ELSEVIER Journal of Econometrics 63 (1994) 105–131

JOURNAL OF Econometrics

Deciding between I(1) and I(0)

James H. Stock

Kennedy School of Government, Harvard University, Cambridge, MA 02138, USA

(Received October 1993)

Abstract

A class of procedures that consistently classify the stochastic component of a time series as being integrated either of order zero [I(0)] or one [I(1)] are proposed for general I(0) or I(1) processes and polynomial or piecewise linear detrending. Large-sample Bayesian inference is free of nuisance parameters describing short-run dynamics and requires specifying priors only on the point hypotheses 'I(0)' and 'I(1)' thereby avoiding problematic choices of parametric priors over roots and nuisance parameters. Applied to the Nelson–Plosser (1982) data with linear detrending, these procedures largely support Nelson and Plosser's original inferences. With piecewise-linear detrending these data are typically uninformative, producing Bayes ratios close to one.

Key words: Unit roots; Integration; Model selection
JEL classification: C22; C11

1. Introduction

It is often desirable to know whether a univariate economic time series is better modeled as being integrated of order one [is I(1)] or of order zero [I(0)]. This distinction can be important when the series in question is subsequently used in multivariate analysis, such as Granger causality tests or cointegration

The author thanks T.W. Anderson, Clive Granger, Lars Hansen, Herman van Dijk, Emanual Parzen, Mark Watson, two anonymous referees, and especially Pierre Perron for helpful discussions on this research. He is also grateful to the seminar participants at the University of Minnesota Time Series Conference, June 1990, and the World Congress of the Econometric Society, August 1990, for comments on an earlier draft of this paper (without Sections 3.2 and 5). Graham Elliott provided excellent research assistance. This research was supported by National Science Foundation grants SES-89-10601 and SES-91-22463 and by the Sloan Foundation.

SSDI 0 3 0 4 4 0 7 6 9 3 0 1 5 6 2 Z

modeling. Alternatively, different economic theories can have different implications for the long-run properties of certain series, in which case it might be desirable to assess the probability that one or the other of these implications is true (see the discussion in Christiano and Eichenbaum, 1990). These two applications suggest the practical value of procedures that produce explicit posterior odds on whether the process that generated the observed data is I(1) or I(0). Such a procedure would permit the formal application of statistical decision theory to the I(1)/I(0) classification problem, with a loss function defined solely in terms of whether the series is I(1) or I(0).

This desire to compute posterior odds for the I(1) and I(0) hypotheses has led to considerable recent empirical and theoretical work on unit roots in economic time series from a Bayesian perspective; see DeJong and Whiteman (1991), Sims (1988), Sims and Uhlig (1991), Sowell (1991), Phillips (1991a), and the discussions of Phillips (1991a) in the October–December 1991 issue of the *Journal of Applied Econometrics*.[1] However, this work relies on finite-dimensional parameterizations of the I(1) and I(0) hypotheses, which in turn requires formulating explicit priors over the values of these parameters. Although various authors have described their priors as flat or uninformative, different 'uninformative' priors yield different inferences, and in any event the geometry of I(1) and I(0) processes is sufficiently complicated that any prior restrictions on parametric approximations are at best difficult to interpret.

This paper proposes a class of Bayesian procedures for deciding whether a process is I(1) or I(0) which avoids the problem of making explicit parametric assumptions about priors within the I(1) or I(0) models. These procedures are developed for two specifications of deterministic trends. The first is general polynomial trends, estimated by OLS, as have been considered elsewhere in the unit roots literature [see, for example, Ouliaris, Park, and Phillips, 1989; Perron, 1991]. Perron (1989) and Rappoport and Reichlin (1989) suggested an alternative to the unit root model for many aggregate economic time series in which the series are stationary around a time trend with a growth rate that changes once during the sample. Although the empirical support for this model is weakened when the break date is treated as unknown (Banerjee, Lumsdaine, and Stock, 1992; Zivot and Andrews, 1992), their results are sufficiently strong to suggest that this model be treated as a plausible specification. The second trend specification therefore is the broken trend model in which the break date is unknown.

The proposed procedures are based on a scaled cumulative sum process of the detrended data, V_T. The process V_T has the following properties: (i) $N_{iT}^{-1/2} V_T$ has a (classical) asymptotic distribution that depends on no unknown nuisance

[1] Since the original draft of this paper was written, three additional closely related papers have appeared, by Kwiatkowski, Phillips, Schmidt, and Shin (1992), Phillips and Ploberger (1992), and Perron (1991); these are discussed below.

parameters, and (ii) for I(0), $N_{iT} = 1$, while for I(1), $N_{iT} \equiv N_T \to \infty$. The first of these properties means that it is possible to compute the asymptotic likelihoods – and thus the likelihood ratio – of observing functionals of V_T alternatively under the I(1) and I(0) hypotheses, and moreover, that in large samples this evaluation does not depend on the nuisance parameters. The second property implies that these asymptotic distributions diverge, so that in large samples the likelihood ratio will tend either to zero or to infinity, thereby providing a consistent procedure for distinguishing I(1) from I(0) processes. The statistics based on the cumulative sum V_T are closely related to statistics studied in three related literatures: tests for structural breaks, LM tests for a random walk drift under the null of stationarity, and recently developed tests of the I(0) null against the I(1) alternative. These links are discussed, and references given, in Section 2.

Because the procedures consistently classify a process as I(0) or I(1), they constitute model selection procedures where the 'model' is broadly interpreted to be one of these two classes. Viewed thus, a variety of simpler approaches already exist, at least in principle, although they are not used in practice. For example, comparing the Dickey–Fuller (1979) t-statistic to a sequence of critical values which tends to $-\infty$ produces a consistent classifier. Similarly, using an existing test of the general I(0) null such as Park's (1990) variable addition tests, or the tests of Bierens (1989) and Bierens and Guo (1993), with an appropriate sequence of critical values will result in consistent classification. However, without further modification these approaches leave unsolved the key practical problem of choosing the sequence of critical values. The procedure here thus can be interpreted from a classical perspective as specifying a data-dependent sequence of critical values, determined up to the choice of point priors. However, from the Bayesian perspective the procedure has the three additional advantages of producing posterior odds ratios, of permitting priors to be specified on the points I(0) and I(1) rather than parametrically, and (unlike parametric Bayesian treatments of the unit roots problem) of producing large-sample inferences which do not depend on the short-run nuisance parameters.

In a paper closely related to this one, Phillips and Ploberger (1991) have also recently used Bayesian posterior odds ratios to construct an I(1)/I(0) decision rule. Like the one proposed here, their procedure permits placing priors only on the I(1) and I(0) 'point' hypotheses. One difference between the Phillips–Ploberger (1991) approach and the approach in this paper is that they study posterior distributions of the data directly, while we consider posterior distributions for a family of statistics whose asymptotic distribution does not depend on the nuisance parameters under the I(0) or I(1) hypothesis.

The remainder of the paper is organized as follows. Theoretical results under general conditions on trends and detrending are presented in Section 2. These general conditions are examined for the leading cases of polynomial trends and piecewise linear trends, detrended by ordinary least squares (OLS), in Section 3.

Section 4 presents Monte Carlo results. Empirical results are given in Section 5, and conclusions are presented in Section 6.

2. The proposed decision rules

2.1. General results

We consider time series which are the sum of a purely deterministic trend, d_t, and a stochastic term, u_t,

$$y_t = d_t + u_t. \tag{1}$$

The I(0) and I(1) hypotheses refer to the order of integration of the stochastic element u_t. Here, the definitions of I(0) and I(1) follow recent convention: a purely stochastic process is said to be I(d) if the process formed from the partial sums of its dth difference, scaled by $T^{-1/2}$, obeys a functional central limit theorem and converges to a constant times a standard Brownian motion. Let $U_{0T}(\lambda) = T^{-1/2}\sum_{s=1}^{[T\lambda]} u_s$ and $U_{1T}(\lambda) = T^{-1/2} u_{[T\lambda]}$, where $[\cdot]$ denotes the greatest lesser integer function, and let $\gamma_x(j) = \mathrm{cov}(x_t, x_{t-j})$ for a second-order stationary process x_t. Also let '$\Rightarrow$' denote weak convergence of random elements of $D[0,1]$ and let $W(\cdot)$ denote a standard Brownian motion process restricted to the unit interval. The I(0) and I(1) hypotheses are given by

$$\text{I(0)}: \quad U_{0T} \Rightarrow \omega_0 W, \qquad \omega_0^2 = \sum_{j=-\infty}^{\infty} \gamma_u(j), \quad 0 < \omega_0 < \infty, \tag{2}$$

$$\text{I(1)}: \quad U_{1T} \Rightarrow \omega_1 W, \qquad \omega_1^2 = \sum_{j=-\infty}^{\infty} \gamma_{\Delta u}(j), \quad 0 < \omega_1 < \infty. \tag{3}$$

Throughout it is assumed that second moments of I(0) random variables exist and that standard estimators of second moments are consistent; in particular, it is assumed that $T^{-1}\sum_{t=|j|+1}^{T} u_t u_{t-j} \xrightarrow{p} \gamma_u(j)$ for u_t I(0) and $T^{-1}\sum_{t=|j|+1}^{T} \Delta u_t \Delta u_{t-j} \xrightarrow{p} \gamma_{\Delta u}(j)$ for u_t I(1), for all fixed j. In stating $\gamma_u(j) = \mathrm{cov}(u_t, u_{t-j})$ [I(0) case] or $\gamma_{\Delta u}(j) = \mathrm{cov}(\Delta u_t, \Delta u_{t-j})$ [I(1) case], we are further assuming that I(0) variables are second-order stationary.

The proposed statistics are based on the scaled partial sum process of the detrended data. Let $\hat{d}_t$ denote the estimate of the trend component and let the detrended process be $y_t^d = y_t - \hat{d}_t$. The decision rules studied are all functionals of the statistic,

$$V_T(\lambda) = \hat{\omega}^{-1} T^{-1/2} \sum_{s=1}^{[T\lambda]} y_s^d, \tag{4}$$

where

$$\hat{\omega}^2 = \sum_{m=-l_T}^{l_T} k(m/l_T)\hat{\gamma}_{yd}(|m|), \qquad \hat{\gamma}_{yd}(m) = T^{-1}\sum_{t=m+1}^{T} y_t^d y_{t-m}^d,$$

where l_T is an increasing sequence of integers and $k(\cdot)$ is a kernel satisfying $k(x) = 0$ for $|x| \geqslant 1$, $k(x) = k(-x)$, $0 < k(x) \leqslant 1$ for $|x| < 1$, $k(0) = 1$, and $l^{-1}\sum_{u=1}^{l} k(u/l) \geqslant k$ for all $l > 1$, where $k > 0$ (see, for example, Andrews, 1991). It is assumed that the sequence l_T is such that $\hat{\omega}^2$ is consistent for the spectral density of u_t under the I(0) hypothesis. With the transformation (4), the task of distinguishing I(0) from I(1) processes is shifted to distinguishing the cumulation of an I(0) process, now I(1), from the cumulation of an I(1) process, now I(2).

