of polynomials in time and of partial sums of $\{\eta_s\}$. Letting $\tilde{F}$ denote the matrix of coefficients expressing these linear combinations, one obtains the representation for the remaining linear combinations of Y_t:

$$B_2 Y_t = \tilde{F}\tilde{\nu}_t \qquad \text{where} \qquad \tilde{\nu}_t = \left(1 \; \xi_t^{1\prime} \; t \; \xi_t^{2\prime} \cdots \xi_t^{g\prime}\right)'. \tag{2.6}$$

Elementary row and column operations need to be performed on (2.6) to put $\tilde{F}$ into the lower reduced echelon form of (2.3). Let $\tilde{D}$ be the $(k-k_1)\times(k-k_1)$ invertible matrix summarizing these row and column operations, so that

$$(2.7)\qquad \begin{bmatrix} Z_t^2 \\ \vdots \\ Z_t^{2g+1} \end{bmatrix} = \tilde{D}B_2Y_t = \begin{bmatrix} F_{22} & \cdots & 0 \\ \vdots & & \vdots \\ F_{2g+1,2} & \cdots & F_{2g+1,2g+1} \end{bmatrix} \tilde{\nu}_t \equiv \tilde{D}\tilde{F}\tilde{\nu}_t.$$

The representation (2.2) and (2.3) obtains from (2.5) and (2.7). Let

$$(2.8)\qquad D = \begin{bmatrix} I_{k_1} & 0 \\ 0 & \tilde{D} \end{bmatrix} B \quad \text{and} \quad F(L) = \begin{bmatrix} F_{11}(L) & 0 \\ 0 & \tilde{D}\tilde{F} \end{bmatrix},$$

where $\nu_t = (\eta_t' \, \tilde{\nu}_t')'$ as in (2.3). Combining (2.5), (2.7), and (2.8) yields

$$(2.9)\qquad Z_t = DY_t = F(L)\nu_t,$$

which is the desired result.

This derivation warrants two remarks. First, when an intercept is included in the regression, D can always be chosen so that $F_{21}(L)=0$ in (2.3). Because excluding an intercept is exceptional in applications, it is assumed throughout that $F_{21}(L)=0$ unless explicitly noted otherwise. Second, it turns out that whether $F_{j1}(L)=0$ for $j>2$ is inessential for our results; what matters is that these lag polynomials decay sufficiently rapidly. When D is obtained using the Jordan form, (2.8) indicates that these terms are zero. Because D is not unique, however, in practical applications (and indeed in the examples presented below) it is often convenient to use a transformation D for which some of these terms are nonzero. We therefore allow for nonzero $F_{j1}(L)$ for $j>2$, subject to a summability condition stated in Section 3.

Stacked Single Equation Form of (2.2) *and* (2.3)

The first order representation (2.1) characterizes the properties of the regressors Y_t. In practice, however, only some of the k equations in (2.1) might be estimated. For example, often some of the elements of Y_t will be nonstochastic and some of the equations will be identities. We therefore consider only $n \leqslant k$ regression equations, which can be represented as the regression of CY_t against Y_{t-1}, where C is a $n\times k$ matrix of constants (typically ones and zeros). With this notation, the n regression equations to be estimated are:

$$(2.10)\qquad CY_t = CAY_{t-1} + CG\Omega^{1/2}\eta_t.$$

Let $S_t = CY_t$, $\tilde{A} = CA$, and $\Sigma^{1/2} = CG\Omega^{1/2}$ (so that $\Sigma^{1/2}$ is $n\times N$). Then these regression equations can be written

$$(2.11)\qquad S_t = \tilde{A}Y_{t-1} + \Sigma^{1/2}\eta_t.$$

The asymptotic analysis of the next two sections examines (2.11) in its stacked single equation form. Let $S=[S_2\, S_3 \,\cdots\, S_T]'$, $\eta = [\eta_2\, \eta_3\, \cdots\, \eta_T]'$, $X=$

$[Y_1\ Y_2 \cdots Y_{T-1}]'$, $s = \mathrm{Vec}(S)$, $v = \mathrm{Vec}(\eta)$, and $\beta = \mathrm{Vec}(\tilde{A}')$, where $\mathrm{Vec}(\cdot)$ denotes the column-wise vectorization. Then (2.11) can be written

$$(2.12) \qquad s = [I_n \otimes X]\beta + [\Sigma^{1/2} \otimes I_{T-1}]v.$$

The coefficient vector δ corresponding to the transformed regressors $Z = XD'$ is $\delta = (I_n \otimes D'^{-1})\beta$. With this transformation, (2.12) becomes

$$(2.13) \qquad s = [I_n \otimes Z]\,\delta + [\Sigma^{1/2} \otimes I_{T-1}]v.$$

Thus (2.13) represents the regression equations (2.10), written in terms of the transformed regressors Z_t, in their stacked single-equation form.

An Example

The framework (2.1)–(2.3) is general enough to include many familiar linear econometric models. As an illustration, a univariate second order autoregression with a unit root is cast into this format, an example which will be taken up again in Section 5. Let the scalar time series variable x_t evolve according to

$$(2.14) \qquad x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \eta_t \qquad (t = 1, \ldots, T),$$

where η_t is i.i.d. $(0,1)$. Suppose that a constant is included in the regression of x_t on its two lags, so that Y_t is given by

$$(2.15) \qquad Y_t = \begin{bmatrix} x_t \\ x_{t-1} \\ 1 \end{bmatrix} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \eta_t.$$

Suppose that $\beta_0 = 0$, $\beta_1 + \beta_2 = 1$, and $|\beta_2| < 1$, so that the autoregressive polynomial in (2.14) has a single unit root. Following Fuller (1976) and Dickey and Fuller (1979), because $\beta_0 = 0$ (2.14) can be rewritten

$$(2.16) \qquad x_t = (\beta_1 + \beta_2)x_{t-1} - \beta_2(x_{t-1} - x_{t-2}) + \eta_t$$

so that, since $\beta_1 + \beta_2 = 1$, x_t has an AR(1) representation in its first difference:

$$(2.17) \qquad \Delta x_t = -\beta_2 \Delta x_{t-1} + \eta_t.$$

Although the transformation to Z_t could be obtained by calculating the Jordan canonical form and eigenvectors of the companion matrix in (2.15), a suitable transformation is readily deduced by inspection of (2.16) and (2.17). Because x_t is integrated and Δx_t is stationary with mean zero, (2.16) suggests letting $Z_t^1 = \Delta x_t$, $Z_t^2 = 1$, and $Z_t^3 = x_t$. Then $k_1 = k_2 = k_3 = 1$ and (2.14) can be rewritten

$$(2.18) \qquad x_t = \delta_1 Z_{t-1}^1 + \delta_2 Z_{t-1}^2 + \delta_3 Z_{t-1}^3 + \eta_t,$$

where $\delta_1 = -\beta_2$, $\delta_2 = \beta_0$, and $\delta_3 = \beta_1 + \beta_2$. In the notation of (2.10), (2.18) is $CY_t = (CAD^{-1})Z_{t-1} + \eta_t$, where $C = (1\ 0\ 0)$, $\Sigma^{1/2} = 1$ and A is the transition matrix in (2.15). The transformed variables are $Z_t = DY_t$, where

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

The transformed coefficient vector is $CAD^{-1} = (\delta_1\ \delta_2\ \delta_3) = \delta'$; note that $\beta = D'\delta$.

A straightforward way to verify that D is in fact a suitable transformation matrix is to obtain the representation of Z_t in terms of ν_t. Write $\Delta x_t = \theta(L)\eta_t$, where $\theta(L) = (1 + \beta_2 L)^{-1}$, and use recursive substitution to express x_t as $x_t = \theta(1)\xi_t^1 + \theta^*(L)\eta_t$, where $\theta_i^* = -\Sigma_{j=i+1}^{\infty}\theta_j$ (i.e. $\theta^*(L) = (1-L)^{-1}[\theta(L) - \theta(1)]$). Thus for this model (2.3) is:

$$(2.19)\qquad \begin{bmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{bmatrix} = \begin{bmatrix} \theta(L) & 0 & 0 \\ 0 & 1 & 0 \\ \theta^*(L) & 0 & \theta(1) \end{bmatrix} \begin{bmatrix} \eta_t \\ 1 \\ \xi_t^1 \end{bmatrix}.$$

3. AN ASYMPTOTIC REPRESENTATION OF THE OLS ESTIMATOR

We now turn to the behavior of the OLS estimator $\hat{\delta}$ of δ in the stacked transformed regression model (2.13),

$$(3.1)\qquad \hat{\delta} = (I_n \otimes Z'Z)^{-1}(I_n \otimes Z')s.$$

The sample moments used to compute the estimator are analyzed in Lemmas 1 and 2. The asymptotic representation of the estimator is then presented as Theorem 1. Some restrictions on the moments of η_t and on the dependence embodied in the lag operator $\{F_{j1}(L)\}$ are needed to assure convergence of the relevant random matrices. For the calculations in the proofs, it is convenient to write these latter restrictions as the assumption that $\{F_{j1}(L)\}$ are g-summable as defined by Brillinger (1981). These restrictions are summarized by:

CONDITION 1: (i) $\exists$ some $\mu_4 < \infty$ such that $E(\eta_{it}^4) < \mu_4$, $i = 1, \ldots, N$. (ii) $\Sigma_{j=0}^{\infty} j^g |F_{m1j}| < \infty$, $m = 1, \ldots, 2g+1$.

Condition 1(ii) is more general than necessary if (2.2) and (2.3) are obtained using the Jordan canonical form of A. Because $F_{11}(L)$ in (2.3) is the inverse of a finite autoregressive lag operator with stable roots, (ii) holds for all finite g for $m = 1$; in addition, (ii) holds trivially for $m > 1$ when Z_t is based on the Jordan canonical representation, since in this case $F_{m1}(L) = 0$. Condition 1(ii) is useful when the transformation D is obtained by other means (for example by inspection), which in general produce nonzero $F_{m1}(L)$. In the proofs, it is also assumed that $\{\eta_s\} = 0$, $s \leqslant 0$. This assumption is a matter of technical convenience. For example, Z_0^i, $i = 1, 3, 5, \ldots, 2g+1$ could be treated as drawn from some distribution without altering the asymptotic results. (For a discussion of various assumptions on initial conditions in the univariate case, see Phillips (1987, Section 2).)

The first lemma concerns sample moments based on the components of the canonical regressors. Let $W(t)$ denote a n-dimensional Wiener process, and let $W^m(t)$ denote its $(m-1)$-fold integral, recursively defined by $W^m(t) = \int_0^t W^{m-1}(s)\,ds$ for $m \geqslant 1$, with $W^1(t) \equiv W(t)$. Also let $\Rightarrow$ denote weak convergence of the associated probability measures in the sense of Billingsley (1968). Thus:

LEMMA 1: *Under Condition* 1, *the following converge jointly*:

(a) $T^{-(m+p+1/2)}\Sigma_1^T t^m \xi_t^{p\prime} \Rightarrow \int_0^1 t^m W^p(t)'\,dt, \qquad m \geqslant 0,\ p \geqslant 1,$

(b) $T^{-(m+p)}\Sigma_1^T \xi_t^m \xi_t^{p\prime} \Rightarrow \int_0^1 W^m(t) W^p(t)'\,dt, \qquad m, p \geqslant 1,$

(c) $T^{-(m+p+1)}\Sigma_1^T t^{m+p} \to (m+p+1)^{-1}, \qquad m, p \geqslant 0,$

(d) $T^{-(p+1/2)}\Sigma_1^{T-1} t^p \eta_{t+1}' \Rightarrow \int_0^1 t^p\,dW(t)', \qquad p \geqslant 0,$

(e) $T^{-p}\Sigma_1^{T-1}\xi_t^p \eta_{t+1}' \Rightarrow \int_0^1 W^p(t)\,dW(t)', \qquad p \geqslant 1,$

(f) $T^{-1}\Sigma_1^T (F_{m1}(L)\eta_t)(F_{p1}(L)\eta_t)' \xrightarrow{p} \Sigma_{j=0}^{\infty} F_{m1j}F_{p1j}',$
$m, p = 1, \ldots, 2g+1,$

(g) $T^{-(p+1/2)}\Sigma_1^T t^p (F_{m1}(L)\eta_t) \Rightarrow F_{m1}(1)\int_0^1 t^p\,dW(t),$
$p = 0, \ldots, g;\ m = 1, \ldots, 2g+1,$

(h) $T^{-p}\Sigma_1^T \xi_t^p (F_{m1}(L)\eta_t)' \Rightarrow K_p + \int_0^1 W^p(t)\,dW(t)' F_{m1}(1)',$
$p = 1, \ldots, g;\ m = 1, \ldots, 2g+1;$

where $K_p = F_{m1}(1)'$ *if* $p = 1$ *and* $K_p = 0$ *if* $p = 2, 3, \ldots, g$.

