Testing for Common Trends

JAMES H. STOCK and MARK W. WATSON*

Cointegrated multiple time series share at least one common trend. Two tests are developed for the number of common stochastic trends (i.e., for the order of cointegration) in a multiple time series with and without drift. Both tests involve the roots of the ordinary least squares coefficient matrix obtained by regressing the series onto its first lag. Critical values for the tests are tabulated, and their power is examined in a Monte Carlo study. Economic time series are often modeled as having a unit root in their autoregressive representation, or (equivalently) as containing a stochastic trend. But both casual observation and economic theory suggest that many series might contain the *same* stochastic trends so that they are cointegrated. If each of n series is integrated of order 1 but can be jointly characterized by $k < n$ stochastic trends, then the vector representation of these series has k unit roots and $n - k$ distinct stationary linear combinations. Our proposed tests can be viewed alternatively as tests of the number of common trends, linearly independent cointegrating vectors, or autoregressive unit roots of the vector process. Both of the proposed tests are asymptotically similar. The first test (q_f) is developed under the assumption that certain components of the process have a finite-order vector autoregressive (VAR) representation, and the nuisance parameters are handled by estimating this VAR. The second test (q_c) entails computing the eigenvalues of a corrected sample first-order autocorrelation matrix, where the correction is essentially a sum of the autocovariance matrices. Previous researchers have found that U.S. postwar interest rates, taken individually, appear to be integrated of order 1. In addition, the theory of the term structure implies that yields on similar assets of different maturities will be cointegrated. Applying these tests to postwar U.S. data on the federal funds rate and the three- and twelve-month treasury bill rates provides support for this prediction: The three interest rates appear to be cointegrated.

KEY WORDS: Cointegration; Factor models; Integrated processes; Multiple time series; Unit roots; Yield curve.

1. INTRODUCTION

There is considerable empirical evidence that many macroeconomic time series are well described by univariate autoregressive integrated moving average (ARIMA) models, so differencing the data produces a series that appears to be covariance stationary. It has been less clear what transformation should be applied to data used in multivariate models, since (loosely speaking) the number of unit roots in a multiple time series may be less than the sum of the number of unit roots in the constituent univariate series. Equivalently, although each univariate series might contain a stochastic trend, in a vector process these stochastic trends might be common to several of the variables. Empirical evidence concerning the number of these common trends is of interest for several reasons. First, an economic or physical theory might predict that the variables contain common trends, and a test for these common trends would be a test of this implication of the theory. Second, one might wish to impose explicitly the number of common trends when making forecasts. Third, it might be desirable to specify a time series model in which all of the variables are stationary, but in which the data are not "overdifferenced." Such overdifferencing would occur were the model specified in terms of the first differences of the variables, because this would ignore the reduced dimensionality of the common trends.

We develop tests of the null hypothesis that an $n \times 1$ time series variable X_t has $k \leq n$ common stochastic trends, against the alternative that it has $m < k$ common trends. It is assumed that each component of X_t is integrated of order 1, but that there are $n - k$ linear combinations of X_t that are integrated of order 0. Engle and Granger (1987) defined such a process to be cointegrated of order (1, 1). If the stationary linear combinations are $\alpha' X_t$, then the columns of α are termed the cointegrating vectors of X_t. Engle and Granger showed that if X_t is cointegrated, then it has a representation in terms of an error-correction model, as developed by Sargan (1964), Davidson, Hendry, Srba, and Yeo (1978), and others.

The concept of cointegration formalizes an older notion that some linear combinations of time series variables appear nonstationary, whereas others appear to be almost white noise. Frisch (1934) referred to those linear combinations of time series data with very small variances as being generated by "true regressions"; one of his primary concerns was with the "multiple colinearity" that arose when there was more than one true regression (cointegrating vector) among the vector of variates. Box and Tiao (1977) associated the least predictable linear combinations (i.e., those with the weakest serial dependence) of X_t with "stable contemporaneous relationships;" they described the most predictable relationships as characterizing dynamic growth common to all of the series.

Cointegrated models can be represented formally in terms of a reduced number of common stochastic trends, plus transitory, or stationary, components. For univariate models, Beveridge and Nelson (1981) showed that any singly integrated ARIMA process has an exactly identified trend plus transitory representation, in which the trend is a random walk and the transitory component is covariance stationary. Fountis and Dickey (1986) extended this de-

* James H. Stock is Assistant Professor of Public Policy, John F. Kennedy School of Government, Harvard University, Cambridge, MA 02138. Mark W. Watson is Associate Professor, Department of Economics, Northwestern University, Evanston, IL 60208. This research was supported in part by National Science Foundation Grants SES-84-08797, SES-85-10289, and SES-86-18984. The authors are grateful to C. Cavanagh, R. F. Engle, J. Huizinga, J. Patel, C. Plosser, P. C. B. Phillips, R. Tsay, the referees, and an associate editor for helpful suggestions.

Journal of the American Statistical Association
December 1988, Vol. 83, No. 404, Theory and Methods

composition to vector autoregressive (VAR) models with $k = 1$. In Section 2, we provide a general representation for $k \leq n$. Because of the equivalence between these models, our proposed tests for k versus m common trends can be thought of as tests for the existence of $n - k$ versus $n - m$ linearly independent cointegrating vectors.

Several special cases of this testing problem have been considered elsewhere. The case that has received the most attention has been testing for 1 versus 0 unit roots in a univariate time series (e.g., see Dickey and Fuller 1979; Fuller 1976; Phillips 1987; Solo 1984). In a multivariate setting, a test of $k = 1$ versus $m = 0$ was developed by Fountis and Dickey (1986) for processes with a VAR representation with iid normal errors. Engle and Granger (1987) proposed and compared a variety of tests when $n = k = 2$ and the hypothesis of interest is $k = 2$ versus $m = 1$. Like the other tests in this literature, our test is based on the roots of the estimated autoregressive representation of the time series.

Section 2 presents the cointegrated and common-trends representations of X_t and summarizes our testing strategy. In Sections 3 and 4, two tests of k versus m common stochastic trends are proposed for the special case that $X_0 = 0$ and the process has no drift. These tests are extended in Section 5 to handle an estimated intercept and drift in the relevant regressions. The asymptotic critical values are tabulated in Section 6, and a small Monte Carlo experiment investigating the size and power of these tests is reported in Section 7. The tests are applied to data on postwar U.S. interest rates in Section 8, and our conclusions are summarized in Section 9.

2. THE MODEL AND TESTING STRATEGY

Let X_t denote an $n \times 1$ time series variable that is cointegrated of order (1, 1). That is, each element of X_t is integrated, but there are r linear combinations of X_t that are stationary. We work with an extension of Engle and Granger's (1987) definition of cointegration that allows for possible drift in X_t. The change in X_t is assumed to have the cointegrated vector moving average representation

$$\Delta X_t = \mu + C(L)\varepsilon_t, \qquad \sum_{j=1}^{\infty} j|C_j| < \infty, \tag{2.1}$$

where $C(z) = \sum_{i=0}^{\infty} C_i z^i$ with $C(0) = I_n$ (the $n \times n$ identity matrix), ε_t is iid with mean 0 and covariance matrix G, L is the lag operator, and $\Delta \equiv 1 - L$. $C(1)$ is assumed to have rank $k < n$, so X_t is cointegrated; that is, there is an $n \times r$ matrix α (where $r = n - k$) such that $\alpha' C(1) = 0$ and $\alpha'\mu = 0$. As Engle and Granger pointed out, this implies that the spectral density matrix of ΔX_t at frequency 0, $(2\pi)^{-1}C(1)GC(1)'$, is singular. The columns of α are the cointegrating vectors of X_t.

A representation for the stationary linear combinations $\alpha' X_t$ is readily obtained from (2.1). Let $v_t = G^{-1/2}\varepsilon_t$ and $\xi_t = \sum_{s=1}^{t} v_s$, adopt the conventional assumption (e.g., Dickey and Fuller 1979) that $\varepsilon_s = 0$ $(s \leq 0)$, and allow X_t to have a nonrandom initial value X_0. Then recursive substitution of (2.1) yields

$$X_t = X_0 + \mu t + C(1)G^{1/2}\xi_t + C^*(L)G^{1/2}v_t, \tag{2.2}$$

where $C^*(L) = (1 - L)^{-1}(C(L) - C(1))$ so that $C_j^* = -\sum_{i=j+1}^{\infty} C_i$. Because $\alpha' C(1) = 0$ and $\alpha'\mu = 0$, it follows that

$$Z_t \equiv \alpha' X_t = \alpha' X_0 + \alpha' C^*(L)G^{1/2}v_t. \tag{2.3}$$

With the additional assumption in (2.1) that $C(L)$ is 1-summable (Brillinger 1981), $C^*(L)$ is absolutely summable and Z_t has bounded variance.