The results are stated under general assumptions on the trend estimation error $\delta_t = \hat{d}_t - d_t$. Let $\|x_t\| = T^{-1}\sum_{t=1}^{T} x_t^2$ for a time series x_t, $D_{0T}(\lambda) = T^{-1/2}\sum_{s=1}^{[T\lambda]}\delta_s$, and $D_{1T}(\lambda) = T^{-1/2}\delta_{[T\lambda]}$. The estimated trend is assumed to satisfy the following conditions:

Detrending Conditions

(A) If u_t is I(0), then:
(i) $(U_{0T}, D_{0T}) \Rightarrow \omega_0(W, D_0)$ where $D_0 \in C[0,1]$.
(ii) $l_T^2\,\|\delta_t\| \to^p 0$.

(B) If u_t *is* I(1), then:
(i) $(U_{1T}, D_{1T}) \Rightarrow \omega_1(W, D_1)$ where $D_1 \in C(0,1]$.
(ii) $\|\Delta\delta_t\| = O_p(1)$.

Specific examples of trends that satisfy these conditions are given in the next section. Because $y_t^d = y_t - \hat{d}_t = u_t - \delta_t$, these conditions lead to general definitions of limiting detrended processes. Let $Y_{0T}^d(\lambda) = T^{-1/2}\sum_{s=1}^{[T\lambda]} y_s^d = U_{0T}(\lambda) - D_{0T}(\lambda)$; for an I(0) process, it follows from condition A(i) that $Y_{0T}^d(\cdot) \Rightarrow \omega_0 W_0^d(\cdot)$, where $W_0^d(\cdot) = W(\cdot) - D_0(\cdot)$. Similarly, let $Y_{1T}^d(\lambda) = T^{-1/2}y_{[T\lambda]}^d = U_{1T}(\lambda) - D_{1T}(\lambda)$; condition B(i) implies that, if u_t is I(1), then $Y_{1T}^d(\cdot) \Rightarrow \omega_1 W_1^d(\cdot)$, where $W_1^d(\cdot) = W(\cdot) - D_1(\cdot)$.

Under these conditions, because $l_T \to \infty$, A(ii) implies that if u_t is I(0), then the estimated trend is consistent (in the L_2 norm). However, if u_t is I(1), $D_{1T}(\cdot) = T^{-1/2}\delta_{[T\cdot]}$ is $O_p(1)$, so the estimated trend is not consistent. In the specifications studied in Section 3, however, $\|\Delta\delta_t\| \to^p 0$ in the I(1) case, so that the first difference of the estimated trend is consistent for the first difference of the true trend.

Limiting representations for the statistic $T^{-1}\sum_{t=1}^{T} V_T(t/T)^2$ under the general I(0) and I(1) hypotheses have been obtained by Kwiatkowski, Phillips, Schmidt, and Shin (1992) (for OLS detrending with a constant or linear time trend) and by Perron (1991) (for general polynomial trends estimated by OLS). Their applications differed from ours, however, respectively testing I(0) vs. I(1) and testing for breaks in deterministic trends. The following theorem extends their results to general trends and provides limiting representations for the statistic $V_T \in D[0,1]$. Let $N_T = T/\sum_{m=-l_T}^{l_T} k(m/l_T)$.

Theorem 1. Suppose $l_T^2 \ln T/T \to 0$, $l_T \to \infty$, and conditions A and B hold.

(a) *If y_t is I(0), then $V_T \Rightarrow W_0^d$.*
(b) *If y_t is I(1), then $N_T^{-1/2} V_T \Rightarrow V_1^d$, where $V_1^d(\lambda) = \int_0^\lambda W_1^d(s)\mathrm{d}s/\{\int_0^1 W_1^d(s)^2 \mathrm{d}s\}^{1/2}$.*

Proofs of theorems are given in the Appendix.

For the detrending procedures studied in Section 3 which satisfy conditions A and B, the distributions of W_0^d and W_1^d depend on the type of detrending but typically do not depend on any unknown parameters [the exception, discussed in detail in the next section, is broken-trend detrending under the I(0) case with an unknown date]. Moreover, V_T has different rates of convergence depending on whether u_t is I(0) or I(1). Thus V_T can be used as the basis of an asymptotic decision rule for categorizing u_t as I(0) or I(1).

2.2. Decision rules based on scalar functionals of V_T

The statistical decision rules considered here are based on scalar functionals of V_T. In particular, we consider functionals $\phi(\cdot)$ that have the properties: (i) ϕ is a continuous mapping from $C[0,1] \to \mathscr{R}^1$; (ii) $\phi(ag) = \phi(g) + 2\ln a$, where a is a scalar and $g \in D[0,1]$;[2] and (iii) $\phi(W_0^d)$ and $\phi(V_1^d)$ respectively have continuous densities f_0 and f_1 with support $(-\infty, \infty)$. Let $\phi_T = \phi(V_T)$. Then the asymptotic approximation to the likelihood ratio (or Bayes factor) B_T of ϕ_T under the I(1) hypothesis, relative to the I(0) hypothesis, is

$$B_T = f_1(\phi_T - \ln N_T)/f_0(\phi_T). \tag{5}$$

It is readily seen that (5) provides a consistent rule for classifying u_t as I(1) or I(0). If the I(0) hypothesis is true, then by the continuous mapping theorem $\phi_T = \phi(V_T) \Rightarrow \phi(W_0^d) = \mathrm{O}_p(1)$, so $f_0(\phi_T) = \mathrm{O}_p(1)$ but $f_1(\phi_T - \ln N_t) \to^p 0$; thus $B_T \to^p 0$ and I(0) is chosen with probability one. On the other hand, if u_t is I(1),

---

[2] The additive property (ii) is achieved in practice by taking logarithmic transformations of a functional $\tilde{\phi}(V_T)$, for which $\tilde{\phi}(ag) = a^2\tilde{\phi}(g)$. Whether ϕ or $\tilde{\phi}$ is used has no theoretical significance. The choice of transformation is instead driven by our computational experience that taking logarithms enhances the numerical stability of the calculations in Section 4.

then $\phi_T - \ln N_T \Rightarrow \phi(V_1^d) = O_p(1)$, but $f_0(\phi_T) \to^p 0$; thus $1/B_T \to^p 0$ and I(1) is chosen with probability one.

Although the focus here is consistent classification rules, we note in passing that the statistics $\phi(V_T)$ can be used to perform classical tests of the I(0) or I(1) null hypotheses. In particular, $\phi(N_T^{-1/2} V_T)$ can be used to test the null hypothesis that u_t is I(1) against the alternative that it is I(0). Critical values are obtained from the density f_1, and consistency of the test follows from the different rate of convergence under the I(0) alternative. Alternatively, $\phi(V_T)$ could be used to construct a consistent test of the I(0) null against the I(1) alternative; see Park (1990), Park and Choi (1988), Saikkonen and Luukkonen (1993), and Kwiatkowski, Phillips, Schmidt, and Shin (1992).

The likelihood ratio (5) permits performing Bayesian inference when priors are specified only on the point hypotheses I(0) and I(1). Let these priors respectively be π_0 and π_1 (so that $\pi_0 + \pi_1 = 1$). Then the posterior odds ratio is the product of these priors and the Bayes factor (5),

$$\Pi_T = (\pi_1/\pi_0) B_T. \tag{6}$$

The consistency of decision rules based on B_T implies that decision rules based on the posterior odds ratio also are consistent.

In the Monte Carlo investigation and empirical analysis of Sections 4 and 5, we will consider three specific functionals ϕ:

$$\phi_1(g) = \ln\left\{\int_0^1 g(s)^2 \mathrm{d}s\right\}, \tag{7a}$$

$$\phi_2(g) = \ln\left\{\left(\sup_{s\in(0,1)} g(s) - \inf_{s\in(0,1)} g(s)\right)^2\right\}, \tag{7b}$$

$$\phi_3(g) = \ln\left\{\sum_{j=1}^{J} \left|\int_0^1 g(s)\mathrm{e}^{-i2\pi js}\mathrm{d}s\right|^2\right\}. \tag{7c}$$

The statistic ϕ_{1T} and close variants have been studied in several related literatures. In terms of the original data, $\phi_{1T} = \phi_1(V_T) = \ln\{\hat{\omega}^{-2}T^{-1} \times \sum_{t=1}^T (T^{-1/2}\sum_{s=1}^t y_s^d)^2\}$. One motivation for using ϕ_{1T} comes from recognizing that, with no trend or detrending, $\tilde{\phi}_{1T} = T^{-1}\sum_{t=1}^T (T^{-1/2}\sum_{s=1}^t u_s)^2/\hat{\gamma}_u(0)$ (which is appropriate if $d_t = 0$ and u_t is serially uncorrelated) is the Sargan–Bhargava (1983)[3] statistic testing the null that $x_t = \sum_{s=1}^t u_s$ has a unit root, which in turn is motivated as being the Durbin–Watson (1950) ratio for the

[3] Also see Bhargava (1986).

Gaussian random walk. If the rejection region is the right tail, $\tilde{\phi}_{1T}$ accordingly can be interpreted as testing the null that u_t is I(0) against the I(1) alternative. The statistic $\tilde{\phi}_{1T}$ *has also been studied by Nabeya and Tanaka* (1988) (to test for random coefficients) and by Saikkonen and Luukkonen (1993) (to test for a unit MA root). Saikkonen and Luukkonen (1993) proposed a generalization of $\tilde{\phi}_{1T}$ to test I(0) vs. I(1) with ARMA errors, although their generalization differs from that examined here. Kwiatkowski, Phillips, Schmidt, and Shin (1992) proposed $\exp(\tilde{\phi}_{1T})$ as a test of the general I(0) null against the I(1) alternative, and the statistic is closely related to Park's (1990) variable addition test. Gardner (1969) and MacNeill (1978) studied $\tilde{\phi}_{1T}$ as a test for a broken time trend (also see Brown, Durbin, and Evans, 1975); Perron (1991) extended their statistic to general error terms and proposed $\exp(\tilde{\phi}_{1T})$. Both Kwiatkowski, Phillips, Schmidt, and Shin (1992) and Perron (1991) derived asymptotic representations for $\exp(\tilde{\phi}_{1T})$ under the general I(0) and I(1) hypotheses.