Similar results for $m, p \leqslant 2$ or for $N = 1$ have been shown elsewhere (Phillips (1986, 1987), Phillips and Durlauf (1986), Solo (1984), and Stock (1987)). The proof of Lemma 1 for arbitrary m, p, N relies on results in Chan and Wei (1988) and is given in the Appendix.

Lemma 1 indicates that the moments involving the different components of Z_t converge at different rates. This is to be expected because of the different orders of these variables; for example, ξ_t^p is of order $O_p(t^{p-1/2})$ for $p = 1, 2, \ldots$. To handle the resultant different orders of the moments of these canonical regressors, define the scaling matrix,

$$\Upsilon_T = \begin{bmatrix} T^{1/2}I_{k_1} & & & & & \\ & T^{1/2}I_{k_2} & & & 0 & \\ & & TI_{k_3} & & & \\ & & & \ddots & & \\ & 0 & & & T^{g-1/2}I_{k_{2g}} & \\ & & & & & T^g I_{k_{2g+1}} \end{bmatrix}.$$

In addition, let H denote the $nk \times nk$ "reordering" matrix such that

$$H(I_n \otimes Z') = \begin{bmatrix} I_n \otimes Z_1' \\ I_n \otimes Z_2' \\ \vdots \\ I_n \otimes Z_{2g+1}' \end{bmatrix}$$

where $Z_j = XD_j'$. Using Lemma 1, we now have the following result about the limiting behavior of the moment matrices based on the transformed regressors, Z_t.

LEMMA 2: *Under Condition* 1, *the following converge jointly*:
(a) $\Upsilon_t^{-1}Z'Z\Upsilon_T^{-1} \Rightarrow V$, *where*

$$V_{11} = \Sigma_{j=0}^{\infty} F_{11j}F_{11j}',$$

$$V_{12} = V_{21}' = \Sigma_{j=0}^{\infty} F_{11j}F_{21j}',$$

$$V_{1p} = V_{p1}' = 0, \qquad p = 3,\ldots,2g+1,$$

$$V_{22} = F_{22}F_{22}' + \Sigma_{j=0}^{\infty} F_{21j}F_{21j}',$$

$$V_{mp} = F_{mm}\int_0^1 W^{(m-2)/2}(t)W^{(p-1)/2}(t)'\,dtF_{pp}',$$

$$m = 3,5,7,\ldots,2g+1;\ p = 3,5,7,\ldots,2g+1,$$

$$V_{mp} = F_{mm}\int_0^1 t^{(m-1)/2}W^{(p-1)/2}(t)'\,dtF_{pp}' = V_{pm}',$$

$$m = 2,4,6,\ldots,2g;\ p = 3,5,7,\ldots,2g+1,$$

$$V_{mp} = \frac{2}{p+m-2}F_{mm}F_{pp}', \qquad m = 2,4,6,\ldots,2g;\ p = m+2,\ldots,2g.$$

(b) $H(I_n \otimes \Upsilon_T^{-1})(I_n \otimes Z')(\Sigma^{1/2} \otimes I_{T-1})v \Rightarrow \phi$, *where* $\phi = (\phi_1'\phi_2' \cdots \phi_{2g+1}')'$, *where*

$$\phi_m = \mathrm{Vec}\left[F_{mm}\int_0^1 W^{(m-1)/2}(t)\,dW(t)'\Sigma^{1/2\prime}\right], \qquad m = 3,5,7,\ldots,2g+1,$$

$$\phi_m = \mathrm{Vec}\left[F_{mm}\int_0^1 t^{(m-2)/2}\,dW(t)'\Sigma^{1/2\prime}\right], \qquad m = 4,6,\ldots,2g,$$

$$\phi_2 = \phi_{21} + \phi_{22}, \qquad \textit{where} \quad \phi_{22} = \mathrm{Vec}\left[F_{22}W(1)'\Sigma^{1/2\prime}\right] \quad \textit{and}$$

$$\begin{bmatrix} \phi_1 \\ \phi_{21} \end{bmatrix} \sim N(0,\Psi), \quad \textit{where} \quad \Psi = \begin{bmatrix} \Sigma \otimes V_{11} & \Sigma \otimes V_{12} \\ \Sigma \otimes V_{21} & \Sigma \otimes (V_{22} - F_{22}F_{22}') \end{bmatrix}$$

and where (ϕ_1, ϕ_{21}) *are independent of* $(\phi_{22}, \phi_3, \ldots, \phi_{2g+1})$. *If* $F_{21}(L) = 0$, $\phi_2 = \phi_{22}$ *and* ϕ_{21} *does not appear in the limiting representation.*

The proof is given in the Appendix.[2]

Lemma 2 makes the treatment of the OLS estimator straightforward. Let M be an arbitrary $n \times n$ matrix, and define the function $\Phi(M, V)$ by

$$\Phi(M,V) = \begin{bmatrix} M \otimes V_{11} & \cdots & M \otimes V_{1,2g+1} \\ \vdots & & \vdots \\ M \otimes V_{2g+1,1} & \cdots & M \otimes V_{2g+1,2g+1} \end{bmatrix}.$$

We now have the following theorem.

THEOREM 1: *Under Condition* 1, $H(I_n \otimes \Upsilon_T)(\hat{\delta} - \delta) \Rightarrow \delta^*$, *where* $\delta^* = \Phi(I_n, V)^{-1}\phi$.

PROOF: Use $\hat{\delta} = (I_n \otimes Z'Z)^{-1}(I_n \otimes Z')s$ and (2.13) to obtain

$$\begin{aligned} H(I_n \otimes \Upsilon_T)(\hat{\delta} - \delta) &= H(I_n \otimes \Upsilon_T)(I_n \otimes Z'Z)^{-1}(I_n \otimes \Upsilon_T) \\ &\quad \times (I_n \otimes \Upsilon_T^{-1})(I_n \otimes Z')(\Sigma^{1/2} \otimes I_{T-1})v \\ &= \left[H\left(I_n \otimes \left[\Upsilon_T^{-1}(Z'Z)\Upsilon_T^{-1}\right]\right)H^{-1}\right]^{-1} \\ &\quad \times \left[H\left(I_n \otimes \Upsilon_T^{-1}\right)(I_n \otimes Z')(\Sigma^{1/2} \otimes I_{T-1})v\right] \\ &\Rightarrow \left[H(I_n \otimes V)H^{-1}\right]^{-1}\phi = \Phi(I_n, V)^{-1}\phi \end{aligned}$$

where Lemma 2 ensures the convergence of the bracketed terms after the second equality. *Q.E.D.*

This theorem highlights several important properties of time series regressions with unit roots. First, $\hat{\delta}$ is consistent when there are arbitrarily many unit roots and deterministic time trends, assuming the model to be correctly specified in the sense that the errors are martingale difference sequences. Because the OLS estimator of β in the untransformed system is $\hat{\beta} = (I_n \otimes X'X)^{-1}(I_n \otimes X')s = (I_n \otimes D')\hat{\delta}$, $\hat{\beta}$ is also consistent.

Second, the estimated coefficients on the elements of Z_t having different orders of probability converge at different rates. When some transformed regressors are dominated by stochastic trends, their joint limiting distribution will be nonnormal, as indicated by the corresponding random elements in V. This observation extends to the model (2.1) results already known in certain univariate and multivariate contexts; for example, Fuller (1976) used a similar rotation and scaling matrix to show that, in a univariate autoregression with one unit root and some stationary roots, the estimator of the unit root converges at rate T, while the estimator of the stationary roots converges at rate $T^{1/2}$. In a somewhat

[2] In independent work, Tsay and Tiao (1990) present closely related results for a vector process with some unit roots but with no deterministic components. While our analysis allows for constants and polynomials in t, not considered in their work, their analysis allows for complex unit roots, not allowed in our model.

more general context, Sims (1978) showed that the estimators of the coefficients on the mean zero stationary variables have normal asymptotic distributions. When the regressions involve X_t rather than Z_t, the rate of convergence of any individual element of $\hat{\beta}$, say $\hat{\beta}_j$, is the slowest rate of any of the elements of $\hat{\delta}$ comprising $\hat{\beta}_j$.

Third, when there are no Z_t regressors dominated by stochastic trends—i.e., $k_3 = k_5 = \cdots = k_{2g+1} = 0$—then $\hat{\delta}$ (and thus $\hat{\beta}$) has an asymptotically normal joint distribution: $H(I_n \otimes \Upsilon_T)(\hat{\delta} - \delta) \xrightarrow{\mathscr{L}} N(0, \Sigma \otimes V^{-1})$, where V is nonrandom because the terms involving the integrals $\int_0^1 W^p(t) W^m(t)'\,dt$ and $\int_0^1 W^p(t)\,dW^m(t)'$ are no longer present. In addition, V is consistently estimated by $\Upsilon_T^{-1} Z'Z \Upsilon_T^{-1}$, from which it follows that the asymptotic covariance matrix of $\hat{\beta}$ is consistently estimated by the usual formula. There are several important cases in which $k_3 = k_5 = \cdots = k_{2g+1} = 0$. For example, if the process is stationary around a nonzero mean or a polynomial time trend, this asymptotic normality is well known. Another example arises when there is a single stochastic trend, but this stochastic trend is dominated by a nonstochastic time trend. This situation is discussed by Dickey and Fuller (1979) for an AR(p) and is studied by West (1988) for a VAR, and we return to it as an example in Section 5.

Fourth, Theorem 1 is also related to discussions of "spurious regressions" in econometrics, commonly taken to mean the regression of one independent random walk with zero drift on another. As Granger and Newbold (1974) discovered using Monte Carlo techniques and as Phillips (1986) showed using functional central limit theory, a regression of one independent random walk on another leads to nonnormal coefficient estimators. A related result obtains here for a single regression ($n = 1$) in a bivariate system ($N = 2$) of two random walks ($k_3 = 2$) with no additional stationary components ($k_1 = 0$) and, for simplicity, no intercept ($k_2 = 0$). Then the regression (2.4) entails regressing one random walk against its own lag and the lag of the second random walk which, if $\Omega = I_2$, would have uncorrelated innovations. The two estimated coefficients are consistent, converging jointly at a rate T to a nonnormal limiting distribution.

4. AN ASYMPTOTIC REPRESENTATION FOR THE WALD TEST STATISTIC

The Wald F statistic, used to test the q linear restrictions on β,

$$H_0\colon R\beta = r \qquad \text{vs.} \qquad H_1\colon R\beta \neq r$$

is

$$F = (R\hat{\beta} - r)'\left[R\left(\hat{\Sigma} \otimes (X'X)^{-1}\right)R'\right]^{-1}(R\hat{\beta} - r)/q. \tag{4.1}$$

In terms of the transformed regressors Z_t the null and alternative hypotheses are

$$H_0\colon P\delta = r \qquad \text{vs.} \qquad H_1\colon P\delta \neq r$$

where $P = R(I_n \otimes D')$ and $\delta = (I_n \otimes D'^{-1})\beta$. In terms of P, δ, and Z, the test statistic (4.1) is

$$(4.2)\qquad F=(P\hat{\delta}-r)'\left[P\left(\hat{\Sigma}\otimes(Z'Z)^{-1}\right)P'\right]^{-1}(P\hat{\delta}-r)/q.$$

The F statistics (4.1), computed using the regression (2.12), and (4.2), computed using the regression (2.13), are numerically equivalent.

