6
A Class of Tests for Integration and Cointegration

JAMES H. STOCK

1 Introduction

The intellectual architecture of cointegration analysis constitutes a watershed accomplishment of time series econometrics in the 1980s. As of 1980, econometricians confronted several apparently conflicting pieces of evidence about long run relations among time series. It was recognized that many macroeconomic time series exhibit trend-like behavior and have considerable persistence. Granger (1966) expressed this as the series having much of their spectral power at low frequencies, and Nelson and Plosser (1982) argued that this persistence was captured by modeling the series as having a unit autoregressive root (being integrated of order one). But how to model multiple time series in levels remained unclear. On the one hand, regressions involving highly persistent, unrelated series can produce spuriously large correlations and thus can incorrectly appear to be related (Yule, 1926; Granger and Newbold, 1974). On the other hand, the view that all levels relations are spurious seemed too severe: some "great ratios," such as the share of consumption in income, appeared stable even though the variables themselves were growing (Kosobud and Klein, 1961), and these great ratios, when incorporated as regressors in otherwise standard "first differences" specifications, can have statistically significant coefficients and can be economically large (Davidson, Hendry, Srba, and Yeo, 1978). The achievement of cointegration analysis, as developed by Granger (1983, 1986), Granger and Weiss (1983), and Engle and Granger (1987), was to provide a unified framework in which to understand and to reconcile the apparent conflict between spurious regressions and economically meaningful long-run relations. Moreover, this early work provided probability models of levels relations which were sufficiently well articulated to form a foundation for the development and analysis of new tools for statistical inference in poten-

The author thanks Don Andrews, Frank Diebold, Bruce Hansen, Jerry Hausman, Danny Quah, Mark Watson, and Jeff Wooldridge for helpful suggestions, and Robin Lumsdaine for research assistance. This paper is a May 1990 revision of a March 1988 paper by the same title, edited for this Festschrift. This work was supported in part by National Science Foundation grant SES-86-18984.

tially cointegrated systems. Among the issues stressed in this early work is the importance of correctly specifying the orders of integration of the component series and testing for possible cointegration among the series, that is, of the construction of powerful and reliable tests for integration and cointegration.[1]

This paper introduces a class of statistics that can be used to test for integration or cointegration in time series data. Many of the existing tests for integration and cointegration are motivated by considering the problem of testing whether an autoregressive root equals one against the alternative that it is not equal to one; see, for example, Dickey and Fuller (1979), Phillips (1987), and Phillips and Perron (1988) for tests for integration, and Engle and Granger (1987), Phillips and Ouliaris (1990), and Stock and Watson (1988) for tests for cointegration. In contrast, statistics in the proposed class test directly the implication that an integrated process has a growing variance, that is, has an order in probability of $T^{1/2}$ (is $O_p(T^{1/2})$).

To motivate these statistics, let y_t, $t = 1, \ldots, T$, be a univariate integrated process with zero drift, $s_{\Delta y}(0)$ be the spectral density of $\Delta y_t \equiv y_t - y_{t-1}$ at frequency zero, $[\cdot]$ denote the greatest lesser integer function, $=>$ denote weak convergence on $D[0, 1]$, and define $v_T^*(\lambda) = (2\pi s_{\Delta y}(0)T)^{-1/2} y_{[T\lambda]}$ for $0 \leq \lambda \leq 1$. Then, under general conditions, $v_T^* => W$, where W is standard Brownian motion on $[0, 1]$. The scaling factor $T^{-1/2}$ is consequence of y_t being I(1). In contrast, if y_t is I(0), y_t is $O_p(1)$.

These observations suggest developing test statistics based on the implication that an I(1) process is $O_p(T^{1/2})$ whereas an I(0) process is $O_p(1)$. The approach pursued in this paper is to consider the class of tests constructed using continuous functionals g: $D[0, 1] \rightarrow \Re^1$, such that the distribution of $g(W)$ has arbitrarily small mass in a small neighborhood of $g(0)$. Heuristically, if y_t is I(0), then by the continuous mapping theorem $g(v_T^*)$ will converge to $g(0)$. This suggests that $g(v_T^*)$ could be used to construct a test of the I(1) null that is consistent against the I(0) alternative. If $s_{\Delta y}(0)$ is suitably estimated, it will have an asymptotic null distribution that does not depend on any nuisance parameters. This approach readily handles general trend specifications, including both the leading cases of demeaning or polynomial detrending as well as more general trend specifications. The details are provided in Section 2.

In addition to suggesting new test statistics, this class of tests provides a unifying framework for many previously proposed tests. It therefore provides a simple way to generalize results for existing tests to different types of trends, for example. Although not all tests for a unit root fit into this framework, three that do are examined in Section 3. These are a modified version of Sargan and Bhargava's (1983) uniformly most powerful (UMP) and Bhargava's (1986) locally most powerful invariant (MPI) tests in the first order case; the Phillips (1987) and Phillips–Perron (1988) Z_α statistic; and Lo's (1991) generalization of Mandelbrot's (1975) rescaled range ("R/S") statistic.

Section 4 extends the results of Section 2 to tests for cointegration. These tests build on Engle and Granger's (1987) suggestion by replacing y_t with the residual from a contemporaneous regression of the level of one element of a detrended multivariate time series on the levels of the other elements. The technical arguments in this section draw on results of Phillips and Ouliaris (1990). Like the univariate tests, these are developed for general trends.

Section 5 considers issues of consistency. Section 6 presents the results of a Monte Carlo experiment that compares selected new $g(\cdot)$ tests with several previously proposed tests. The experimental design includes several models estimated for postwar US data. Section 7 concludes.

2 Tests for Integration

Suppose that, under the null hypothesis, the univariate time series variable y_t can be written as the sum of a purely deterministic component $d_t(\beta)$ and a component that is integrated of order one:

$$y_t = d_t(\beta) + \sum_{s=1}^{t} u_s, \quad t = 1,\dots,T. \tag{2.1}$$

The finite dimensional parameter vector β is estimated by $\hat{\beta}$. Let $\omega = 2\pi s_u(0)$, where $s_u(0)$ is the spectral density of u_t at frequency zero, let $\nu_T(\lambda) = T^{-1/2}\sum_{s=1}^{[T\lambda]} u_s$, and let $D_T(\lambda; \beta) = d_{[T\lambda]}(\beta)$. The two processes ν_T and $D_T(\cdot; \beta)$ are assumed to satisfy:

Assumption 1. The following hold jointly:

(a) $\nu_T => \sqrt{\omega} W$, where $0 < \omega < \infty$, and

(b) $T^{-1/2}\{D_T(\cdot; \hat{\beta}) - D_T(\cdot; \beta)\} => \sqrt{\omega} D$, where $D \in D[0,1]$ has a distribution that does not depend on β or on the nuisance parameters describing the distribution of $\{u_t\}$.

This covers many special cases of practical interest. Consider first the partial sum process ν_T. For most of what follows, the general Assumption 1(a) will suffice, although at times it is convenient to consider the special case in which u_t is the linear process,

$$u_t = c(L)\epsilon_t, \quad \sum_{j=0}^{\infty} j\left|c_j\right| < \infty, \quad c(1) \neq 0 \tag{2.2}$$

where ϵ_t is a martingale difference sequence with $E[\epsilon_t^2|\epsilon_{t-1}, \epsilon_{t-2}, \dots] = \sigma^2$ and $\sup_t E[|\epsilon_t|^{2+\delta}|\epsilon_{t-1}, \epsilon_{t-2}, \dots] < \infty$ for some $\delta > 0$. The one-summability of $c(L)$, along with these moment conditions, implies Assumption 1(a) (where $\omega = c(1)^2\sigma^2$) using a standard functional central limit theorem (e.g., Chan and Wei, 1988, theorem 2.2; Hall and Heyde, 1980, theorem 4.1).[2]

Assumption 1(b) is satisfied by a wide variety of trend functions. To establish notation, some important special cases follow.

A. *Constant.* In this case $d_t(\beta) = \beta_0$. A natural estimator of β_0 is $\hat{\beta}_0 = \bar{y} \equiv T^{-1}\sum_{t=1}^{T} y_t$; the demeaned series is

$$y_t^{\mu} = y_t - \bar{y}. \tag{2.3a}$$

B. *Linear time trend.* For reasons discussed below, it is convenient to normalize the known parts of the trend specification to be bounded. Thus the linear trend is written $d_t(\beta) = \beta_0 + \beta_1(t/T)$. If the unknown parameters β_0 and β_1 are estimated by the ordinary least squares (OLS) estimators $(\hat{\beta}_0, \hat{\beta}_1)$, then the detrended series is

$$y_t^{\tau} = y_t - \hat{\beta}_0 - \hat{\beta}_1(t/T). \tag{2.3b}$$

Bhargava (1986) derives an alternative detrending procedure for his locally MPI tests in the first order Gaussian case with drift; his estimators are $\tilde{\beta}_0 = \bar{y} - (T+1)/(2(T-1))(y_T - y_1)$ and $\tilde{\beta}_1 = (T/(T-1))(y_T - y_1)$. The detrended series is

$$y_t^{B} = y_t - \tilde{\beta}_0 - \tilde{\beta}_1(t/T). \tag{2.3c}$$

C. *General trends that are linear in* β. This framework nests the more general time trend $d_t(\beta) = \beta' x_t$, where x_t is a $J \times 1$ vector of known deterministic terms. Let $\tau_T(\lambda) = x_{[T\lambda]}$ and assume that:

$$\tau_T \to \tau, \quad \tau_i \in D[0,1], \quad i = 1, \ldots, J, \quad \text{where } M = \int_0^1 \tau(s)\tau(s)' ds \tag{2.4}$$

is nonsingular and $\sup_{\lambda \in [0,1]} |\tau_{iT}(\lambda)| \leq \bar{\tau}$ for all T, $i = 1, \ldots, J$.

A natural estimator of β is the OLS estimator $\hat{\beta} = (\sum_{t=1}^{T} x_t x_t')^{-1} \sum_{t=1}^{T} x_t y_t$. The normalization $|\tau_{iT}(\lambda)| \leq \bar{\tau}$ obtains by selecting the proper scaling factor, which can be done without loss of generality. In this normalization the true coefficient might depend on T, e.g. a trend γt with γ fixed is rewritten as $\beta_1(t/T)$, where $\beta_1 = \gamma T$. This complicates the interpretation of the coefficients but simplifies treatment of the projections, the latter being of interest here. This is clarified by two examples.

C(i). *Polynomial time trends.* In this case, $d_t(\beta) = \sum_{j=0}^{J-1} \beta_j (t/T)^j$, so that $\tau_T(\lambda) = (1, [T\lambda]/T, \ldots, ([T\lambda]/T)^{J-1})'$. Evidently $\tau_T \to \tau = (1, \lambda, \ldots, \lambda^{J-1})'$; direct calculation indicates that M is nonsingular, with $M_{ij} = 1/(i+j-1)$.

C(ii). *Segmented time trend.* Perron (1989) has proposed tests for a unit root against the alternative hypothesis that the deterministic component of the series contains a segmented trend. An example is the "mean shift" model,

$$d_t(\beta) = \beta_0 + \beta_1 \mathbf{1}(t > [T\lambda_0]) + \beta_2(t/T), \quad 0 < \lambda_0 < 1,$$

where $\mathbf{1}(\cdot)$ denotes the indicator function. In this model the intercept term increases from β_0 to $\beta_0 + \beta_1$ after a fraction λ_0 of the sample has passed; generalizations to other shifts or breaks in the trend are discussed in Perron (1989).

In the mean shift model, $\tau_T(\lambda) = (1, 1([T\lambda] > [T\lambda_0]), [T\lambda]/T)$, and (2.4) is satisfied with $\tau(\lambda) = (1, \mathbf{1}(\lambda > \lambda_0), \lambda)$.

The limiting representations in Assumption 1 depend on ω. To eliminate this dependence, assume for now that there is a consistent estimator for ω:

Assumption 2. Under the null hypothesis $\hat{\omega} \xrightarrow{p} \omega$.

A variety of such estimators exist; an autoregressive estimator is considered in Section 4.

Define v_T^d to be the scaled stochastic process formed using the detrended series:

$$v_T^d(\lambda) \equiv (T\hat{\omega})^{-\frac{1}{2}}\{y_{[T\lambda]} - d_{[T\lambda]}(\hat{\beta})\}. \tag{2.5}$$

If Assumptions 1 and 2 hold, then

$$v_T^d => V^d \tag{2.6}$$

where V^d is the stochastic process $V^d = W - D$.

In general, Assumption 1 must be verified for the form of detrending and the null stochastic process at hand. The condition in Assumption 1(b) is demonstrated here for the leading cases. Let $v_T(\lambda) = (T\hat{\omega})^{-½}y_{[T\lambda]}$, $v_T^\mu(\lambda) = (T\hat{\omega})^{-½}y_{[T\lambda]}^\mu$, $v_T^\tau(\lambda) = (T\hat{\omega})^{-½}y_{[T\lambda]}^\tau$, and $v_T^B(\lambda) = (T\hat{\omega})^{-½}y_{[T\lambda]}^B$. We have:

Theorem 1. Assume that Assumptions 1(a) and 2 hold.

(a) If $d_t(\beta) = 0$, then $v_T => W$;

(b) If $d_t(\beta) = \beta_0$, then $v_T^\mu => V^\mu$, where $V^\mu(\lambda) = W(\lambda) - \int_0^1 W(s)ds$;

(c) If $d_t(\beta) = \beta_0 + \beta_1 t$, then $v_T^\tau => V^\tau$ and $v_T^B => V^B$, where $V^\tau(\lambda) = W(\lambda) - a_1(\lambda)\int_0^1 W(s)ds - a_2(\lambda)\int_0^1 sW(s)ds$ and $V^B(\lambda) = W(\lambda) - (\lambda - ½)W(1) - \int_0^1 W(s)ds$, where $a_1(\lambda) = 4 - 6\lambda$ and $a_2(\lambda) = -6 + 12\lambda$.

(d) For general trends that are linear in β, if (2.4) holds and β is estimated by OLS, then $v_T^d => V^d$, where $V^d(\lambda) = W(\lambda) - \{\int_0^1 W(s)\tau(s)'dsM^{-1}\}\tau(\lambda)$.