The statistic $\tilde{\phi}_{2T}$ is based on the range of the cumulative process, scaled by an estimator of the spectral density of y_t at frequency zero. Scaled by its variance rather than $\hat{\omega}$, this was proposed by Mandelbrot and Van Ness (1968) and Mandelbrot (1975); Lo (1991) studied the generalization (7b) and applied it to financial time series data.

The statistic $\tilde{\phi}_{3T}$ has a somewhat different motivation: if y_t is I(1), then the cumulative process will have more mass in its spectral density at low (but nonzero) frequencies than it will if y_t is I(0). Although the population spectral density of V_T is not well-defined for frequencies near zero, $\tilde{\phi}_{3T}$ nonetheless has a well-behaved asymptotic distribution for fixed integer J.

3. Examples of estimated trends

This section provides specific results for two types of trends and detrending procedures, polynomial time trends detrended by OLS and piecewise linear or broken trends, also with OLS detrending. Both are shown to satisfy the detrending conditions A and B in Section 2.

3.1. Polynomial detrending by OLS

Consider the polynomial time trend,

$$d_t = z_t'\beta, \tag{8}$$

where $z_t = (1, t, t^2, \ldots, t^q)'$ and where the unknown parameters β are estimated by regressing y_t onto z_t to obtain the OLS estimator $\hat{\beta}$ of β. Thus $q = 0$ corresponds to subtracting from y_t its sample mean and $q = 1$ corresponds to linear detrending by OLS. For general q, under (8) the detrended data are $y_t^d = y_t - z_t'\hat{\beta} = u_t - \delta_t$, where $\delta_t = z_t'(\sum_{t=1}^T z_t z_t')^{-1} \sum_{t=1}^T z_t u_t$.

To verify that Theorem 1 applies when this detrending procedure is used, it is sufficient to verify that condition A and B hold. The relevant properties of the detrending process are summarized in Theorem 2.

Theorem 2. If d_t is given by (8) and β is estimated by OLS, then:

(a) *If y_t is I(0), then:*
   (i) $(U_{0T}, D_{0T}) \Rightarrow \omega_0(W, D_0)$, *where* $D_0(\lambda) = v(\lambda)' M^{-1}\Phi$, *where* Φ, M, *and* v *are respectively* $(q+1)\times 1$, $(q+1)\times(q+1)$, *and* $(q+1)\times 1$, *and* $\Phi_i = W(1) - (i-1)\int_0^1 s^{i-2}W(s)\mathrm{d}s$, $M_{ij} = 1/(i+j-1)$, *and* $v_i(\lambda) = \lambda^i/i$.
   (ii) $T\|\delta_t\| \Rightarrow \omega_0^2 \Phi' M^{-1}\Phi$.

(b) *If y_t is I(1) then:*
   (i) $(U_{1T}, D_{1T}) \Rightarrow \omega_1(W, D_1)$, *where* $D_1(\lambda) = \xi(\lambda)' M^{-1}\psi$, *where* $\xi_i(\lambda) = \lambda^{i-1}$ *and* $\psi = \int_0^1 \xi(s) W(s)\mathrm{d}s$.
   (ii) $T\|\Delta\delta_t\| \Rightarrow \omega_1^2 \psi' M^{-1} M^{\dagger} M^{-1}\psi$, *where* $M_{ij}^{\dagger} = (i-1)(j-1)/\max(1, i+j-3)$.

Parts a(i) and b(i) of Theorem 2 have previously been obtained by Ouliaris, Park, and Phillips (1989) and Perron (1991). Theorem 2 implies that conditions A and B hold when y_t is detrended using a polynomial deterministic trend, estimated by OLS. Parts a(i) and b(i) verify conditions A(i) and B(i), respectively. Condition A(ii) follows from part a(ii) under the rate condition in Theorem 1 (that is, $l_T^2 \ln T/T \to 0$ and $T\|\delta_t\| \Rightarrow \omega_0^2 \Phi' M^{-1}\Phi$ implies $l_T^2\|\delta_t\| \xrightarrow{p} 0$). Part b(ii) implies $\|\Delta\delta_t\| \xrightarrow{p} 0$, which verifies condition B(ii).

3.2. *Piecewise-linear ('broken-trend') detrending*

The piecewise-linear trend consists of two connected linear time trends which break at the fraction τ_0 of the sample that corresponds to a break in period $k_0 = [T\tau_0]$. We assume that τ_0 is unknown within a range $\tau_{\min} \leqslant \tau_0 \leqslant \tau_{\max}$. The trend term is

$$d_t(k_0) = \alpha + \beta t + \gamma_T(t - k_0)\mathbf{1}(t > k_0) = z_t(k_0)'\theta, \tag{9}$$

where $z_t(k) = (1, t, (t-k)\mathbf{1}(t > k))'$ and $\theta = (\alpha, \beta, \gamma_T)'$, where $\mathbf{1}(\cdot)$ is the indicator function.

If the break coefficient γ_T is fixed, the break point will be estimated consistently. In practical applications, there is uncertainty about the break date, however, so we adopt a nesting for γ_T under which the rescaled estimator of τ has asymptotic sampling uncertainty. We do this by adopting Picard's (1985) conditions on the local-to-zero break model:

$$T^{1/2}|\gamma_T| \to 0, \qquad T^{3/2}|\gamma_T| \to \infty. \tag{10}$$

The estimated trend is

$$\hat{d}_t(\hat{k}) = z_t(\hat{k})'\,\hat{\theta}(\hat{k}), \tag{11}$$

where

$$\hat{\theta}(k) = \left(\sum_{t=1}^{T} z_t(k)z_t(k)'\right)^{-1} \sum_{t=1}^{T} z_t(k)\,y_t, \tag{12}$$

and where $\hat{k}$ is the value of k that minimizes the sum of squared residuals, that is, $\hat{k}$ solves

$$\min_{k=k_{\min},\ldots,k_{\max}} \sum_{t=1}^{T} \hat{u}_t(k)^2, \tag{13}$$

where $\hat{u}_t(k) = y_t - z_t(\mathrm{k})'\hat{\theta}(k)$, $k_{\min} = [T\tau_{\min}]$, and $k_{\max} = [T\tau_{\max}]$.

The asymptotic behavior of the trend estimation error, $\delta_t = \hat{d}_t(k) - d_t(k_0)$, is summarized in Theorem 3 for the local break model.

Theorem 3. Let d_t be given by (9), where γ_T satisfies (10), and let $d_t(\hat{k})$ be given by (11)–(13).

(a) *If y_t is I(0), then:*
  (i) $(U_{0T}, D_{0T}) \Rightarrow \omega_0(W, D_0)$, *where* $D_0(\lambda) = v(\lambda,\tau_0)'\,\Phi(\tau_0)$, *where* $v(\lambda,\tau) = (-(\lambda-\tau)\mathbf{1}(\lambda>\tau), \lambda, \frac{1}{2}\lambda^2, \frac{1}{2}(\lambda-\tau)^2\mathbf{1}(\lambda>\tau))'$, *and the* 4×1 *random vector* $\Phi(\tau_0)$ *is a functional of* W *that is distributed* $\mathrm{N}(0,\Omega(\tau_0)^{-1})$, *where* $\Omega_{11} = 1-\tau$, $\Omega_{12} = -(1-\tau)$, $\Omega_{13} = -\frac{1}{2}(1-\tau^2)$, $\Omega_{14} = -\frac{1}{2}(1-\tau^2)$, $\Omega_{22} = 1$, $\Omega_{23} = \frac{1}{2}$, $\Omega_{24} = \frac{1}{2}(1-\tau)^2$, $\Omega_{33} = \frac{1}{3}$, $\Omega_{34} = (2-3\tau+\tau^3)/6$, *and* $\Omega_{44} = (1-\tau)^3/3$.
  (ii) $T\|\delta_t\| \Rightarrow \omega_0^2\Phi(\tau_0)'\{\int_0^1 \xi^\dagger(s,\tau_0)\,\xi^\dagger(s,\tau_0)'\mathrm{d}s\}\,\Phi(\tau_0)$, *where* $\xi^\dagger(s,\tau) = (-\mathbf{1}(s>\tau), \xi(s,\tau)')'$, *and where* $\xi(s,\tau) = (1, s, (s-\tau)\mathbf{1}(s>\tau))'$.

(b) *If y_t is I(1), then:*
  (i) $(U_{1T}, D_{1T}) \Rightarrow \omega_1(W, D_1)$, *where* $D_1(\lambda) = F_1(\lambda,\tau^*)$, *where* $F_1(\lambda,\tau) = \xi(\lambda,\tau)'M(\tau)^{-1}\Psi(\tau)$, *where* $\Psi(\tau) = \int_0^1 \xi(s,\tau)W(s)\mathrm{d}s$, $M(\tau) = \int_0^1 \xi(s,\tau)\times \xi(s,\tau)'\mathrm{d}s$, *and* $\xi(\lambda,\tau)$ *is defined in a(ii) of this theorem, and* τ^* *has the distribution*, $\operatorname{argmin}_{\tau\in[\tau_{\min},\tau_{\max}]}\int_0^1\{W(s) - F_1(s,\tau)\}^2\mathrm{d}s$.
  (ii) *Let* $\eta_T(\lambda,\tau) = \delta_{[T\lambda]}([T\tau]) - \delta_{[T\lambda]-1}([T\tau])$. *Then* $\eta_T(\cdot,\cdot) \rightarrow^p 0$.

It follows from Theorem 3 that the detrending error δ_t satisfies conditions A and B. Parts a(i) and b(i) respectively verify conditions A(i) and B(i). Condition A(ii) follows from part a(ii) of the theorem as long as $l_T^2/T \rightarrow 0$, which holds by assumption. Condition B(ii), the mean-square consistency for zero of η_T, follows from the sup-norm consistency result in part b(ii).

One possibility is that the series is detrended using the broken trend model (9), but in fact there is no break in the trend, that is, $\gamma_T = 0$. In this case τ_0 is not identified, and the conditions of Theorem 3 no longer hold. The next theorem summarizes the properties of the trend estimation error when in fact $\gamma_T = 0$.

Theorem 4. Let d_t be given by (9), let $d_t(\hat{k})$ be given by (11)–(13), and suppose that the true value of γ_T is zero.