As in Section 3, it is convenient to rearrange the restrictions from the equation-by-equation ordering implicit in P to an ordering based on the rates of convergence of the various estimators. Accordingly, let $P=\tilde{P}H$, where H is the reordering matrix defined in Section 3, so that $\tilde{P}$ contains the restrictions on the reordered parameter vector $H\delta$. Without loss of generality, $\tilde{P}$ can be chosen to be upper triangular, so that the (i, j) block of $\tilde{P}$, $\tilde{P}_{ij}$ is zero for $i>j$, where $i, j=1,\dots,2g+1$. Let the dimension of $\tilde{P}_{ij}$ be $q_i\times nk_j$, so that q_1 is the number of restrictions being tested that involve the nk_1 coefficients on the transformed variables Z_t^1; these restrictions can potentially involve coefficients on other transformed variables as well. Similarly, q_2 is the number of restrictions involving the nk_2 coefficients on Z_t^2 (and perhaps also $Z_t^3,\dots, Z_t^{2g+1}$), and so forth, so that $q=\Sigma_{j=1}^{2g+1}q_j$.

In the previous section, it was shown that the rates of convergence of the coefficients on the various elements of Z_t differ, depending on the order in probability of the regressor. The implication of this result for the test statistic is that, if a restriction involves estimated coefficients that exhibit different rates of convergence, then the estimated coefficient with the slowest rate of convergence will dominate the test statistic. This is formalized in the next theorem.

THEOREM 2: *Under Condition* 1, $qF\Rightarrow\delta^{*\prime}P^{*\prime}[P^*\Phi(\Sigma^{-1},V)^{-1}P^{*\prime}]^{-1}P^*\delta^*$, *where* δ^* *is defined in Theorem* 1 *and*

$$P^*=\left[\begin{array}{cc|ccc}\tilde{P}_{11} & \tilde{P}_{12} & & 0 & \\ 0 & \tilde{P}_{22} & & & \\ \hline & & \tilde{P}_{33} & & 0\\ & 0 & & \ddots & \\ & & 0 & & \tilde{P}_{mm}\end{array}\right].$$

PROOF: First note that, since $P=\tilde{P}H$ by definition, $P(I_n\otimes\Upsilon_T^{-1})=\tilde{P}H(I_n\otimes\Upsilon_T^{-1})=\tilde{P}(\Upsilon_T^{-1}\otimes I_n)H=\Upsilon_T^{*-1}P_T^*H$, where Υ_T^* is the $q\times q$ scaling matrix

$$\Upsilon_T^*=\begin{bmatrix}T^{1/2}I_{q_1} & & & & & \\ & T^{1/2}I_{q_2} & & & 0 & \\ & & TI_{q_3} & & & \\ & & & \ddots & & \\ & 0 & & & T^{g-1/2}I_{q_{2g}} & \\ & & & & & T^{g}I_{q_{2g+1}}\end{bmatrix},$$

and where $\{P_T^*\}$ is a sequence of $q \times nk$ nonstochastic matrices with the limit $P_T^* \to P^*$ as $T \to \infty$, where P^* is given in the statement of the theorem. (The matrix H reorders the coefficients, while $(I_n \otimes \Upsilon_T^{-1})$ scales them; $H(I_n \otimes \Upsilon_T^{-1}) = (\Upsilon_T^{-1} \otimes I_n)H$ states that these operators commute.) Thus, under the null hypothesis,

$$\begin{aligned}
qF &= (P\hat{\delta} - r)'\left[P\left(\hat{\Sigma} \otimes (Z'Z)^{-1}\right)P'\right]^{-1}(P\hat{\delta} - r) \\
&= \left[P(I_n \otimes \Upsilon_T^{-1})(I_n \otimes \Upsilon_T)(\hat{\delta} - \delta)\right]' \\
&\quad \times \left[P(I_n \otimes \Upsilon_T^{-1})(I_n \otimes \Upsilon_T)\left(\hat{\Sigma} \otimes (Z'Z)^{-1}\right)(I_n \otimes \Upsilon_T)(I_n \otimes \Upsilon_T^{-1})P'\right]^{-1} \\
&\quad \times \left[P(I_n \otimes \Upsilon_T^{-1})(I_n \otimes \Upsilon_T)(\hat{\delta} - \delta)\right] \\
&= \left[\Upsilon_T^{*-1}P_T^*H(I_n \otimes \Upsilon_T)(\hat{\delta} - \delta)\right]' \\
&\quad \times \left[\Upsilon_T^{*-1}P_T^*H\left(\hat{\Sigma} \otimes \left(\Upsilon_T(Z'Z)^{-1}\Upsilon_T\right)\right)H'P_T^{*\prime}\Upsilon_T^{*-1}\right]^{-1} \\
&\quad \times \left[\Upsilon_T^{*-1}P_T^*H(I_n \otimes \Upsilon_T)(\hat{\delta} - \delta)\right] \\
&= \left[P_T^*H(I_n \otimes \Upsilon_T)(\hat{\delta} - \delta)\right]' \\
&\quad \times \left[P_T^*H\left(\hat{\Sigma} \otimes \left(\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1}\right)\right)^{-1}H'P_T^{*\prime}\right]^{-1} \\
&\quad \times \left[P_T^*H(I_n \otimes \Upsilon_T)(\hat{\delta} - \delta)\right] \\
&\Rightarrow (P^*\delta^*)'\left[P^*H(\Sigma \otimes V^{-1})H'P^{*\prime}\right]^{-1}(P^*\delta^*)
\end{aligned}$$

where the last line uses Lemma 2, Theorem 1, $P_T^* \to P^*$, and $\hat{\Sigma} \xrightarrow{p} \Sigma$ (where the consistency of $\hat{\Sigma}$ follows from the stated moment conditions). The result obtains by noting that $H(\Sigma \otimes V^{-1})H' = H(\Sigma^{-1} \otimes V)^{-1}H' = \Phi(\Sigma^{-1}, V)^{-1}$. *Q.E.D.*

Before turning to specific examples, it is possible to make three general observations based on this result. First, in the discussion of Theorem 1 several cases were listed in which, after rescaling, the estimator $\hat{\delta}$ will have a nondegenerate jointly normal distribution and V will be nonrandom. Under these conditions, qF will have the usual χ_q^2 asymptotic distribution.

Second, suppose that only one restriction is being tested, so that $q = 1$. If the test involves estimators that converge at different rates, only that part of the restriction involving the most slowly converging estimator(s) will matter under the null hypothesis, at least asymptotically. This holds even if the limit of the moment matrix $Z'Z$ is not block diagonal or the limiting distribution is nonnormal. This is the analogue in the testing problem of the observation made in the previous section that a linear combination of estimators that individually converge at different rates has a rate of convergence that is the slowest of the various constituent rates. In the proof of Theorem 2, this is an implication of the block diagonality of P^*.

Third, there are some special cases in which the usual χ^2 theory applies to Wald tests of restrictions on coefficients on integrated regressors, say Z_t^3. An example is when the true system is $Y_{1t} = \rho y_{1t-1} + \eta_{1t}$, $|\rho| < 1$, and $Y_{jt} = Y_{jt-1} + \eta_{jt}$, $j = 2, \ldots, k$, where $E\eta_t\eta_t' = \text{diag}(\sigma_1^2, \Sigma_{22})$. In the regression of Y_{1t} on $\gamma_0 + \beta_1 Y_{1t-1} + \beta_2 Y_{2t-1} + \cdots + \beta_k Y_{kt-1}$, $(\hat{\beta}_2, \ldots, \hat{\beta}_k)$ have a joint asymptotic distribution that is a random mixture of normals and the Wald test statistic has an asymptotic χ^2 distribution; for more extensive discussions, see for example Johansen (1988) and Phillips (1988). The key condition is that the integrated regressors and partial sums of the regression error be asymptotically independent stochastic processes. This circumstance seems exceptional in conventional VAR applications and we do not pursue it here.

5. UNIVARIATE AUTOREGRESSIONS WITH UNIT ROOTS

Theorems 1 and 2 provide a simple derivation of the Fuller (1976) and Dickey-Fuller (1979) statistics used to test for unit roots in univariate autoregressions. These results are well known, but are presented here as a straightforward illustration of the more general theorems. We consider a univariate second order autoregression, first without and then with a drift.

EXAMPLE 1—An AR(2) with One Unit Root: Suppose that x_t is generated by (2.14) with one unit root ($\beta_1 + \beta_2 = 1$) and with no drift ($\beta_0 = 0$), the case described in the example concluding Section 2. Because a constant term is included in the regression, $F_{21}(L) = 0$ and V is block diagonal. Combining the appropriate elements from Lemma 1 and using $F(L)$ from (2.19),

$$\Upsilon_T(\hat{\delta} - \delta)$$

$$\Rightarrow \left\{ \begin{array}{c} \delta_1^* \\ \begin{bmatrix} 1 & \theta(1)\int W(t)\,dt \\ \theta(1)\int W(t)\,dt & \theta(1)^2\int W(t)^2\,dt \end{bmatrix}^{-1} \begin{bmatrix} \int dW(t) \\ \theta(1)\int W(t)\,dW(t) \end{bmatrix} \end{array} \right\},$$

where $\delta_1^* \sim N(0, V_{11}^{-1})$, with $V_{11} = \Sigma_{j=0}^{\infty}\theta_j^2 = (1 - \beta_2^2)^{-1}$ and $\theta(1) = (1 + \beta_2)^{-1}$. Thus the coefficients on the stationary terms (Z_{1t} and Z_{2t}) converge at the rate $T^{1/2}$, while $\hat{\delta}_3$ converges at the rate T. In terms of the coefficients of the original regression, $\beta_2 = -\delta_1$ and $\beta_1 = \delta_1 + \delta_3$. Since $T^{1/2}(\hat{\delta}_3 - \delta_3) \xrightarrow{p} 0$, both $\hat{\beta}_1$ and $\hat{\beta}_2$ have asymptotically normal marginal distributions, converging at the rate $T^{1/2}$; however, $\hat{\beta}_1$ and $\hat{\beta}_2$ have a degenerate joint distribution when standardized by $T^{1/2}$.

While the marginal distribution of $\hat{\delta}_1$ is normal, the marginal distribution of $\hat{\delta}_2$ (the intercept) is not, since the "denominator" matrix in the limiting representation of $(\hat{\delta}_2, \hat{\delta}_3)$ is not diagonal and contains random elements, and since $\int W(t)\,dW(t)$ has a nonnormal distribution. Thus tests involving $\hat{\delta}_1$, or $\hat{\delta}_1$ in combination with $\hat{\delta}_3$, will have the usual χ^2 distribution, while tests on any other coefficient (or combination of coefficients) will not.

In the special case that $\beta_0 = 0$ is imposed, so that an intercept is not included in the regression, V is a diagonal 2×2 matrix (here $F_{21}(L) = 0$ even though there is no intercept since ΔX_t has mean zero). Thus, using Theorem 1, the limiting distribution of the OLS estimator of δ_3 has the particularly simple form

$$T(\hat{\delta}_3 - \delta_3) \Rightarrow \int W(t)\, dW(t) \Big/ \left[\theta(1) \int W(t)^2\, dt\right],$$

which reduces to the standard formula when $\beta_2 = 0$, so that $\theta(1) = 1$.

The limiting representation of the square of the Dickey-Fuller t ratio testing the hypothesis that x_t has a unit root, when the drift is assumed to be zero, can be obtained from Theorem 2. When there is no estimated intercept the F statistic testing the hypothesis that $\delta_3 = 1$ has the limit,

$$F \Rightarrow \left[\int W(t)\, dW(t)\right]^2 \Big/ \int W(t)^2\, dt. \tag{5.1}$$

As Solo (1984) and Philips (1987) have shown, (5.1) is the Wiener process limiting representation of the square of the Dickey-Fuller "$\hat{\tau}$" statistic, originally analyzed by Fuller (1976) using other techniques.