Proofs of theorems are in Appendix A.

Theorem 1 provides expressions for V^d, the limiting detrended process, in the leading cases. The processes V^μ and V^τ, derived in Park and Phillips (1988, 1989) and Stock and Watson (1988), have natural Hilbert space interpretations as the continuous time analogues of the discrete time detrended processes; for a discussion, see Park and Phillips (1988, 1989). To simplify notation the result (d) is presented explicitly only for the OLS estimator. However, it generalizes directly to other estimators of β that are linear in y_t as long as the weights satisfy a condition analogous to (2.4).

It follows from (2.6) and the Continuous Mapping Theorem (e.g. Hall and Heyde, 1980, theorem A.3) that if $g(\cdot)$ is a continuous function from $D[0, 1]$ to $\mathfrak{R}^1$, then

$$g(v_T^d) => g(V^d) \tag{2.7}$$

Theorem 1 implies that this holds for each of the leading cases.

The asymptotic representations (2.6) and (2.7) form the basis for the proposed class of tests of integration. Specifically, let $G^d = \{g : D[0, 1] \to \Re^1\}$ be the collection of functionals that satisfy:

(a) g is continuous; (2.8)

(b) $\exists\ c_\alpha$, $|c_\alpha| < \infty$, such that $\Pr[g(V^d) \le c_\alpha] = \alpha$ for all α, $0 < \alpha < 1$;

(c) $g(0) < c_\alpha$ for all α, $0 < \alpha < 1$.

Several remarks are in order.

(a) The family G^d provides a natural class of test statistics for the null hypothesis that $y_t - d_t(\beta)$ is I(1) against the alternative that it is I(0). Under the null, $g(v_T^d)$ has an asymptotic distribution on the real line, with critical values c_α that depend on the functional g. Under a fixed stationary alternative, $y_t - d_t(\beta)$ is $O_p(1)$. This suggests constructing tests of level α of the form, reject if $g(v_T^d) \le c_\alpha$. (The use of the left tail as the rejection region is without loss of generality, since the $g(\cdot)$ can be replaced by $-g(\cdot)$.) Conditions for consistency, including choice of $\hat{\omega}$, are discussed in Section 5.

(b) This approach to constructing tests suggests working backwards from the desired asymptotic representations to the actual test statistic. For example, suppose one wished to develop a unit root test with a χ_1^2 asymptotic null distribution. Because W (and in general V^d) has Gaussian marginals, it is simple to write down continuous functionals of these processes with χ_1^2 distributions. For example, $W(1)^2$ and $V^d(1)^2$ have distributions that are multiples of a χ_1^2, so $g(v_T^d) = v_T^d(1)^2/E(V^d(1)^2)$ provides a unit root test statistic that has a χ_1^2 distribution. Kahn and Ogaki's (1988) J_T test statistic is of this form: $J_T = g(v_T) + o_p(1)$, where $g(f) = f(1)^2$, so that J_T has an asymptotic χ_1^2 distribution.

(c) Other examples of existing tests that fall into this framework are a modification of the Sargan–Bhargava (1983) and Bhargava (1986) tests, the Phillips Perron Z_α statistic, and Lo's (1991) modification of Mandelbrot's (1975) R/S statistic. These are discussed in more detail in the next section.

(d) Condition (2.8)(b) indexes G^d by the type of detrending. In general the type of detrending restricts the functionals g that can be used to construct consistent tests. For example, consider $g(f) = \{\int_0^1 sf(s)\,ds\}^2$. From Theorem 1, $g(V^\mu) = \{\int_{s=0}^1 s(W(s) - \int_{r=0}^1 W(r)\,dr)\,ds\}^2 = \{\int_0^1 (s - ½)\,W(s)\,ds\}^2$, which is distributed as a constant times a χ_1^2. Because g is also continuous, $g \in G^\mu$. However, straightforward algebra shows that $g(V^\tau) = 0$ a.s., so $g \notin G^\tau$. This result has a simple interpretation in light of the discussion of V^τ following Theorem 1: because V^τ can be thought of as detrended Brownian motion, its Hilbert space projection against a linear trend is zero.

(e) The definition of G^d might seem so broad as to include all consistent unit root tests, but this is not so; the class refers only to continuous functionals of v_T. An example of a test not in this class is Park and Choi's (1988) $J(p, q)$ statistic (also see Park, 1990). To be concrete, suppose that $d_t = 0$, so the

$J(0, 1)$ statistic is appropriate; this is $J(0,1) = T^{-1}\sum_{t=1}^{T} y_t^2 / T^{-1}\sum_{t=1}^{T}(y_t - \bar{y})^2 - 1 \Rightarrow \{\int_0^1 W(s)^2 ds / (\int_0^1 W(s)\,ds)^2 - 1\}^{-1}$. The asymptotic representation suggests considering the functional, $g(f) = \{\int_0^1 f^2 / (\int_0^1 f)^2 - 1\}^{-1}$. However, this functional is not continuous at $f = 0$.[3] This is not just a technicality: when $v_T \Rightarrow 0$ under the alternative, g must be continuous at 0 to ensure consistency.

The reason that the J statistic does not fall into this class yet is consistent in that it exploits a different property to obtain consistency. The G class exploits the different orders in probability of the underlying time series, while the J statistic exploits the relative differences in rates of convergence of sums of $\Delta^{-d}u_t$ and $(\Delta^{-d}u_t)^2$ for $d = -1, 0, 1$, where u_t is I(0).

(f) The fact that the form of the $g(\cdot)$ functional does not depend on the type of detrending (subject to remark (d) above) emphasizes that the steps of eliminating the deterministic components and testing for unit roots are conceptually distinct. This is sometimes obscured in procedures in which detrending and testing are performed in the same step. It is also evident that detrending a series when it is not required does not change the asymptotic size of the tests (although it can adversely affect power), because D does not depend on β. In contrast, failing to detrend a series that contains a trend typically leads to a loss of consistency and an incorrect asymptotic size.

3 Relation to Previously Proposed Tests for Integration

This section shows that several previously proposed tests for integration can be expressed, or are closely related to, $g(\cdot)$ tests. The main result is that there are simple $g(\cdot)$ statistics that are asymptotically equivalent to Sargan and Bhargava's (1983) and Bhargava's (1986), where the asymptotic equivalence occurs both under the null and under a local alternative. It is also shown that the Phillips–Perron Z_α statistic can be cast as a $g(\cdot)$ statistic, as can Lo's (1991) modification of Mandelbrot's (1975) R/S statistic.

3.1 The Sargan–Bhargava and Bhargava Tests

Sargan and Bhargava (1983) considered the problem of testing for integration in the first order Markov model with i.i.d. Gaussian errors:

$$y_t = \beta_0 + \sum_{s=1}^{t} \rho^{t-s}\epsilon_s, \quad \epsilon_s \text{ i.i.d. } N(0,\sigma^2), \quad t = 1,\ldots,T, \tag{3.1}$$

where $(\rho, \beta_0, \sigma^2)$ are unknown parameters. They showed that the statistic

$$R = \sum_{t=2}^{T}(\Delta y_t)^2 \Big/ \sum_{t=1}^{T}(y_t^{\mu})^2 \tag{3.2}$$

provides the basis for a test of the random walk null ($\rho = 1$) against the stationary alternative ($|\rho| < 1$). (Note that R is the Durbin–Watson statistic for a regression of y_t against a constant.) Similarly, in the case of a time trend, Bhargava (1986) considered the maintained model,

$$y_t = \beta_0 + \beta_1 t + \sum_{s=1}^{t} \rho^{t-s}\epsilon_s, \quad \epsilon_s \text{ i.i.d. } N(0,\sigma^2), \quad t = 1,\ldots,T, \tag{3.3}$$

where $(\rho, \beta_0, \beta_1, \sigma^2)$ are unknown; he showed that

$$R = \sum_{t=2}^{T}(\Delta y_t^B)^2 / \sum_{t=1}^{T}(y_t^B)^2 \tag{3.4}$$

can be used to construct a test of $\rho = 1$ against $|\rho| < 1$, which he claimed to be MPI in the neighborhood of $\rho = 1$, where y_t^B is given in (2.3c).

Sargan–Bhargava and Bhargava did not provide asymptotic distribution theory for their tests. However, it obtains as a direct consequence of Theorem 1.

Corollary 1.1. Assume that y_t is generated by (2.1). Then

(a) $T^{-1}R^{-1} => m\int_0^1 V^{\mu}(s)^2 ds$

(b) $T^{-1}R_2^{-1} => m\int_0^1 V^{B}(s)^2 ds$

where $m = \omega/\text{var}(\Delta y_t)$.

Thus the asymptotic null distribution of the R and R_2 statistics depends on the nuisance parameters through the spectral density of Δy_t at the origin. However, there is a simple $g(\cdot)$ statistic that is essentially the same as the R and R_2 statistic, but which provides the basis for an asymptotically similar test under the null model (2.1). Let

$$g_{SB}(f) = \int_0^1 f(s)^2 ds, \tag{3.5}$$

so that $g_{SB}(v_T^{\mu}) = T^{-1}\sum_{t=1}^{T}(v_t^{\mu})^2 = T^{-2}\sum_{t=1}^{T}(y_t^{\mu})^2/\hat{\omega}$ and $g_{SB}(v_T^{B}) = T^{-1}\sum_{t=1}^{T}(v_t^{B})^2 = T^{-2}\sum_{t=1}^{T}(y_t^{B})^2/\hat{\omega}$. Thus $g_{SB}(v_T^{\mu})$ and $g_{SB}(v_T^{B})$ are respectively $T^{-1}R^{-1}$ and $T^{-1}R_2^{-1}$, except that $\hat{\omega}$ appears in the denominator rather than $T^{-1}\sum_{t=2}^{T}(\Delta y_t)^2$. It follows from Theorems 1 and 2 that, if Assumptions 1 and 2 hold and $\omega = \text{var}(\Delta y_t)$ (as is implied by (3.1) with $\rho = 1$), then $g_{SB}(v_T^{\mu})$ and $g_{SB}(v_T^{B})$ respectively are asymptotically equivalent to the Sargan–Bhargava and Bhargava tests under the null hypothesis.

The primary merit of the R and R_2 statistics is Sargan–Bhargava's and Bhargava's claims of finite sample optimality properties under (3.1) and (3.3). It is therefore of interest to compare the power of R and R_2 to that of the $g_{SB}(\cdot)$ statistics. This can be done by deriving a representation for these statistics under a local alternative, that is, under an alternative that approaches the unit root null as the sample size increases. Since the Sargan–Bhargava and Bhargava exact tests assume a first order model, suppose that y_t is generated by

$$y_t = \rho_T y_{t-1} + \epsilon_t, \quad t = 1,\dots,T, \quad \epsilon_t \text{ i.i.d. } (0,\sigma_\epsilon^2), \tag{3.6}$$

where $E|\epsilon_t|^4 < \infty$. As the local alternative, adopt Phillips' (1987) formulation, so that $\rho_T = e^{c/T}$ where c is a constant (also see Cavanagh, 1985; Chan and Wei, 1987). Following Phillips (1987) and Phillips and Perron (1988), define $W_c(\lambda) = \int_{s=0}^{\lambda} e^{(\lambda-s)c} dW(s)$.

Computation of the $g(\cdot)$ statistics requires selecting a specific estimator of ω. For this discussion, suppose that ω is estimated by an average of sample covariances:

$$\tilde{w} = \{T^{-1}\sum_{t=2}^{T} \hat{u}_t^2 + 2\sum_{j=1}^{\ell} T^{-1} \sum_{t=j+2}^{T} \hat{u}_t \hat{u}_{t-j}\}, \tag{3.7}$$

where $\hat{u}_t = y_t^\mu - \hat{\rho} y_{t-1}^\mu$, where $\hat{\rho} = \sum_{t=2}^T y_t^\mu y_{t-1}^\mu / \sum_{t=2}^T (y_{t-1}^\mu)^2$. Note that, as is pointed out in Phillips and Ouliaris (1990) and Stock and Watson (1988), it is important to use the estimated $\hat{\rho}$ rather than the null value $\rho = 1$ in constructing $\hat{\omega}$ to ensure that the resulting test is consistent.

The next result uses Theorem 6.2 of Phillips and Perron (1988) to provide a limiting representation for the $g(\cdot)$, R, and R_2 statistics under this local alternative.

Theorem 2. Let y_t be generated by (3.6), and let the estimator $\tilde{\omega}$ in (3.7) be used to construct v_T^μ and v_T^B, where $\ell \to \infty$ as $\mathrm{T} \to \infty$ and $\ell = o(\mathrm{T}^{1/4})$. Then:

(a) $g_{SB}(v_T^\mu) => \int_0^1 V_c^\mu(s)^2 ds$,

(b) $g_{SB}(v_T^B) => \int_0^1 V_c^B(s)^2 ds$,

(c) $T^{-1}R^{-1} => \int_0^1 V_c^\mu(s)^2 ds$, and

(d) $T^{-1}R_2^{-1} => \int_0^1 V_c^B(s)^2 ds$,

where $V_c^\mu(\lambda) = W_c(\lambda) - \int_0^1 W_c(s) ds$ and $V_c^B(\lambda) = W_c(\lambda) - (\lambda - ½) W_c(1) - \int_0^1 W_c(s) ds$.

In summary the $g_{SB}(\cdot)$ statistics have asymptotic null distributions that do not depend on the nuisance parameters for general I(1) processes; Theorem 2 implies that, for the AR(1) case, they have the same local asymptotic representations and therefore the same local power as the original Sargan–Bhargava and Bhargava statistics.