(a) *If y_t is I(0), then:*
 (i) $(U_{0T}, D_{0T}) \Rightarrow \omega_0(W, D_0)$, *where* $D_0(\lambda) = \tilde{v}(\lambda, \tau^{\dagger})' \tilde{\Phi}(\tau^{\dagger})$, *where* $\tilde{\Phi}(\tau) = \int_0^1 \xi(s,\tau) \times \mathrm{d}W(s)$, $\tilde{v}(\lambda,\tau) = (\lambda, \frac{1}{2}\lambda^2, \frac{1}{2}(\lambda-\tau)^2 \mathbf{1}(\lambda > \tau))'$, *and* $\tau^{\dagger}$ *has the distribution* $\operatorname{argmax}_{\tau \in [\tau_{\min}, \tau_{\max}]} \tilde{\Phi}(\tau)' M(\tau)^{-1} \tilde{\Phi}(\tau)$, *where* $\xi(\lambda,\tau)$ *and* $M(\tau)$ *are defined in Theorem 3.*
 (ii) $T \| \delta_t \| \Rightarrow \tilde{\Phi}(\tau^{\dagger})' M(\tau^{\dagger})^{-1} \tilde{\Phi}(\tau^{\dagger})$.

(b) *If y_t is I(1), then the results of Theorem 3(b) continue to hold.*

In the I(1) case, the distribution of the detrended process does not depend on whether γ_T is zero or local to zero, an intuitively plausible result because the trend process and τ_0 are not estimated consistently in the I(1) case even for nonzero γ_T. In the I(0) case, however, the detrended process has a different distribution if $\gamma_T = 0$ than if γ_T is local to zero. This differs from the results for polynomial detrending, and raises the practical problem that the distribution f_0 in (5) will depend on γ and, if $\gamma \neq 0$, on τ_0. Our proposed solution is to estimate $f_0(x)$ by $f_0(x; \hat{b}/\hat{\sigma}_u, \hat{\tau})$, where $f_0(x; \hat{b}/\hat{\sigma}_u, \hat{\tau})$, is the density of $W_0^d(\cdot; \hat{\gamma}/\hat{\sigma}, \hat{\tau})$, where $\hat{b} = T\hat{\gamma}$ is computed from the OLS estimator of $\hat{\gamma}$ with sample size T. Then $W_0^d(\cdot; \hat{\gamma}/\hat{\sigma}, \hat{\tau})$, is computed as the limit of the partial sum process constructed from the residuals from the broken-trend detrending of a time series with an i.i.d. N(0, 1) stochastic component and with trend $d_t = (\hat{b}/T\hat{\sigma}_u)(t - [T\hat{\tau}])\mathbf{1}(t > [T\hat{\tau}])$, where $\hat{b}, \hat{\sigma}_u$, and $\hat{\tau}$ are held fixed.

Assuming that u_t is in fact I(0), this procedure is justified in two steps, first for γ_T local to zero, next for $\gamma_T = 0$. First suppose that γ is local to zero as in Theorem 3; to be concrete, let $\gamma_T = b/T$, where b is a constant, a sequence that satisfies (10). Under this local nesting $\hat{b} \rightarrow^p b$, $\hat{\sigma}_u \rightarrow^p \sigma_u$, and $\hat{\tau} \rightarrow^p \tau_0$ (Picard, 1985; Bai, 1992; see the proof to Theorem 3), $W_1^d(\lambda; \hat{b}, \tau)$ is continuous in τ, and W_1^d does not depend explicitly on b beyond the maintained assumption that $b \neq 0$. Thus the distribution of $W_1^d(\cdot, \tau_0)$ in Theorem 3(a) can be approximated with an asymptotically negligible error by the distribution of $W_1^d(\cdot; \hat{b}/\hat{\sigma}_u, \hat{\tau})$. Next suppose that $\gamma = 0$, so that τ_0 is unidentified and $\hat{\tau}$ has the limiting distribution of $\tau^{\dagger}$ given in Theorem 4(a). Then results in the proofs of Theorems 3 and 4 suggest that the distribution of D_0 (and thus of f_0) is continuous in b as $b \rightarrow 0$ and moreover $\hat{b} \rightarrow^p 0$.[4] It follows that $W_1^d(\cdot; \hat{b}/\hat{\sigma}_u, \hat{\tau})$ has the limiting distribution in Theorem 4, so that for $b = 0$ or $b \neq 0$ this procedure yields a consistent estimator of f_0.

If u_t is in fact I(1), then for this procedure to yield a consistent decision rule it is sufficient to show that the proposed procedure produces a limiting I(0)

[4] From Bai (1992), under the local nesting $\gamma_T = b/T$, $T^{1/2}(\hat{b} - b)$ has an asymptotic normal distribution if $b \neq 0$, and from the proof of Theorem 4, $T^{1/2}\hat{b} = O_p(1)$ if $b = 0$, so $\hat{b} \rightarrow^p b$ for general b.

distribution f_0 that has support on the real line. In fact, a stronger result holds, namely that if u_t is I(1) and γ_T is local to zero with the nesting $\gamma_T = b/T$ (where b might be zero), then the distribution of $\phi(W_0^d(\cdot; \hat{b}/\hat{\sigma}_u, \hat{\tau}))$ converges to the distribution resulting from Theorem 3(a), with the random variable τ^* [defined in Theorem 3(b)] replacing τ_0. This follows from the fact that, if u_t is I(0), $\hat{b} = O_p(1)$ under the local assumption [an implication of the proofs of Theorems 3(b) and 4(b)], so the coefficient on the trend used to generate W_0^d will with probability one satisfy (10) and therefore will satisfy the conditions of Theorem 3(a). Although $f_0(x; \hat{b}/\hat{\sigma}_u, \hat{\tau})$ for fixed x will asymptotically be a random variable in this case (because $\hat{\tau} \Rightarrow \tau^*$), the posterior odds and Bayes ratios will still yield consistent decision procedures.

4. Numerical issues and finite sample performance

4.1. Computation of densities and likelihood ratios

Because the limiting distributions of the statistics ϕ_T are nonstandard, $f_1(\phi_T)$, $f_0(\phi_T)$, B_T, and Π_T were evaluated numerically using a kernel density estimator of the likelihood. The approach was first to produce a matrix of pseudo-random realizations of the limiting random variables $\phi(W_0^d)$ and $\phi(V_1^d)$, and second to use these realizations to evaluate the likelihoods $f_0(\phi_T)$ and $f_1(\phi_T)$ for observed ϕ_T. Specifically, series of length 100 were drawn according to the I(0) model $u_t = \varepsilon_t$, ε_t i.i.d. N(0, 1); these data were then used to construct V_T (imposing $l_T = 0$) and ϕ_T, and the realization of ϕ_T was saved. This was repeated using the I(1) model $\Delta u_t = \varepsilon_t$, ε_t i.i.d. N(0, 1), $T = 100$, and $\phi(N_T^{-1/2} V_T)$ was computed (with $l_T = 0$) and saved. Both cases entailed 8000 Monte Carlo replications. Given a realization $\hat{\phi}_T$, the densities $f_0(\hat{\phi}_T)$ and $f_1(\hat{\phi}_T)$ were then computed by kernel density estimation.[5] Five trend specifications are considered: no detrending, demeaning, linear detrending by OLS, and two versions of broken trends

---

From Theorem 3(a) and its proof, for $b \neq 0$, $D_{0T}(\lambda) = \tilde{\Phi}(\lambda, \hat{\tau})' \tilde{\Phi}(\hat{\tau}) + bh(\lambda)\tau^* + o_p(1)$ uniformly in λ, where τ^* is an $O_p(1)$ random variable that does not depend on λ. It follows that the limit as $b \to 0$ of the distribution of D_0 is the distribution in Theorem 4(a) if the distribution of $\hat{\tau}$ is continuous in b. To argue this, note that for $b \neq 0$, $\hat{\tau}$ solves $\max_{\tau \in [\tau_{min}, \tau_{max}]} \tilde{\Phi}(\tau)' M(\tau)^{-1} \Phi(\tau) + bQ(\tau)$, where $Q(\tau) = O_p(1)$ uniformly in a $T^{-1/2}$ neighborhood of τ_0. As $b \to 0$, the objective function converges to the objective function in Theorem 4(a), suggesting the continuity of the distribution of $\hat{\tau}$ as $b \to 0$.

[5] The Monte Carlo and empirical results were computed using a flat kernel with bandwidth $\kappa\sigma_\phi$, where σ_ϕ is the standard deviation of the asymptotic distribution of the statistic in question and $\kappa = 0.1$. Fig. 1 was computed using a Gaussian kernel with the same bandwidth. The density estimates and likelihood ratios are numerically stable in the sense that the I(1)/I(0) decision rates are insensitive to the choice of bandwidth over the range $\kappa = 0.02$–0.20. Programs in GAUSS to perform these evaluations are available from the author upon request.

estimated by OLS, first with $\gamma = 0$ and second with $\gamma = 0.5$ and $\tau_0 = 0.5$. In this final case, the pseudo-data were generated using $y_t = \gamma(t - [T\tau_0])\,\mathbf{1}(t > [T\tau_0]) + u_t$ because of the dependence of the limiting I(0) distribution on τ_0 for γ_T local to zero. Because $T = 100$ was used to generate the null distributions, in the nesting $\gamma_T = b/T$ this final trend specification corresponds to $b = 50$. (Looking ahead to the empirical results, this value of b is large relative to empirical estimates using annual time series data for the United States, so we would expect the $\gamma = 0$ and $\gamma_T = 50/T$ cases to bracket a wide range of cases of empirical interest.) This nesting is used to set γ as a function of the sample size in the Monte Carlo simulations reported in the Section 4.2 below. The Monte Carlo and empirical work with the trend-break specifications all used $\tau_{\min} = 0.15$ and $\tau_{\max} = 0.85$.

The densities f_0 and f_1 for ϕ_1 [that is, the densities of $\phi_1(W_0^d)$ and $\phi_1(V_1^d)$] and the corresponding cdf's are plotted in Fig. 1 for the demeaned case, which is typical of the five trend specifications. In each trend case, the I(1) distribution lies to the left of the I(0) distribution. Although the I(0) distribution does not change its shape substantially as the detrending procedure changes, the I(1) distribution does, becoming more bell-shaped the greater is the amount of detrending.