EXAMPLE 2—AR(2) with One Unit Root and Nonzero Drift: Suppose that x_t evolves according to (2.14), except that $\beta_0 \neq 0$. If $\beta_1 + \beta_2 = 1$ then the companion matrix A in (2.15) has 2 unit eigenvalues which appear in a single Jordan block. Again, D and $F(L)$ are most easily obtained by rearranging (2.14), imposing the unit root, and solving for the representation of x_t in terms of the canonical regressors. This yields

$$x_t = \delta_1 Z_{t-1}^1 + \delta_2 Z_{t-1}^2 + \delta_4 Z_{t-1}^4 + \eta_t$$

where $Z_t^1 = \Delta x_t - \mu$, $Z_t^2 = 1$, $Z_t^4 = x_t$, $\delta_1 = -\beta_2$, $\delta_2 = \beta_0 - \mu\beta_2$, and $\delta_4 = \beta_1 + \beta_2$, where $\mu = \beta_0/(1 + \beta_2)$ is the mean of Δx_t. In addition,

$$\begin{bmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^4 \end{bmatrix} = \begin{bmatrix} \theta(L) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \theta^*(L) & 0 & \theta(1) & \mu \end{bmatrix} \begin{bmatrix} \eta_t \\ 1 \\ \xi_t^1 \\ t \end{bmatrix},$$

where $\theta(L) = (1 + \beta_2 L)^{-1}$ and $\theta^*(L) = (1 - L)^{-1}[(1 + \beta_2 L)^{-1} - (1 + \beta_2)^{-1}]$, so $k_1 = 1$, $k_2 = 1$, $k_3 = 0$, and $k_4 = 1$. Because there are no elements of Z_t dominated by a stochastic integrated process, $\hat{\delta}$ has an asymptotically normal distribution after appropriate scaling, from which it follows that $\hat{\beta}$ has an asymptotically normal distribution.

If a time trend is included as a regressor, qualitatively different results obtain. Appropriately modifying the state vector and companion matrix in (2.15) to include t, the transformed regressors become:

$$\begin{bmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \\ Z_t^4 \end{bmatrix} = \begin{bmatrix} \theta(L) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \theta^*(L) & 0 & \theta(1) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ 1 \\ \xi_t^1 \\ t \end{bmatrix}.$$

In this case, $\hat{\delta}_2$, $\hat{\delta}_3$ and $\hat{\delta}_4$ have nonnormal distributions. The F statistic testing $\delta_3 = 1$ is the square of the Dickey-Fuller "$\hat{\tau}_\tau$" statistic testing the hypothesis that x_t has a unit root, when it is maintained that x_t is an AR(2) and allowance is made for a possible drift; its limiting representation is given by direct calculation using Theorem 2, which entails inverting the lower 3×3 diagonal block of V. The (nonstandard) limiting distribution of the F statistic testing the joint hypothesis that $\delta_3 = 1$ and $\beta_0 = 0$ can also be obtained directly using this framework.

6. VAR'S WITH SOME UNIT ROOTS

Many hypotheses of economic interest can be cast as linear restrictions on the parameters of VAR's. This section examines F tests of two such hypotheses. The first concerns the lag length in the VAR, and the second is a test for Granger causality. These tests are presented for a trivariate VAR in which each variable has a unit root with nonzero drift in its univariate representation. Four different cases are considered, depending on whether the variables are cointegrated and whether time is included as a regressor. We first present the transformation (2.2) and (2.3) for these different cases, then turn to the analysis of the two tests.

Suppose that the 3×1 vector X_t obeys

$$X_t = \gamma_0 + A(L)X_{t-1} + \eta_t \qquad (t = 1, \dots, T), \tag{6.1}$$

where $n = N = 3$, where $A(L)$ is a matrix polynomial of order p and where it is assumed that γ_0 is nonzero. When time is not included as a regressor, (6.1) constitutes the regression model as well as the true model assumed for X_t; when time is included as a regressor, the regression model is

$$X_t = \gamma_0 + \gamma_1 t + A(L)X_{t-1} + \eta_t \tag{6.2}$$

where the true value of γ_1 is zero.

Suppose that there is at least one unit root in $A(L)$ and that, taken individually each element of X_t is integrated of order one. Then ΔX_t is stationary and can be written,

$$\Delta X_t = \mu + \theta(L)\eta_t \tag{6.3}$$

where by assumption $\mu_i \neq 0$, $i = 1, 2, 3$. This implies that X_t has the representation

$$X_t = \mu t + \theta(1)\xi_t^1 + \theta^*(L)\eta_t. \tag{6.4}$$

Thus each element of X_t is dominated by a time trend. When time is not included as a regressor, Y_t is obtained by stacking $(X_t', X_{t-1}', \dots, X_{t-p+1}', 1)$; when time is included, this stacked vector is augmented by t. Note that, if X_t is not cointegrated, then $\theta(1)$ is nonsingular and $A(L)$ contains 3 unit roots. However, if X_t has a single cointegrating vector α (so that $\alpha'\mu = \alpha'\theta(1) = 0$), then $\theta(1)$ does not have full rank and $A(L)$ has only two unit roots.

As in the previous examples, it is simplest to deduce a suitable transformation matrix D by inspection. If the regression equations do not include a time trend, then (6.1) can be written

$$(6.5)\qquad X_t = A^*(L)(\Delta X_{t-1} - \mu) + [\gamma_0 + A^*(1)\mu] + A(1)X_{t-1} + \eta_t,$$

where $A_j^* = -\sum_{i=j+1}^{p} A_i$, so that $A^*(L)$ has order $p-1$. If the regression contains a time trend, then (6.2) can be written as

$$(6.6)\qquad \begin{aligned} X_t = {} & A^*(L)(\Delta X_{t-1} - \mu) + [\gamma_0 + \gamma_1 + A^*(1)\mu] \\ & + A(1)X_{t-1} + \gamma_1(t-1) + \eta_t. \end{aligned}$$

Note that, if X_t is not cointegrated, then $A(1) = I$, $\theta(L) = [I - A^*(L)L]^{-1}$, and $\mu = [I - A^*(1)]\gamma_0$, while if X_t is cointegrated $A(1) - I$ has rank 1.

The part of the transformation from Y_t to Z_t involving X_t depends on whether the system is cointegrated and on whether a time trend is included in the regression. Using (6.5) and (6.6) as starting points, this transformation, the implied $F(L)$ matrix, and the coefficients in the transformed system are now presented for each of the four cases.

CASE 1—No Cointegration, Time Trend excluded from the Regression: Each element of X_t is (from (6.4)) dominated by a deterministic rather than a stochastic time trend. However, because μ is 3×1 there are two linear combinations of X_t that are dominated not by a time trend, but rather by the stochastic trend component ξ_t^1. Thus Z_t^4 can be chosen to be any single linear combination of X_t; any two linearly independent combinations of X_t that are generalized cointegrating vectors with respect to the time trend can be used as a basis for Z_t^3. To be concrete, let:

$$Z_t^1 = \begin{bmatrix} \Delta X_t - \mu \\ \vdots \\ \Delta X_{t-p+2} - \mu \end{bmatrix}, \qquad Z_t^2 = 1,$$

$$Z_t^3 = \begin{bmatrix} X_{1t} - (\mu_1/\mu_3)X_{3t} \\ X_{2t} - (\mu_2/\mu_3)X_{3t} \end{bmatrix}, \qquad Z_t^4 = X_{3t}.$$

Using (6.4), the two nonstationary components can be expressed as

$$(6.7a)\qquad Z_{it}^3 = \phi_i(1)\xi_t + \phi_i^*(L)\eta_t, \qquad i = 1,2,$$

$$(6.7b)\qquad Z_t^4 = \mu_3 t + e_3'\theta(1)\xi_t^1 + e_3'\theta^*(L)\eta_t,$$

where $\phi_i(1) = [e_i - (\mu_i/\mu_3)e_3]'\theta(1)$, $\phi_i^*(L) = [e_i - (\mu_i/\mu_3)e_3]'\theta^*(L)$, and where

e_j denotes the jth 3-dimensional unit vector. The $F(L)$ matrix is thus given by

$$\text{(6.8)} \qquad \begin{bmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \\ Z_t^4 \end{bmatrix} = \begin{bmatrix} F_{11}(L) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \phi^*(L) & 0 & \phi(1) & 0 \\ e_3'\theta^*(L) & 0 & e_3'\theta(1) & \mu_3 \end{bmatrix} \begin{bmatrix} \eta_t \\ 1 \\ \xi_t^1 \\ t \end{bmatrix}$$

where $\phi^*(L) = [\phi_1^*(L) \;\; \phi_2^*(L)]'$ and $\phi(1) = [\phi_1(1) \;\; \phi_2(1)]'$, and where $k_1 = 3(p-1)$, $k_2 = 1$, $k_3 = 2$, and $k_4 = 1$.

To ascertain which coefficients in the original regression model (6.1) correspond to which block of Z_t, it is convenient to let ζ_t denote the transformed variables in (6.7), so that

$$\text{(6.9)} \qquad \zeta_t = WX_t, \qquad W = \begin{bmatrix} 1 & 0 & -\mu_1/\mu_3 \\ 0 & 1 & -\mu_2/\mu_3 \\ 0 & 0 & 1 \end{bmatrix}.$$

Because $A(1)X_t = A(1)W^{-1}\zeta_t$, the coefficients on Z_t^3 and Z_t^4 can be obtained by calculating $A(1)W^{-1}$. Upon doing so, the regression equation (6.5) becomes,

$$\text{(6.10)} \qquad X_t = A^*(L)(\Delta X_{t-1} - \mu) + [\gamma_0 + A^*(1)\mu] Z_{t-1}^2 + [A(1)e_1 \; A(1)e_2] Z_{t-1}^3 + A(1)[(\mu_1/\mu_3)e_1 + (\mu_2/\mu_3)e_2 + e_3] Z_{t-1}^4 + \eta_t$$

which gives an explicit representation of the coefficients in the transformed regression model in terms of the coefficients in the original model (6.1).

CASE 2—No Cointegration, Time Trend Included in the Regression: When time is included as a regressor, a natural choice for Z_t^4 is t; for F_{44} to have full rank, all elements of X_t will appear, after a suitable transformation, in Z_t^3. Thus Z_t^1 and Z_t^2 are the same as in Case 1, $Z_t^3 = X_t - \mu t = \theta(1)\xi_t + \theta^*(L)\eta_t$, and $Z_t^4 = t$ so that

$$\text{(6.11)} \qquad \begin{bmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \\ Z_t^4 \end{bmatrix} = \begin{bmatrix} F_{11}(L) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \theta^*(L) & 0 & \theta(1) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ 1 \\ \xi_t^1 \\ t \end{bmatrix}.$$

In contrast to the previous case, now $k_3 = 3$ and $k_4 = 1$. Solving for the implied coefficients on the transformed variates as was done in Case 1, the regression equation (6.6) becomes

$$\text{(6.12)} \qquad X_t = A^*(L)(\Delta X_{t-1} - \mu) + [\gamma_0 + A^*(1)\mu] Z_{t-1}^2 + A(1) Z_{t-1}^3 + \mu Z_{t-1}^4 + \eta_t.$$

CASE 3—Cointegration, Time Trend Excluded from the Regression: When X_t is cointegrated with a single cointegrating vector, the 2×3 matrix $F_{33} = \phi(1)$ in

(6.8) no longer has full row rank, so an alternative representation must be developed. Informally, this can be seen by recognizing that, if X_t is cointegrated and if there is a single cointegrating vector, then there can only be two distinct trend components since there is some linear combination of X_t that has no stochastic or deterministic trend. Thus $k_3 + k_4$ will be reduced by one when there is a single cointegrating vector, relative to Case 1. A formal proof of this proposition proceeds by contradiction. Suppose that X_t has a single cointegrating vector α (so that $\alpha' X_t$ is stationary) and that $k_3 = 2$. Let $\alpha = (1\ \alpha_1\ \alpha_2)'$, so that $\alpha'\mu = 0$ implies that α can be rewritten as $\alpha = [1\ \alpha_1\ (-\alpha_1\mu_2 - \mu_1)/\mu_3]'$. Now consider the linear combination of Z_t^3:

$$\begin{aligned} Z_{1t}^3 + \beta Z_{2t}^3 &= [X_{1t} - (\mu_1/\mu_3)X_{3t}] + \beta[X_{2t} - (\mu_2/\mu_3)X_{3t}] \\ &= [1\ \beta\ (-\beta\mu_2 - \mu_1)/\mu_3]\, X_t = \alpha' X_t \end{aligned}$$

where α_1 has been set to β in the final equality. Since $\alpha' X_t$ is stationary by assumption, $Z_{1t}^3 + \alpha_1 Z_{2t}^3$ is stationary; thus $(1\ \alpha_1)\phi(1) = 0$. Since $\phi(1) = F_{33}$, this violates the condition that F_{33} must have full row rank.