3.2 The Phillips and Phillips–Perron Tests

The Z_α statistics proposed by Phillips (1987) and Phillips and Perron (1988) also are closely related to a $g(\cdot)$ statistic. Assume here that $d_t(\beta) = 0$ and $y_0 = 0$, so that their statistic is

$$Z_\alpha = T(\hat{\rho} - 1) - \tfrac{1}{2}\{\tilde{\omega} - T^{-1}\sum_{t=1}^{T}(y_t - \hat{\rho} y_{t-1})^2\} / T^{-2}\sum_{t=1}^{T} y_{t-1}^2, \tag{3.8}$$

where $\tilde{\omega}$ and $\hat{\rho}$ are defined in and following (3.7), respectively. Because $\sum_{t=1}^{T} y_{t-1}\Delta y_t = \frac{1}{2}\{y_T^2 - \sum_{t=1}^{T}(\Delta y_t)^2\}$ (where $y_0 = 0$), (3.8) can be rearranged to yield

$$Z_\alpha = \tfrac{1}{2}\{\tilde{v}_T(1)^2 - 1\} / T^{-1}\sum_{t=1}^{T-1} \tilde{v}_T(t/T)^2 - \tfrac{1}{2}T(\hat{\rho}-1)^2, \tag{3.9}$$

where $\tilde{v}_T(\lambda) = (T\tilde{\omega})^{-\frac{1}{2}} y_{[T\lambda]}$. Thus the Z_α statistic can be rewritten as $Z_\alpha = g_z(\tilde{v}_T) - \frac{1}{2}T(\hat{\rho}-1)^2$, where:

$$g_z(f) = \tfrac{1}{2}\{f(1)^2 - 1\} / \int_0^1 f(s)^2 \mathrm{d}s. \tag{3.10}$$

Under the null and the local alternative (3.6), $T(\hat{\rho}-1)^2 \xrightarrow{p} 0$ so $Z_\alpha - g_z(\tilde{v}_T) \xrightarrow{p} 0$, so in this case Z_α and $g_z(\tilde{v}_T)$ are asymptotically equivalent. Tests formed using $g_z(f)$ in (3.10) will be referred to below as modified Z_α (MZ_α) tests. In addition, g_z is the functional associated with the Dickey–Fuller root test in the first order model.

3.3 The R/S Statistic

Mandelbrot (1975) proposed the "rescaled range," or "range over standard deviation" (R/S), statistic to detect fractionally differenced processes. Mandelbrot's motivation is essentially that used here: the order in probability of a fractionally differenced process is other than $T^{\frac{1}{2}}$ (the order depends on the fractional differencing parameter). Suppose here that $d_t(\beta) = \beta_0$. Then the R/S statistic is,

$$\tilde{Q}_T = T^{-\frac{1}{2}}(\max_{1\le t\le T} y_t - \min_{1\le t\le T} y_t)/(T^{-1}\sum_{t=2}^{T} \Delta y_t^2)^{\frac{1}{2}}.$$

Because $\sup_{\lambda\in[0,1]} f(\lambda) - \inf_{\lambda\in[0,1]} f(\lambda)$ is continuous, it follows from Assumption 1 that $\tilde{Q}_T \Rightarrow \sqrt{m}\{\sup_{\lambda\in[0,1]} W(\lambda) - \inf_{\lambda\in[0,1]} W(\lambda)\}$, where $m = \omega/\mathrm{var}(\Delta y_t)$. As do the Sargan–Bhargava and Bhargava statistics, the asymptotic distribution of $\tilde{Q}_T$ depends on the nuisance parameter m.

Lo (1991) suggested eliminating this dependence on m by considering a modification of the R/S statistic:

$$Q_T = (T\hat{\omega})^{-\frac{1}{2}}(\max_{1\le t\le T} y_t - \min_{1\le t\le T} y_t). \tag{3.11}$$

Evidently this is a g-statistic: $Q_T = g_{RS}(v_T)$, where $g_{RS}(f) = \sup_{\lambda\in[0,1]} f(\lambda) - \inf_{\lambda\in[0,1]} f(\lambda)$. Under Assumptions 1 and 2, $g_{RS}(v_T) \Rightarrow g_{RS}(W)$.

4 Testing for Cointegration

Engle and Granger (1987) proposed testing for whether a multivariate time series is cointegrated of order (1, 1) by regressing one of the individual series on the remaining series and testing for a unit root in the residual. The reasoning behind this procedure is that, if the multivariate series are individually integrated and are not jointly cointegrated, then any linear combination of the series will be integrated; if, however, the series are cointegrated, then the linear combination formed by the cointegrating vector will be stationary. Since the OLS estimator of the cointegrating vector is consistent (Stock, 1987), the residuals from a levels cointegrating regression can proxy for this linear combination.

Engle and Yoo (1987) derived the asymptotic distribution of the Dickey–Fuller (1979) $\hat{\tau}_\mu$ t-statistic, computed using the residuals from a levels regression, under the null hypothesis that the constituent time series are independent random walks. Phillips and Ouliaris (1990) derived the asymptotic distributions of this and other residual-based cointegration test statistics under more general conditions. Their approach is used here to generalize the univariate $g(\cdot)$ tests to residual-based tests for cointegration.

Let Y_t be a $n \times 1$ time series variable generated by the multivariate counterpart of (2.1):

$$Y_t = d_t(\beta) + \sum_{s=1}^{t} U_s, \quad t = 1,\ldots,T, \tag{4.1}$$

where $d_t(\beta)$ and U_t are $n \times 1$ deterministic and stochastic terms, respectively. The parameter vector β is typically unknown and is estimated by $\hat{\beta}$. The assumption made on $\{d_t(\hat{\beta}),\ U_t\}$ in the multivariate case is analogous to Assumptions 1(a) and 1(b). Let $\nu_T^d(\lambda; \hat{\beta}) = T^{-1/2}\{Y_{[T\lambda]} - d_{[T\lambda]}(\hat{\beta})\}$. Then:

Assumption 3. $\nu_T^d(\cdot,\ \hat{\beta}) => B^d$, where B^d is an $n \times 1$ Gaussian process on $D[0,\ 1]$ with mean 0 and $E\{B^d(\lambda)B^d(\lambda)'\} = \Omega f(\lambda)$, where f is a scalar function on $[0,\ 1]$ that does not depend on β or the distribution of $\{U_t\}$, and where Ω is $n \times n$.

Because Ω can be factored as $\Omega = HH'$, Assumption 3 implies that

$$\nu_T^d(\cdot, \hat{\beta}) => HV^d, \tag{4.2}$$

where V^d is a $n \times 1$ Gaussian process with $E(V^d(\lambda)V^d(\lambda)') = f(\lambda)I_n$, so the distribution of V^d does not depend on β or on the nuisance parameters describing the distribution of $\{U_t\}$.

Like Assumption 1, Assumption 3 is satisfied under general conditions on U_t. For example, if $U_t = C(L)\epsilon_t$, where $C(L)$ is 1-summable and ϵ_t is an $n \times 1$ martingale difference sequence satisfying multivariate extensions of the moment conditions stated after (2.2), then Assumption 3 follows from Chan and Wei (1988), Theorem 2.2, with $\Omega = C(1)\Sigma_\epsilon C(1)'$, where $\Sigma_\epsilon = E\epsilon_t\epsilon_t'$. As is discussed below, the assumption is also satisfied by a large number of estimated trend specifications.

Under the null hypothesis, Y_t is not cointegrated. It is assumed here that interest is in testing whether Ω has full versus reduced rank; β is treated as a nuisance parameter. (An alternative, not pursued here, would be to consider cointegration that involves the trend terms as well.) Thus it is natural to consider residual-based tests in which the series are initially detrended.

Partition Y_t^d as $[Y_{1t}^d | Y_{2t}^{d\prime}]'$, where $Y_t^d = Y_t - d_t(\hat{\beta})$ and Y_{1t}^d is a scalar process. Suppose that the cointegrating vector α is estimated by regressing Y_{1t}^d on Y_{2t}^d:

$$\hat{\alpha}^d = \begin{bmatrix} 1 \\ -(\sum Y_{2t}^d Y_{2t}^{d\prime})^{-1}(\sum Y_{2t}^d Y_{1t}^d) \end{bmatrix}.$$

The residuals are $\hat{z}_t^d = \hat{\alpha}^{d\prime} Y_t^d$. Let

$$\tilde{\omega}_z = \{T^{-1}\sum_{t=2}^{T}\tilde{u}_t^2 + 2\sum_{j=1}^{\ell} T^{-1}\sum_{t=j+2}^{T}\tilde{u}_t\tilde{u}_{t-j}\}, \tag{4.3}$$

where $\tilde{u}_t = \hat{z}_t - \tilde{\rho}\hat{z}_{t-1}$, with $\tilde{\rho} = \sum_{t=2}^{T}\hat{z}_t\hat{z}_{t-1}/\sum_{t=2}^{T}\hat{z}_{t-1}^2$.

The strategy is to apply the univariate $g(\cdot)$ tests to the residuals $\hat{z}_t^d$. The standardized detrended residuals are:

$$\hat{v}_T^d(\lambda) = (T\tilde{\omega}_z)^{-\frac{1}{2}}\hat{z}_{[T\lambda]}^d. \tag{4.4}$$

To handle the leading special cases, analogously to Section 2 adopt the notation $\hat{v}_T$, $\hat{v}_T^\mu$, $\hat{v}_T^\tau$, and $\hat{v}_T^B$. As in the univariate case, these standardized residuals have limiting representations in terms of functionals of Brownian motion. Let W denote standard $n \times 1$ Brownian motion, and partition the $n \times 1$ stochastic processes V, W, etc. conformably with Y_t. The following theorem summarizes results for the leading cases.

Theorem 3. Assume that Assumption 3 holds, that Ω has full rank, and that $\sup_t E(U_{it})^4 < \infty$, $i = 1, \ldots, n$. Then $\hat{v}_T^d => (\tilde{\alpha}^{d\prime}\tilde{\alpha}^d)^{-1/2}\tilde{\alpha}^{d\prime}V^d$, where

$$\tilde{\alpha}^d = [1 \,|\, -(\int_0^1 V_1^d(s)V_2^d(s)'ds)(\int_0^1 V_2^d(s)V_2^d(s)'ds)^{-1}]'.$$

In the leading special cases, this result holds with:

(a) for $d_t(\beta) = 0$, $V^d = W$;

(b) for $d_t(\beta) = \beta_0$, $V^\mu(\lambda) = W(\lambda) - \int_0^1 W(s)ds$;

(c) for $d_t(\beta) = \beta_0 + \beta_1 t$, $V^\tau(\lambda) = W(\lambda) - a_1(\lambda)\int_0^1 W(s)ds - a_2(\lambda)\int_0^1 sW(s)ds$ and $V^B(\lambda) = W(\lambda) - (\lambda - 1/2)W(1) - \int_0^1 W(s)ds$, where $a_1(\lambda) = 4 - 6\lambda$ and $a_2(\lambda) = -6 + 12\lambda$.

Theorem 3 states that, under the null hypothesis of no cointegration, the residuals from the levels regressions have an asymptotic distribution that is the same as the indicated functionals of multivariate Brownian motion.

The multivariate version of trends that are linear in β is $d_t(\beta) = \beta x_t$, where β is $n \times J$ and x_t is $J \times 1$. The OLS estimator is $\hat{\beta} = (\sum_{t=1}^{T} Y_t x_t')(\sum_{t=1}^{T} x_t x_t')^{-1}$. If $\tau_T(\lambda) = x_{[T\lambda]}$ satisfies (2.4), a direct calculation as in the proof of Theorem 1(d) shows that $\nu_T^d(\cdot,\hat{\beta}) => HV^d$, where $V^d(\lambda) = W(\lambda) - \{\int_0^1 W(s)\tau(s)'dsM^{-1}\}\tau$.

Thus, like the leading cases addressed in Theorem 4, this more general trend specification satisfies Assumption 3 and (4.2).

As in the univariate case, if $g(\cdot)$ is continuous, then $g(\hat{v}_T^d) => g(V^d)$. Because the limiting representations in Theorem 3 do not depend on any nuisance parameters other than n, this can form the basis for constructing tests with critical values that can be tabulated as a function of n. This therefore generalizes Phillips and Ouliaris's (1990) results for selected residual-based tests for cointegration to the entire family of tests defined by G^d.

As a specific example, consider the modification of the Sargan–Bhargava statistic formed using (3.5):

$$g_{SB}(\hat{v}_T) = T^{-2}\sum_{t=1}^{T}\hat{z}_t^2 \,/\, \tilde{\omega} => \tilde{\alpha}'\{\int_0^1 V(s)V(s)'ds\}\tilde{\alpha}\,/(\tilde{\alpha}'\tilde{\alpha}). \qquad (4.5)$$

This statistic is related to the $\hat{P}_u$ statistic proposed by Phillips and Ouliaris (1990); the difference is that, in (4.5), the cointegrating residuals are scaled by $\tilde{\omega}$, while in the $\hat{P}_u$ statistic these residuals are scaled by an estimate of the conditional variance of Y_{1t} given Y_{2t} constructed using an estimate of the spectral density matrix $S_{\Delta Y}(0)$.

5 Consistency

This section examines the consistency of univariate g-tests against I(0) alternatives both in general and in some special cases. The general consistency result is given in the following theorem. Recall that, in the notation of Section 2, $D_T(\lambda, \beta) \equiv d_{[T\lambda]}(\beta)$ and, for trends that are linear in β, $d_t(\beta) = \beta' x_t$, with $\tau_T(\lambda) \equiv x_{[T\lambda]}$.

Theorem 4. Assume that $y_t = d_t(\beta) + w_t$, where $\sup_t E|w_t|^{2+\delta} < \infty$ for some $\delta > 0$, and that one of the following holds:

(a) $T^{-1/2}\{D_T(\cdot, \hat{\beta}) - D_T(\cdot, \beta)\} => 0$; or

(b) $D_T(\lambda, \beta) = \beta'\tau_T(\lambda)$, where τ_T satisfies (2.4) and β is estimated by OLS.

If $\hat{\omega} \xrightarrow{p} k \neq 0$, then $g(v_T^d)$ is consistent for all $g \in G^d$.

This provides sufficient conditions for the consistency of all $g \in G^d$. However, some $g \in G^d$ will be consistent under weaker conditions, and rate results can be obtained for specific tests if different conditions are assumed. In addition, this theorem assumes that $\hat{\omega}$ has a nonzero limit under the alternative, something that must be shown on an estimator-by-estimator basis.