The cointegrated process X_t has an alternative representation in terms of a reduced number of common random walks plus a stationary component. This "common trends" representation is readily derived from (2.2). Because $C(1)$ has rank $k < n$, there is an $n \times r$ matrix H_1 with rank r such that $C(1)H_1 = 0$. Furthermore, if H_2 is an $n \times k$ matrix with rank k and columns orthogonal to the columns of H_1, then $A \equiv C(1)H_2$ has rank k. The $n \times n$ matrix $H = (H_1\ H_2)$ is nonsingular and $C(1)H = (0\ A) = AS_k$, where S_k is the $k \times n$ selection matrix $[0_{k\times(n-k)}\ I_k]$, where $0_{k\times(n-k)}$ is a $k \times (n - k)$ matrix of zeros. In addition, because $\alpha' C(1) = 0$ and $\alpha'\mu = 0$, μ lies in the column space of $C(1)$ and can be written $\mu = C(1)\tilde{\mu}$, where $\tilde{\mu}$ is an $n \times 1$ vector. Thus (2.2) yields the common-trends representation for X_t:

$$\begin{aligned} X_t &= X_0 + C(1)[\tilde{\mu}t + G^{1/2}\xi_t] + C^*(L)G^{1/2}v_t \\ &= X_0 + C(1)H[H^{-1}\tilde{\mu}t + H^{-1}G^{1/2}\xi_t] + a_t \\ &= X_0 + A\tau_t + a_t, \qquad \tau_t = \pi + \tau_{t-1} + v_t, \end{aligned} \tag{2.4}$$

where $a_t = C^*(L)G^{1/2}v_t$, $\tau_t = S_k H^{-1}\tilde{\mu}t + S_k H^{-1}G^{1/2}\xi_t$, $\pi = S_k H^{-1}\tilde{\mu}$, and $v_t = S_k H^{-1}G^{1/2}v_t$. [For a different derivation of the common-trends representation (2.4) and further discussion, see King, Plosser, Stock, and Watson (1987).] The common-trends representation expresses X_t as a linear combination of k random walks with drift π, plus some transitory components, a_t, that are integrated of order 0.

The common-trends representation provides a convenient framework in which to motivate our proposed tests. Putting aside for the moment the complications that arise from a nonzero intercept and time trend in (2.2), a natural approach to testing k versus m common stochastic trends would be to examine the first-order serial correlation matrix of X_t. Because X_t is composed of both integrated and nonintegrated components, however, its estimated first-order serial correlation matrix has a nonstandard limiting distribution that generally depends on nuisance parameters in complicated ways. To mitigate this difficulty we examine functions of regression statistics of a linear transformation of X_t, denoted by Y_t, chosen so that under the null hypothesis the first $n - k$ elements of Y_t are not integrated, whereas the final k elements of Y_t can be expressed in terms of the k separate trends. More precisely, let $Y_t = DX_t$, where $D = [\alpha\ \alpha^\dagger]'$, where $\alpha^\dagger$ is an $n \times k$ matrix of constants chosen so that $\alpha^{\dagger\prime}\alpha = 0$ and $\alpha^{\dagger\prime}\alpha^\dagger = I_k$. The first $n - k$ elements of Y_t are Z_t in (2.3). Let W_t denote the final k integrated elements of Y_t. It follows from (2.1) that

$$\Delta W_t = \alpha^{\dagger\prime}\mu + u_t, \tag{2.5}$$

where $u_t = \tilde{C}(L)v_t$, with $\tilde{C}(L) = \alpha^{\dagger\prime}C(L)G^{1/2}$. Combining (2.3) and (2.5),

$$\Delta_k Y_t = \delta + F(L)v_t, \tag{2.6}$$

where

$$\Delta_k = \begin{bmatrix} I_{n-k} & 0 \\ 0 & \Delta I_k \end{bmatrix}, \qquad \delta = \begin{bmatrix} \alpha' X_0 \\ \alpha^{\dagger\prime}\mu \end{bmatrix},$$

$$F(L) = \begin{bmatrix} \alpha' C^*(L)G^{1/2} \\ \tilde{C}(L) \end{bmatrix}.$$

Recursive substitution of (2.6) shows that Y_t can be represented as

$$\begin{aligned} Y_t &= \begin{bmatrix} \alpha' X_0 \\ \alpha^{\dagger\prime} X_0 \end{bmatrix} + \begin{bmatrix} 0_{(n-k)\times 1} \\ \alpha^{\dagger\prime}\mu \end{bmatrix} t \\ &\quad + \begin{bmatrix} 0_{(n-k)\times n} \\ \tilde{C}(1) \end{bmatrix} \xi_t + \begin{bmatrix} \alpha' C^*(L)G^{1/2} \\ \tilde{C}^*(L) \end{bmatrix} v_t \\ &= \beta_1 + \beta_2 t + \beta_3 \xi_t + \beta_4(L)v_t, \end{aligned} \tag{2.7}$$

where $\tilde{C}^*(L) = (1 - L)^{-1}(\tilde{C}(L) - \tilde{C}(1))$.

In terms of W_t, a test of k versus m common trends becomes a test of whether $\tilde{C}(1)$ has rank k against the alternative, that it has rank m. To motivate the proposed tests, suppose that $X_0 = \mu = 0$, and consider the result of regressing W_t onto W_{t-1}. Under the null hypothesis, W_t is a linear combination of k integrated processes, so Φ, the probability limit of

$$\tilde{\Phi} = [\textstyle\sum W_t W_{t-1}'][\sum W_{t-1}W_{t-1}']^{-1}, \tag{2.8}$$

has k real unit roots. Under the alternative W_t includes m integrated variables and $k - m$ nonintegrated variables, or equivalently W_t has $k - m$ linearly independent cointegrating vectors. Thus under the alternative Φ has only m unit eigenvalues corresponding to the m integrated variables, and $k - m$ eigenvalues with modulus (and therefore with real parts) less than 1. Letting λ_{m+1} denote the eigenvalue of Φ with the $(m + 1)$th-largest real part, our null and alternative hypotheses are H_0: real$(\lambda_{m+1}) = 1$ versus H_1: real$(\lambda_{m+1}) < 1$.

Much is known about the properties of $\tilde{\Phi}$ when Φ has some unit roots. When $n = 1$ and u_t is serially uncorrelated, $\tilde{\Phi}$ has the distribution studied by White (1958), Fuller (1976), Dickey and Fuller (1979), and others. Phillips (1987) examined the distribution of $T(\tilde{\Phi} - 1)$ under less restrictive conditions on the errors; this analysis was generalized to the multivariate case by Phillips and Durlauf (1986). Unfortunately, when u_t is serially correlated the distribution of $\tilde{\Phi}$ and its eigenvalues $\tilde{\lambda}$ depends on the autocovariances of u_t. This dependence makes it impossible to tabulate the asymptotic critical values of a statistic based on $\tilde{\lambda}$ in a practical way. Strategies for circumventing this problem and developing asymptotically similar tests are presented in Sections 3 and 4 for the case ($\beta_1 = 0$, $\beta_2 = 0$) and are extended to the cases ($\beta_1 \neq 0$, $\beta_2 = 0$) and ($\beta_1 \neq 0$, $\beta_2 \neq 0$) in Section 5.

3. A TEST BASED ON FILTERING THE DATA

This section presents a test statistic in which the nuisance parameters of the process are eliminated by assuming a parametric representation for the process generating W_t. The development of this test parallels Dickey and Fuller's (1979) approach to testing for a unit root in a univariate time series. Specifically, suppose that ΔW_t has a finite-order VAR representation so that (2.5) can be rewritten as

$$\Pi(L)\Delta W_t = \gamma + \eta_t, \tag{3.1}$$

where $\Pi(L)$ is a matrix lag polynomial of known order p with all roots outside the unit circle, η_t is iid with mean 0, and $\Pi(0)$ is normalized so that $E\eta_t\eta_t' = I_k$. In this section it is assumed that $W_0 = \gamma = 0$.

First, suppose that D and $\Pi(L)$ are known and let $\zeta_t = \Pi(L)W_t$. Under (3.1), $\Pi(L)\Delta W_t = \Delta[\Pi(L)W_t] = \eta_t$, so under the null hypothesis the elements of $\Pi(L)W_t$ are random walks. In contrast, under the alternative of $m < k$ common trends, only m components of $\Pi(L)W_t$ are random walks, whereas the remaining elements are integrated of order 0. This suggests testing for k versus m common trends by examining the roots of the first sample autocorrelation matrix formed using ζ_t,

$$\tilde{\Phi}_f = [\textstyle\sum \zeta_t\zeta_{t-1}'][\sum \zeta_{t-1}\zeta_{t-1}']^{-1}.$$

Rewriting $\tilde{\Phi}_f$, we have

$$T[\tilde{\Phi}_f - I_k] = \Psi_{kT}'(\Gamma_{kT})^{-1}, \tag{3.2}$$

where $\Psi_{kT} = T^{-1}\sum \zeta_{t-1}\eta_t'$ and $\Gamma_{kT} = T^{-2}\sum \zeta_{t-1}\zeta_{t-1}'$.