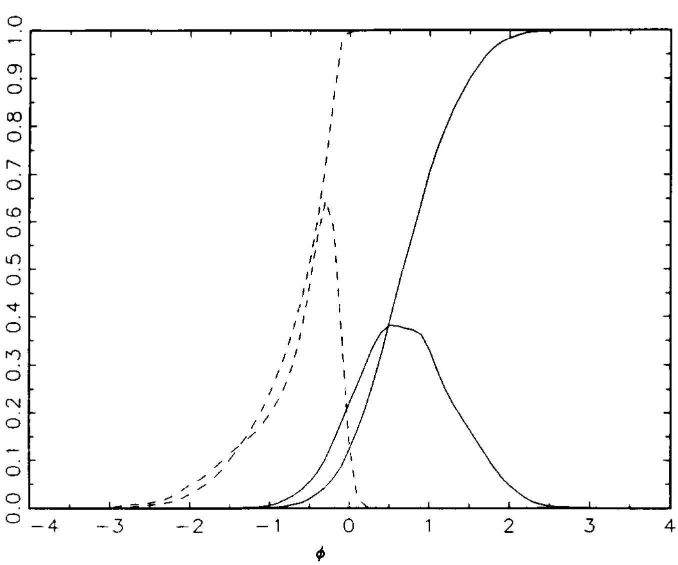

Fig. 1. Asymptotic cdf and pdf of ϕ_{1T} under I(0) (solid line) and I(1) (dashed line), demeaned case.

4.2. Finite-sample performance: Monte Carlo results

This subsection reports the results of a Monte Carlo experiment which studied the ability of the proposed procedure to classify correctly Gaussian ARMA (1,1) processes. The spectral density estimator is a truncated version of one recommended by Andrews (1991). Specifically, the Parzen kernel was used and l_T was chosen as $l_T = \min(\hat{l}_T, l_{T,\max})$, where $\hat{l}_T$ is Andrews' (1991) automatic bandwidth selector. Because $\hat{l}_T$ is unbounded in the I(1) case, it was truncated at $l_{T,\max} = [10(T/100)^{0.2}]$, where the rate is taken from Andrews (1991) and satisfies the condition of Theorem 1, and where 10 was picked arbitrarily.

Monte Carlo rates at which the series are classified as I(0) based on the posterior odds ratios, for various prior odds, sample sizes, and nuisance parameters, are summarized in Table 1 for the ϕ_{2T} statistic. One way to make comparisons across panels is to consider the performance of the classifiers, standardized so that their error rate is constant for a certain model; this is analogous to comparing size-adjusted power of tests. Such a comparison shows that increasing the extent of detrending reduces the discriminatory power of the statistics. For example, for $\pi_1 = \pi_0$ and $T = 100$ for ϕ_{2T}, the random walk error rates for the demeaned, detrended, and broken trend-detrended ($\gamma = 0$) cases are comparable, respectively 0.13, 0.12, and 0.09, but the I(0) correct classification rates for $\rho = 0.9$, $\theta = 0$ drop sharply to 0.47, 0.27, and 0.17, respectively. Moreover, the IMA error rates increase with the extent of detrending, respectively rising from 0.36 to 0.57 to 0.73 for $\rho = 1$, $\theta = -0.875$ for the three detrending cases. In short, detrending leads to large-root I(0) AR models being increasingly classified as I(1) and large-root I(1) MA models being increasingly classified as I(0). This parallels the well-known result that the power of tests of a unit AR root against a given alternative declines with the extent of the detrending. Also, the classification rates are sensitive to the prior odds for moderate sample sizes. For example, for $T = 100$, $\rho = 1$, and $\theta = 0$, decreasing the prior odds ratio in favor of I(1) from 1 to 0.25 increases the false classification rate for ϕ_{2T} (linearly detrending) from 12% to 38%.

These results, along with Monte Carlo results for ϕ_{1T} and ϕ_{3T} reported in the working paper version of this paper (Stock, 1992) permit comparisons across the ϕ_{1T}, ϕ_{2T}, and ϕ_{3T} statistics. First consider the leading case of linear detrending. Holding the random walk classification rate constant, ϕ_{2T} has higher correct classification rates than ϕ_{1T} for the AR models with large roots, and it has comparable error rates for the IMA models. Similarly, holding constant the random walk correct classification rate, ϕ_{2T} outperforms ϕ_{3T} for the I(0) AR(1) models. The results for ϕ_{3T} indicate a large incorrect classification rate of the random walk with even prior odds (41% for $T = 100$), creating an additional difficulty with interpreting this statistic. In contrast, the $T = 100$ error rates for ϕ_{2T} are 12% in the random walk case and 6% for the i.i.d. process. For linear detrending and typical macroeconomic sample sizes, this evidence suggests that,

Table 1
Monte Carlo results: I(0) classification rates for the ϕ_{2T} statistic
Model: $(1-\rho L)x_t = (1+\theta L)\varepsilon_t$, ε_t i.i.d.N(0, 1)

		$\theta=0$, $\rho=$							$\rho=1.0$, $\theta=$		
T	π_1/π_0	0	0.6	0.8	0.9	0.95	0.975	1.0	−0.875	−0.75	−0.5
(A) Demeaned											
100	1.00	0.96	0.78	0.67	0.47	0.29	0.25	0.13	0.36	0.17	0.13
100	0.50	0.98	0.90	0.82	0.67	0.48	0.42	0.25	0.43	0.32	0.23
100	0.25	0.99	0.96	0.91	0.81	0.64	0.55	0.36	0.51	0.42	0.33
100	0.10	1.00	0.99	0.96	0.88	0.75	0.66	0.43	0.59	0.51	0.42
200	1.00	0.98	0.86	0.74	0.61	0.39	0.26	0.09	0.15	0.10	0.09
200	0.50	0.99	0.93	0.84	0.72	0.52	0.39	0.16	0.20	0.15	0.17
200	0.25	0.99	0.98	0.92	0.84	0.67	0.57	0.25	0.27	0.27	0.25
200	0.10	0.99	0.99	0.97	0.92	0.78	0.68	0.31	0.34	0.35	0.33
(B) Linear trend/OLS detrending											
100	1.00	0.94	0.72	0.45	0.27	0.15	0.14	0.12	0.57	0.25	0.15
100	0.50	0.97	0.85	0.65	0.40	0.26	0.22	0.22	0.65	0.32	0.25
100	0.25	0.98	0.95	0.83	0.69	0.50	0.49	0.38	0.71	0.43	0.44
100	0.10	0.99	0.99	0.95	0.82	0.71	0.67	0.58	0.77	0.60	0.62
200	1.00	0.99	0.81	0.66	0.40	0.23	0.11	0.07	0.29	0.11	0.06
200	0.50	0.99	0.91	0.80	0.59	0.41	0.23	0.16	0.33	0.18	0.13
200	0.25	0.99	0.96	0.86	0.68	0.49	0.32	0.20	0.41	0.25	0.17
200	0.10	1.00	0.99	0.92	0.79	0.63	0.45	0.30	0.50	0.34	0.29
(C) Broken trend/OLS detrending, $\gamma=0$											
100	1.00	0.91	0.39	0.25	0.17	0.13	0.10	0.09	0.73	0.38	0.14
100	0.50	0.95	0.81	0.69	0.60	0.48	0.42	0.39	0.81	0.50	0.45
100	0.25	0.98	0.97	0.92	0.87	0.82	0.76	0.77	0.86	0.65	0.77
100	0.10	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.83	0.99
200	1.00	0.97	0.63	0.44	0.24	0.12	0.08	0.07	0.54	0.17	0.09
200	0.50	0.97	0.82	0.69	0.45	0.30	0.21	0.19	0.59	0.24	0.22
200	0.25	0.99	0.92	0.84	0.68	0.51	0.38	0.38	0.66	0.35	0.40
200	0.10	1.00	0.98	0.96	0.92	0.82	0.74	0.71	0.75	0.52	0.74
(D) Broken trend/OLS detrending, $\gamma=50/T$, $\lambda=0.5$											
100	1.00	0.96	0.38	0.24	0.11	0.08	0.10	0.08	0.78	0.40	0.09
100	0.50	0.98	0.93	0.85	0.68	0.61	0.59	0.55	0.85	0.54	0.54
100	0.25	0.99	1.00	1.00	0.98	0.95	0.92	0.90	0.92	0.70	0.93
100	0.10	1.00	1.00	1.00	1.00	0.99	0.99	0.97	0.95	0.85	0.99
200	1.00	0.99	0.71	0.47	0.26	0.12	0.07	0.07	0.58	0.13	0.08
200	0.50	1.00	0.86	0.73	0.45	0.28	0.17	0.15	0.63	0.19	0.18
200	0.25	1.00	0.97	0.90	0.70	0.51	0.36	0.30	0.67	0.28	0.36
200	0.10	1.00	0.99	0.96	0.84	0.68	0.52	0.46	0.73	0.40	0.56

Entries are the fraction of times that the posterior odds ratio favors I(0) over I(1) for the indicated prior odds ratio π_1/π_0. The spectral density ω^2 was estimated using the Parzen kernel with bandwidth truncation parameter l_T, computed using Andrews' (1991) automatic procedure, truncated at $10(T/100)^{0.2}$, as discussed in the text. Based on 500 Monte Carlo replications for each entry.

of the three tests, ϕ_{2T} has the best performance, followed by ϕ_{1T}. Turning to the broken-trend case, ϕ_{3T} and especially ϕ_{1T} have large incorrect classification rates for the random walk with even prior odds. Thus posterior odds based on ϕ_{1T} and ϕ_{3T} are likely to be misleading. Holding the random walk correct classification rate constant, ϕ_{2T} generally exhibits the lowest error rates of the three statistics in the other (ρ, θ) combinations. In summary, the ϕ_{1T} and ϕ_{2T} statistics exhibit the best overall performance in the linear trend case, while ϕ_{2T} is preferred in the broken trend case. These statistics are therefore used in the empirical analysis in the next section.

5. Empirical results

Table 2 reports posterior odds ratios for Nelson and Plosser's (1982) annual data on 14 aggregate economic time series for the United States. First consider the results for linear detrending. The ϕ_{1T} and ϕ_{2T} posterior odds ratios yield the same I(1)/I(0) classifications of 13 of the 14 statistics; for these 13 series, 12 of the classifications agree with Nelson and Plosser's (1982) results based on Dickey–Fuller t-statistics, that the series are consistent with the I(1) model. The only series on which the ϕ_{1T} and ϕ_{2T} statistics disagree is the unemployment rate, for which the ϕ_{1T} posterior odds ratio just favors I(1). Because the unemployment rate is bounded below and above, it is arguably more appropriate to demean than to linearly detrend this series. For demeaned unemployment, the posterior odds ratios (even prior odds) are 0.44 and 0.11 for the ϕ_{1T} and ϕ_{2T} statistics, respectively, both favoring the I(0) hypothesis, with the evidence using the ϕ_{2T} statistic being rather strong. These two observations suggest classifying the unemployment rate as I(0).