We therefore examine the consistency of four specific $g(\cdot)$ tests, constructed using an autoregressive estimator of ω, in the case that $d_t(\beta) = 0$. The four functionals are:

$$g_1(f) = \int_0^1 |f(s)|^r ds, \qquad (5.1a)$$

$$g_2(f) = \int_0^1 \ln|f(s)|\,ds, \tag{5.1b}$$
$$g_3(f) = [\int_0^1 s^r f(s)\,ds]^2, \tag{5.1c}$$
$$g_{RS}(f) = \sup_{\lambda\in[0,1]} f(s) - \inf_{\lambda\in[0,1]} f(s), \tag{5.1d}$$

where $f(s)$ is a function on the unit interval.

The g_1 functional is examined because, for $r = 2$, it leads to the generalization of the Sargan–Bhargava and Bhargava statistics. As mentioned in Section 2, the g_3 functional has the useful property that, in the univariate context, its distribution is a constant times a χ_1^2. (Note however that for $n > 1$ the distribution of g_3 is not χ^2, but rather involves the random processes in Theorem 3.) The g_{RS} functional leads to Lo's (1991) modification (3.11) of the R/S statistic. The g_2 statistic is included to demonstrate the variety of conditions that lead to consistency of $g(\cdot)$ tests.

The calculated test statistics are the sample analogs of (5.1):

$$g_1(v_T) = T^{-1}\textstyle\sum_{t=1}^T |v_t|^r, \tag{5.2a}$$
$$g_2(v_T) = T^{-1}\textstyle\sum_{t=1}^T \ln|v_t|, \tag{5.2b}$$
$$g_3(v_T) = \{T^{-1}\textstyle\sum_{t=1}^T (t/T)^r v_t\}^2, \tag{5.2c}$$
$$g_{RS}(v_T) = \max_{1\le t\le T} v_t - \min_{1\le t\le T} v_t = Q_T. \tag{5.2d}$$

Here, ω is estimated by a sequence of autoregressive spectral density estimators. That is, let $\hat{a}(1)$ be the OLS estimator of $a(1)$ in the regression,

$$\Delta y_t = \beta_0 + \beta_1 y_{t-1} + a(L)\Delta y_{t-1} + e_t, \tag{5.3}$$

where $a(L)$ is a lag polynomial of length p and e_t is the regression error. The autoregressive spectral density estimator is $\hat{\omega} = \hat{\sigma}_\epsilon^2/(1 - \hat{a}(1))^2$.

Were y_{t-1} not in the regression (5.3), $\hat{\omega}$ would be the autoregressive spectral density estimator considered by Berk (1974), who proved its consistency under general conditions on $c(L)$ in (2.2) for $p \to \infty$ and $p = o(T^{1/3})$. Said and Dickey (1984) extended this result to the regression coefficients of (5.3); the consistency of $\hat{\omega}$ under the null is an implication of their Theorem 6.1. Thus $\hat{\omega}$ satisfies Assumption 2.

The properties of the univariate tests are investigated under the alternative that y_t is a stationary linear process with nonzero mean:

$$y_t = \beta_0 + w_t, \text{ where } w_t = b(L)\epsilon_t,\ \epsilon_t \text{ i.i.d. } (0,\sigma_\epsilon^2),\ \mathrm{E}\epsilon_t^4 < \infty, \tag{5.4}$$
$$\text{where } 0 < \left|b(e^{i\omega})\right|^2 < \infty\ \ \forall\omega\in(-\pi,\pi],\ \sum_{j=0}^{\infty}\left|b_j\right| < \infty,$$
$$\text{and } \sum_{j=0}^{\infty} j\left|d_j\right| < \infty, \text{ where } d(z) = b(z)^{-1}.$$

In general $\hat{\omega}$ converges to a nonzero constant under this alternative:

Lemma 1. Suppose that y_t is generated by (5.4), $p \to \infty$, $p = o(T^{1/3})$, $p^{1/2}\sum_{k=p+1}^{\infty}|d_k| \to 0$, and $\sum_{j=0}^{\infty}(j-1)d_j \neq 0$. Then $\hat{\omega} \xrightarrow{p} \sigma_\epsilon^2/(\sum_{j=0}^{\infty}(j-1)d_j)^2 \equiv \kappa$.

The next theorem provides rates of consistency against (5.4).

Theorem 5. Assume that the conditions of Lemma 1 hold.

(a) If $E(\epsilon_t^{2r}) < \infty$, then $(T^{½r}/\ln T)g_1(v_T) \xrightarrow{p} 0$ for all $r > 0$.

(b) If $E(\ln|y_t|)^2 < \infty$, then $g_2(v_T) + (½ - \delta)\ln T \xrightarrow{p} -\infty$ for all $\delta > 0$.

(c) $Tg_3(v_T) \xrightarrow{p} \beta_0^2/\kappa(1 + r)^2$ for all r ≥ 0.

(d) If $\epsilon_j = 0$, $j \leq 0$, and if there exist nonstochastic sequences $\{a_T, b_T\}$ such that $b_T + a_T\max_{1\leq t\leq T}|\epsilon_t| \xrightarrow{d} \epsilon^*$, where ϵ^* has a distribution on the real line, then $g_{RS}(v_T) = O_p(\max(|b_T|, 1)/\sqrt{T}a_T)$.

Two remarks are in order.

(a) The rates at which the statistics converge under the alternative is strikingly different. Indeed, the rate of convergence of the test $g_1(\cdot)$ can be made arbitrarily large; the cost of a faster rate is an increased number of moments of ϵ_t assumed to exist under the alternative. It is instructive to compare these to other rate results in the literature. In particular, Phillips and Ouliaris (1990) show that the Z_α statistic has a rate of T under the fixed alternative, whereas the Dickey–Fuller t-statistic has a rate of $T^{½}$. All else equal, these results suggest that the Z_α, MZ_α, and the modified Sargan–Bhargava statistics might be expected to exhibit better power against I(0) alternatives than the Dickey–Fuller t-statistic.

(b) The rate of convergence of the R/S statistic depends on the rate of convergence of the extreme order statistics of $\{\epsilon_t\}$. In general, if $b_T + a_T\max_{1\leq t\leq T}|\epsilon_t|$ has a limiting distribution, then it will be of the Fréchet, Weibull, or Gumbel form; the sequences $\{a_T, b_T\}$, if they exist, depend on the distribution of $\{\epsilon_t\}$ (Reiss, 1989, p. 152). For example, if $\{\epsilon_t\}$ are i.i.d. $N(0, \sigma^2)$, then $a_T = (2\ln T)^{½}$ and $b_T = -2\ln T$. Thus, for Gaussian ϵ_t, $Q_T = O_p((2\ln T)^{½}/\sqrt{T})$ under H_1. It can also be shown that $v_T => 0$ under the conditions of Theorem 5(d). Thus these conditions are sufficient to show the consistency not just of the R/S test, but of all g-statistics. In this sense, the treatment of the R/S statistic in theorem 5 relies on the strongest conditions for consistency.

6 Monte Carlo Results

This section presents the results of a Monte Carlo study of several of the tests discussed in the previous sections. The design includes two types of models: stylized models found elsewhere in the literature, included here to permit comparisons across studies, and empirical models based on postwar US time series data. The motivation for using the empirical models is that, to make recommendations for empirical practice, it is important to study the finite sample behavior of the statistics in probability models typical of those found in applications.

6.1 *Univariate Tests*

The experiment examines the behavior of seven tests for a unit root under eight different probability models (data generating processes). The leading case in practice is an I(1) null with unknown drift versus the alternative that the process is stationary around a linear time trend; see, for example, the extensive discussion in Christiano and Eichenbaum (1989). Thus the Monte Carlo experiment focuses on the linear trend case, $d_t(\beta) = \beta_0 + \beta_1(t/T)$.

Of the seven unit root tests, five are $g(\cdot)$ tests discussed above, of which four are new, and two are standard tests. These two are the Dickey–Fuller t-statistic and the Phillips–Perron Z_α statistic. These statistics have been studied extensively elsewhere and therefore provide a basis for comparison of these results with Monte Carlo examinations of other tests. The unit root tests considered are:

1. The modified Sargan–Bhargava (MSB) statistic, based on the functional

$$g_{MSB}(f) = \{\int_0^1 f(s)^2 ds\}^{\frac{1}{2}}; \tag{6.1}$$

2. The Dickey–Fuller t-statistic (DF_τ);
3. The Phillips–Perron Z_α statistic given in (3.8);
4. The modified Phillips–Perron Z_α statistic (MZ_α), $g_z(v_T^\tau)$, where $g_z(f) = ½(f(1)^2 - 1)/\int_0^1 f(s)^2 ds$;
5. The modified R/S statistic Q_t given in (3.11);
6. The g_2 statistic in (5.2b), in which $g_2(f) = \int_0^1 \ln|f(s)|ds$;
7. The g_3 statistic in (5.2c) with $r = 2$, so that $g_3(f) = \{\int_0^1 s^2 f(s)ds\}^2$.

The MSB statistic (6.1) is evaluated using v_T^B, while the MZ_α, R/S, g_2, and g_3 statistics are evaluated using v_T^τ; for these tests, the spectral density was estimated by the autoregressive spectral estimator $\hat{\omega}$ defined following (5.3), with 4 lags ($T = 100$) and 5 lags ($T = 200$). The DF_τ autoregression was evaluated with 4 lags ($T = 100$) and 5 lags ($T = 200$). The Z_α statistic was evaluated using $\ell = 4$ ($T = 100$) and $\ell = 5$ ($T = 200$).

Asymptotic critical values for the $g(\cdot)$ tests were computed by Monte Carlo simulation using a standard Gaussian random walk of length $T = 500$. Critical values for the univariate DF_τ and Z_α statistics were taken from Fuller (1976, pp. 373 and 371, respectively). As a reference, asymptotic critical values for the MSB statistic are provided in Table 1.

Of the eight probability models considered, the first three are standard and provide a basis for comparison against other results in the literature (in particular Schwert, 1989). These take the form:

$$(1 - \rho L)y_t = u_t, \quad u_t = \epsilon_t + \theta\epsilon_{t-1}, \quad \epsilon_t \text{ i.i.d. } N(0,1), \quad t = 1,\ldots,T, \tag{6.2}$$

where $\theta = 0$ (model 1), $\theta = 0.5$ (model 2), and $\theta = -0.5$ (model 3).

The remaining probability models were estimated using five quarterly US

Table 1. Asymptotic critical percentiles for the univariate MSB statistic

Percentile	Demeaned	Detrended
.025	0.17405	0.15250
.05	0.19144	0.16449
.10	0.21426	0.18050
.20	0.24894	0.20415
.30	0.27957	0.22418
.50	0.34302	0.26235
.70	0.42787	0.30843
.80	0.49094	0.34229
.90	0.58267	0.39049
.95	0.66777	0.43341
.975	0.74723	0.47113

Notes: The MSB statistic is $g_{\mathrm{MSB}}(v_T^{\mu})$ (demeaned) and $g_{\mathrm{MSB}}(v_T^{B})$ (detrended), where $g_{\mathrm{MSB}}(f) = \{\int_0^1 f(s)^2 ds\}^{1/2}$ and where v_T^{μ} and v_T^{B} are defined in Section 2. Based on 20,000 Monte Carlo replications with $T = 500$.

time series over the period 1948:I to 1988:IV. The series were selected using two criteria: first, that the series either has been examined in the literature for its unit root properties or is closely related to series that have been so studied; and second, it has a spectral shape that is representative of those found in postwar US data. The evaluation of the second criterion drew on Stock and Watson (1990), in which various time series properties of 163 monthly US time series, including plots of their spectra, are cataloged.

The five series chosen are: model 4: the real money supply (M2), in logarithms; model 5: the inventory to sales ratio for manufacturing and trade (IVT82); model 6: the number of new business incorporations, in logarithms (INC); model 7: the 90-day US Treasury bill rate (FYGM3); and model 8: total real personal income less transfer payments (GMYXP8). The data (and mnemonics) are from the CITIBASE database. Most empirical macroeconomic research uses quarterly series, often quarterly averages of monthly values, so these monthly series were aggregated to the quarterly level before transformations or estimation.

The empirical models were obtained by estimating

$$a(L)(1-\rho L)y_t = \beta_0 + \epsilon_t, \tag{6.3}$$

where $a(L)$ has order 6 and, for estimation, $\rho = 1$ is imposed.[4] A natural way to assess the fit of these models is to consider the qualitative adequacy of their approximations in the frequency domain. To this end, the spectra of these series, estimated using the AR(6) approximation (solid line) and using a smoothed periodogram with a Fejer kernel with bandwidth 10 (dashed line) are graphed in Figures 1–5. In each case, there appears to be no qualitatively

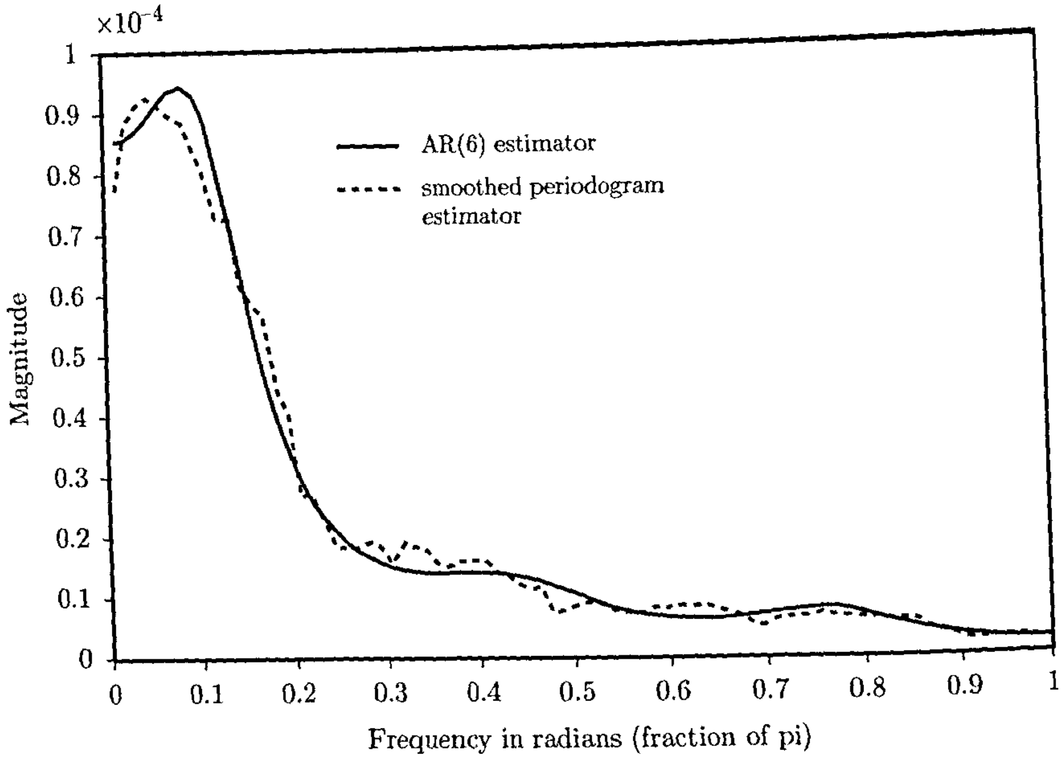

Figure 1. Real money supply (M2), USA, 1948:I–1988:IV: estimated spectral density (growth rates)