The limiting behavior of Ψ_{kT} and Γ_{kT} has been treated in the univariate case by (for example) White (1958), Solo (1984), and Phillips (1987), and in the multivariate case by Phillips and Durlauf (1986) and Chan and Wei (1988). These random matrices converge weakly to functionals of the k-dimensional Wiener process $B_k(t)$: $\Gamma_{kT} \Rightarrow \Gamma_k \equiv \int_0^1 B_k(t)B_k(t)'\,dt$ and $\Psi_{kT} \Rightarrow \Psi_k \equiv \int_0^1 B_k(t)\,dB_k(t)'$, where $\Rightarrow$ denotes weak convergence on the space of continuous functions on $[0, 1]^k$ in the sense of Billingsley (1968). Thus from (3.2), $T[\tilde{\Phi}_f - I_k] \Rightarrow \Psi_k'\Gamma_k^{-1}$; that is, $T[\tilde{\Phi}_f - I_k]$ converges weakly to a random variable that has the same distribution as $\Psi_k'\Gamma_k^{-1}$. It follows that $T(\tilde{\lambda}_f - \iota) \Rightarrow \lambda_*$, where λ_* denotes the vector of ordered eigenvalues of $\Psi_k'\Gamma_k^{-1}$, $\tilde{\lambda}_f$ denotes the vector of ordered eigenvalues of $\tilde{\Phi}_f$, and $\iota = (1\ 1 \cdots 1)'$.

If D and $\Pi(L)$ were known a test statistic could be constructed using $\tilde{\lambda}_f$. In applications, however, D and $\Pi(L)$ are typically unknown. This deficiency can be remedied by using estimators $\hat{D}$ and $\hat{\Pi}(L)$ of D and $\Pi(L)$, respectively. For the moment, assume that $\hat{D}$ and $\hat{\Pi}(L)$ exist and that (a) $\hat{D} \xrightarrow{p} RD$ under both H_0 and H_1, where $R = \operatorname{diag}(R_1, R_2)$, where R_1 and R_2 are, respectively, nonsingular $(n - k) \times (n - k)$ and $k \times k$ matrices under the null and $(n - m) \times (n - m)$ and $m \times m$ matrices under the alternative, and (b) $\hat{\Pi}(L) \xrightarrow{p} R_2\Pi(L)R_2^{-1}$ under H_0. Let $\hat{W}_t = S_k\hat{D}X_t$ and $\hat{\zeta}_t = \hat{\Pi}(L)\hat{W}_t$. Then one could consider the ordinary least squares (OLS) estimator

$$\hat{\Phi}_f = [\textstyle\sum \hat{\zeta}_t\hat{\zeta}_{t-1}'][\sum \hat{\zeta}_{t-1}\hat{\zeta}_{t-1}']^{-1}.$$

This modified version of $\tilde{\Phi}$, computed using the filtered series $\hat{\zeta}_t$, has a limiting representation in which the nuisance parameters in (3.1) do not appear. Letting $\hat{\lambda}_f$ denote the vector of ordered eigenvalues of $\hat{\Phi}_f$, we have Theorem 3.1.

Theorem 3.1. Suppose that $\hat{D} \xrightarrow{p} RD$, W_t is generated by (3.1) with $W_0 = \gamma = 0$, $\hat{\Pi}(L) \xrightarrow{p} R_2\Pi(L)R_2^{-1}$, and $\max_i E(\eta_{it}^4) \le \mu_4 < \infty$. Then, (a) $T(\hat{\Phi}_f - I_k) \Rightarrow R_2\Psi_k'\Gamma_k^{-1}R_2^{-1}$, (b) $T(\hat{\lambda}_f - \iota) \Rightarrow \lambda_*$, and (c) $T(|\hat{\lambda}_f| - \iota) \Rightarrow \text{real}(\lambda_*)$.

Proof. At the suggestion of the editor the proofs of all lemmas and theorems are omitted but provided in Stock and Watson (1988).

Theorem 3.1 suggests testing for k versus m common trends—or equivalently for k versus m real unit roots in Φ—using the statistic

$$q_f(k, m) = T[\text{real}(\hat{\lambda}_{f,m+1}) - 1],$$

where $\hat{\lambda}_{f,m+1}$ is the $(m + 1)$th element of $\hat{\lambda}_f$. Under the null hypothesis, from Theorem 3.1(b) $q_f(k, m)$ asymptotically has the same distribution as $\text{real}(\lambda_{*m+1})$.

The construction of q_f requires the estimation of RD and the autoregressive matrix polynomial $\Pi(L)$ in (3.1). The $n \times n$ matrix RD can be estimated in a variety of ways. The first $n - k$ rows of D (and thus of RD) are a basis for the space spanned by the cointegrating vectors of X_t under the null. Because the cointegrating vectors form linear combinations of X_t that have bounded variance from the otherwise integrated elements of X_t, they (like the autoregressive coefficient in the univariate unit-root problem or its multivariate analog, discussed in the preceding sections) can be estimated consistently without specifying a particular parametric process for the additional stationary components. As demonstrated in Stock (1987, theorem 2), if X_t has the representation (2.1) with $n - k$ cointegrating vectors and $\max_i E(\varepsilon_{it}^4) \le \mu_4 < \infty$, then the cointegrating vectors consisting of the columns of α can be estimated by contemporaneous OLS regressions of one element of X_t on the others, after an arbitrary normalization to ensure that the estimates are linearly independent.

We adopt a modification of this approach, in which the cointegrating vectors are constructed to be orthonormal with the first cointegrating vector forming the linear combination of X_t having the smallest variance, the second cointegrating vector having the next smallest variance, and so on. Implementing this procedure simply entails estimating the principal components of X_t; α is estimated by those linear combinations corresponding to the smallest $n - k$ principal components, and $\alpha^\dagger$ is estimated by the linear combinations corresponding to the largest k principal components. Since $\hat{\alpha}$ consistently estimates the cointegrating vectors up to an arbitrary linear transformation, $\hat{D} \xrightarrow{p} RD = [\alpha R_1' \ \alpha^\dagger R_2']'$ for some R_1 and R_2.

Since $\hat{\Phi} \xrightarrow{p} I_k$ under the null [where $\hat{\Phi} = \sum \hat{W}_t\hat{W}_{t-1}'(\sum \hat{W}_{t-1}\hat{W}_{t-1}')^{-1}$], the parameters of $R_2\Pi(L)R_2^{-1}$ can be estimated consistently by a VAR(p) regression using either $\Delta\hat{W}_t$ or $\hat{u}_t$, where $\hat{u}_t$ are the residuals from a regression of $\hat{W}_t$ onto $\hat{W}_{t-1}$. In either case, normalizing the VAR coefficient matrices so that the VAR residuals have an identity-contemporaneous covariance matrix ensures that $\hat{\Pi}(L) \xrightarrow{p} R_2\Pi(L)R_2^{-1}$.

This test is consistent against the alternative that there are m rather than k common trends using either estimator of $\Pi(L)$, even if the process is not autoregressive of order p but satisfies (2.1) with $n - m$ cointegrating vectors. Under the alternative, $\hat{D}$ (constructed using principal components) converges in probability to some matrix D_a, the first $n - m$ rows of which contain the cointegrating vectors of X_t and the final m rows of which are orthogonal to the cointegrating vectors. In addition, under the alternative $\hat{\Pi}(L)$ converges to some (finite-order) matrix lag polynomial $\Pi_a(L)$ even if ΔW_t does not have a VAR(p) representation. From (2.1) and the definition of $\hat{\zeta}_t$, $\Delta\hat{\zeta}_t = \hat{\Pi}(L)S_k\hat{D}C(L)\varepsilon_t$, where $\hat{\Pi}(L) \xrightarrow{p} \Pi_a(L)$ and $\hat{D} \xrightarrow{p} D_a$. Since $\Pi_a(L)$ has finite order and $C(L)$ is absolutely summable under both the null and the alternative, $\Pi_a(L)S_kD_aC(L)$ is absolutely summable. Furthermore, under the alternative, $\text{rank}(\Pi_a(1)S_kD_aC(1)) \equiv m' \le \text{rank}(C(1)) = m < k$. Using a construction like (2.6), it can be shown that as the sample size tends to infinity, ζ_t [and, by the convergence of $\hat{\Pi}(L)$ and $\hat{D}$, $\hat{\zeta}_t$] has m' unit roots in its sample first-order autoregressive matrix and $k - m'$ roots less than 1 in modulus and therefore with real parts that are less than 1. In particular, $\text{real}(\hat{\lambda}_{f,m+1}) - 1$ converges in probability to a negative number, so the test is consistent. Note that a consistent test obtains whether the filter is estimated using either $\Delta\hat{W}_t$ or $\hat{u}_t$, assuming that the order of the filter is fixed.

4. A TEST BASED ON CORRECTING THE OLS AUTOREGRESSIVE MATRIX

Our second proposed statistic tests for k versus $k - 1$ common trends using a corrected version of $\tilde{\Phi}$, the sample first-order autocorrelation matrix for W_t in (2.8), under the assumption that $\beta_1 = \beta_2 = 0$ in (2.7). In this case, $\tilde{\Phi}$ has the asymptotic representation given in Lemma 4.1.

Lemma 4.1. If $\max_i E(v_{it}^4) \le \mu_4 < \infty$ and $\beta_1 = \beta_2 = 0$ in (2.7), then

$$T(\tilde{\Phi} - I_k) - [\tilde{C}(1)\Psi_{nT}'\tilde{C}(1)' + M'][\tilde{C}(1)\Gamma_{nT}\tilde{C}(1)']^{-1} \xrightarrow{p} 0,$$

where $\Psi_{nT} = T^{-1}\sum \xi_{t-1}v_t'$, $\Gamma_{nT} = T^{-2}\sum \xi_t\xi_t'$, and $M = [\sum_{j=0}^{\infty} (\tilde{C}_j^* - \tilde{C}_j)\tilde{C}_j' + \tilde{C}(1)\tilde{C}(1)'] = \sum_{j=1}^{\infty} Eu_{t-j}u_t'$.