To obtain a valid transformation, W must be chosen so that $\zeta_t = WX_t$ has one stationary element, one element dominated by the stochastic trend, and one element dominated by the time trend. To be concrete, let

$$W = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 \\ 1 & 0 & -\mu_1/\mu_3 \\ 0 & 0 & 1 \end{bmatrix},$$

where it is assumed that $\alpha_1 \neq 0$ so that X_{2t} enters the cointegrating relation. Accordingly, let

$$\begin{aligned} \zeta_{1t} &= \alpha' X_t = \alpha'\theta^*(L)\eta_t, \\ \zeta_{2t} &= X_{1t} - (\mu_1/\mu_3)X_{3t} = [e_1 - (\mu_1/\mu_3)e_3]'\theta^*(L)\eta_t \\ &\quad + [e_1 - (\mu_1/\mu_3)e_3]'\theta(1)\xi_t, \\ \zeta_{3t} &= X_{3t} = \mu_3 t + e_3'\theta(1)\xi_t + e_3'\theta^*(L)\eta_t. \end{aligned}$$

Now let

$$Z_t^1 = \begin{bmatrix} \Delta X_t - \mu \\ \vdots \\ \Delta X_{t-p+2} - \mu \\ \zeta_{1t} \end{bmatrix}, \qquad Z_t^2 = 1, \qquad Z_t^3 = \zeta_{2t}, \qquad Z_t^4 = \zeta_{3t},$$

and use the notation $\phi_1^*(L)$ and $\phi_1(1)$ from (6.7a) to obtain:

$$\begin{bmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \\ Z_t^4 \end{bmatrix} = \begin{bmatrix} F_{11}(L) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \phi_1^*(L) & 0 & \phi_1(1) & 0 \\ e_3'\theta^*(L) & 0 & e_3'\theta(1) & \mu_3 \end{bmatrix} \begin{bmatrix} \eta_t \\ 1 \\ \xi_t \\ t \end{bmatrix} \tag{6.13}$$

so that $k_1 = 3(p-1)+1$, $k_2 = 1$, $k_3 = 1$, and $k_4 = 1$. Expressed in terms of the transformed regressors, the regression equation becomes

$$(6.14) \quad \begin{aligned} X_t &= A^*(L)(\Delta X_{t-1} - \mu) + A(1)(e_2/\alpha_1)\zeta_{1t-1} + [\gamma_0 + A^*(1)\mu] Z^2_{t-1} \\ &\quad + A(1)(e_1 - e_2/\alpha_1) Z^3_{t-1} \\ &\quad + A(1)[(\mu_1/\mu_3)e_1 - ([\alpha_2 + (\mu_1/\mu_3)]/\alpha_1)e_2 + e_3] Z^4_{t-1} + \eta_t. \end{aligned}$$

The $F(L)$ matrix in (6.13) and the transformed regression equation (6.14) have been derived under the assumption that there is only one cointegrating vector. If instead there are two cointegrating vectors, so that there is only one unit root in $A(L)$, then $k_3 = 0$ so that there is no Z^3_t transformed regressor. In this case (studied in detail by West (1988)) all the coefficient estimators will be asymptotically normally distributed, and all test statistics will have the usual asymptotic χ^2_q distributions.

CASE 4—Cointegration, Time Trend Included in the Regression: The representation for this case follows by modifying the representation for Case 3 to include a time trend. Let

$$\begin{aligned} \zeta_{1t} &= \alpha' X_t = \alpha'\theta^*(L)\eta_t, \\ \zeta_{2t} &= X_{1t} - \mu_1 t = e_1'\theta^*(L)\eta_t + e_1'\theta(1)\xi_t, \\ \zeta_{3t} &= X_{3t} - \mu_3 t = e_3'\theta^*(L)\eta_t + e_3'\theta(1)\xi_t, \end{aligned}$$

and let

$$Z^1_t = \begin{bmatrix} \Delta X_t - \mu \\ \vdots \\ \Delta X_{t-p+2} - \mu \\ \zeta_{1t} \end{bmatrix}, \qquad Z^2_t = 1, \qquad Z^3_t = \begin{bmatrix} \zeta_{2t} \\ \zeta_{3t} \end{bmatrix}, \qquad Z^4_t = t.$$

Letting $\pi^*(L) = [e_1'\theta^*(L) \ e_3'\theta^*(L)]'$ and $\pi(1) = [e_1'\theta(1) \ e_3'\theta(1)]'$, one obtains:

$$(6.15) \quad \begin{bmatrix} Z^1_t \\ Z^2_t \\ Z^3_t \\ Z^4_t \end{bmatrix} = \begin{bmatrix} F_{11}(L) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \pi^*(L) & 0 & \pi(1) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ 1 \\ \xi_t \\ t \end{bmatrix}$$

so that $k_1 = 3(p-1)+1$, $k_2 = 1$, $k_3 = 2$, and $k_4 = 1$. The transformed regression equation is

$$(6.16) \quad \begin{aligned} X_t &= [A^*(L)(\Delta X_{t-1} - \mu) + A(1)(e_2/\alpha_1)\zeta_{1t-1}] \\ &\quad + [\gamma_0 + A^*(1)\mu] Z^2_{t-1} \\ &\quad + [A(1)(e_1 - e_2/\alpha_1)\zeta_{2t-1} + A(1)(-\alpha_2 e_2/\alpha_1 + e_3)\zeta_{3t-1}] \\ &\quad + A(1)\mu Z^4_{t-1} + \eta_t. \end{aligned}$$

These transformations facilitate the analysis of the two hypothesis tests.

EXAMPLE 1—Tests of Lag Length: A common problem in specifying VAR's is determining the correct lag length, i.e., the order p in (6.1). Consider the null hypothesis that $A(L)$ has order $m \geqslant 1$ and the alternative that $A(L)$ has order $p > m$:

$$H_1\colon A_j = 0,\ j = m+1, \dots, p, \quad \text{vs.} \quad K_1\colon A_j \neq 0, \text{some } j = m+1, \dots, p.$$

The restrictions on the parameters of the transformed regression model could be obtained by applying the rotation form Y_t to Z_t, as discussed in general in Section 4. However, in these examples the restrictions are readily deduced by comparing (6.10), (6.12), (6.14), and (6.16) with (6.5) and (6.6). By definition $A_{j+1} = A^*_{j+1} - A^*_j$ and $A^*_p = 0$; thus $A_j = 0$ for $j \geqslant m+1$ implies and is implied by $A^*_j = 0$, $j \geqslant m$. In terms of the transformed regression model, H_1 and K_1 therefore become

$$H_1^*\colon A^*_j = 0,\ j = m, \dots, p-1, \quad \text{vs.}$$

$$K_1^*\colon A^*_j \neq 0, \text{some } j = m, \dots, p-1.$$

In each of the four cases, the restrictions embodied in H_1^* are linear in the coefficients of the Z_t^1 regressors. Since the regression is assumed to include a constant term, $F_{21}(L) = 0$ in each of the four cases. Thus in each case V is block diagonal, with the first block corresponding to the stationary mean zero regressors. It follows directly from Theorem 2 that (q times) the corresponding test statistic will have the usual $\chi^2_{3(p-m)}$ distribution.

EXAMPLE 2—Granger Causality Tests: The second hypothesis considered is that lags of X_{2t} do not help predict X_{1t} given lagged X_{1t} and X_{3t}:

$$H_2\colon A_{12j} = 0,\ j = 1, \dots, p, \quad \text{vs.} \quad K_2\colon A_{12j} \neq 0, \text{some } j = 1, \dots, p.$$

In terms of the transformed regression models, this becomes

$$H_2^*\colon A_{12}(1) = 0 \quad \text{and} \quad A^*_{12j} = 0,\ j = 1, \dots, p-1, \quad \text{vs.}$$

$$K_2^*\colon A_{12}(1) \neq 0 \quad \text{or} \quad A^*_{12j} \neq 0, \text{some } j = 1, \dots, p-1.$$

As in the previous example, the second set of restrictions in H_2^* are linear in the coefficients of the Z_t^1 regressors in each of the four cases. However, H_2^* also includes the restriction that $A_{12}(1) = 0$. Thus whether the F statistic has a nonstandard distribution hinges on whether $A_{12}(1)$ can be written as a coefficient on a mean zero stationary regressor.

In Case 1, $A_{12}(1)$ is the (1,2) element of the matrix of coefficients on Z_t^3 in (6.10), and it does not appear alone as a coefficient, or a linear combination of coefficients, on the stationary mean zero regressors. It follows that in Case 1 the restriction on $A_{12}(1)$ imparts a nonstandard distribution to the F statistic, even though the remaining restrictions involve coefficients on Z_t^1. In Case 2, inspection of (6.12) leads to the same conclusion: $A_{12}(1)$ appears as a coefficient on Z_t^3, and $A_{12}(1) = 0$ implies and is implied by the corresponding coefficient on Z_t^3 equaling zero. Thus the test statistics will have a nonstandard limiting distribution. However, because F_{22}, F_{33}, and F_{44} differ between Cases 1 and 2, the distributions of the F statistic will differ.

In both Case 1 and Case 2, the distribution of the F test depends on nuisance parameters and thus cannot conveniently be tabulated. However, since these nuisance parameters can be estimated consistently, the limiting distribution of the test statistic can be computed numerically. Since V is block diagonal, in both cases the statistic takes on a relatively simple (but nonstandard) asymptotic form. Let $\tilde{V}$ denote the (random) lower $(k_2+k_3+k_4)\times(k_2+k_3+k_4)$ block of V and let $\tilde{P}_{11}$ and $\tilde{P}_{33}$ denote the restrictions on the coefficients corresponding to the transformed regressors Z_1 and Z_3 in the stacked single-equation form, respectively, as detailed in Section 4. Then a straightforward calculation using Theorem 2 shows that $pF \Rightarrow F_1 + F_2$ where $F_1 = (\tilde{P}_{11}\delta_1^*)'[\tilde{P}_{11}(\Sigma \otimes V_{11}^{-1})\tilde{P}_{11}']^{-1}(\tilde{P}_{11}\delta_1^*) \sim \chi^2_{p-1}$ and $F_2 = (\tilde{P}_{33}\delta_3^*)'[\tilde{P}_{33}\Phi(\Sigma^{-1},\tilde{V})^{-1}\tilde{P}_{33}']^{-1}(\tilde{P}_{33}\delta_3^*)$, where the elements of F_2 consist in part of the functionals of Wiener processes given in Lemma 2 and where the block diagonality of V and Lemma 2(b) imply that F_1 and F_2 are independent.