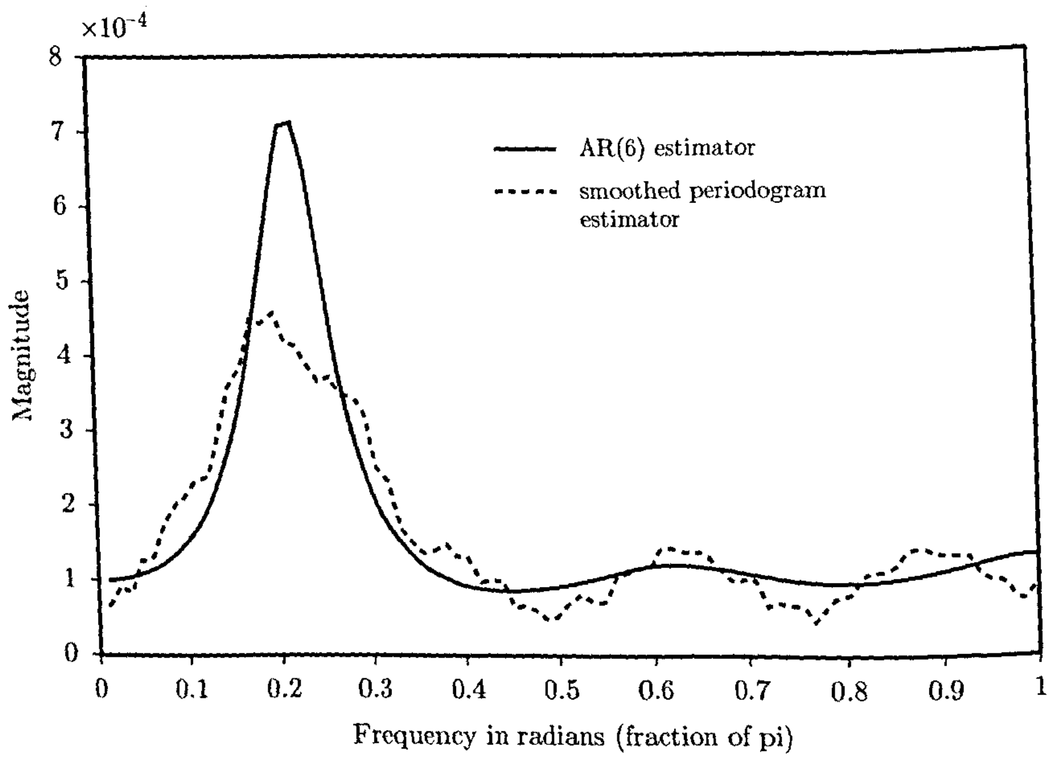

Figure 2. Inventory to sales ratio, manufacturing and trade, USA, 1948:I–1988:IV: estimated spectral density (changes)

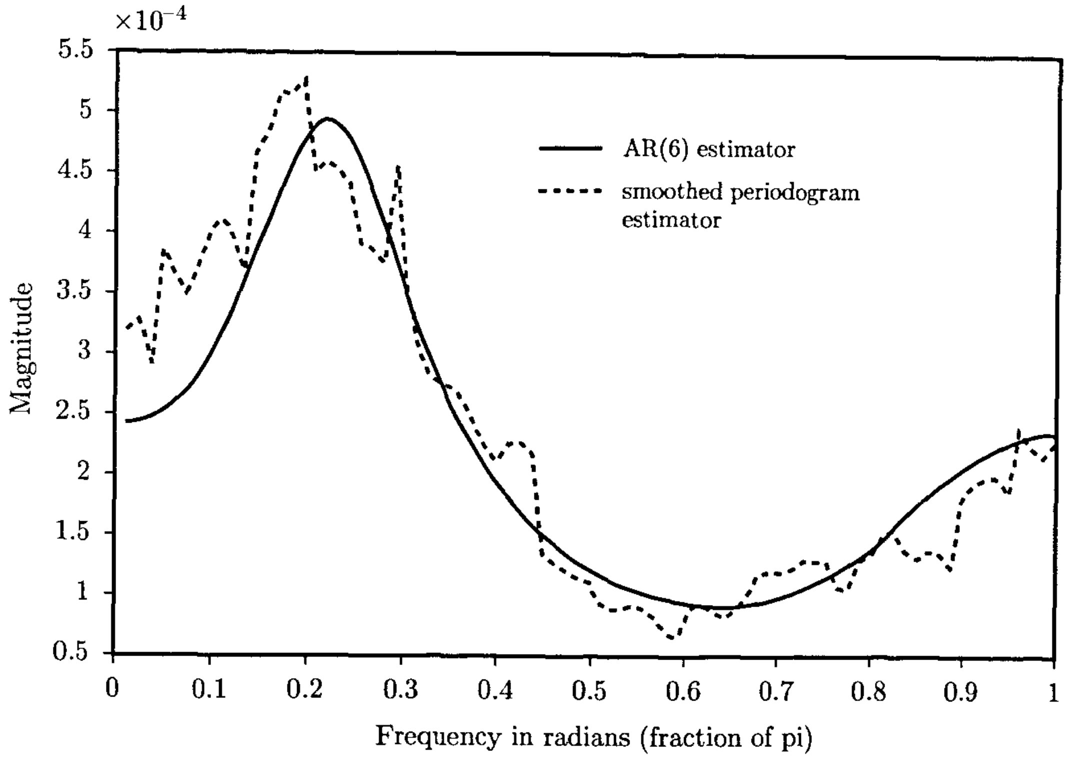

Figure 3. Number of new business incorporations, 1948:I–1988:IV: estimated spectral density (growth rates)

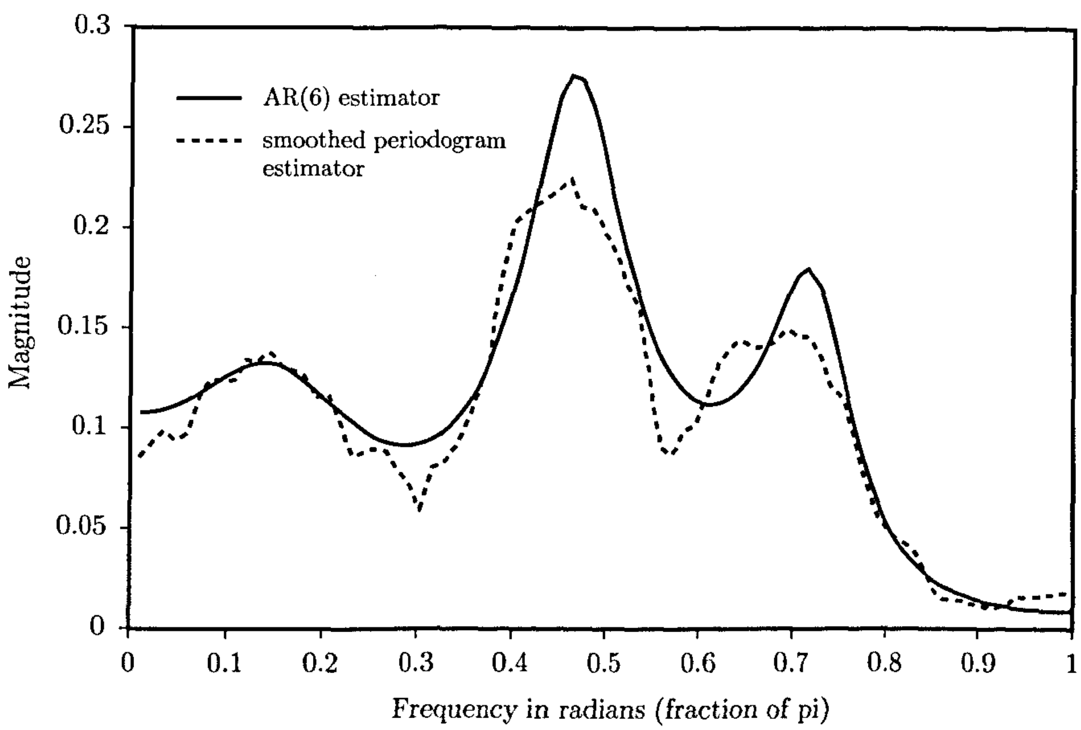

Figure 4. 90-day Treasury bill rate, USA, 1948:I–1988:IV: estimated spectral density (changes)

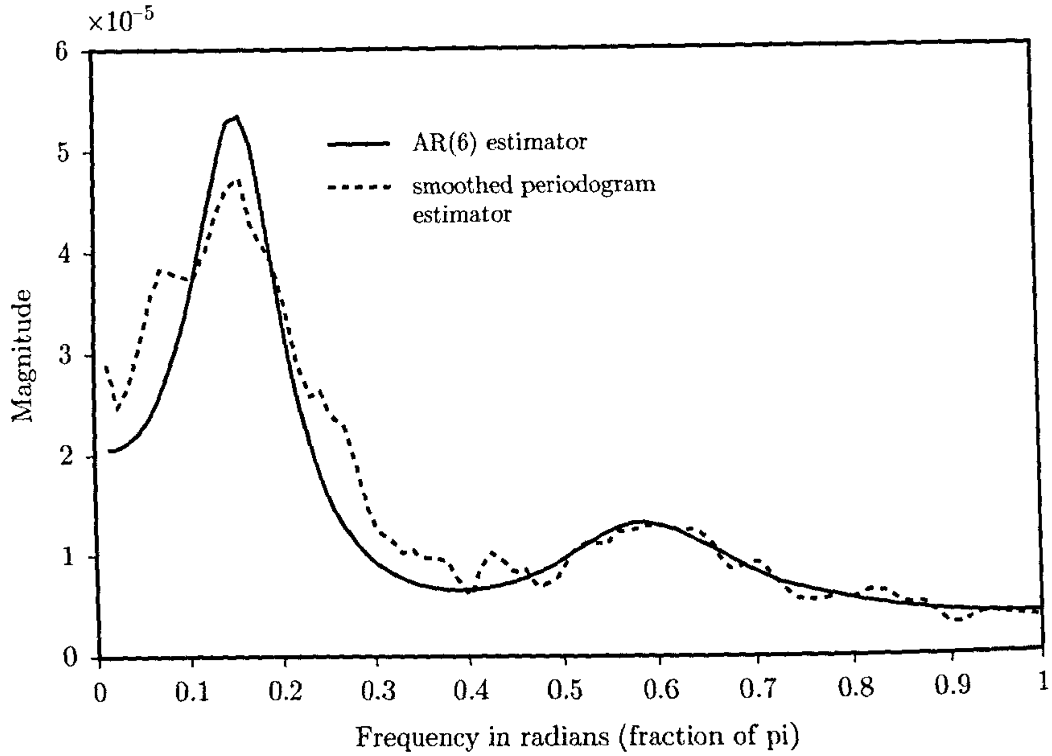

Figure 5. Total real personal income less transfer payments, USA, 1948:I–1988:IV: estimated spectral density (growth rates)

important feature of the spectrum missed by the AR(6) estimates. One notable feature is that although the 90-day Treasury bill rate and new business incorporations exhibit substantial high frequency power, none of the spectra increase sharply with frequency. This is typical of the 163 series in Stock and Watson (1990). In only a few cases, such as import and export figures, are the series dominated by their high frequency movements. Even in these cases, however, aggregation to the quarterly level substantially reduces the high frequency component. The estimated parameters are provided in Appendix B.

Because these models are based on autoregressive approximations, there is the possibility that the performance of autoregression-based tests will be overstated because the model approximation error is slight. For each series, the calculations were therefore repeated using an ARMA(1, 5) model, $(1 - \rho L)y_t^T = b(L)\epsilon_t$ with ρ and $b(L)$ estimated by maximum likelihood. By varying ρ in simulations (keeping $b(L)$ fixed), this model is capable of generating data under both the null and the alternative.

For each of the eight models, the tests were studied using Gaussian errors under the null ($\rho = 1$) and under three alternatives: for $T = 100$, $\rho = .95$, $.90$, and $.80$; for $T = 200$, $\rho = .975$, $.95$, and $.90$. These choices of ρ permit examining the predicted stability of the power function under the local alternative $\rho_T = e^{c/T} = 1 + c/T + o(c/T)$.

6.2 Cointegration Tests

The test statistics examined in the univariate experiments were also examined using four bivariate models, except that the g_3 statistic was evaluated using $r = 1$ in (5.1c), i.e. $g_3(f) = \{\int_0^1 sf(s)\,ds\}^2$. The test statistics were computed as described for the univariate tests, except using the cointegrating residuals based on demeaned series. The choice was made to demean rather than detrend for comparability with earlier studies. Critical values for the DF_τ ("ADF" in Engle and Granger, 1987; Phillips and Ouliaris, 1990) and Z_α tests were taken from Phillips and Ouliaris (1990); for the other statistics, they were computed by Monte Carlo simulation using the asymptotic representations of these statistics with 5,000 Monte Carlo replications.