This lemma indicates that $T(\tilde{\Phi} - I_k)$ asymptotically consists of two parts. The first, $[\tilde{C}(1)\Psi_{nT}\tilde{C}(1)']'[\tilde{C}(1)\Gamma_{nT}\tilde{C}(1)']^{-1}$, is T times the error in the estimate of Φ obtained by regressing the random walk $\tilde{C}(1)\xi_t$ onto its lagged value. The second, $M'[\tilde{C}(1)\Gamma_{nT}\tilde{C}(1)']^{-1}$, is analogous to the $O(T^{-1})$ bias in contemporaneous regressions of cointegrated variables. This bias arises from the correlation between the regressor W_{t-1} and u_t in (2.5). This term is related to the bias in OLS regression estimates when there are stationary lagged dependent variables and serially correlated errors. In the present context, since u_t is not integrated (but is serially correlated) and W_t is integrated, this correlation produces not inconsistency but a component of $\tilde{\Phi}$ that is $O_p(T^{-1})$.

The bias term M is problematic, since its presence means that the distribution of $\tilde{\Phi}$ (and its eigenvalues) depends on M and thus on $\tilde{C}(L)$. Nevertheless, the limiting representation in Lemma 4.1 suggests a solution to this problem: Modify the OLS estimator $\tilde{\Phi}$ using an estimator of M so that the asymptotic distribution of the eigenvalues of the modified OLS estimator depends only on Γ_n and Ψ_n. This approach generalizes to the multivariate-setting Phillips (1987, theorem 5.1) test for a single unit root in a univariate process. Specifically, were W_t observed and M known, a corrected estimator $\tilde{\Phi}_c$ could be computed by subtracting off the troublesome term:

$$\tilde{\Phi}_c = [T^{-2} \sum W_t W'_{t-1} - T^{-1}M'][T^{-2} \sum W_{t-1}W'_{t-1}]^{-1}.$$

Letting $\tilde{\lambda}_c$ denote the vector of the k ordered eigenvalues of $\tilde{\Phi}_c$, we have Lemma 4.2.

Lemma 4.2. Let Ω by a $k \times k$ matrix such that $\Omega\Omega' = \tilde{C}(1)\tilde{C}(1)'$. Then, under the conditions of Lemma 4.1, (a) $T(\tilde{\Phi}_c - I_k) \Rightarrow \Omega\Psi'_k\Gamma_k^{-1}\Omega^{-1}$ and (b) $T(\tilde{\lambda}_c - \iota) \Rightarrow \lambda_*$.

According to Lemma 4.2, the distribution of the standardized eigenvalues of $\tilde{\Phi}_c$ do not depend on any nuisance parameters and thus can be tabulated. But $\tilde{\Phi}_c$ cannot itself form the basis for a test because it involves W_t, which is not directly observed, and M, which depends on unknown parameters. As we discuss later, however, M can be estimated; suppose that the estimator of M, $\hat{M}$, is such that $\hat{M} \xrightarrow{p} R_2MR'_2$. Let $\hat{Y}_t = \hat{D}X_t$, and use $\hat{W}_t = S_k\hat{Y}_t$ and $\hat{M}$ to form the analog of $\tilde{\Phi}_c$,

$$\hat{\Phi}_c = [T^{-2} \sum \hat{W}_t\hat{W}'_{t-1} - T^{-1}\hat{M}'][T^{-2} \sum \hat{W}_{t-1}\hat{W}'_{t-1}]^{-1}.$$

The consistency of $\hat{D}$ and $\hat{M}$ ensure that the eigenvalues of $\hat{\Phi}_c$, $\hat{\lambda}_c$, are asymptotically equivalent to the eigenvalues of $\tilde{\Phi}_c$.

Theorem 4.1. Suppose that $\hat{D} \xrightarrow{p} RD$ and $\hat{M} \xrightarrow{p} R_2MR'_2$. Then, under the assumptions of Lemma 4.1, (a) $T(\hat{\Phi}_c - I_k) \Rightarrow R_2\Omega\Psi'_k\Gamma_k^{-1}\Omega^{-1}R_2^{-1}$ and (b) $T(\hat{\lambda}_c - \iota) \Rightarrow \lambda_*$.

Part (a) of this theorem presents a limiting representation for the ordered eigenvalues of $\hat{\Phi}_c$. We therefore define the test statistic

$$q_c(k, k - 1) = T[\text{real}(\hat{\lambda}_{c,k}) - 1],$$

where $\hat{\lambda}_{c,k}$ is the kth element of $\hat{\lambda}_c$. Under the null hypothesis, $q_c(k, k - 1)$ converges to the real part of the smallest eigenvalue of the random matrix $\Psi'_k\Gamma_k^{-1}$.

The construction of the q_c statistic requires estimators $\hat{D}$ and $\hat{M}$. Construction of $\hat{D}$ was discussed in Section 3. The second expression for M in Lemma 4.1 suggests an estimator of M based on the sample covariances of the $k \times 1$ vector of residuals $\hat{u}_t = \hat{W}_t - \hat{\Phi}\hat{W}_{t-1}$ from the regression of $\hat{W}_t$ onto $\hat{W}_{t-1}$. The estimation of M is clearly related to the problem of estimating the spectral density matrix of u_t at frequency 0, $(2\pi)^{-1} \sum_{j=-\infty}^{\infty} V_j = (2\pi)^{-1}(V_0 + M + M')$ (where $V_j = Eu_tu'_{t-j}$), so techniques developed for its estimation can be applied here. Let $\hat{V}_j = T^{-1} \sum_{t=j+1}^{T} \hat{u}_t\hat{u}'_{t-j}$. Then M can be estimated by

$$\hat{M} = \sum_{j=1}^{J} K(j)\hat{V}'_j, \tag{4.1}$$

where $K(j)$ is a (time domain) kernel. For a proof of the consistency of $\hat{M}$ in the univariate case for $K(j) = 1$ and $J = o(T^{1/4})$, see Phillips (1987).

The test based on $q_c(k, k - 1)$ is consistent if $\hat{M} \xrightarrow{p} \sum_{j=1}^{\infty} Eu_{t-j}u'_t$ under the alternative, where $u_t = W_t - \Phi W_{t-1}$. To demonstrate this consistency, write $\Phi = Q\Lambda Q^{-1}$ under the alternative, where Λ is a diagonal matrix with the roots of Φ on the diagonal so that the first $k - 1$ diagonal elements are 1 and the final element is less than 1 in modulus, and where Q is the $k \times k$ matrix of eigenvectors of Φ. Under the alternative, the last element of the transformed variate QW_t [say $(QW_t)_k$] is a stationary process. Let ρ_j and $f_{(QW_t)_k}(\omega)$ denote the jth autocorrelation and the (scalar) spectral density of $(QW_t)_k$, respectively. A calculation using the techniques in Stock and Watson (1988) shows that under the fixed alternative, $\hat{\lambda}_{ck} \xrightarrow{p} 1 - (1 - \rho_1)^2(1 + \sum_1^{\infty} \rho_j) \equiv \lambda_{ck}$. Because $f_{(QW_t)_k}(0) = c(1 + 2 \sum_1^{\infty} \rho_j) \geq 0$ under the alternative (where c is a positive constant), $\sum_1^{\infty} \rho_j \geq -\frac{1}{2}$, so $\lambda_{ck} \leq 1 - \frac{1}{2}(1 - \rho_1)^2 < 1$ for $|\rho_1| < 1$. Thus $T(\hat{\lambda}_{ck} - 1)$ tends to $-\infty$ under the fixed alternative, demonstrating that the test is consistent.

Not all candidate estimators of the correction term M result in a consistent test. In particular, suppose that $\Delta\hat{W}_t$ rather than $\hat{u}_t$ is used to construct an estimator $\tilde{M}$ so that $\tilde{M} \xrightarrow{p} \sum_{j=1}^{\infty} E[\Delta W_{t-j}\Delta W'_t]$ under both the null and the fixed alternative. Under the null, the tests formed using $\tilde{M}$ and $\hat{M}$ are asymptotically equivalent. Under the alternative, however, if corrected using $\tilde{M}$, $\hat{\Phi}_c \xrightarrow{p} I_k$, so in particular $\text{real}(\hat{\lambda}_{c,k}) \xrightarrow{p} 1$ and a one-sided test based on this root is not consistent.