It is interesting to note that, at the level of the I(0)/I(1) classification, the only difference between the posterior odds ratio results and conventional Dickey–Fuller tests is for the money stock. However, for this series neither the classical nor the Bayesian results are clear-cut: the classical 90% asymptotic confidence interval based on inverting the ADF statistic is wide (0.687, 1.030), and barely contains 1, while the two ϕ_T posterior odds ratios exceed 0.8, providing only weak evidence in favor of the I(0) model.

From a Bayesian perspective, the posterior odds ratios provide information about the relative likelihood of the I(1) and I(0) models. For some series, in particular industrial production, consumer prices and stock prices, the evidence strongly favors the I(1) model. However, for most series the evidence is much less strong. For example, a researcher with a prior odds ratio of 1/2 in favor of the I(0) hypothesis would reach the conclusion that the GNP deflator is I(0) using the ϕ_{2T} statistic, and that seven additional series are I(0) using the ϕ_{1T} statistic.

Posterior odds ratios for broken-trend detrended statistics are also presented in Table 2. As discussed in Section 3, in the I(0) case the asymptotic distribution

Table 2
Posterior odds ratios that a series is I(1), Nelson–Plosser data set
Prior odds ratio = $\pi_1/\pi_0 = 1.0$

		Classical ADF statistics			Bayesian classifier Linear detrended		Bayesian classifier Broken-trend detrended				
		ADF statistics			Post. odds for		Post. odds for $B(\phi_2)$				
Series	T	P	$\hat{\tau}_\tau$	90% conf. interval	$B(\phi_1)$	$B(\phi_2)$	(a)	(b)	(c)	$T\hat{\gamma}/\hat{\sigma}_u$	$\hat{k}$
Real GNP	62	1	− 2.994	(0.604, 1.042)	1.44	3.89	1.01	0.85	0.90	12.67	1934
Nominal GNP	62	1	− 2.321	(0.757, 1.060)	1.54	4.06	1.03	0.79	0.85	11.66	1936
Real per capita GNP	62	1	− 3.045	(0.591, 1.041)	1.35	2.43	1.03	0.85	0.92	11.06	1933
Industrial production	111	5	− 2.529	(0.836, 1.031)	5.10	5.88	3.63	7.33	3.89	− 8.42	1901
Employment	81	2	− 2.655	(0.757, 1.039)	1.62	2.14	1.46	1.54	1.08	− 17.51	1907
Unemployment rate	81	3	− 3.552	(0.577, 0.950)	1.07	0.44	1.39	1.48	1.67	− 2.81	1926
GNP deflator	82	1	− 2.516	(0.787, 1.041)	1.05	1.15	1.85	1.83	2.02	7.37	1940
Consumer prices	111	3	− 1.972	(0.901, 1.037)	7.75	44.64	1.21	1.53	1.35	24.36	1899
Wages	71	2	− 2.236	(0.800, 1.054)	1.37	2.31	1.40	1.15	1.21	8.51	1939
Real wages	71	1	− 3.049	(0.644, 1.035)	2.07	12.71	0.97	0.97	1.06	14.85	1933
Money stock	82	1	− 3.078	(0.687, 1.030)	0.89	0.84	1.47	1.57	1.57	− 6.88	1918
Velocity	102	0	− 1.663	(0.929, 1.042)	2.18	38.00	4.87	6.41	3.05	22.52	1944
Bond yield	71	2	0.686	(1.032, 1.075)	2.10	6.38	1.34	2.00	0.01–0.09[a]	27.42	1955
S&P 500	100	2	− 2.122	(0.873, 1.039)	4.84	22.67	0.92	1.11	1.07	30.98	1945

T is the total number of observations on each series, including observations used for initial conditions. The columns headed p and $\hat{\tau}_\tau$, respectively, give the number of lagged first differences in the ADF regressions and the ADF t-statistic. To facilitate comparisons of results, p was taken from Nelson and Plosser (1982). The 90% confidence interval column, computed following Stock (1991), is the confidence interval that results from inverting the ADF t-statistics. The two 'linear detrended' columns present the Bayes ratios for ϕ_{1T} and ϕ_{2T}, where $\hat{\omega}$ was computed as described in the notes to Table 1. For the remaining 'broken-trend detrended' columns, the entries are the Bayes ratios for ϕ_{2T}, computed using asymptotic I(1) distribution and (a) the $\gamma = 0$ asymptotic I(0) distribution, (b) the ($\gamma = 50/T$, $\tau = 0.5$) I(0) asymptotic distribution, and (c) the distribution of $\phi_2(W_1^d(\cdot; \hat{b}/\hat{\sigma}_u, \hat{\tau}))$ where $\hat{b} = T\hat{\gamma}$, computed as described in Section 3 using 4000 Monte Carlo replications and where $\tau_{min} = 0.15$ and $\tau_{max} = 0.85$. The final column gives the estimated break date $\hat{k}$. Data sources: See Nelson and Plosser (1982). The data are annual, with all series ending in 1970. All series except the bond yield were analyzed in logarithms.

[a] The kernel density estimate of f_0 for the bond yield, but not for the other series, is sensitive to the bandwidth choice; the reported range for the bond yield in the final column is for bandwidths from 0.1 to 0.25.

depends on the break parameters γ and τ (if $\gamma \neq 0$), so the Bayes ratios are evaluated for $\gamma = 0$, for ($\gamma = 50/T$, $\tau = 0.5$), and for estimated γ, τ, where the limiting distribution under I(0) was approximated by the distribution of $\phi_2(W_0^d((\cdot; \hat{b}/\hat{\sigma}_u, \hat{\tau}))$ as described in Section 3. The likelihood ratio statistic for the estimated γ, τ case was computed as described in Section 4.1, except that the kernel density evaluations were based on 4000 Monte Carlo replications.

The striking feature of the broken trend results is that most of the Bayes ratios are near one. In several cases, the I(1)/I(0) classification is sensitive to which I(0) distribution is used to compute B. However, with the exception of the bond yield, in these cases the Bayes ratio typically ranges from 0.8 to 1.1, so that small shifts from even prior odds would change the classification. In this sense, for all series except industrial production, the GNP deflator, velocity and perhaps the bond yield, the data are uninformative about the I(0)/I(1) classification under the broken trend model. For industrial production and velocity, the reported Bayes ratios favor the I(1) model. The Bayes ratios also provide moderately strong evidence in favor of the I(1) model for the GNP deflator.

For 13 of the series, the Bayes ratio computed using the $f_0(\cdot; \hat{b}/\hat{\sigma}_u, \hat{\tau})$ distribution either falls within, or is close to, the range of $B(\phi_2)$ in the $b = 0$ and $b = 50$ cases. This is unsurprising, in the sense that in absolute values the estimates of b/σ_u, $T\hat{\gamma}/\hat{\sigma}_u$, are small and always less than 50. The one series for which inferences differ is the bond yield. For this series, the likelihood ratios are also unstable to changes in the kernel density estimator and bandwidth used to evaluate f_0. The source of this instability is that the point estimate of ϕ_2 for the bond yield falls in the tails of both the I(0) and I(1) distributions; that is, after broken-trend detrending the empirical realization of ϕ_2 for the bond yield is unlikely to have been generated by either an I(0) or I(1) process. This suggests exploring other characterizations of the long-run properties of the bond yield, such as fractionally integrated models.

In summary, if linear detrending is used, the Nelson–Plosser (1982) I(1)/I(0) classifications are supported by the proposed decision-theoretic procedures, with the sole exception of the money supply for which the posterior odds slightly favor I(0). For several series the empirical evidence is weak, in the sense that moderately strong priors that a series is I(0) would change the posterior conclusion. When the series are detrended using piecewise-linear trends, the evidence in these data is much weaker, with most Bayes ratios in the range 0.8–1.3.

6. Conclusions

Although the procedures proposed here have an explicitly Bayesian motivation, they alternatively can be given a classical interpretation, and indeed classical asymptotic arguments have been key to eliminating the dependence of the procedures on the short-run nuisance parameters. From a frequentist perspective, the procedures are simply model selection techniques adapted to the I(1)/I(0) classification problem, and the prior odds ratio π_1/π_0 is a parameter to be chosen by the econometrician. This choice could, for example, be made by controlling the rejection rate (the 'size') for a specific leading case. Indeed, according to the Monte Carlo results, our use of even prior odds for the ϕ_{2T} statistic in the empirical analysis corresponds to sizes of just under 10% for the $T = 100$, random walk model, in both the linear and piecewise-linear trend specifications.

These results suggest several directions for further theoretical development. Primary among these is the desirability of constructing optimal classifiers among the set considered here. Because the Bayes ratio is just the likelihood ratio, the construction of an optimal ϕ_T is related to the existence of sufficient statistics for the I(0) and I(1) models. In the related problem of testing the Gaussian I(1) null against the I(0) alternative, parameterized in terms of whether the largest autoregressive root is one or less than one, the asymptotic minimal sufficient statistic has dimension two, so no uniformly most powerful test of the unit AR root null exists (Elliott, Rothenberg, and Stock, 1992). Similarly, there does not exist a uniformly most powerful test of the I(0) null against the I(1) alternative, even in the simplest Gaussian parameterizations (e.g., Shively, 1988). This suggests that any single ϕ_T will not be uniformly best for all true models. Whether a one-dimensional ϕ_T can come close to being uniformly optimal, as it can in the unit AR root testing problem, is a subject for future research.

It would also be of interest to compare these classifiers to other approaches, such as I(0) or I(1) tests with critical values that depend on the sample size or the Phillips–Ploberger (1991) posterior odds ratio approach. Another question is the calibration of this classifier in the context of specific loss functions, which presumably would depend on the application at hand. These problems are left for future research.

Appendix A: Proofs of Theorems 1–4

The proofs are applications of the functional central limit theorem (FCLT); see, for example, Hall and Heyde (1980), Ethier and Kurtz (1986), or Herrndorf (1984). Throughout, set $K_T(m) = k(m/l_T)/\sum_{j=-l_T}^{l_T} k(j/l_T)$.