The simulation examined four models. The first three are bivariate extensions of the univariate model (6.2), namely,

$$y_{1t} = \tfrac{1}{2}\sum_{s=1}^{t} u_{1t} + \tfrac{1}{2}\sum_{s=1}^{t} \rho^{t-s} u_{2t}, \quad y_{2t} = \tfrac{1}{2}\sum_{s=1}^{t} u_{1t} - \tfrac{1}{2}\sum_{s=1}^{t} \rho^{t-s} u_{2t}, \tag{6.4}$$

where $u_t = \epsilon_t + \theta\epsilon_{t-1}$, $t = 1, \ldots, T$, with ϵ_t i.i.d. $N(0, I_2)$ and $\theta = 0$ (model 1), 0.5 (model 2), and −0.5 (model 3). Under the alternative, $\rho < 1$, the cointegrating vector is $(1\ -1)$ and $z_t = \rho z_{t-1} + u_{2t}$. This simple model provides a useful starting point: because z_t is generated by the same process under the alternative in (6.4) as is y_t in the stationary univariate case in (6.2), the differences in size and power between the univariate and bivariate tests can largely be attributed to random variations of the estimated cointegrating vector and to the different types of detrending, rather than to differences in the specification of nuisance parameters.

The experiment also examined an empirical model (model 4) of the form:

$$\begin{aligned} \Delta y_t &= \mu_1 + u_{1t}, \\ (1 - \rho L) z_t &= \mu_2 + u_{2t}, \end{aligned} \tag{6.5}$$

where $U_t = (u_{1t}\, u_{2t})'$ was approximated by a VAR(6). The model was estimated with y_{1t} and y_{2t} respectively being inventories (IVMT82) and sales (MT82) in manufacturing and trade, aggregated from the monthly to the quarterly level, in logarithms, from 1952:I to 1988:IV. The motivation for choosing this pair is that the inventory-to-sales ratio was one of the series examined in the previous Monte Carlo simulation. The cointegrating vector $(1\ -1)$ was imposed in estimation, and the possibility of cointegration was admitted in estimation by specifying the right-hand side of the z_t equation in (6.5) in terms of z_{t-1} and lags of Δz_{t-1}, with the dependent variable being z_{t-1}.

In all four models, the pseudo-data were generated using $T = 200$, Gaussian errors, and zero drift ($\mu_1 = \mu_2 = 0$ in (6.5)). Under the null of no cointegration, $\rho = 1$; the alternatives were $\rho = .95$, .9, and .85.

6.3 Results and Discussion

The univariate results are summarized in Tables 2–4. For each model, the first row gives the size of the test based on the asymptotic critical value. Because the size typically differs from the asymptotic level, the reported power is size-

Table 2. Size and power of tests for integration, detrended statistics, $T = 100$
Rejection probabilities for tests with asymptotic level 5%

Model	ρ	Test statistic						
		MSB	DF_τ	Z_α	MZ_α	R/S	g_2	g_3
1	1.00	.081	.070	.062	.084	.157	.083	.054
	.95	.164	.070	.102	.162	.153	.144	.093
	.90	.306	.128	.235	.308	.274	.275	.143
	.80	.574	.309	.669	.572	.477	.524	.236
2	1.00	.075	.062	.009	.072	.174	.076	.053
	.95	.168	.072	.081	.163	.158	.144	.089
	.90	.308	.123	.159	.306	.272	.272	.126
	.80	.564	.282	.374	.564	.478	.503	.211
3	1.00	.043	.082	.802	.042	.055	.043	.063
	.95	.170	.074	.091	.170	.162	.158	.087
	.90	.327	.149	.198	.329	.298	.304	.139
	.80	.606	.441	.491	.628	.571	.595	.259
4	1.00	.028	.105	.008	.029	.187	.037	.052
	.95	.231	.071	.080	.260	.231	.222	.070
	.90	.487	.117	.168	.524	.466	.436	.123
	.80	.737	.210	.359	.760	.662	.679	.201
5	1.00	.168	.203	.072	.171	.257	.167	.090
	.95	.181	.079	.082	.175	.171	.166	.086
	.90	.384	.165	.147	.380	.359	.352	.150
	.80	.781	.476	.335	.773	.733	.736	.313
6	1.00	.003	.074	.018	.003	.026	.009	.055
	.95	.253	.083	.071	.284	.258	.242	.073
	.90	.580	.151	.157	.607	.560	.526	.126
	.80	.874	.373	.402	.891	.814	.838	.251
7	1.00	.047	.046	.065	.046	.099	.047	.054
	.95	.160	.072	.103	.165	.145	.146	.085
	.90	.297	.119	.223	.305	.249	.275	.128
	.80	.532	.272	.621	.532	.417	.478	.209
8	1.00	.042	.223	.079	.037	.174	.054	.069
	.95	.048	.041	.036	.068	.052	.066	.023
	.90	.465	.128	.128	.555	.478	.474	.081
	.80	.915	.350	.434	.942	.903	.894	.291

Notes: The pseudo-data were generated using (6.2) (models 1–3) and (6.3) (models 4–8). The first row for each model (with $\rho = 1$) presents size, based on the asymptotic critical values; the remaining rows present the size-adjusted power. The MSB statistic was evaluated using v_T^B, the remaining g-statistics and Z_α using v_T^τ, and $(1, t)$ were included as regressors in the DF_τ autoregression. The asymptotic critical values were computed by Monte Carlo simulation. The results are based on 5,000 Monte Carlo replications.

Table 3. Size and power of tests for integration, detrended statistics, $T = 200$
Rejection probabilities for tests with asymptotic level 5%

Model	ρ	Test statistic						
		MSB	DF_τ	Z_α	MZ_α	R/S	g_2	g_3
1	1.00	.060	.060	.072	.058	.085	.062	.064
	.975	.141	.084	.093	.143	.140	.134	.068
	.95	.302	.155	.210	.307	.272	.273	.100
	.90	.681	.436	.651	.689	.592	.608	.165
2	1.00	.047	.052	.025	.049	.095	.047	.055
	.975	.148	.073	.089	.143	.134	.136	.069
	.95	.310	.138	.205	.318	.273	.286	.107
	.90	.683	.399	.561	.687	.572	.620	.187
3	1.00	.032	.080	.721	.025	.020	.027	.061
	.975	.165	.078	.100	.165	.154	.158	.071
	.95	.348	.163	.238	.360	.319	.336	.109
	.90	.727	.509	.699	.769	.670	.727	.190
4	1.00	.053	.103	.033	.052	.165	.059	.062
	.975	.091	.068	.067	.109	.098	.095	.045
	.95	.321	.148	.184	.375	.324	.302	.091
	.90	.727	.367	.521	.773	.684	.666	.175
5	1.00	.157	.200	.243	.158	.178	.147	.070
	.975	.185	.097	.102	.181	.179	.172	.077
	.95	.408	.202	.220	.403	.373	.371	.135
	.90	.843	.600	.592	.849	.785	.807	.234
6	1.00	.018	.073	.065	.018	.032	.020	.056
	.975	.078	.079	.056	.092	.085	.093	.047
	.95	.334	.168	.179	.385	.343	.337	.090
	.90	.813	.473	.559	.862	.785	.791	.199
7	1.00	.032	.043	.095	.027	.046	.031	.057
	.975	.149	.078	.102	.158	.151	.141	.074
	.95	.322	.153	.233	.334	.300	.288	.108
	.90	.682	.412	.669	.699	.584	.628	.184
8	1.00	.107	.214	.189	.113	.199	.110	.060
	.975	.002	.033	.006	.003	.001	.008	.005
	.95	.052	.143	.070	.126	.051	.154	.032
	.90	.683	.516	.501	.848	.729	.801	.167

Notes: See the notes to Table 2.

adjusted: the power was computed using the critical values that produce 5 percent rejections under the null hypothesis for the indicated number of observations for the given model. This size-adjusted power is reported in the remaining three rows for each model.

In model 1, all the statistics have sizes close to their 5 percent level, even for $T = 100$ (except for the R/S statistic). The size results for the DF_τ and Z_α statistics for models 2 and 3 are similar to those that have appeared elsewhere, with the size of the DF_τ statistic being stable but the size of the

Table 4. Size and power of tests for integration, ARMA(1, 5) models, detrended statistics, $T = 200$
Rejection probabilities for tests with asymptotic level 5%

Model	ρ	Test statistic						
		MSB	DF_τ	Z_α	MZ_α	R/S	g_2	g_3
4	1.00	.076	.068	.023	.079	.188	.078	.060
	.975	.159	.079	.090	.161	.150	.149	.077
	.95	.308	.134	.183	.312	.273	.277	.102
	.90	.618	.327	.468	.619	.501	.550	.168
5	1.00	.023	.037	.016	.022	.060	.029	.051
	.975	.151	.070	.086	.157	.151	.138	.079
	.95	.307	.124	.172	.313	.284	.269	.124
	.90	.637	.320	.420	.625	.526	.554	.188
6	1.00	.089	.079	.019	.093	.175	.092	.057
	.975	.150	.079	.094	.162	.154	.145	.075
	.95	.308	.144	.184	.316	.279	.274	.118
	.90	.638	.362	.467	.642	.542	.573	.186
7	1.00	.190	.149	.032	.208	.345	.187	.065
	.975	.169	.083	.091	.168	.162	.149	.078
	.95	.343	.158	.208	.336	.303	.304	.119
	.90	.665	.399	.585	.667	.568	.610	.202
8	1.00	.079	.076	.022	.088	.209	.089	.057
	.975	.152	.073	.090	.154	.147	.131	.076
	.95	.308	.135	.186	.310	.283	.253	.110
	.90	.633	.338	.504	.625	.537	.535	.180

Notes: The data were generated according to estimated ARMA(1, 5) models described in the text. See the notes to Table 2.

Z_α statistic changing substantially with θ. The tradeoff between this stability of the size and the size-adjusted power is evident in these models, with the Z_α statistic having better power than DF_τ. Interestingly, in models 1–3 the MZ_α statistic has power properties comparable to the Z_α statistic, but has size that is closer to the level. Indeed, the MZ_α statistic has essentially the same size as the MSB statistic under model 1, the case in which the MSB statistic is asymptotically equivalent to Bhargava's exact test. The g_2 statistic also exhibits good power, better than DF_τ and almost as good as MZ_α, and stable size in these simple models.

The size distortions in models 2 and 3 are not in the same direction for all the statistics. In particular, the size of the MZ_α statistic in model 3 is substantially below its level, while the opposite is true for the Z_α statistic. Because the main difference between these statistics is the spectral estimator at frequency zero, this emphasizes its key role in determining the sampling properties of the statistics.

The results for the empirically derived models differ from those for the simpler models and warrant several observations. First, the size distortions for the DF_τ statistic are as large as for the Z_α statistic (models 5 and 8). Second, the size distortions, while substantial, are not nearly as severe as in model 3. A possible explanation is that none of the models exhibit spectra with the preponderance of power at high frequencies. However, this possibility was admitted by estimating relatively long autoregressions on potentially overdifferenced series. To the extent that these series are typical of those used in empirical analyses, this negative MA problem might not be as severe in practice as might be suggested by the results for model 3.

Third, the R/S statistic exhibits substantial size distortions, even for $T = 200$ and even for models in which the other statistics do not exhibit this problem. This suggests that the sampling distribution of the finite-sample extreme order statistic in series with complicated dependence is not well-approximated by its asymptotic limit.

Fourth, although the size results for the AR and ARMA simulations typically differ, the size performance for the autoregressions-based tests is not clearly better in the AR simulations. The implication, reinforced by the results for models 2 and 3, is that specification bias is not key in determining the sampling distribution of the spectral density estimator. However, the size distortion of the Z_α statistic is substantially less when the ARMA model is used; in this case, the Z_α statistic (which uses 5 lags for $T = 200$) is correctly specified.

Fifth, the size-adjusted power of the autoregressive tests is reduced when the data are generated according to the MA specification. However, the overall ranking of the powers is largely unaffected, with the Dickey–Fuller t-statistic having lower power than the MSB, Z_α, or MZ_α statistics. This is consistent with the theoretical predictions of Theorem 4 (and related results in the literature) that indicate relatively faster rates of convergence for these statistics under the null.

Sixth, the MSB statistic performs well. Aside from the g_3 statistic (which has the best size performance and by far the worst power), the MSB statistic has the least size distortion, followed by g_2. The power of the MSB statistic is also quite good. For example, in models 5 and 6, its power against $\rho = .9$ with $T = 200$ exceeds 80 percent in the AR simulations. The power also compares well to the power of Bhargava's MPI R_2 statistic in the first order model (model 1). For example, the power of the R_2 test was separately computed to be .76 for $\rho = .9$ and $T = 200$, while the MZ_α and MSB statistics respectively have power .68 and .69. In several models, MZ_α performs as well as MSB.

The cointegration results are summarized in Table 5. The most powerful tests are the Z_α, MZ_α, and MSB tests. The univariate size difficulties of the Z_α statistic are mirrored in bivariate model 3. For the models considered here, only the DF_τ and g_3 statistics have size consistently close to their level, and the power of the g_3 statistic is so low as to eliminate it as a serious contender. For all tests, in models 1–3 the power is lower than in the corresponding uni-

Table 5. Size and power of cointegration tests, demeaned statistics, $n = 2$, $T = 200$
Rejection probabilities for tests with asymptotic level 5%

Model	ρ	Test statistic						
		MSB	DF_τ	Z_α	MZ_α	R/S	g_2	g_3
1	1.00	.115	.052	.069	.104	.157	.112	.064
	.95	.180	.127	.192	.179	.154	.170	.055
	.90	.426	.329	.553	.433	.338	.392	.081
	.85	.669	.583	.878	.673	.530	.620	.112
2	1.00	.124	.054	.024	.110	.185	.121	.066
	.95	.160	.121	.154	.162	.152	.163	.061
	.90	.388	.305	.413	.391	.322	.380	.090
	.85	.620	.556	.684	.621	.509	.593	.115
3	1.00	.070	.050	.663	.059	.060	.068	.060
	.95	.143	.119	.184	.147	.133	.146	.062
	.90	.358	.336	.579	.365	.317	.351	.072
	.85	.564	.599	.901	.574	.481	.553	.083
4	1.00	.127	.078	.056	.112	.148	.126	.068
	.95	.108	.113	.077	.110	.093	.103	.080
	.90	.173	.193	.100	.174	.147	.165	.105
	.85	.251	.298	.124	.253	.206	.243	.138

Notes: The pseudo-data were generated using (6.4) (models 1–3) and (6.5) (model 4). The first row for each model (with $\rho = 1$) presents size, based on the asymptotic critical values; the remaining rows present the size-adjusted power. All statistics were evaluated using the residuals from cointegrating regressions of y^μ_{1t} on y^μ_{2t}. For the MSB, MZ_α, R/S, g_2, and g_3 tests, these residuals were standardized by an autoregressive estimate of the spectral density with 5 lags. The asymptotic critical values were computed by Monte Carlo simulation. The results are based on 5,000 Monte Carlo replications.

variate case. A striking feature of model 4 is that the power of all of the tests is very low, much lower than in the stylized models 1–3, although the size distortion is no more severe than under the base model 1. Despite the limited nature of these results, they suggest that these tests can exhibit low power and substantial size distortions, indicating room for further work.