In Cases 3 and 4, X_t is cointegrated and the situation changes. In both (6.14) and (6.16), $A_{12}(1)$ appears as a coefficient on ζ_{1t}, the "equilibrium error" formed by the cointegrating vector. Since ζ_{1t} is stationary with mean zero, the estimator of $A_{12}(1)$ will thus be asymptotically normal, converging at the rate $T^{1/2}$, and the F-test will have an asymptotic χ^2_p/p distribution.[3]

At first glance, the asymptotic results seem to depend on the arbitrarily chosen transformations (6.8), (6.11), (6.13), and (6.15). This is, however, not so: while these transformations have been chosen to make the analysis simple, the same results would obtain for any other transformation of the form (2.2) and (2.3). One implication of this observation is that, since X_{1t}, X_{2t}, and X_{3t} can be permuted arbitrarily in the definitions of ζ_t used to construct D and $F(L)$ in the four cases, the F statistic testing the exclusion of any one of the regressors and its lags will have the same properties as given here for X_{2t-1} and its lags.

The intuition behind these results is simple. Each element of X_t has a unit root —and thus a stochastic trend—in its univariate autoregressive representation. In Cases 1 and 2, these stochastic trends are not cointegrated and dominate the long run relation among the variables (after eliminating the effect of the deterministic time trend) so that a test of $A_{12}(1)=0$ is like a test of one of the coefficients in a regression of one random walk on two others and its lags. In contrast, when the system is cointegrated, there are only two nondegenerate stochastic trends. Including X_{1t-1} and X_{3t-1} in the regression "controls for" these trends, so that a test of $A_{12}(1)=0$ (and the other Granger noncausality restrictions) behaves like a test of coefficients on mean zero stationary regressors.

[3] This assumes that $\alpha_1 \neq 0$, so that there is a linear combination involving X_{2t} which is stationary. If $\alpha_1 = 0$, there is no such linear combination, in which case the test statistic will have a nonstandard asymptotic distribution.

7. PRACTICAL IMPLICATIONS AND CONCLUSIONS

Application of the theory developed in this paper clearly is computationally demanding. Application of the corresponding Bayesian theory, conditional on initial observations and Gaussian disturbances, can be simpler and in any case is quite different. Because the Bayesian approach is entirely based on the likelihood function, which has the same Gaussian shape regardless of the presence of nonstationarity, Bayesian inference need take no special account of nonstationarity. The authors of this paper do not have a consensus opinion on whether the Bayesian approach ought simply to replace classical inference in this application. But because in this application, unlike most econometric applications, big differences between Bayesian and classical inference are possible, econometricians working in this area need to form an opinion as to why they take one approach or the other.

This work shows that the common practice of attempting to transform models to stationary form by difference or cointegration operators whenever it appears likely that the data are integrated is in many cases unnecessary. Even with a classical approach, the issue is not whether the data are integrated, but rather whether the estimated coefficients or test statistics of interest have a distribution which is nonstandard if in fact the regressors are integrated. It will often be the case that the statistics of interest have distributions unaffected by the nonstationarity, in which case the hypotheses can be tested without first transforming to stationary regressors. It remains true, of course, that the usual asymptotic distribution theory generally is not useful for testing hypotheses that cannot entirely be expressed as restrictions on the coefficients of mean zero stationary linear combinations of Y_t. These "forbidden" linear combinations can thus be characterized as those which are orthogonal to the generalized cointegrating vectors comprising the row space of D_1, i.e. to those generalized cointegrating vectors that reduce Y_t to a stationary process with mean zero. In particular, individual coefficients in the estimated autoregressive equations are asymptotically normal with the usual limiting variance, unless they are coefficients of a variable which is nonstationary and which does not appear in any of the system's stationary linear combinations.

Whether to use a transformed model when the distribution of a test of the hypothesis of interest depends on the presence of nonstationarity is a difficult question. A Bayesian approach finds no reason ever to use a transformed model, except possibly for computational simplicity. Under a classical approach, if one has determined the form of the transformed model on the basis of preliminary tests for cointegration and unit roots, use of the untransformed model does not avoid pretest bias because the distribution theory for the test statistics will depend on the form of the transformation. One consideration is that tests based on the transformed model will be easier to compute. Tests based on the two versions of the model will, however, be different even asymptotically, and might have different power, small-sample accuracy, or degree of pretest bias. We regard comparison of classical tests based on the transformed and untransformed models as an interesting open problem.

To use classical procedures based on the asymptotic theory, one must address the discontinuity of the distribution theory. It can and will occur that a model has all its estimated roots less than one and the stationary asymptotic theory (appropriate if all roots are in fact less than one) rejects the null hypothesis of the maximal root being greater than, say, .98, yet the nonstationary asymptotic theory (appropriate if the maximal root is one) fails to reject the null hypothesis of a unit root. In practice it may be usual to treat something like the convex closure of the union of the stationary-theory confidence region and the unit root null hypothesis as if it were the actual confidence region. Is this a bad approximation to the true confidence region based on exact distribution theory? We should know more than we do about this.

When nonstationarity itself is not the center of interest or when the form and degree of nonstationarity is unknown, the discontinuity of the asymptotic theory raises serious problems of pretesting bias. As we have already noted, in order to test a null hypothesis of Granger causal priority with the classical theory one must first decide on whether nonstationarity is present and, if so, its nature. To the extent that the results of preliminary tests for nonstationarity and cointegration are correlated with results of subsequent tests for causal priority, interpretation of the final results is problematic. When the preliminary tests suggest a particular nonstationary form for the model but at a marginal p-value of, say, .10 or .15, one could consider tests of the hypotheses of interest both under the integrated and nonintegrated maintained hypotheses. Results are likely often to differ, however, and this asymptotic theory offers no guidance as to how to resolve the differences with formal inference.

This paper provides the asymptotic distribution theory for statistics from autoregressive models with unit roots. Now that these difficulties are resolved, it appears that a new set of issues—related to the logical foundations of inference and the handling of pretest bias—arise to preserve this area as an arena of controversy.

Dept. of Economics, University of Minnesota, Minneapolis, MN 55455, U.S.A.
Kennedy School of Government, Harvard University, Cambridge, MA 02138, U.S.A.
Dept. of Economics, Northwestern University, Evanston, IL 60208, U.S.A.

Manuscript received January, 1987; final revision received March, 1989.

APPENDIX

PROOFS OF LEMMAS 1 AND 2

PROOF OF LEMMA 1: The proof of this lemma uses results developed in Chan and Wei (1988, Theorem 2.4), who consider the convergence of related terms to functionals of Wiener processes and to stochastic integrals based on Wiener processes. Throughout we condition on $\{\eta_s = 0\}$, $s \leqslant 0$. This is done for convenience, and could, for example, be weakened to permit the initial conditions for Z_t^1 to be drawn from a stationary distribution.

(a) First consider $m=0$:

$$T^{-(p+1/2)}\sum_1^T \xi_t^p = T^{-(p+1/2)}\sum_{t=1}^T \sum_{s=1}^t \xi_s^{p-1}$$
$$= T^{-1}\sum_{t=1}^T\left[T^{-(p-1/2)}\sum_{s=1}^t \xi_s^{p-1}\right].$$

Thus, if $T^{-(p-1/2)}\Sigma_{s=1}^t \xi_s^{p-1} \to W^p(\tau)$, where $\tau = \lim(t/T)$, then

$$\text{(A.1)} \qquad T^{-(p+1/2)}\sum_{s=1}^t \xi_s^p \to W^{p+1}(\tau) \quad \text{or} \quad T^{-(p+1/2)}\xi_t^{p+1} \to W^{p+1}(\tau)$$

by the Continuous Mapping Theorem (for the univariate case, see Hall and Heyde (1980, Appendix II); for the multivariate case, see Chan and Wei (1988)). Letting $\xi_t^0 \equiv \eta_t$ and using Chan and Wei's (1988) results, $T^{-1/2}\xi_t^1 = T^{-1/2}\Sigma_{s=1}^t \xi_s^0 \Rightarrow W(\tau)$ so (A.1) follows by induction.

For $m>0$,

$$T^{-(m+p+1/2)}\sum_1^T t^m \xi_t^p = T^{-1}\sum_1^T (t/T)^m\left[T^{-(p-1/2)}\xi_t^p\right] \Rightarrow \int_0^1 \tau W^p(\tau)\,d\tau.$$

$$\text{(b)} \qquad T^{-(m+p)}\sum_1^T \xi_t^m \xi_t^{p\prime} = T^{-1}\sum_1^T\left(T^{-(m-1/2)}\xi_t^m\right)\left(T^{-(p-1/2)}\xi_t^p\right)'$$
$$\Rightarrow \int_0^1 W^m(\tau)W^p(\tau)'\,d\tau$$

where we use (a) for the convergence of $T^{-(p-1/2)}\xi_t^p$.

(c) Obtains by direct calculation.

$$\text{(d)} \qquad T^{-(p+1/2)}\sum_1^T t^p \eta_{t+1}' = T^{-1/2}\sum_1^T (t/T)^p \eta_{t+1}' \Rightarrow \int_0^1 t^p\,dW(t)',$$

$$\text{(e)} \qquad T^{-p}\sum_1^T \xi_t^p \eta_{t+1}' = T^{-1/2}\sum_1^T\left(T^{-(p-1/2)}\xi_t^p\right)\eta_{t+1}' \Rightarrow \int_0^1 W^p(t)\,dW(t)'$$

where the convergence follows using Theorem 2.4 (ii) of Chan and Wei (1988) for $p=1$ and using their Theorem 2.4 (i) for $p>1$.

(f) This follows, using Chebyschev's inequality, from $\Sigma|F_{11j}| < \infty$ and bounded 4th moments.

(g) The approach used to prove (g) and (h) extends the argument used in Solo (1984). (We thank an anonymous referee for substantially simplifying our earlier proofs of (g) and (h).) First consider the case $p=0$. From Lemma 1(d) and Condition 1(ii) it follows that:

$$T^{-1/2}\sum_1^T F_{m1}(L)\eta_t = F_{m1}(1)\left(T^{-1/2}\xi_T\right) + T^{-1/2}F_{m1}^*(L)\eta_T \Rightarrow F_{m1}(1)\int_0^1 dW(t)$$

where $F_{m1j}^* = -\Sigma_{i=j+1}^\infty F_{m1i}$, yielding the desired result.

The general case is proven by induction. Let $H(L)$ be a matrix lag polynomial. Assume that if

$$\sum_0^\infty j^k|H_j| < \infty \qquad (k=0,\dots,p-1),$$

then

$$T^{-(k+1/2)}\sum_1^T t^k H(L)\eta_t \Rightarrow H(1)\int_0^1 t^k\,dW(t) \qquad (k=0,\dots,p-1).$$

Now note that $F_{m1}(L)\eta_t = F_{m1}(1)\eta_t + F^*_{m1}(L)\Delta\eta_t$, so that the term in question can be written,

$$\text{(A.2)}\qquad T^{-(p+1/2)}\sum_1^T t^p F_{m1}(L)\eta_t = F_{m1}(1)\left[T^{-(p+1/2)}\sum_1^T t^p\eta_t\right] + T^{-(p+1/2)}\sum_1^T t^p F^*_{m1}(L)\Delta\eta_t$$

$$= F_{m1}(1)\left[T^{-(p+1/2)}\sum_1^T t^p\eta_t\right] + T^{-1/2}F^*_{m1}(L)\eta_T$$

$$+ T^{-(p-1/2)}\sum_1^{T-1}\left[t^p - (t+1)^p\right]F^*_{m1}(L)\eta_t$$

$$= F_{m1}(1)\left[T^{-(p+1/2)}\sum_1^T t^p\eta_t\right] + T^{-1/2}F^*_{m1}(L)\eta_T$$

$$+ \sum_{k=0}^{p-1} d_k T^{-(p-k)}\left[T^{-(k+1/2)}\sum_1^{T-1} t^k F^*_{m1}(L)\eta_t\right]$$

where $\{d_k\}$ are the constants from the binomial expansion, $t^p - (t+1)^p = \sum_{k=0}^{p-1} d_k t^k$.