Proof of Theorem 1

(a) By Assumption A(i),

$$T^{-1/2}\sum_{t=1}^{[T\lambda]} y_t^d = T^{-1/2}\sum_{t=1}^{[T\lambda]} (u_t - \delta_t)$$
$$= U_{0T}(\lambda) - D_{0T}(\lambda) \Rightarrow \omega_0(W(\lambda) - D_0(\lambda)) = \omega_0 W_0^d(\lambda). \qquad \text{(A.1)}$$

Let $\tilde{\omega}^2 = \sum_{m=-l_T}^{l_T} k(m/l_T)T^{-1}\sum_{t=|m|+1}^{T} u_t u_{t-|m|}$. Under the stated assumptions on l_T and the kernel k, $\tilde{\omega}^2 \to^p \omega_0^2$. Thus it is sufficient to show that $|\hat{\omega}^2 - \tilde{\omega}^2| \to^p 0$. Now,

$$|\hat{\omega}^2 - \tilde{\omega}^2| = \sum_{m=-l_T}^{l_T} k(m/l_T)T^{-1} \sum_{t=|m|+1}^{T} (\delta_t \delta_{t-|m|} - u_t \delta_{t-|m|} - u_{t-|m|}\delta_t)$$

$$\leqslant \sum_{m=-l_T}^{l_T} k(m/l_T)\{2\|u_t\|^{1/2}\|\delta_t\|^{1/2} + \|\delta_t\|\}$$

$$\leqslant (2l_T + 1)\{2\|u_t\|^{1/2}\|\delta_t\|^{1/2} + \|\delta_t\|\},$$

where $\|\delta_t\| = T^{-1}\sum_{t=1}^{T}\delta_t^2$. Because $l_T \to \infty$ and $\|u_t\| \to^p \gamma_u(0)$, $|\hat{\omega}^2 - \tilde{\omega}^2| \to^p 0$ if $l_T^2\|\delta_t\| \to^p 0$, which is assumed as condition A(ii).

(b) Write

$$N_T^{-1/2}V_T(\lambda) = N_T^{-1/2}\hat{\omega}^{-1}T^{-1/2}\sum_{s=1}^{[T\lambda]} y_s^d = B_T^{-1/2}A_T(\lambda),$$

where

$$A_T(\lambda) = T^{-3/2}\sum_{s=1}^{[T\lambda]} y_s^d, \qquad B_T = T^{-1}\sum_{m=-l_T}^{l_T} K_T(m)\hat{\gamma}_{yd}(|m|).$$

By Assumption B(i),

$$A_T(\cdot) \Rightarrow \omega_1 \int_{s=0}^{\cdot} W_1^d(s)\mathrm{d}s.$$

In the case of no detrending, it was shown by Phillips (1991b, App.) that if u_t is I(1), then $B_T \Rightarrow \omega_1^2\int_0^1 W_1(s)^2\mathrm{d}s$. This result was extended to linear trends (OLS detrending) by Kwiatkowski, Phillips, Schmidt, and Shin (1992) and to general polynomial trends (OLS detrending) by Perron (1991). Lemma A.1 (below) extends this result extended to the general trends satisfying conditions A and B. It is shown in the lemma that $\Xi_T = T^{-2}\sum_{t=1}^{T}(y_t^d)^2 - B_T \to^p 0$, so that by Assumption B(i) and the continuous mapping theorem, $B_T \Rightarrow \omega_1^2\int_0^1 W_1^d(s)^2\,\mathrm{d}s$. Combining the limiting representations for $A_T(\cdot)$ and B_T yields the desired result. □

Lemma A.1. Let $\Xi_T = T^{-2}\sum_{t=1}^{T}(y_t^d)^2 - B_T$, *where* $B_T = T^{-1}\sum_{m=-l_T}^{l_T}K_T(m) \times \hat{\gamma}_{yd}(|m|)$. *If* y_t *is I(1),* $l_T^2\ln T/T \to 0$, *and Assumption B holds, then* $\Xi_T \to^p 0$.

Proof. Use $\sum_{m=-l_T}^{l_T} K_T(m) = 1$ and $K_T(m) = K_T(-m)$ to write $\Xi_T = 2T^{-1}\sum_{m=1}^{l_T} K_T(m)(\hat{\gamma}_{yd}(0) - \hat{\gamma}_{yd}(m))$. For $m \geqslant 1$,

$$\hat{\gamma}_{yd}(0) - \hat{\gamma}_{yd}(m) = T^{-1} \sum_{t=m+1}^{T} y_t^d \Delta_m y_t^d + T^{-1} \sum_{t=1}^{m} (y_t^d)^2,$$

where $\Delta_m = 1 - L^m$. Thus,

$$\begin{aligned} |\Xi_T| &\leqslant 2T^{-1} \sum_{m=1}^{l_T} K_T(m) |T^{-1} \sum_{t=m+1}^{T} y_t^d \Delta_m y_t^d| \\ &\quad + 2T^{-1} \sum_{m=1}^{l_T} K_T(m) |T^{-1} \sum_{t=1}^{m} (y_t^d)^2| \\ &= A_{1T} + A_{2T}, \end{aligned}$$

say. These two terms are handled in turn.

(i) A_{1T}. Note that $|T^{-1}\sum_{t=m+1}^{T} y_t^d \Delta_m y_t^d| \leqslant \{T^{-1}\sum_{t=1}^{T}(y_t^d)^2\}^{1/2}\{T^{-1} \times \sum_{t=m+1}^{T}(\Delta_m y_t^d)^2\}^{1/2}$.

The definition of $K_T(m)$ and the assumption, made following (4) in the text, that $l^{-1}\sum_{u=1}^{l} k(u/\lambda) \geqslant \underline{k}$, where $\underline{k}$ is a positive constant, imply that $K_T(m) \leqslant l_T^{-1}\underline{k}^{-1}$. Using this inequality and $x^{1/2} < 1 + x$ for all $x \geqslant 0$, we have

$$\begin{aligned} A_{1T} &\leqslant 2T^{-1} \sum_{m=1}^{l_T} K_T(m) \{T^{-1} \sum_{t=1}^{T} (y_t^d)^2\}^{1/2} \{T^{-1} \sum_{t=m+1}^{T} (\Delta_m y_t^d)^2\}^{1/2} \\ &\leqslant 2\{T^{-2} \sum_{t=1}^{T} (y_t^d)^2\}^{1/2} \{l_T^{-1} \underline{k}^{-1} q_T^{-1/2} \sum_{m=1}^{l_T} (q_T T^{-2} \sum_{t=m+1}^{T} (\Delta_m y_t^d)^2)^{1/2}\} \\ &\leqslant 2\{T^{-2} \sum_{t=1}^{T} (y_t^d)^2\}^{1/2} \{\underline{k}^{-1} l_T^{-1} q_T^{-1/2} \sum_{m=1}^{l_T} [1 + q_T T^{-2} \sum_{t=m+1}^{T} (\Delta_m y_t^d)^2]\} \\ &= 2\{T^{-2} \sum_{t=1}^{T} (y_t^d)^2\}^{1/2} \{\underline{k}^{-1} q_T^{-1/2} + \underline{k}^{-1} l_T^{-1} q_T^{1/2} T^{-2} \sum_{m=1}^{l_T} \sum_{t=m+1}^{T} (\Delta_m y_t^d)^2\} \\ &= 2D_{1T}^{1/2}\{D_{2T} + D_{3T}\}, \end{aligned}$$

where $D_{1T} = T^{-2}\sum_{t=1}^{T}(y_t^d)^2$, $D_{2T} = \underline{k}^{-1}q_T^{-1/2}$, $D_{3T} = \underline{k}^{-1}l_T^{-1}q_T^{1/2}T^{-2} \times \sum_{m=1}^{l_T}\sum_{t=m+1}^{T}(\Delta_m y_t^d)^2$, and q_T is a positive nonstochastic sequence such that $q_T \to \infty$. To be concrete, set $q_T = (\ln T)^2$.

Now $D_{1T} \Rightarrow \omega_1^2 \int_0^1 W_1^d(s)^2 \mathrm{d}s$ by Assumption B(i). Also, $D_{2T} = 1/(\underline{k} \ln T) \to 0$. Thus $A_{1T} \to^p 0$ if $D_{3T} \to^p 0$. Also, $D_{3T} = (\ln T/(\underline{k} l_T T))\sum_{m=1}^{l_T} \|\Delta_m y_t^d\| \leqslant$

$(\ln T/(\underline{k} l_T T))\sum_{m=1}^{l_T}\{\|\Delta_m u_t\|^{1/2} + \|\Delta_m \delta_t\|^{1/2}\}^2$. Also, $\|\Delta_m u_t\| = \|\sum_{j=0}^{m-1}\Delta u_{t-j}\| \leqslant m^2\|\Delta u_t\|$ and $\|\Delta_m \delta_t\| \leqslant m^2\|\Delta \delta_t\|$. Thus,

$$D_{3T} \leqslant (\ln T/(\underline{k} l_T T)) \sum_{m=1}^{l_T} \{m\|\Delta u_t\|^{1/2} + m\|\Delta \delta_t\|^{1/2}\}^2$$

$$\leqslant (c/\underline{k})(l_T^2 \ln T/T)\{\|\Delta u_t\| + \|\Delta \delta_t\|\}^{1/2}, \tag{A.2}$$

for some constant c. Under the I(1) assumption, $\|\Delta u_t\| \to^p \gamma_{\Delta u}(0)$, and under Assumption B(ii) $\|\Delta\delta_t\| = \mathrm{O}_p(1)$; with the rate condition $l_T^2 \ln T/T \to 0$, it follows that $D_{3T} \to^p 0$.

(ii) A_{2T}. $A_{2T} = 2T^{-1}\sum_{m=1}^{l_T} K_T(m)T^{-1}\sum_{t=1}^{m}(y_t^d)^2 \leqslant 2T^{-1}\sum_{m=1}^{l_T} K_T(m) \times (T^{-1}\sum_{t=1}^{l_T}(y_t^d)^2) \leqslant T^{-2}\sum_{t=1}^{l_T}(y_t^d)^2 = \int_0^{l_T/T}\{U_{1T}(\lambda) - D_{1T}(\lambda)\}^2 \mathrm{d}\lambda \to^p 0$.

by Assumption B(i) and because $l_T/T \to 0$. □

Proof of Theorem 2

Throughout, let $\Upsilon_T = \operatorname{diag}(1, T, \ldots, T^q)$ and let $M_T = T^{-1}\Upsilon_T^{-1}\sum_{t=1}^T z_t z_t' \Upsilon_T^{-1}$. The nonstochastic $(q+1)\times(q+1)$ matrix M_T has typical element $M_{T,ij} = T^{-1}\sum_{t=1}^T (t/T)^{i+j-2}$, which has the limit $M_{T,ij} \to 1/(i+j-1) = M_{ij}$ whether y_t is I(0) or I(1).