7 Summary and Conclusions

The statistics developed here provide a new class of tests for integration and cointegration. The idea behind these tests is simple: that an integrated process grows at rate $T^{1/2}$, whereas a stationary process does not. This formulation also provides a unifying framework for many previous tests found in the literature and extends to general trend processes.

Once the class of tests is formulated, a natural question is which member of the class will have the greatest power against a specific model, and whether the finite-sample properties of the test are satisfactory. Although it is possible to obtain general results for some processes, such as local-to-unit root processes, and to characterize the rate of convergence of specific members of the class under a fixed alternative, such results provide only a partial answer to the question of determining the optimal test in this class. The approach taken here has been to investigate size and power numerically. The main novelty of the Monte Carlo simulations is the use of empirical models to generate the pseudo-data. In the univariate case, these models are representative of a broad set of US time series, and include series that have substantial power at high frequencies.

While the Monte Carlo results examine only a limited number of tests, they nonetheless provide several conclusions for empirical practice. First, the MSB and MZ_α tests have good power properties, relative to the other tests. The size of both tests is also reasonably well controlled. Second, the size distortions of the Z_α statistic suggested elsewhere might be overstated in terms of their practical importance: in the empirical models, both in their AR and MA formulation, the size distortions are less than in the stylized pure MA(1) models (but see Schwert, 1989, for an argument for using more extreme MA representations). Third, simulations using the empirical models suggest that the size distortion for the Dickey–Fuller t-statistic can be more severe than is suggested by simulations based on the simple MA(1) models; in addition, the size-adjusted power of the Dickey–Fuller statistics is typically substantially lower than that of several of the other statistics. Fourth, the modified R/S statistic has consistently poor size properties.

The main conclusion from the more limited Monte Carlo analysis of the residual-based cointegration tests is that, of those studied, none performs particularly well. Although the size of some of the tests, for example the Dickey–Fuller t-statistic, is close to its asymptotic level, the power of all of the tests is poor in the one empirical model examined.

Appendix A: Proofs of Theorems

Proof of Theorem 1. (a)–(c). The results follow from Assumption 1 and Theorem 2.2 of Chan and Wei (1988) after straightforward calculations. The calculations for $y^{\tau}_{[T\lambda]}$ are given in the proof of Theorem 5.1 of Stock and Watson (1988).

(d). Write

$$T^{-\frac{1}{2}}\{D_T(\lambda;\hat{\beta}) - D_T(\lambda;\beta)\} = \{(T^{-1}\sum_{t=1}^{T} x_t x_t')^{-1}(T^{-\frac{3}{2}}\sum_{t=1}^{T} x_t \sum_{s=1}^{t} u_s)\}'\tau_T([T\lambda]/T).$$

It follows from (2.4) that $T^{-1}\sum_{t=1}^{T} x_t x_t' \rightarrow M$ and $T^{-3/2}\sum_{t=1}^{T} x_t \sum_{s=1}^{t} u_s \Rightarrow \sqrt{\omega}\int_0^1 \tau(s)\, W(s)\, ds$, so $T^{-1/2}\{D_T(\lambda;\hat{\beta}) - D_T(\lambda;\beta)\} \Rightarrow \sqrt{\bar{\omega}}\{M^{-1}\int_0^1 \tau(s)\, W(s)\, ds\}'\tau(\lambda)$.

Because τ and M do not depend on unknown parameters, Assumption 1(b) is satisfied. It follows that $v_T^d => W - \beta^{*\prime}\tau$, where $\beta^* = M^{-1}\int_0^1 \tau(s)W(s)ds$. ■

Proof of Theorem 2. The results (a) and (b) follow from Theorem 6.2 of Phillips and Perron (1988), since under the local alternative $T^{1/2}(\hat{\rho} - \rho_T) \xrightarrow{p} 0$. The results (c) and (d) follow by direct calculation, noting that $T(\rho_T - 1) \to c$. ■

Proof of Theorem 3. Consider the terms $T^{-1/2}\hat{\alpha}^{d\prime}Y_t^d$ and $\tilde{\omega}_z$ separately. First, following Phillips and Ouliaris (1990), partition H conformably with Y_t so that

$$H = \begin{bmatrix} h_{11} & H_{12} \\ 0 & H_{22} \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix},$$

where h_{11} is a scalar. Then $T^{-1/2}\hat{\alpha}' Y_{[T\lambda]}^d$ can be written as:

$$\begin{aligned} T^{-\frac{1}{2}}\hat{\alpha}'Y_{[T\lambda]}^d &= [1 \mid -(\textstyle\sum Y_{1t}^d Y_{2t}^{d\prime})(\sum Y_{2t}^d Y_{2t}^{d\prime})^{-1}](T^{-\frac{1}{2}}Y_{[T\lambda]}^d) \\ &=> [1 \mid -(\textstyle\int H_1 V^d V_2^{d\prime} H_{22}')(\int H_{22} V_2^d V_2^{d\prime} H_{22}')^{-1}]\begin{bmatrix} h_{11} & H_{12} \\ 0 & H_{22} \end{bmatrix} V^d(\lambda) \\ &= [h_{11} \mid H_{12} - H_1(\textstyle\int V^d V_2^{d\prime})(\int V_2^d V_2^{d\prime})^{-1}]V^d(\lambda) \\ &= [h_{11} \mid H_{12} - h_{11}(\textstyle\int V_1^d V_2^{d\prime})(\int V_2^d V_2^{d\prime})^{-1} - H_{12}]V^d(\lambda) \\ &= h_{11}\tilde{\alpha}^{d\prime}V^d(\lambda) \end{aligned} \tag{A.1}$$

where $\tilde{\alpha}^{d\prime} = [1 | -(\int V_1^d V_2^{d\prime})(\int V_2^d V_2^{d\prime})^{-1}]$, and, to simplify notation, $\int V_1^d V_2^{d\prime}$ represents $\int_0^1 V_1^d(s)\, V_2^d(s)' ds$, etc.

Next consider $\tilde{\omega}_z$ and let $\alpha^{*d} = [1 | -H_1 \int V^d V_2^{d\prime})(\int V_2^d V_2^{d\prime})^{-1} H_{22}^{-1}]$. Then

$$\begin{aligned} \tilde{\omega}_z &= \sum_{j=-\ell}^{\ell} T^{-1} \sum_{t=j+2}^{T} (\hat{z}_t^d - \tilde{\rho}\hat{z}_{t-1}^d)(\hat{z}_{t-j}^d - \tilde{\rho}\hat{z}_{t-j-1}^d) \\ &= \hat{\alpha}^{d\prime}\left[\sum_{j=-\ell}^{\ell} T^{-1} \sum_{t=j+2}^{T} (Y_t^d - \tilde{\rho}Y_{t-1}^d)(Y_{t-j}^d - \tilde{\rho}Y_{t-j-1}^d)'\right]\hat{\alpha}^d \\ &=> \alpha^{*d\prime}\Omega\alpha^{*d}. \end{aligned}$$

The argument that the term in brackets converges to Ω is made in Phillips and Ouliaris (1990), proof of Theorem 4.1(a). It relies on $\ell \to \infty$, $\ell = o(T^{1/4})$, and $\tilde{\rho} - 1 = o_p(1)$; see Phillips and Ouliaris (1990) for the details. Using Ω = HH′ and (A.1), one obtains:

$$\tilde{\omega}_z => \alpha^{*d\prime}\Omega\alpha^{*d} = \alpha^{*d\prime}HH'\alpha^{*d} = h_{11}^2\tilde{\alpha}^{d\prime}\tilde{\alpha}^d. \tag{A.2}$$

It follows from (A.1) and (A.2) that $(T\tilde{\omega}_z)^{-1/2}\hat{\alpha}^{d\prime}Y_{[T\lambda]}^d => (\tilde{\alpha}^{d\prime}\tilde{\alpha}^d)^{-1/2}\tilde{\alpha}^{d\prime}V^d(\lambda)$.

(a), (b), (c).The proof of (a) is immediate with $Y_t^d = Y_t$, $V^d = W$, etc. To prove (b) and (c) it only needs to be shown that $D_T(\lambda; \hat{\beta})$ is such that (A3) is satisfied. This follows by direct calculation. For example, $T^{-1/2}(Y_{[T\lambda]}^\mu - T^{-1}\sum_{t=1}^T Y_t) => B(\lambda) - \int_0^1 B(s)ds \equiv B^\mu(\lambda)$, where B has covariance matrix Ω

$= 2\pi S_U(0)$; this can be factored as $B^\mu = HV^\mu$, where V^μ is as given in the statement of the theorem. ■

Proof of Theorem 4. (a) Write $v_T^d(\lambda) = (T\hat{\omega})^{-1/2}\{y_{[T\lambda]} - d_{[T\lambda]}(\hat{\beta})\} = (T\hat{\omega})^{-1/2}\{y_{[T\lambda]} - d_{[T\lambda]}(\beta)\} - (T\hat{\omega})^{-1/2}\{D_T(\lambda, \hat{\beta}) - D_T(\lambda, \beta)\}$.

By assumption, $\hat{\omega} \xrightarrow{p} k > 0$ and $T^{-1/2}\{D_T(\cdot, \hat{\beta}) - D_T(\cdot, \beta)\} => 0$. Thus the result follows if $(T\hat{\omega})^{-1/2}w_T^d => 0$, where $w_T^d(\lambda) \equiv y_{[T\lambda]} - d_{[T\lambda]}(\beta) = w_{[T\lambda]}$.

To show that $(T\hat{\omega})^{-1/2}w_T^d => 0$, it is convenient to work in the restriction of $D[0, 1]$ to $C[0, 1]$, with the metric $\rho_C(f_1, f_2) = \sup_{\lambda\in[0,1]}|f_1(\lambda) - f_2(\lambda)|$ (see Hall and Heyde, 1980). Let $w_0 = 0$ and define the interpolation of $w_{[T\lambda]}$ to be $w_T^\dagger(\lambda) = \{w_{[T\lambda]} + (T\lambda - [T\lambda])(w_{[T\lambda]+1} - w_{[T\lambda]})\}/(T\hat{\omega})^{1/2}$ so that $w_T^\dagger(\lambda) \in C[0, 1]$. Note that $\max_{1\le t\le T}|w_t| = \sup_{\lambda\in[0,1]}|w_T^\dagger(\lambda)|$. Then, for all $\epsilon > 0$,

$$\begin{aligned}\Pr[\rho_C(w_T^\dagger, 0) > \epsilon] &= \Pr[\sup_{0\le\lambda\le1}\left|w_T^\dagger(\lambda)\right| > \epsilon]\\ &= \Pr[\max_{1\le t\le T}\left|w_t/(T\hat{\omega})^{\frac{1}{2}}\right| > \epsilon]\\ &\le \mathrm{E}\{\max_{1\le t\le T}\left|w_t\right|/\{(T\hat{\omega})^{\frac{1}{2}}\epsilon\}\}^{2+\delta}\\ &\le T^{-\frac{1}{2}\delta}\epsilon^{-(2+\delta)}T^{-1}\sum_{t=1}^{T}E\{\left|w_t\right|^{2+\delta}\hat{\omega}^{-(1+\frac{1}{2}\delta)}\},\end{aligned}$$

which tends to zero under the assumptions $\hat{\omega} \xrightarrow{p} k \ne 0$ and $\sup_t E|w_t|^{2+\delta} < \infty$. Thus $w_T^\dagger => 0$ so $v_T^d => 0$. Because g is continuous at 0, $g(v_T^d) => g(0)$, so in particular, by the definition of G^d, $\Pr[g(v_T^d) < c_\alpha] \to \Pr[g(0) < c_\alpha] = 1$ for all α.

(b) Write $D_T(\lambda, \hat{\beta}) - D_T(\lambda, \beta) = \tau_T(\lambda)'(\hat{\beta} - \beta)$. By assumption, $\tau_T \to \tau$, where $\sup_{\lambda\in[0,1]}|\tau_i(\lambda)| \le \bar{\tau}$, $i = 1, \ldots, J$. Thus $D_T(\lambda, \hat{\beta}) - D_T(\lambda, \beta) => 0$ if $\hat{\beta} - \beta \xrightarrow{p} 0$. But this follows from noting first that $T^{-1}\sum x_t x_t' \to M$, where M is nonsingular by assumption; and second that:

$$\mathrm{E}T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}x_t x_s' w_t w_s \le \bar{\tau}^2 T^{-1}\sum_{u=-\infty}^{\infty}\left|\mathrm{cov}(w_t, w_{t-u})\right| < \infty.$$

With this result and Chebyshev's inequality, $\hat{\beta} - \beta \xrightarrow{p} 0$, so $D_T(\lambda, \hat{\beta}) - D_T(\lambda, \beta) => 0$, so condition (i) in the statement of the theorem is satisfied. ■

Proof of Lemma 1. Under (5.4), y_t has the autoregressive representation, $d(L)y_t = \beta_0 d(1) + \epsilon_t$, or $y_t = \beta_0 d(1) + \delta(L)y_{t-1} + \epsilon_t$. Thus:

$$\Delta y_t = \beta_0 d(1) + (\delta(1) - 1)y_{t-1} + \delta^*(L)\Delta y_{t-1} + \epsilon_t, \tag{A.3}$$

where $d_0 = 1$, $\delta(L) = L^{-1}(1 - d(L))$ and $\delta_j^* = -\sum_{i=j+1}^{\infty}\delta_i$. Thus (A.3) is the population regression, the parameters of which are estimated by the sequence of pth order autoregressions (5.3). The stated assumptions satisfy the conditions of Berk's (1974) Theorem 1 for the ℓ_1-consistency of $\hat{\delta}(z)$. Thus the OLS estimator of a linear combination of these parameters will be consistent. Thus $\hat{\omega} \xrightarrow{p} \sigma_\epsilon^2/(1 - \delta^*(1))^2$. By direct calculation, $\delta^*(1) = \sum_{j=0}^{\infty}(-\sum_{i=j+1}^{\infty}\delta_i) = 1$

$-\sum_{j=0}^{\infty}(1-j)d_j$. Thus $\hat{\omega} \xrightarrow{p} \sigma_\epsilon^2/[\sum_{j=0}^{\infty}(j-1)d_j]^2 \equiv \kappa$, where $\sum_{j=0}^{\infty}(j-1)d_j$ is nonzero by assumption. ■

Proof of Theorem 5. (a) $(T^{½r}/\ln T)g_1(v_T) = T^{-1}\sum_{t=1}^{T}T^{½r}|(T\hat{\omega})^{-½}y_t|^r/\ln T = \hat{\omega}^{-½r}(T^{-1}\sum_{t=1}^{T}|y_t|^r)/\ln T$.