The first term in (A.2) has the desired limit by Lemma 1(d). The second term in (A.2) vanishes in probability by condition 1(ii). The final p terms in (A.2) converge to zero by the inductive assumption if $\sum_{j=0}^{\infty} j^k |F^*_{m1j}| < \infty$ for all $k = 0,\ldots,p-1$. To verify this final condition, note that

$$\text{(A.3)}\qquad \sum_{j=0}^{\infty} j^k |F^*_{m1j}| \leqslant \sum_{i=0}^{\infty}\sum_{j=0}^{i-1} j^k |F_{m1i}| \leqslant C_{k+1}\sum_{j=0}^{\infty} j^{k+1}|F_{m1j}|,$$

where C_{k+1} is a finite constant. The final expression in (A.3) is bounded by assumption (Condition 1(ii)) so the result obtains under the inductive assumption. Since the inductive assumption is satisfied for $p = 0$, the result follows for $p = 0, 1, \ldots, g$.

(h) We prove the lemma for $F_{11}(L)$, showing the result first for $p = 1$. Following Stock (1987), write

$$T^{-1}\sum_1^T \xi_t (F_{11}(L)\eta_t)' = H_3 - H_4 + \left(T^{-1}\sum_1^T \xi_{t-1}\eta_t' + H_5\right)\sum_0^{T-1} F'_{11j}$$

where

$$H_3 = T^{-1}\sum_0^{T-1}\rho_j F'_{11j}, \qquad \rho_j = \sum_{j+1}^{T}(\xi_t - \xi_{t-j})\eta'_{t-j},$$

$$H_4 = T^{-1}\sum_0^{T-1}\phi_j F'_{11j}, \qquad \phi_j = \sum_{T-j+1}^{T}\xi_t\eta_t',$$

$$H_5 = T^{-1}\sum_1^T \eta_t\eta_t'.$$

Now, $H_5 \xrightarrow{p} I$ and $T^{-1}\sum\xi_{t-1}\eta_t' \Rightarrow \int_0^1 W(t)\,dW(t)'$. In addition, Stock (1987) shows that $H_3 \xrightarrow{p} 0$ and $H_4 \xrightarrow{p} 0$ if $\sum|F_{11j}| < \infty$, which is true under Condition 1. Thus, for $p = 1$,

$$T^{-1}\sum_1^T \xi_t (F_{11}(L)\eta_t)' \Rightarrow F_{11}(1)' + \int_0^1 W(t)\,dW(t)' F_{11}(1)',$$

which is the desired result.

The case of $p \geqslant 2$ is proven by induction. Let $H(L)$ be a lag polynomial matrix with $\sum_0^\infty j^k|H_j| < \infty$, $k = 1,\ldots,p-1$, and assume that

$$T^{-k}\sum_1^T \xi_t^k (H(L)\eta_t)' \Rightarrow K_k + \int_0^1 W^k(t)\,dW(t)' H(1)', \qquad k = 1,\ldots,p-1,$$

where K_k is given in the statement of the lemma with $H(1)$ replacing $F_{m1}(1)$. Now write

$$\text{(A.4)}\qquad T^{-p}\sum_1^T \xi_t^p[F_{m1}(L)\eta_t]' = T^{-p}\sum_1^T \xi_t^p\eta_t' F_{m1}(1)' + T^{-p}\sum_1^T \xi_t^p[F_{m1}^*(L)\Delta\eta_t]'$$

where $F_{m1\,j}^* = -\sum_{i=j+1}^{\infty} F_{m1i}$.

Consider the first term in (A.4). Noting that $\xi_t^p = \eta_t + \sum_{k=1}^{p} \xi_{t-1}^k$, one obtains:

$$\text{(A.5)}\qquad T^{-p}\sum_1^T \xi_t^p\eta_t' F_{m1}(1)' = T^{-p}\sum_1^T \xi_{t-1}^p\eta_t' F_{m1}(1)' + T^{-p}\sum_1^T \eta_t\eta_t' F_{m1}(1)'$$
$$+\sum_{k=1}^{p-1} T^{-(p-k)}\left[T^{-k}\sum_1^T \xi_{t-1}^k\eta_t'\right]F_{m1}(1)'.$$

Since $p \geqslant 2$, Lemma 1(e) and Condition 1(ii) ensure that all but the first terms in (A.5) vanish in probability. Applying Lemma 1(e) to the first term in (A.5), one obtains (for $p \geqslant 2$):

$$T^{-p}\sum_1^T \xi_t^p\eta_t' F_{m1}(1)' \Rightarrow \int_0^1 W(t)\, dW(t)' F_{m1}(1)'.$$

All that remains is to show that the second term in (A.4) vanishes in probability. Now

$$\text{(A.6)}\qquad T^{-p}\sum_1^T \xi_t^p[F_{m1}^*(L)\Delta\eta_t]' = T^{-p}\sum_{t=1}^T\sum_{s=1}^t \xi_s^{p-1}[F_{m1}^*(L)\Delta\eta_t]'$$
$$= T^{-p}\sum_{s=1}^T \xi_s^{p-1}\left[\sum_{t=s}^T F_{m1}^*(L)\Delta\eta_t\right]'$$
$$= T^{-p}\sum_{s=1}^T \xi_s^{p-1}[F_{m1}^*(L)\eta_T]' - T^{-p}\sum_{s=1}^T \xi_s^{p-1}[F_{m1}^*(L)\eta_{s-1}]'$$
$$= \left[T^{-(p-1/2)}\sum_1^T \xi_t^{p-1}\right]\left[T^{-1/2}F_{m1}^*(L)\eta_T\right]'$$
$$- T^{-p}\sum_1^T \eta_t[F_{m1}^*(L)\eta_{t-1}]'$$
$$- \sum_{k=1}^{p-1} T^{-(p-k)}\left[T^{-k}\sum_{t=1}^T \xi_{t-1}^k(F_{m1}^*(L)\eta_{t-1})'\right].$$

The first term in (A.6) vanishes by Lemma 1(a), the inequality (A.3), and condition 1(ii). The second term vanishes by Chebyschev's inequality and Condition 1(ii). The remaining $p-1$ terms in (A.6) vanish by the inductive assumption if $\sum j^k|F_{m1\,j}^*| < \infty$, but this final condition is implied by (A.3) and Condition 1(ii). Since the inductive assumption was shown to hold for $p=1$, the result for $p \geqslant 2$ follows.

PROOF OF LEMMA 2: We calculate the limits of the various blocks separately. The joint convergence of these blocks is assured by Theorem 2.4 of Chan and Wei (1988).

(a) Consider $(\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1})_{pm}$ for

(i) $p = m = 1$,

(ii) $p = 1, \quad m = 2,4,6,\ldots, M$,

(iii) $p = 1, \quad m = 3,5,7,\ldots, M-1$,

(iv) $p = 3,5,7,\ldots, M-1, \quad m = 3,5,7,\ldots, M-1$,

(v) $p = 2,4,6,\ldots, M, \quad m = 3,5,7,\ldots, M-1$,

(vi) $p = 2,4,6,\ldots, M, \quad m = 2,4,6,\ldots, M$,

where M is an even integer.

(i) $p=m=1$:

$$\left(\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1}\right)_{11}=T^{-1}\Sigma\left(F_{11}(L)\eta_t\right)\left(F_{11}(L)\eta_t\right)'$$
$$\xrightarrow{p}\sum_0^\infty F_{11_j}F_{11_j}'\equiv V_{11}\qquad\text{by Lemma 1(f).}$$

(ii) $p=1$, $m=2$:

$$\left(\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1}\right)_{12}=T^{-1}\Sigma\left(F_{11}(L)\eta_t\right)\left(F_{21}(L)\eta_t+F_{22}\right)'$$
$$=T^{-1}\Sigma\left(F_{11}(L)\eta_t\right)\left(F_{21}(L)\eta_t\right)'+T^{-1}\Sigma F_{11}(L)\eta_t'F_{22}'$$
$$\xrightarrow{p}\sum_0^\infty F_{11_j}F_{21_j}'\equiv V_{12}$$

by Lemma 1(f) and (g), using $\Sigma_0^\infty|F_{21_j}|<\infty$.

$p=1$, $m=4,6,\ldots,M$:

$$\text{(A.7)}\qquad \left(\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1}\right)_{1m}=T^{-m/2}\Sigma\left(F_{11}(L)\eta_t\right)\left(F_{mm}t^{(m-2)/2}+F_{mm-1}\xi_t^{(m-2)/2}\right.$$
$$\left.+\cdots+F_{m2}+F_{m1}(L)\eta_T\right)'$$
$$=T^{-(m-2)/2-1}\Sigma\left(F_{11}(L)\eta_t\right)t^{(m-2)/2}F_{mm}'$$
$$+T^{-(m-2)/2}T^{-1}\Sigma\left(F_{11}(L)\eta_t\right)\xi_t^{(m-2)/2'}F_{mm-1}'$$
$$+\cdots+T^{-(m-2)/2}T^{-1}\Sigma\left(F_{11}(L)\eta_t\right)F_{m2}'$$
$$+T^{-(m-2)/2}T^{-1}\Sigma\left(F_{11}(L)\eta_t\right)\left(F_{m1}(L)\eta_t\right)'.$$

Each of the terms in (A.7) converges to zero in probability by Lemma 1(g), (h), and (f) (using $\Sigma|F_{m1_j}|<\infty$) respectively, for $m>2$. That the omitted intermediate terms converge to zero in probability follows by induction. Thus $(\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1})_{1m}\xrightarrow{p}0$, $m=4,6,8,\ldots,M$.

(iii) $p=1$, $m=3,5,7,\ldots,M-1$:

$$\text{(A.8)}\qquad \left(\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1}\right)_{1m}=T^{-m/2}Z_1'Z_m$$
$$=T^{-m/2}\Sigma\left(F_{11}(L)\eta_t\right)\left(F_{mm}\xi_t^{(m-1)/2}+F_{mm-1}t^{(m-3)/2}\right.$$
$$\left.+\cdots+F_{m2}+F_{m1}(L)\eta_t\right)'$$
$$=T^{-1/2}T^{-(m-1)/2}\Sigma\left(F_{11}(L)\eta_t\right)\xi_t^{(m-1)/2'}F_{mm}'$$
$$+T^{-1/2}T^{-((m-3)/2+1)}\Sigma\left(F_{11}(L)\eta_t\right)t^{(m-3)/2}F_{mm-1}'$$
$$+\cdots+T^{-(m-2)/2}T^{-1}\Sigma\left(F_{11}(L)\eta_t\right)F_{m2}'$$
$$+T^{-(m-2)/2}T^{-1}\Sigma\left(F_{11}(L)\eta_t\right)\left(F_{m1}(L)\eta_t\right)'.$$

Each of the terms in (A.8) vanish asymptotically by application of Lemma 1(h), (g), and (f) (using $\Sigma|F_{m1_j}|<\infty$), for $m\geqslant 3$. By induction, the intermediate terms also vanish thus $(\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1})_{1m}\xrightarrow{p}0$, $m=3,5,7,\ldots,M-1$.