5. MODIFICATIONS FOR ESTIMATED INTERCEPTS AND DRIFTS

In practice it is desirable to allow for nonzero X_0, and in many applications a more appropriate model might be one in which X_t has a nonzero drift as well as a cointegrated stochastic structure. This section addresses the problem of testing the null hypothesis that the rank of $C(1)$ is k, against the alternative that it is $m < k$ when the intercept and drift may be nonzero. In terms of (2.7), this entails testing that the rank of $\tilde{C}(1)$ is k versus m when either (a) $\beta_2 = 0$ but β_1 might be nonzero (but is nonrandom) or (b) β_1 and β_2 are nonrandom but might both be nonzero. In case (b) under the null X_t has k linear combinations that are random walks with nonzero drifts, whereas under the alternative X_t has m such linear combinations. In a univariate setting macroeconomic data are often modeled as stationary in first differences around a constant nonzero mean; as Beveridge and Nelson (1981) showed, this implies that the process can be written as the sum of a random walk with nonzero drift and a mean-0 stationary component. Letting β_1 and β_2 be nonzero both generalizes this univariate specification to the multivariate case and permits testing for common trends against an alternative, in which up to $n - m$ components are stationary around a linear time trend. For a discussion of the macroeconomic implications of stochastic versus deterministic trends in economic time series, see Nelson and Plosser (1982); for an alternative approach in which the drift in the stochastic trend is itself modeled as a random walk (so that the series is stationary only after taking second differences), see Harvey (1985).

We follow Fuller's (1976) and Dickey and Fuller's (1979) univariate treatment of intercepts and time trends and modify the previous test statistics so that an intercept or an intercept and a drift are estimated. Accordingly, let $\hat{Y}_t^\mu = \hat{Y}_t - T^{-1} \sum \hat{Y}_t$ and $\hat{Y}_t^\tau = \hat{Y}_t - \hat{\beta}_1 - \hat{\beta}_2 t$, where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the OLS estimates of β_1 and β_2 obtained by regressing $\hat{Y}_t$ on a constant and t, and let $\hat{W}_t^\mu = S_k \hat{Y}_t^\mu$ and $\hat{W}_t^\tau = S_k \hat{Y}_t^\tau$. The modification to the filtering test entails estimating the autoregressive polynomial $\Pi(L)$ using $\hat{W}_t^\mu$ or $\hat{W}_t^\tau$ rather than $\hat{W}_t$ (as in Sec. 3). Let $\hat{\zeta}_t^\mu = \hat{\Pi}(L)\hat{W}_t^\mu$ and $\hat{\zeta}_t^\tau = \hat{\Pi}(L)\hat{W}_t^\tau$, and define

$$\hat{\Phi}_f^\mu = [\sum \hat{\zeta}_t^\mu \hat{\zeta}_{t-1}^{\mu\prime}][\sum \hat{\zeta}_{t-1}^\mu \hat{\zeta}_{t-1}^{\mu\prime}]^{-1}$$

and

$$\hat{\Phi}_f^\tau = [\sum \hat{\zeta}_t^\tau \hat{\zeta}_{t-1}^{\tau\prime}][\sum \hat{\zeta}_{t-1}^\tau \hat{\zeta}_{t-1}^{\tau\prime}]^{-1}.$$

Let $\Psi_k^\mu = \int_0^1 B_k^\mu(t)\, dB_k^\mu(t)'$, $\Gamma_k^\mu = \int_0^1 B_k^\mu(t) B_k^\mu(t)'\, dt$, $\Psi_k^\tau = \int_0^1 B_k^\tau(t)\, dB_k^\tau(t)'$, and $\Gamma_k^\tau = \int_0^1 B_k^\tau(t) B_k^\tau(t)'\, dt$, where $B_k^\mu(t) = B_k(t) - \int_0^1 B_k(s)\, ds$ and $B_k^\tau(t) = B_k(t) - \int_0^1 a_1(s) B_k(s)\, ds - t \int_0^1 a_2(s) B_k(s)\, ds$, where $a_1(s) = 4 - 6s$ and $a_2(s) = -6 + 12s$. Also, let $\hat{\lambda}_f^\mu$, $\hat{\lambda}_f^\tau$, λ_*^μ, and λ_*^τ, respectively, denote the ordered eigenvalues of $\hat{\Phi}_f^\mu$, $\hat{\Phi}_f^\tau$, $\Psi_k^{\mu\prime}(\Gamma_k^\mu)^{-1}$, and $\Psi_k^{\tau\prime}(\Gamma_k^\tau)^{-1}$. We now have Theorem 5.1.

Theorem 5.1. Suppose that $\hat{D} \xrightarrow{p} RD$, W_t is generated by (3.1), $\hat{\Pi}(L) \xrightarrow{p} R_2 \Pi(L) R_2^{-1}$, and $\max_i E(\eta_{it}^4) \le \mu_4 < \infty$. (a) If $\gamma = 0$ and W_0 is an arbitrary constant, then (i) $T(\hat{\Phi}_f^\mu - I_k) \Rightarrow R_2 \Psi_k^{\mu\prime}(\Gamma_k^\mu)^{-1} R_2^{-1}$, (ii) $T(\hat{\lambda}_f^\mu - \iota) \Rightarrow \lambda_*^\mu$, and (iii) $T(|\hat{\lambda}_f^\mu| - \iota) \Rightarrow \text{real}(\lambda_*^\mu)$. (b) If γ and W_0 are arbitrary constants, then (i) $T(\hat{\Phi}_f^\tau - I_k) \Rightarrow R_2 \Psi_k^{\tau\prime}(\Gamma_k^\tau)^{-1} R_2^{-1}$, (ii) $T(\hat{\lambda}_f^\tau - \iota) \Rightarrow \lambda_*^\tau$, and (iii) $T(|\hat{\lambda}_f^\tau| - \iota) \Rightarrow \text{real}(\lambda_*^\tau)$.

The counterparts of $q_f(k, m)$ when there might be a nonzero intercept or a nonzero intercept and drift are

$$q_f^\mu(k, m) = T[\text{real}(\hat{\lambda}_{f,m+1}^\mu) - 1]$$

and

$$q_f^\tau(k, m) = T[\text{real}(\hat{\lambda}_{f,m+1}^\tau) - 1],$$

which (respectively) have the same limiting distribution under the null as $\text{real}(\lambda_{*m+1}^\mu)$ and $\text{real}(\lambda_{*m+1}^\tau)$, where $\hat{\lambda}_{f,m+1}^\mu$ is the $(m + 1)$th-largest eigenvalue of $\hat{\Phi}_f^\mu$, and so on.

The modification to the q_c statistic for a nonzero intercept or drift proceeds similarly. Suppose that D were known, let $Y_t^\mu = Y_t - T^{-1} \sum Y_t$ and $Y_t^\tau = Y_t - \tilde{\beta}_1 - \tilde{\beta}_2 t$, where $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are the coefficients from regressing Y_t onto $(1, t)$. Let $W_t^\mu = S_k Y_t^\mu$ and $W_t^\tau = S_k Y_t^\tau$. By analogy to $\tilde{\Phi}$, define

$$\tilde{\Phi}^\mu = [T^{-2} \sum W_t^\mu W_{t-1}^{\mu\prime}][T^{-2} \sum W_{t-1}^\mu W_{t-1}^{\mu\prime}]^{-1}$$

and

$$\tilde{\Phi}^\tau = [T^{-2} \sum W_t^\tau W_{t-1}^{\tau\prime}][T^{-2} \sum W_{t-1}^\tau W_{t-1}^{\tau\prime}]^{-1}.$$

The treatment of $\tilde{\Phi}^\mu$ and $\tilde{\Phi}^\tau$ parallels the treatment of $\tilde{\Phi}$ in Section 4. It is first shown that the asymptotic distributions of $\tilde{\Phi}$, $\tilde{\Phi}^\mu$, and $\tilde{\Phi}^\tau$ depend on the same nuisance parameters, although the random components in the asymptotic representations differ. This makes it possible to construct corrected matrices $\tilde{\Phi}_c^\mu$ and $\tilde{\Phi}_c^\tau$, the eigenvalues of which have a distribution that is independent of the nuisance parameters.

Lemma 5.1. Suppose that $\max_i E(v_{it}^4) \le \mu_4 < \infty$. (a) If $\beta_2 = 0$ in (2.7), and β_1 is an arbitrary constant, then

$$T(\tilde{\Phi}^\mu - I) - [\tilde{C}(1)\Psi_{nT}^{\mu\prime}\tilde{C}(1)' + M'][\tilde{C}(1)\Gamma_{nT}^\mu \tilde{C}(1)']^{-1} \xrightarrow{p} 0.$$

(b) If β_1 and β_2 in (2.7) are arbitrary constants, then

$$T(\tilde{\Phi}^\tau - I) - [\tilde{C}(1)\Psi_{nT}^{\tau\prime}\tilde{C}(1)' + M'][\tilde{C}(1)\Gamma_{nT}^\tau \tilde{C}(1)']^{-1} \xrightarrow{p} 0,$$

where $\Psi_{nT}^\mu = T^{-1} \sum \xi_{t-1}^\mu \Delta \xi_t^{\mu\prime}$, $\Gamma_{nT}^\mu = T^{-2} \sum \xi_t^\mu \xi_t^{\mu\prime}$, $\Psi_{nT}^\tau = T^{-1} \sum \xi_{t-1}^\tau \Delta \xi_t^{\tau\prime}$, and $\Gamma_{nT}^\tau = T^{-2} \sum \xi_t^\tau \xi_t^{\tau\prime}$, where $\xi_t^\mu = \xi_t - T^{1/2}\Theta_{0T}$ and $\xi_t^\tau = \xi_t - T^{1/2}\Theta_{1T} - T^{-1/2}\Theta_{2T} t$, where $\Theta_{iT} = T^{-3/2} \sum_{t=1}^T a_{it} \xi_t$ $(i = 0, 1, 2)$, with $a_{0t} = 1$, $a_{1t} = 4 - 6(t/T)$, and $a_{2t} = -6 + 12(t/T)$.