By Lemma 1, $\hat{\omega}^{½r} \xrightarrow{p} \kappa^{½r}$. By Chebyschev's inequality, it therefore suffices to show that $E(T^{-1}\sum_{t=1}^{T}|y_t|^r/\ln T)^2 \to 0$. Now:

$$E(T^{-1}\sum_{t=1}^{T}|y_t|^r/\ln T)^2 = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}E|\beta_0+w_t|^r|\beta_0+w_s|^r/(\ln T)^2,$$

which converges to zero if $E|w_tw_s|^r \le c < \infty$ for all t, s. Now:

$$\begin{aligned}E|w_tw_s|^r &\le Ew_t^{2r} = E(b(L)\epsilon_t)^{2r}\\ &\le \sum_{j_1=0}^{\infty}\cdots\sum_{j_{2r}=0}^{\infty}|b_{j_1}|\cdots|b_{j_{2r}}|E(|\epsilon_{t-j_1}|\cdots|\epsilon_{t-j_{2r}}|)\\ &\le (\sum_{j=0}^{\infty}|b_j|)^{2r}E(\epsilon_t^{2r}),\end{aligned}$$

for all t, whence the result follows.

(b) $\Pr[g_2(v_T) + (½-\delta)\ln T > c] = \Pr[T^{-1}\sum_{t=1}^{T}\ln|y_t| + ½\ln\hat{\omega} > c + \delta\ln T] \le E\{T^{-1}\sum_{t=1}^{T}\ln|y_t| + ½\ln\hat{\omega}\}^2/(c+\delta\ln T)^2$.

Because $\ln\hat{\omega} \xrightarrow{p} \ln\kappa$, the final expression tends to 0 for any fixed c, $|c|<\infty$, if $E(T^{-1}\sum_{t=1}^{T}\ln|y_t|)^2$ is bounded. But $E(T^{-1}\sum_{t=1}^{T}\ln|y_t|)^2 \le E(\ln|y_t|)^2$, which is finite by assumption, so $\Pr[g_2(v_T)+(½-\delta)\ln T > c] \to 0$ for all $\delta > 0$, $|c|<\infty$.

(c) $Tg_3(v_T) = T\{T^{-1}\sum(t/T)^r y_t/[T\hat{\omega}]^{½}\}^2 = \{T^{-1}\sum(t/T)y_t\}^2/\hat{\omega}$. The result follows from $T^{-1}\sum(t/T)^r y_t \xrightarrow{p} \beta_0/(1+r)$.

(d)
$$\begin{aligned}g_{RS}(v_T) &= (T\hat{\omega})^{-½}\{\max_{1\le t\le T}(\beta_0+w_t) - \min_{1\le t\le T}(\beta_0+w_t)\}\\ &\le 2(T\hat{\omega})^{-½}\max_{1\le t\le T}|w_t|\\ &\le 2(T\hat{\omega})^{-½}\max_{1\le t\le T}(\sum_{j=0}^{t-1}|b_j||\epsilon_{t-j}|)\\ &\le (2\hat{\omega}^{-½}\sum_{j=0}^{\infty}|b_j|)T^{-½}\max_{1\le t\le T}|\epsilon_t|.\end{aligned}$$

Let $\hat{q} \equiv 2\hat{\omega}^{-½}\sum_{j=0}^{\infty}|b_j|$ and $|\epsilon|_{(T)} \equiv \max_{1\le t\le T}|\epsilon_t|$. Then:

$$\begin{aligned}a_TT^{\frac{1}{2}}/\max(|b_T|,1))g_{RS}(v_T) &\le \hat{q}(a_T/\max(|b_T|,1))\{(b_T+a_T|\epsilon|_{(T)})/a_T - b_T/a_T\}\\ &= \hat{q}(b_T+a_T|\epsilon|_{(T)})/\max(|b_T|,1) - \hat{q}b_T/\max(|b_T|,1) \equiv \Psi_T.\end{aligned}$$

If $\max(|b_T|,1)\to\infty$, then $\Psi_T \xrightarrow{p} 2\sum_{j=0}^{\infty}|b_j|/\kappa = O(1)$. If $\max(|b_T|,1)\to 1$, then $\Psi_T => (2\sum_{j=0}^{\infty}|b_j|/\kappa)(\epsilon^* + \lim_{T\to\infty}b_T) = O_p(1)$. In either case, $(a_TT^{½}/\max(|b_T|,1))g_{RS}(v_T) = O_p(1)$. ■

Appendix B: Parameter Values for Univariate Empirical Models

Table B.1. Parameter values for autoregressive models
$$\Delta y_t = \mu + \sum_{j=1}^{6} a_j \Delta y_{t-j} + \epsilon_t,\ \mathrm{var}(\epsilon_t) = \sigma^2$$

	FM2D82	IVT82	INC	FYGM3	GMYXP8
μ	.00258	.00014	.01161	.03745	.00604
a_1	.67637	.20533	.23772	.38856	.41444
a_2	−.08429	.07999	.17634	−.50840	−.01607
a_3	.15530	−.09785	−.23341	.42858	.19326
a_4	−.12098	−.15694	−.04855	−.28511	−.09893
a_5	.13364	−.18534	.00143	.28979	−.14897
a_6	−.13687	−.02694	−.02781	−.20488	−.05931
σ^2	.00871	.02996	.03466	.72906	.00810

Table B.2. Parameter values for moving average models
$$y_t = \rho y_{t-1} + \sum_{j=1}^{5} b_j \epsilon_{t-j},\ \mathrm{var}(\epsilon_t) = \sigma^2$$

	FM2D82	IVT82	INC	FYGM3	GMYXP8
μ	−.01259	−.00119	−.00403	−.06702	.00293
ρ	.94752	.70443	.83050	.82045	.91451
b_1	.69039	.44577	.35072	.51793	.44412
b_2	.38310	.42911	.37087	−.28322	.22312
b_3	.32080	.13251	−.00319	.23636	.34864
b_4	.16604	.03387	−.03071	.36649	.28337
b_5	.11714	−.17188	−.00009	−.08113	.12500
σ^2	.00870	.02897	.03373	.70532	.00804

Notes: The mnemonics refer to the Citibase name of the series; the series and transformations are discussed in the text. The moving average parameters were estimated by maximum likelihood using detrended series.

Notes

1. My personal introduction to these issues came when John Geweke asked me to discuss some paper by Granger on modeling time series in levels (Granger, 1983) at the winter meetings of the Econometric Society in December 1983. I was a first year assistant professor and this was my first opportunity to be a discussant. In that discussion I showed that if two variables are cointegrated of order (1,1) in Granger's (1983) terminology, if they have a bivariate finite order moving average representation in first differences, and if there are sufficiently many moments, then the estimator of the cointegrating coefficients will be consistent at rate T. Professor Granger approached me afterward and encouraged me to write up my notes. The result was eventually published (Stock, 1987), and I have always been grateful for his intellectual generosity, encouragement, and support.
2. Alternative conditions under which Assumption 1(a) is satisfied are provided by Herrndorf (1984) (also see Phillips, 1987; Ethier and Kurtz, 1986).

3. To see this, let $f_1(\lambda) = \frac{1}{2}\lambda\delta$, $f_2(\lambda) = \frac{1}{2}\delta$, so $f_1, f_2 \in C[0, 1]$. Under the sup norm, $\rho_C(f_1, f_2) = \sup_{\lambda\in[0,1]}|f_1(\lambda) - f_2(\lambda)|$, both f_1 and f_2 are in a δ-neighborhood of 0. But $g(f_1) = 3$ and $g(f_2) = \infty$, so there is no open δ-neighborhood of 0 that maps into an arbitrarily small ϵ-neighborhood.
4. The first differences were used to produce null models that would err on the side of inducing substantial power at high frequencies, a situation in which unit root tests have performed poorly (see Schwert, 1989). The same AR(6) approximation was used for all series in part because of its theoretical justification (Berk, 1974), in part because the objective here is not to develop optimal forecasting models of these particular series but rather to have a conveniently parameterized set of time series models with representative spectra.

References

Berk, K. N. (1974), "Consistent Autoregressive Spectral Estimates." *Annals of Statistics*, 2: 489–502.

Bhargava, A. (1986), "On the Theory of Testing for Unit Roots in Observed Time Series." *Review of Economic Studies*, 53: 369–384.

Cavanagh, C. L. (1985), "Roots Local to Unity." Mimeograph, Harvard University.

Chan, N. H. and C. Z. Wei (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes." *Annals of Statistics*, 15: 1050–63.

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

Christiano, L. C. and M. Eichenbaum (1989), "Unit Roots in GNP: Do We Know and Do We Care?" *Carnegie–Rochester Conference Series on Public Policy.*

Davidson, J. E. H., D. F. Hendry, F. Srba, and S. Yeo (1978), "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumer's Expenditure and Income in the United Kingdom." *Economic Journal*, 86: 661–92.

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: 427–31.

Engle, R. F. and C. W. J. Granger (1987), "Dynamic Model Specification with Equilibrium Constraints: Co-Integration and Error-Correction." *Econometrica*, 55: 251–76.

Engle, R. F., and B. S. Yoo (1987), "Forecasting and Testing in Co-integrated Systems." *Journal of Econometrics*, 35: 143–59.

Ethier, S. N. and T. G. Kurtz (1986), *Markov Processes.* New York: Wiley.

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

Granger, C. W. J. (1966), "The Typical Spectral Shape of an Economic Variable," *Econometrica*, 34: 150–61.

Granger, C. W. J. (1983), "Co-integrated Variables and Error Correcting Models," UCSD Discussion Paper number 83-13a.

Granger, C. W. J. (1986), "Developments in the Study of Co-integrated Economic Variables." *Oxford Bulletin of Economics and Statistics*, 48: 213–28.

Granger, C. W. J. and P. Newbold (1974), "Spurious Regressions in Econometrics." *Journal of Econometrics*, 2: 111–20.

Granger, C. W. J. and A. Weiss (1983), "Time Series Analysis of Error-Correction Models." In S. Karlin, T. Amemiya, and L. A. Goodman (eds.), *Studies in Econo-*

metrics, Time Series and Multivariate Statistics, in honor of T. W. Anderson. Academic Press.

Hall, P. and C. C. Heyde (1980), *Martingale Limit Theory and its Applications.* New York: Academic Press.

Herrndorf, N. (1984), "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables." *Annals of Probability*, 12: 141–53.

Kahn, J. A. and M. Ogaki (1988), "A Chi-Square Test for a Unit Root." University of Rochester, Discussion Paper no. 212.

Kosobud, R. and L. Klein (1961), "Some Econometrics of Growth: Great Ratios of Economics." *Quarterly Journal of Economics* 25 (May): 173–98.

Lo, A. (1991), "Long Term Memory in Stock Market Prices." *Econometrica*, 59: 1279–1314.

Mandelbrot, B. (1975), "Limit Theorems on the Self-Normalized Range for Weakly and Strongly Dependent Processes." *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 31: 271–85.

Nelson, C. and C. Plosser (1982), "Trends and Random Walks in Macroeconomic Time Series." *Journal of Monetary Economics*, 10: 139–62.

Park, J. Y. (1990), "Testing for Unit Roots and Cointegration by Variable Addition." In G. F. Rhodes and T. B. Fomby (eds.), *Advances in Econometrics: Co-Integration, Spurious Regressions and Unit Roots.* Greenwich, CT: JAI Press.

Park, J. Y. and B. Choi (1988), "A New Approach to Testing for a Unit Root." CAE Working Paper no. 88–23, Cornell University.

Perron, P. (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis." *Econometrica*, 57: 1361–1402.

Phillips, P. C. B. (1987), "Time Series Regression with a Unit Root." *Econometrica*, 55: 277–302.

Phillips, P. C. B. and S. Ouliaris (1990), "Asymptotic Properties of Residual-Based Tests for Cointegration." *Econometrica*, 58: 165–94.

Phillips, P. C. B. and P. Perron (1988), "Testing for a Unit Root in Time Series Regression." *Biometrika*, 75: 335–46.

Reiss, R. D. (1989), *Approximate Distributions of Order Statistics.* New York: Springer-Verlag.

Said, S. E. and D. A. Dickey (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order." *Biometrika*, 71: 599–607.

Sargan, J. D. and A. Bhargava (1983), "Testing for Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk." *Econometrica*, 51: 153–74.

Schwert, W. G. (1989), "Tests for Unit Roots: A Monte Carlo Investigation." *Journal of Business and Economic Statistics*, 7: 147–59.

Stock, J. H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors." *Econometrica*, 55: 1035–56.

Stock, J. H. and M. W. Watson (1988), "Testing for Common Trends." *Journal of the American Statistical Association*, 83: 1097–1107.

Stock, J. H. and M. W. Watson (1990). "Business Cycle Properties of Selected U.S. Economic Time Series." Mimeograph, Harvard University.

Yule, G. U. (1926), "Why Do We Sometimes Get Nonsense Correlations Between Time Series?" *Journal of the Royal Statistical Society B*, 89: 1–64.