(iv) $p,m=3,5,7,\ldots,M-1$:

$$\text{(A.9)}\qquad \left(\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1}\right)_{mp}=T^{-(m-1)/2}T^{-(p-1)/2}Z_p'Z_m$$
$$=T^{-(m+p-2)/2}\Sigma\left(F_{mm}\xi_t^{(m-1)/2}+F_{mm-1}t^{(m-3)/2}+\cdots+F_{m1}(L)\eta_t\right)$$
$$\times\left(F_{pp}\xi_t^{(p-1)/2}+F_{pp-1}t^{(p-3)/2}+\cdots+F_{p1}(L)\eta_t\right)'.$$

The leading term in (A.9) converges to a random variable:

$$\text{(A.10)}\qquad F_{mm}T^{-(m+p-2)/2}\Sigma\xi_t^{(m-1)/2}\xi_t^{(p-1)/2'}F_{pp}'\Rightarrow F_{mm}\int_0^1W^{(m-1)/2}(t)W^{(p-1)/2}(t)'\,dtF_{pp}'$$

by Lemma 1(b). We now argue that the remaining terms in (1.3) converge to zero in probability. First, it follows from (A.10) that the cross terms in $(\xi_t^{(i-1)/2}, \xi_t^{(j-1)/2}) \xrightarrow{p} 0$ for $i \leq m$, $j \leq p$, and $i+j < m+p$. Second, the terms in $(\xi_t^{(i-1)/2}, t^{(j-3)/2})$ and $(t^{(i-3)/2}, \xi_t^{(j-1)/2})$ vanish for $i \leq m$, $j \leq p$. For example,

$$T^{-(m+p-2)/2}\Sigma F_{mm}\xi_t^{(m-1)/2}t^{(p-3)/2}F'_{pp-1}$$
$$= F_{mm}T^{-1/2}T^{-[(m-1)/2+(p-3)/2+1/2]}\Sigma\xi_t^{(m-1)/2}t^{(p-3)/2}F'_{pp-1} \xrightarrow{p} 0$$

by Lemma 1(a). Finally, the cross terms of $(\xi_t^{(i-1)/2}, F_{p1}(L)\eta_t)$ and $(t^{(i-3)/2}, F_{p1}(L)\eta_t)$, $i \leq m$, all converge in probability to zero using the arguments in (ii) and (iii) above. Thus, for $p, m = 3, 5, 7, \ldots, M-1$,

$$\left(\Upsilon_T^{-1}Z'Z\Upsilon_t^{-1}\right)_{pm} \to F_{pp}\int_0^1 W^{(p-1)/2}(t)W^{(m-1)/2}(t)'\,dtF'_{mm}.$$

(v) $p = 2, 4, 6, \ldots, M$; $m = 3, 5, 7, \ldots, M-1$:

$$\text{(A.11)}\qquad \left(\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1}\right)_{pm} = T^{-(p+m-2)/2}\Sigma\left(F_{pp}t^{(p-2)/2} + F_{pp-1}\xi_t^{(p-2)/2} + \cdots + F_{p1}(L)\eta_t\right)$$
$$\times\left(F_{mm}\xi_t^{(m-1)/2} + F_{mm-1}t^{(m-3)/2} + \cdots + F_{m1}(L)\eta_t\right)'$$
$$= T^{-(p+m-2)/2}\Sigma F_{pp}t^{(p-2)/2}\xi_t^{(m-1)/2\prime}F'_{mm} + \text{cross terms.}$$

The arguments in (ii)–(iv) imply that the cross terms in (A.11) converge to zero in probability. Applying Lemma 1(a) to the leading term in (A.11), $(\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1})_{pm} \Rightarrow F_{pp}\int_0^1 t^{(p-2)/2}W^{(m-1)/2}(t)'\,dtF'_{mm}$.

(vi) $p, m = 2, 4, 6, \ldots, M$, $m+p > 4$:

$$\text{(A.12)}\qquad \left(\Upsilon_T^{-1}Z'Z\Upsilon_T^{-1}\right)_{pm} = T^{-(p+m-2)/2}\Sigma\left(F_{pp}t^{(p-2)/2} + F_{pp-1}\xi_t^{(p-2)/2} + \cdots + F_{p1}(L)\eta_t\right)$$
$$\times\left(F_{mm}t^{(m-2)/2} + F_{mm-1}\xi_t^{(m-2)/2} + \cdots + F_{m1}(L)\eta_t\right)'$$
$$= T^{-(p+m-2)/2}\Sigma F_{pp}t^{(p-2)/2+(m-2)/2}F'_{mm} + \text{cross terms}$$
$$\xrightarrow{p} [(p+m-4)/2+1]^{-1}F_{pp}F'_{mm} = 2/(p+m-2)F_{pp}F'_{mm}$$

where the leading term in (A.12) converges nonstochastically using Lemma 1(c) and the remaining cross terms $\xrightarrow{p} 0$ by repeated application of Lemma 1, $\Sigma|F_{m1j}| < \infty$, and the arguments in (ii)–(iv) above. The expression for V_{22} obtains directly.

(b) Let $v^+ = [\Sigma^{1/2} \otimes I_{T-1}]v$ and let $M = 2g+1$, so that

$$H\left(I_n \otimes \Upsilon_T^{-1}\right)(I_n \otimes Z')v^+ = \begin{bmatrix} T^{-1/2}(I_n \otimes Z_1')v^+ \\ T^{-1/2}(I_n \otimes Z_2')v^+ \\ \vdots \\ T^{-(M-1)/2}(I_n \otimes Z_M')v^+ \end{bmatrix}$$

where $(I_n \otimes Z_m')v^+ = \mathrm{Vec}(\Sigma Z_t^m\eta'_{t+1}\Sigma^{1/2\prime})$. Thus consider:

(i) $\qquad T^{-(m-1)/2}\Sigma Z_t^m\eta'_{t+1}\Sigma^{1/2\prime} \qquad m = 3, 5, 7, \ldots, M,$

(ii) $\qquad T^{-(m-1)/2}\Sigma Z_t^m\eta'_{t+1}\Sigma^{1/2\prime} \qquad m = 2, 4, 6, \ldots, M-1,$

and $T^{-1/2}(I_n \otimes Z_1')v^+$.

(i) $m = 3, 5, 7, \ldots, M$:

$$\text{(A.13)}\qquad T^{-(m-1)/2}\Sigma Z_t^m\eta'_{t+1}\Sigma^{1/2\prime} = T^{-(m-1)/2}\Sigma\left(F_{mm}\xi_t^{(m-1)/2} + F_{mm-1}t^{(m-3)/2}\right.$$
$$\left. + \cdots + F_{m1}(L)\eta_t\right)\eta'_{t+1}\Sigma^{1/2\prime}$$
$$= F_{mm}T^{-(m-1)/2}\Sigma\xi_t^{(m-1)/2}\eta'_{t+1}\Sigma^{1/2\prime}$$
$$+ F_{mm-1}T^{-1/2}T^{-[(m-3)/2+1/2]}\Sigma t^{(m-3)/2}\eta'_{t+1}\Sigma^{1/2\prime}$$
$$+ \cdots + T^{-(m-3)/2}T^{-1}\Sigma\left(F_{m1}(L)\eta_t\right)\eta'_{t+1}\Sigma^{1/2\prime}.$$

The leading term in (A.13) converges to a nondegenerate random variable by Lemma 1(e), while the remaining terms vanish asymptotically by Lemma 1(d), (e), and (f), and by induction. Thus, for $m = 3, 5, 7, \ldots, M$,

$$T^{-(m-1)/2}\Sigma Z_t^m \eta_{t+1}' \Sigma^{1/2\prime} \Rightarrow F_{mm}\int_0^1 W^{(m-1)/2}(t)\, dW(t)'\Sigma^{1/2\prime}.$$

(ii) $m = 4, 6, \ldots M-1$:

$$\text{(A.14)}\qquad \begin{aligned} T^{-(m-1)/2}\Sigma Z_t^m \eta_{t+1}'\Sigma^{1/2\prime} &= T^{-(m-1)/2}\Sigma\Big(F_{mm}t^{(m-2)/2} + F_{mm-1}\xi_t^{(m-2)/2} \\ &\qquad + \cdots + F_{m1}(L)\eta_t\Big)\eta_{t+1}'\Sigma^{1/2\prime} \\ &= F_{mm}T^{-(m-1)/2}\Sigma t^{(m-2)/2}\eta_{t+1}'\Sigma^{1/2\prime} + \text{cross terms} \\ &\Rightarrow F_{mm}\int_0^1 t^{(m-2)/2}\, dW(t)'\Sigma^{1/2\prime} \end{aligned}$$

where the cross terms in (A.14) vanish using the result in (ii) above and the g-summability of $F_{m1}(L)$ for $m = 4, 6, \ldots, M-1$.

For $m = 2$, the expression in (A.14) is:

$$\text{(A.15)}\qquad T^{-1/2}\Sigma Z_t^2\eta_{t+1}'\Sigma^{1/2\prime} = T^{-1/2}\Sigma\big(F_{21}(L)\eta_t\big)\eta_{t+1}'\Sigma^{1/2\prime} + F_{22}T^{-1/2}\Sigma\eta_{t+1}'\Sigma^{1/2\prime}.$$

Suppose that both terms in (A.15) have well-defined limits, so that $\mathrm{Vec}[T^{-1/2}\Sigma Z_t^2\eta_{t+1}'\Sigma^{1/2\prime}] \Rightarrow \phi_2 = \phi_{21} + \phi_{22}$, where ϕ_{21} and ϕ_{22} correspond to the two terms in (A.15). Since the second term in (A.15) converges to $F_{22}W(1)'\Sigma^{1/2\prime}$, $\phi_{22} = \mathrm{Vec}[F_{22}W(1)'\Sigma^{1/2\prime}]$. Thus it remains only to examine ϕ_{21} and ϕ_1.

The first term in (A.15) has a limiting distribution that is jointly normal with the term for $m = 1$. Using the CLT for stationary processes with finite fourth moments,

$$\begin{bmatrix} \mathrm{vec}\big[T^{-1/2}\Sigma\big(F_{11}(L)\eta_t\big)\eta_{t+1}'\Sigma^{1/2\prime}\big] \\ \mathrm{vec}\big[T^{-1/2}\Sigma\big(F_{21}(L)\eta_t\big)\eta_{t+1}'\Sigma^{1/2\prime}\big] \end{bmatrix} \Rightarrow \begin{bmatrix} \phi_1 \\ \phi_{21} \end{bmatrix} \sim N(0, \Psi),$$

where

$$\begin{aligned} \Psi &= \begin{bmatrix} \Sigma\otimes\Sigma_j F_{11j}F_{11j}' & \Sigma\otimes\Sigma_j F_{11j}F_{21j}' \\ \Sigma\otimes\Sigma_j F_{21j}F_{11j}' & \Sigma\otimes\Sigma_j F_{21j}F_{21j}' \end{bmatrix} \\ &= \begin{bmatrix} \Sigma\otimes V_{11} & \Sigma\otimes V_{12} \\ \Sigma\otimes V_{21} & \Sigma\otimes(V_{22} - F_{22}F_{22}') \end{bmatrix}. \end{aligned}$$

Theorem 2.2 of Chan and Wei (1988) implies that (ϕ_1, ϕ_{21}) are independent of $(\phi_{22}, \phi_3, \ldots, \phi_{2g+1})$.

REFERENCES

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

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

CHAN, N. H., AND C. Z. WEI (1988): "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics*, 16, 367–401.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 366, 427–431.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Co-Integration and Error-Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

FULLER, W. A. (1976): *Introduction to Statistical Time Series*. New York: Wiley.

GRANGER, C. W. J. (1983): "Co-Integrated Variables and Error-Correcting Models," Discussion Paper #83-13, University of California—San Diego.

GRANGER, C. W. J., AND P. NEWBOLD (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120.

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Applications*. New York: Academic Press.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.
PHILLIPS, P. C. B. (1986): "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–340.
——— (1987): "Time Series Regression With a Unit Root," *Econometrica*, 55, 277–302.
——— (1988): "Optimal Inference in Cointegrated Systems," Cowles Foundation Discussion Paper No. 866, Yale University.
PHILLIPS, P. C. B., AND S. N. DURLAUF (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–496.
SIMS, C. A. (1978): "Least Squares Estimation of Autoregressions with Some Unit Roots," Center for Economic Research, University of Minnesota, Discussion Paper No. 78-95.
SOLO, V. (1984): "The Order of Differencing in ARIMA Models," *Journal of the American Statistical Association*, 79, 916–921.
STOCK, J. H. (1987): "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, 55, 1035–1056.
TSAY, R. S., AND G. C. TIAO (1990): "Asymptotic Properties of Multivariate Nonstationary Processes with Applications to Autoregressions," forthcoming, *Annals of Statistics*, 18, March.
WEST, K. D. (1988): "Asymptotic Normality, When Regressors Have a Unit Root," *Econometrica*, 56, 1397–1418.
WHITE, J. S. (1958): "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," *Annals of Mathematical Statistics*, 29, 1188–1197.