These limiting representations depend on M, given in Lemma 4.1. This dependence can be eliminated by correcting $\tilde{\Phi}^\mu$ and $\tilde{\Phi}^\tau$ using an estimator of M, $\hat{M}$, as suggested in Section 4. In addition, since D and therefore W_t are unknown, replace W_t with $\hat{W}_t = S_k \hat{D} X_t$. Accordingly, let

$$\hat{\Phi}_c^\mu = [T^{-2} \sum \hat{W}_t^\mu \hat{W}_{t-1}^{\mu\prime} - T^{-1}\hat{M}'][T^{-2} \sum \hat{W}_{t-1}^\mu \hat{W}_{t-1}^{\mu\prime}]^{-1}$$

and

$$\hat{\Phi}_c^\tau = [T^{-2} \sum \hat{W}_t^\tau \hat{W}_{t-1}^{\tau\prime} - T^{-1}\hat{M}'][T^{-2} \sum \hat{W}_{t-1}^\tau \hat{W}_{t-1}^{\tau\prime}]^{-1},$$

and let $\hat{\lambda}_c^\mu$ and $\hat{\lambda}_c^\tau$, respectively, denote the vector of ordered eigenvalues of $\hat{\Phi}_c^\mu$ and $\hat{\Phi}_c^\tau$. We now have Theorem 5.2.

Theorem 5.2. Suppose that $\hat{D} \xrightarrow{p} RD$, $\hat{M} \xrightarrow{p} R_2 M R_2'$, and the assumptions of Lemma 5.1 hold. (a) If $\beta_2 = 0$ in (2.7) and β_1 is an arbitrary constant, then (i) $T(\hat{\Phi}_c^\mu - I) \Rightarrow R_2 \Omega \Psi_k^{\mu\prime}(\Gamma_k^\mu)^{-1}\Omega^{-1} R_2^{-1}$, (ii) $T(\hat{\lambda}_c^\mu - \iota) \Rightarrow \lambda_*^\mu$, and (iii) $T(|\hat{\lambda}_c^\mu| - \iota) \Rightarrow \text{real}(\lambda_*^\mu)$. (b) If β_1 and β_2 in (2.7) are arbitrary constants, then (i) $T(\hat{\Phi}_c^\tau - I) \Rightarrow R_2 \Omega \Psi_k^{\tau\prime}(\Gamma_k^\tau)^{-1}\Omega^{-1} R_2^{-1}$, (ii) $T(\hat{\lambda}_c^\tau - \iota) \Rightarrow \lambda_*^\tau$, and (iii) $T(|\hat{\lambda}_c^\tau| - \iota) \Rightarrow \text{real}(\lambda_*^\tau)$.

This theorem makes it possible to construct test statistics analogous to $q_c(k, k - 1)$ accounting for either an estimated intercept or an estimated intercept and drift. The only modification is that the tests are, respectively, computed using deviations of $\hat{W}_t$ around its average or the residuals from a regression of $\hat{W}_t$ onto a constant and a linear time trend. Therefore, let

$$q_c^\mu(k, k - 1) = T[\text{real}(\hat{\lambda}_{c,k}^\mu) - 1]$$

and

$$q_c^\tau(k, k - 1) = T[\text{real}(\hat{\lambda}_{c,k}^\tau) - 1],$$

Table 1. Quantiles of real(λ_*)

Dimension of λ_*	Significance level	Eigenvalue number					
		1	2	3	4	5	6
1	1%	−13.8					
	2.5%	−10.6					
	5%	−8.0					
	10%	−5.6					
	15%	−4.36					
	50%	−.87					
	90%	.94					
	95%	1.30					
2	1%	−6.7	−24.4				
	2.5%	−5.1	−20.4				
	5%	−3.78	−17.5				
	10%	−2.71	−14.3				
	15%	−2.10	−12.3				
	50%	−.21	−5.8				
	90%	1.15	−1.30				
	95%	1.50	−.62				
3	1%	−4.24	−15.0	−34.6			
	2.5%	−3.23	−12.9	−29.7			
	5%	−2.53	−11.1	−26.0			
	10%	−1.82	−9.2	−22.2			
	15%	−1.4	−8.1	−19.9			
	50%	.02	−3.97	−11.6			
	90%	1.24	−.56	−4.97			
	95%	1.58	−.08	−3.83			
4	1%	−3.19	−11.5	−22.6	−43.3		
	2.5%	−2.5	−9.9	−20.1	−38.3		
	5%	−1.95	−8.5	−18.0	−34.4		
	10%	−1.4	−7.2	−15.6	−30.0		
	15%	−1.07	−6.4	−14.1	−27.2		
	50%	.14	−3.13	−8.4	−17.5		
	90%	1.29	−.21	−3.57	−9.4		
	95%	1.62	.20	−2.74	−7.8		
5	1%	−2.67	−9.6	−18.3	−30.1	−51.6	
	2.5%	−2.04	−8.3	−16.1	−27.1	−46.2	
	5%	−1.64	−7.2	−14.5	−24.7	−41.9	
	10%	−1.17	−6.1	−12.6	−22.0	−37.4	
	15%	−.88	−5.5	−11.4	−20.2	−34.5	
	50%	.22	−2.66	−6.9	−13.4	−23.6	
	90%	1.32	−.02	−2.93	−7.3	−14.1	
	95%	1.66	.36	−2.25	−6.1	−12.3	
6	1%	−2.25	−8.3	−15.5	−24.5	−38.1	−60.2
	2.5%	−1.75	−7.3	−13.8	−22.3	−34.3	−54.6
	5%	−1.40	−6.4	−12.4	−20.4	−31.5	−49.8
	10%	−1.00	−5.4	−10.9	−18.2	−28.3	−44.8
	15%	−.74	−4.84	−9.9	−16.8	−26.3	−41.7
	50%	.28	−2.28	−6.0	−11.3	−18.6	−29.7
	90%	1.36	.13	−2.5	−6.2	−11.4	−19.1
	95%	1.69	.49	−1.89	−5.3	−9.9	−16.8

where $\hat{\lambda}^{\mu}_{c,k}$ (or $\hat{\lambda}^{\tau}_{c,k}$) is the kth-largest eigenvalue of $\hat{\Phi}^{\mu}_{c}$ (or $\hat{\Phi}^{\tau}_{c}$). Theorems 5.1 and 5.2 imply that under the null hypothesis $q^{\mu}_{f}(k, m) \Rightarrow \text{real}(\lambda^{\mu}_{*m+1})$, $q^{\tau}_{f}(k, m) \Rightarrow \text{real}(\lambda^{\tau}_{*m+1})$, $q^{\mu}_{c}(k, k-1) \Rightarrow \text{real}(\lambda^{\mu}_{*k})$, and $q^{\tau}_{c}(k, k-1) \Rightarrow \text{real}(\lambda^{\tau}_{*k})$.

6. CRITICAL VALUES

Although the preceding asymptotic representations do not provide explicit distributions of the proposed test statistics, they do suggest a simple procedure for computing the asymptotic distributions using Monte Carlo techniques. For example, from Theorem 3.1 (b) and Theorem 4.1 (b), the asymptotic distributions of $q_f(k, k-1)$ and $q_c(k, k-1)$ are the same as the asymptotic distribution of the real part of the smallest root of $\Psi'_{kT}\Gamma^{-1}_{kT}$, which in turn has the same asymptotic distribution as real(λ_{*k}), the real part of the smallest root of $\Psi'_{k}\Gamma^{-1}_{k}$. Theorems 5.1 and 5.2 imply that similar remarks apply for the q^{μ}_{f}, q^{τ}_{f}, q^{μ}_{c}, and q^{τ}_{c} test statistics. Accordingly, the distributions of the real parts of the ordered roots of $\Psi'_{kT}\Gamma^{-1}_{kT}$, $\Psi^{\mu\prime}_{kT}(\Gamma^{\mu}_{kT})^{-1}$, and $\Psi^{\tau\prime}_{kT}(\Gamma^{\tau}_{kT})^{-1}$ were computed using 30,000 Monte Carlo replications with $T = 1{,}000$. (As a check of whether $T = 1{,}000$ is sufficiently large, the $k = 3$ entries in the tables were recomputed using 10,000 replications with $T = 2{,}000$. The discrepancies between the two distributions were negligible.)

Selected quantiles of the distribution of real(λ_{*j}) are tabulated in Table 1 for $k = 1, \ldots, 6$ and $j = 1, \ldots,$

Table 2. Quantiles of real(λ_*^μ)

Dimension of λ_*^μ	Significance level	Eigenvalue number					
		1	2	3	4	5	6
1	1%	−20.6					
	2.5%	−16.8					
	5%	−14.1					
	10%	−11.2					
	15%	−9.5					
	50%	−4.36					
	90%	−.82					
	95%	−.11					
2	1%	−12.3	−30.9				
	2.5%	−10.3	−26.4				
	5%	−8.8	−23.0				
	10%	−7.2	−19.5				
	15%	−6.2	−17.2				
	50%	−3.03	−9.7				
	90%	−.29	−4.05				
	95%	.33	−3.10				
3	1%	−9.1	−20.1	−40.2			
	2.5%	−7.9	−17.7	−35.4			
	5%	−6.8	−15.7	−31.5			
	10%	−5.7	−13.5	−27.3			
	15%	−4.99	−12.1	−24.8			
	50%	−2.53	−7.1	−15.6			
	90%	−.07	−2.91	−8.1			
	95%	.53	−2.19	−6.8			
4	1%	−7.6	−15.9	−27.7	−49.2		
	2.5%	−6.6	−14.1	−24.8	−43.6		
	5%	−5.8	−12.6	−22.5	−39.3		
	10%	−4.91	−10.9	−19.8	−35.0		
	15%	−4.33	−9.9	−18.2	−32.1		
	50%	−2.33	−5.8	−11.9	−21.6		
	90%	.04	−2.35	−6.3	−12.7		
	95%	.64	−1.73	−5.3	−11.0		
5	1%	−6.7	−13.5	−22.7	−35.5	−57.1	
	2.5%	−5.8	−12.1	−20.3	−32.0	−51.5	
	5%	−5.2	−10.8	−18.4	−29.2	−47.0	
	10%	−4.45	−9.5	−16.5	−26.4	−42.1	
	15%	−3.94	−8.6	−15.2	−24.4	−39.1	
	50%	−2.05	−5.1	−10.0	−17.0	−27.8	
	90%	.13	−2.02	−5.4	−10.3	−17.6	
	95%	.75	−1.43	−4.59	−8.9	−15.5	
6	1%	−6.1	−12.2	−19.7	−29.1	−42.5	−65.5
	2.5%	−5.4	−10.9	−17.8	−26.5	−39.1	−59.7
	5%	−4.75	−9.7	−16.2	−24.5	−36.1	−54.9
	10%	−4.09	−8.6	−14.4	−22.1	−32.8	−49.7
	15%	−3.64	−7.8	−13.3	−20.6	−30.7	−46.3
	50%	−1.9	−4.62	−8.9	−14.5	−22.3	−34.0
	90%	.2	−1.77	−4.83	−8.9	−14.5	−22.8
	95%	.79	−1.26	−4.06	−7.8	−12.8	−20.3

k; the quantiles for real(λ_{*j}^μ) and real(λ_{*j}^τ) are given in Tables 2 and 3, respectively. Referring to Table 1, the blocks of rows represent the dimension of λ, or equivalently k, the dimension of W_t used to construct the q_f or q_c tests. The columns of the table denote the jth-largest eigenvalue, corresponding to the eigenvalue on which the test is based when there are $m = j - 1$ unit roots under the alternative. For example, in a test of $k = 4$ versus $m = 3$ unit roots, the $q_f(4, 3)$ or $q_c(4, 3)$ tests would be based on the fourth-largest eigenvalue, so the 5% critical value for the test (taken from Table 1) is -34.4 and the 1% critical value is -43.3. For a test of $k = 4$ versus $m = 1$ unit roots, the $q_f(4, 1)$ test would be based on the second-largest eigenvalue, for which the 5% and 1% critical values are -8.5 and -11.5, respectively. If the q_f^μ or q_c^μ tests are used, the critical values come from Table 2. If the q_f^τ or q_c^τ tests are used, the critical values come from Table 3.

The asymptotic null distribution of the $q_f^\mu(k, k - 1)$ and $q_c^\mu(k, k - 1)$ statistics [i.e., the distribution of the real part of the smallest eigenvalue of $\Psi_k^{\mu\prime}(\Gamma_k^\mu)^{-1}$] is plotted in Figure 1 for $k = 1, \ldots, 6$. The figure emphasizes how severely the cdf's of the smallest eigenvalues are shifted below 0, even when $k = 1$ or 2.

7. SIZE AND POWER COMPUTATIONS

This section reports the results of a small Monte Carlo experiment that investigates the size and power of the tests in samples of sizes typically encountered in applied work.

Table 3. Quantiles of real(λ^{τ}_{*})

Dimension of λ^{τ}_{*}	Significance level	Eigenvalue number 1	2	3	4	5	6
1	1%	−29.2					
	2.5%	−24.8					
	5%	−21.7					
	10%	−18.2					
	15%	−16.1					
	50%	−9.0					
	90%	−3.8					
	95%	−2.7					
2	1%	−19.1	−39.2				
	2.5%	−16.8	−34.6				
	5%	−14.9	−30.8				
	10%	−12.9	−26.7				
	15%	−11.6	−24.2				
	50%	−7.0	−15.1				
	90%	−2.94	−7.6				
	95%	−1.97	−6.3				
3	1%	−15.2	−27.1	−48.7			
	2.5%	−13.4	−24.3	−43.5			
	5%	−12.1	−22.1	−39.0			
	10%	−10.7	−19.5	−34.6			
	15%	−9.7	−17.8	−31.8			
	50%	−6.2	−11.3	−21.4			
	90%	−2.52	−5.9	−12.5			
	95%	−1.55	−4.91	−10.7			
4	1%	−13.2	−22.0	−35.3	−57.2		
	2.5%	−11.8	−19.8	−31.6	−51.7		
	5%	−10.7	−18.0	−28.9	−47.0		
	10%	−9.5	−16.0	−25.9	−42.0		
	15%	−8.7	−14.7	−24.0	−38.9		
	50%	−5.7	−9.5	−16.7	−27.6		
	90%	−2.23	−5.1	−10.0	−17.4		
	95%	−1.32	−4.19	−8.7	−15.3		
5	1%	−12.2	−19.0	−28.7	−42.4	−64.6	
	2.5%	−10.9	−17.2	−26.3	−38.8	−59.2	
	5%	−9.8	−15.7	−24.2	−35.9	−54.5	
	10%	−8.7	−14.0	−21.9	−32.6	−49.2	
	15%	−8.0	−12.8	−20.4	−30.4	−46.0	
	50%	−5.3	−8.4	−14.3	−22.1	−33.7	
	90%	−2.12	−4.50	−8.8	−14.2	−22.5	
	95%	−1.20	−3.72	−7.7	−12.6	−20.1	
6	1%	−11.2	−17.0	−25.1	−35.3	−49.7	−73.2
	2.5%	−10.1	−15.4	−23.1	−32.7	−45.7	−67.1
	5%	−9.1	−14.1	−21.3	−30.2	−42.5	−62.4
	10%	−8.1	−12.6	−19.3	−27.7	−38.9	−56.8
	15%	−7.5	−11.6	−18.0	−26.0	−36.7	−53.2
	50%	−4.99	−7.5	−12.9	−19.1	−27.6	−39.9
	90%	−2.00	−4.05	−8.0	−12.5	−18.8	−27.8
	95%	−1.06	−3.36	−7.0	−11.0	−16.9	−25.2

The $q_f^{\mu}(2, 1)$ and $q_c^{\mu}(2, 1)$ tests were studied using two different models for Y_t. In the first, Y_t was generated by the VAR(2)

$$(1 - \phi L)(1 - \Phi L)Y_t = \varepsilon_t, \tag{7.1}$$

and in the second by the mixed-vector (autoregressive moving average) ARMA (1, 1) process

$$(1 - \Phi L)Y_t = (1 + \theta L)\varepsilon_t, \tag{7.2}$$

where in (7.1) and (7.2) $E\varepsilon_t\varepsilon_t' = G$, and where

$$\Phi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & .5 \end{bmatrix}, \qquad G = \begin{bmatrix} 1 & .5 & -.25 \\ .5 & 1 & .5 \\ -.25 & .5 & 1 \end{bmatrix}.$$

Both ϕ and θ are scalars that are less than 1 in absolute value. Under the null hypothesis, $\rho = 1$, so there are two common trends; under the alternative, $|\rho| < 1$, and there is only one common trend. The tests were computed as described in the previous sections, using principal components to construct $\hat{D}$ from the generated Y_t. Although Y_t as generated by (7.1) or (7.2) is not cointegrated (since Y_{3t} is not integrated), because $\hat{D}$ is computed by principal components numerically equivalent test statistics would be obtained using $X_t = PY_t$, where P is any nonsingular matrix. In particular, P could be chosen so that X_t is cointegrated.

The experiments were performed using 2,000 replications with a sample size of $T = 200$. This sample is typical of that found in macroeconomic research; for example, the postwar quarterly National Income and Product Ac-

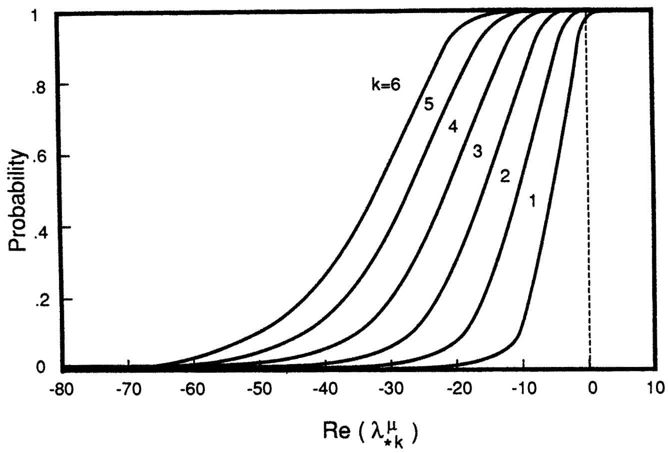

Figure 1. Cumulative Distribution Function of real(λ^{μ}_{*k}).

counts data set from 1947:1 to 1986:4 contains 160 observations, and the monthly financial data set examined in Section 8 has 236 observations. Initial values of $Y_0 = \varepsilon_0 = 0$ were used, and the tests were computed using the generated data $Y_1, \ldots, Y_{200}$. Fewer observations were used to compute the VAR's and covariance matrices entering the correction terms as necessary. The $q_c^\mu(2, 1)$ test statistics were computed using a rectangular window of order J, so $K(j) = 1$ for $|j| \leq J$ and 0 otherwise. When the data were generated by (7.1), the $q_f^\mu(2, 1)$ statistic was computed by filtering the first differences of the integrated (under the null) components using an estimated VAR(1); the correction term M in the $q_c^\mu(2, 1)$ statistic was estimated using a window of order $J = 3$. When the data were generated by (7.2), the $q_f^\mu(2, 1)$ filter was estimated using a VAR(3); the $q_c^\mu(2, 1)$ correction was estimated using $J = 1$. Thus the order of the filter in the $q_f^\mu(2, 1)$ statistic was correct under the null when the data were generated by (7.1), and the order of the window in the $q_c^\mu(2, 1)$ statistic was correct under the null when the data were generated by (7.2). In the other cases, a longer VAR (or additional covariance terms) was incorporated to approximate the covariance structure implied by the vector MA (or AR) in first differences under the null.

Columns A and B of Table 4 contain results for the VAR model (7.1) with $\phi = 4$, and columns C and D contain results for the vector ARMA model (7.2) with $\theta = .4$. The nominal sizes of the q_f^μ test (columns A and C) are somewhat above their actual level, whereas the nominal size of the q_c^μ test is above its level when the data are generated by a VAR (column B) and somewhat below its level when the data are generated by a vector MA (column D). In addition, the $q_f^\mu(2, 1)$ test exhibits greater nominal power than the approximate $q_c^\mu(2, 1)$ test with the VAR data-generation process, whereas the reverse is true when the data are generated by the vector ARMA process.

Table 4. Monte Carlo Experiment Results: Rejection Probabilities

		Data-generating process			
		(7.1), with $\phi = .4$		(7.2), with $\theta = .4$	
ρ	Level	A, $q_f^\mu(2, 1)$	B, $q_c^\mu(2, 1)$	C, $q_f^\mu(2, 1)$	D, $q_c^\mu(2, 1)$
1.00	5%	.03	.03	.03	.07
	10%	.07	.06	.06	.13
.95	5%	.11	.10	.08	.22
	10%	.21	.18	.19	.35
.90	5%	.40	.34	.30	.60
	10%	.59	.50	.51	.74
.80	5%	.92	.82	.86	.99
	10%	.97	.90	.95	.99

NOTE: The results were computed using 2,000 Monte Carlo draws with a sample size of $T = 200$.

8. COMMON TRENDS IN POSTWAR U.S. INTEREST RATES

In this section we test for the number of common trends among three U.S. interest rates with different maturities. The data are 236 monthly observations from January 1960 to August 1979 on the federal funds rate (FF) (an overnight interbank loan rate), the 90-day treasury bill rate (TB3), and the one-year treasury bill rate (TB12). The treasury bill rates are secondary market rates, and all rates are on an annualized basis. All three rates were obtained from the Citibase financial data base.

The theory of the term structure of interest rates suggests that there is at most one common stochastic trend underlying these three rates: Because the expected return on a multiperiod instrument in theory equals the expected return obtained from rolling over a sequence of one-period instruments, a stochastic trend in the short-term rate is inherited by the longer-term rate.

Table 5 presents various tests for unit roots in these interest rates. Although all three rates appear to contain a unit root, the differences among them (the spreads) seem to be stationary. [Application of the Dickey–Fuller $\hat{\tau}_\mu(4)$ test to the first difference of each interest rate rejects the null of a second unit root at the 1% level.] This suggests that there is a single common trend. The multivariate results confirm this suspicion. Testing for 3 versus 1 common

Table 5. Integration and Cointegration Tests on Three Monthly Interest Rates, 1960:1–1979:8

	Univariate results			
	Sample autocorrelations			
Series	Lag 1	Lag 2	Lag 3	$\hat{\tau}_\mu(4)$
FF	.975	.941	.899	−1.79
TB3	.971	.936	.899	−1.44
TB12	.972	.935	.898	−1.34
FF–TB3	.902	.825	.741	−3.17[b]
FF–TB12	.932	.863	.801	−2.82[c]
TB3–TB12	.851	.727	.614	−4.09[a]
Common trend tests				
$q_f^\mu(3, 1)$; $p = 2$, -22.6^a; $p = 4$, -21.6^a				
$q_f^\mu(2, 1)$; $p = 2$, -23.2^b; $p = 4$, -23.5^b				
$q_c^\mu(2, 1)$; $J = 2$, -24.7^b; $J = 4$, -29.6^b				

NOTE: $\hat{\tau}_\mu(4)$ denotes the Dickey–Fuller (1979) t test for a unit root in a univariate series including an estimated constant with an AR(4) correction. The q_c^μ statistics were computed using a flat kernel for $K(j)$ in (4.1) to weight the J estimated autocovariances. FF denotes the federal funds rate, TB3 denotes the 90-day treasury bill rate, and TB12 denotes the one-year treasury bill rate.

[a] Significant at the 1% level.
[b] Significant at the 5% level.
[c] Significant at the 10% level.

trend using the $q_f^\mu(3, 1)$ statistic, from Table 2 the 5% critical value is -15.7 and the 1% critical value is -20.1. The reported test statistics are more negative than both of these critical values, indicating rejection at the 1% level. Tests of the more refined hypothesis of 2 versus 1 also reject the null in favor of a model in which these three rates contain a single common trend.

9. CONCLUSIONS

The procedures proposed in this article provide a way to test for a reduced number of common trends in a multivariate time series model. Although the tests developed apply to real unit roots, they can be applied to certain cointegrated seasonal models. In particular, suppose that $(1 - L)$ in (2.1) is replaced by a seasonal difference $(1 - L^d)$, where d is some integer. Since $(1 - L^d) = (1 - L)(1 + L + \cdots + L^{d-1})$, the tests and asymptotic theory apply directly to the transformed series $(1 + L + \cdots + L^{d-1})X_t$. This approach only tests for cointegration at frequency 0; however, it is possible that alternative tests could be developed for cointegration at seasonal frequencies.

The derivation of the tests and the Monte Carlo results suggest that the q_f test might perform better than the q_c test if under the null the data are generated by a VAR, whereas the reverse is true if the data are generated by a vector moving average process. Further simulation studies are needed to characterize more fully the circumstances in which the tests are likely to perform well.

[Received June 1986. Revised May 1988.]

REFERENCES

Beveridge, S., and Nelson, C. R. (1981), "A New Approach to Decomposition of Economic Time Series Into Permanent and Transitory Components With Particular Attention to Measurement of the Business Cycle," *Journal of Monetary Economics*, 7, 151–174.

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

Box, G. E. P., and Tiao, G. C. (1977), "A Canonical Analysis of Multiple Time Series," *Biometrika*, 64, 355–365.

Brillinger, D. P. (1981), *Time Series Analysis and Theory*, San Francisco: Holden-Day.

Chan, N. H., and Wei, C. Z. (1988), "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *The Annals of Statistics*, 16, 367.

Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, S. (1978), "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumer's Expenditure and Income in the United Kingdom," *Economic Journal*, 88, 661–692.

Dickey, D. A., and Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

Engle, R. F., and Granger, C. W. J. (1987), "Cointegration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276.

Fountis, N. G., and Dickey, D. A. (1986), "Testing For a Unit Root Nonstationarity in Multivariate Autoregressive Time Series," mimeo, North Carolina State University, Dept. of Statistics.

Frisch, R. (1934), *Statistical Confluence Analysis by Means of Complete Regression Systems*, Oslo: Universitets Økonomiske Institutt.

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

Harvey, A. C. (1985), "Trends and Cycles in Macroeconomic Time Series," *Journal of Business & Economic Statistics*, 3, 216–227.

King, R., Plosser, C. I., Stock, J. H., and Watson, M. W. (1987), "Stochastic Trends and Economic Fluctuations," Working Paper 2229, National Bureau of Economic Research, Cambridge, MA.

Nelson, C. R., and Plosser, C. I. (1982), "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139–162.

Phillips, P. C. B. (1987), "Time Series Regression With Unit Roots," *Econometrica*, 55, 277–302.

Phillips, P. C. B., and Durlauf, S. N. (1986), "Multiple Time Series Regression With Integrated Processes," *Review of Economic Studies*, 53, 473–496.

Sargan, J. D. (1964), "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in *Econometric Analysis for National Economic Planning*, eds. P. E. Hart, G. Mills, and J. N. Whittaker, London: Butterworth, pp. 25–63.

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.

Stock, J. H., and Watson, M. W. (1988), "Testing for Common Trends: Technical Appendix," Discussion Paper 167D, Harvard University, Kennedy School of Government.

White, J. S. (1958), "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," *Annals of Mathematical Statistics*, 29, 1188–1197.