(a)(i) Direct calculation shows that $D_{0T}(\lambda) = T^{-1/2}\sum_{s=1}^{[T\lambda]}\delta_s = v_T(\lambda)' M_T^{-1}\Phi_T$, where $v_T(\lambda) = T^{-1}\sum_{s=1}^{[T\lambda]}\Upsilon_T^{-1}z_s$ and $\Phi_T = T^{-1/2}\sum_{t=1}^T \Upsilon_T^{-1}z_t u_t$. The $(q+1)\times 1$-dimensional process $v_T(\lambda)$ is nonstochastic and has the limit $v_T(\lambda) \to v(\lambda)$, where the ith element is $v_i(\lambda) = \lambda^i/i$. Under the I(0) assumption (2), the random $(q+1)$-vector Φ_T has the limit, $\Phi_T \Rightarrow \omega_0\Phi$, where $\Phi_1 = W(1)$ and $\Phi_i = W(1) - (i-1)\int_0^1 s^{i-2}W(s)\mathrm{d}s$ for $i = 2, \ldots, q+1$. Thus $(U_{0T}, D_{0T}(\cdot)) \Rightarrow \omega_0(W(\cdot), D_0(\cdot))$ where $D_0(\lambda) = \Phi' M^{-1} v(\lambda)$, which verifies condition A(i).

(a)(ii) Similar calculations demonstrate that $T\|\delta_t\| = \Phi_T' M_T^{-1}\Phi_T \Rightarrow \omega_0^2 \Phi' M^{-1}\Phi = \mathrm{O}_p(1)$.

(b)(i) $D_{1T}(\lambda) = T^{-1/2}\delta_{[T\lambda]} = \xi_T(\lambda)' M_T^{-1}\Psi_T$, where $\xi_T(\lambda) = \Upsilon_T^{-1}z_{[T\lambda]}$ and $\Psi_T = T^{-3/2}\sum_{t=1}^T \Upsilon_T^{-1}z_t u_t = \int_0^1 \xi_T(s)U_{1T}(s)\mathrm{d}s$. The nonstochastic $(q+1)\times 1$ vector has the limit $\xi_T(\cdot) \to \xi(\cdot)$, where $\xi_i(\lambda) = \lambda^{i-1}$. As a consequence of this result, the I(0) assumption (2), and the continuous mapping theorem, $\Psi_T \Rightarrow \omega_1\Psi$, where $\Psi = \int_0^1 \xi(s)W(s)\mathrm{d}s$. Thus $D_{1T}(\cdot) \Rightarrow \omega_1 \Psi' M^{-1}\xi(\cdot) = \omega_1 D_1(\cdot)$.

(b)(ii) Write $\Delta\delta_t = (\hat\beta - \beta)'\Delta z_t = (\hat\beta - \beta)' z_t^-$, where $z_{tj}^- = t^{j-1} - (t-1)^{j-1}$. Also define $S_T(\lambda) = \{T^{1/2}(\delta_{[T\lambda]} - \delta_{[T\lambda]-1})\}^2$. Then $S_T(t/T) = (T^{1/2}\Delta\delta_t)^2 = \beta_T^{*\prime} M_T^-(t/T)\beta_T^*$, where $M_T^-(t/T) = T^2\Upsilon_T^{-1}z_t^- z_t^{-\prime}\Upsilon_T^{-1}$ and $\beta_T^* = T^{-1/2}\Upsilon_T(\hat\beta - \beta)$. Using the results in the proof of part (b)(i), $\beta_T^* = M_T^{-1}\Psi_T \Rightarrow \omega_1 M^{-1}\Psi$. In addition, let $\xi_T^-(\lambda) = T\Upsilon_T^{-1}z_{[T\lambda]}^-$; then $\xi_T^-(\lambda) \to \xi^-(\lambda) = (0, 1, 2\lambda, \ldots, q\lambda^{q-1})$ and

$M_T^-(\lambda) = \xi_T^-(\lambda)\xi_T^-(\lambda)' \to \xi^-(\lambda)\xi^-(\lambda)'$, both uniformly in λ. It follows that $\|T^{1/2}\Delta\delta_t\| = \int_0^1 S_T(\lambda)\mathrm{d}\lambda \Rightarrow \omega_1^2 \Psi' M^{-1} M^\dagger M^{-1}\Psi$, where $M^\dagger = \int_0^1 \xi^-(\lambda)\xi^-(\lambda)'\mathrm{d}\lambda$. Direct evaluation of $M^\dagger$ shows that $M_{i1}^\dagger = M_{1i}^\dagger = 0$ and $M_{ij}^\dagger = (i-1)\times(j-1)/(i+j-3)$, $i,j \geqslant 2$. □

Proof of Theorem 3

(a) The proof uses Corollary 1 in Bai (1992, Ch. 2). Under the conditions of Theorem 3(a), Bai (1992) shows that $(T^{1/2}\gamma_T(\hat{k}-k),\ T^{1/2}\Upsilon_T(\hat{\theta}-\theta)')' \Rightarrow \omega_0\Phi$, where $\Upsilon_T = \mathrm{diag}(1, T, T)$ and $\Phi = (k^{*\prime}\ \ \theta^{*\prime})'$ is distributed $\mathrm{N}(0,\ \Omega(\tau_0)^{-1})$, where $\Omega(\tau)$ is given in the statement of Theorem 3(a). Under this nesting, $\hat{k} - k \neq \mathrm{O}_p(1)$, but $\hat{\tau} = \hat{k}/T$ is consistent for τ_0: because $T^{1/2}\gamma_T(\hat{k}-k_0) \Rightarrow \omega_0 k^* = \mathrm{O}_p(1)$, $T^{3/2}\gamma_T(\hat{\tau}-\tau_0) \Rightarrow \omega_0 k^*$, but $T^{3/2}\gamma_T \to \infty$ by assumption, so $\hat{\tau} \to^p \tau_0$.

(i) It is useful to express the trend estimation error as the sum of two components, a term arising from the error in estimating θ and a term arising from the error in estimating k: $\delta_t(\hat{k}) = \hat{d}_t(\hat{k}) - d_t(k_0) = z_t(\hat{k})'(\hat{\theta}-\theta) + (z_t(\hat{k}) - z_t(k_0))'\theta$. Thus,

$$D_{0T}(\lambda,\hat{\tau}) = T^{-1/2}\sum_{s=1}^{[T\lambda]} \delta_s(\hat{k}) = \tilde{v}_T(\lambda,\hat{\tau})'\hat{\theta}_T(\hat{\tau}) + \rho_T(\lambda,\hat{\tau}), \tag{A.3}$$

where $\tilde{v}_T(\lambda,\tau) = T^{-1}\sum_{s=1}^{[T\lambda]}\Upsilon_T^{-1} z_s([T\tau]) = \int_{s=0}^{\lambda}\xi_T(s,\tau)\mathrm{d}s$, where $\xi_T(\lambda,\tau) = \Upsilon_T^{-1} z_{[T\lambda]}([T\tau]) = \{1, [T\lambda]/T, ([T\lambda]-[T\tau])\mathbf{1}(\lambda > \tau)/T\}'$, $\hat{\theta}_T(\hat{\tau}) = T^{1/2}\Upsilon_T(\hat{\theta}-\theta)$, and $\rho_T(\lambda,\tau) = T^{-1/2}\sum_{s=1}^{[T\lambda]}(z_s([T\tau]) - z_s([T\tau_0]))'\theta = \int_{s=0}^{\lambda} e_T(s,\tau)\mathrm{d}s$, where $e_T(\lambda,\tau) = T^{1/2}(z_{[T\lambda]}([T\tau]) - z_{[T\lambda]}([T\tau_0]))'\theta$. The three terms $\tilde{v}_T$, $\hat{\theta}_T$, and ρ_T are considered in turn.

$\tilde{v}_T(\lambda,\tau)$. This is a deterministic function of λ and τ. Note that ξ_T is deterministic and has the limit,

$$\xi_T(\cdot,\cdot) \to \xi(\cdot,\cdot), \qquad \xi(\lambda,\tau) = (1, \lambda, (\lambda-\tau)\mathbf{1}(\lambda > \tau))'. \tag{A.4}$$

Because $\tilde{v}_T$ is a continuous functional of ξ_T, it has the limit, $\tilde{v}_T(\cdot,\cdot) \to \tilde{v}(\cdot,\cdot)$, where $\tilde{v}(\lambda,\tau) = (\lambda, \frac{1}{2}\lambda^2, \frac{1}{2}(\lambda-\tau)^2\mathbf{1}(\lambda > \tau))'$. Because ξ_T, and therefore $\tilde{v}_T$, and their limits are continuous in τ, and because $\hat{\tau} \to^p \tau_0$, $\tilde{v}_T(\cdot,\hat{\tau}) \Rightarrow \tilde{v}(\cdot,\tau_0)$.

$\hat{\theta}_T(\hat{\tau})$. From Bai (1992), $\hat{\theta}_T(\hat{\tau}) \Rightarrow \theta^*$ as defined previously.

$\rho_T(\lambda,\tau)$. A direct calculation shows that

$$\begin{aligned} & e_T(t/T, k/T) \\ &= T^{1/2}\gamma_T \operatorname{sign}(k_0 - k)(t - \min(k,k_0))\mathbf{1}(\min(k,k_0) \leqslant t < \max(k,k_0)) \\ &\quad - T^{1/2}\gamma_T(k-k_0)\mathbf{1}(t \geqslant \max(k,k_0)). \end{aligned} \tag{A.5}$$

Although $e_T(\lambda,\hat{\tau})$ is discontinuous in λ in the limit, $\int_{s=0}^{\lambda} e_T(s,\hat{\tau})\mathrm{d}s$ is continuous in λ. The consistency of $\hat{\tau}$, the continuity of $\int_{s=0}^{\lambda} e_T(s,\tau)\mathrm{d}s$, and a straightforward calculation imply that $\rho_T(\lambda,\hat{\tau}) \Rightarrow -\int_0^{\lambda}\omega_0 k^*\mathbf{1}(s > \tau_0)\mathrm{d}s = -\omega_0 k^*(\lambda-\tau_0)\mathbf{1}(\lambda > \tau_0)$.

Combining these three results, we have that $D_{0T}(\cdot,\hat{\tau}) \Rightarrow \omega_0 \tilde{v}(\cdot,\tau_0)'\theta^* - \omega_0 k^*(\lambda-\tau_0)\mathbf{1}(\lambda > \tau_0) = \omega_0 v(\cdot,\tau_0)'\Phi$, where $v(\lambda,\tau)$ and Φ are defined in the statement of the theorem.

(ii) Define $\zeta_T(\lambda) = T^{1/2}\delta_{[T\lambda]}$, so that $T\|\delta_t\| = \int_0^1 \zeta_T(\lambda)^2 \mathrm{d}\lambda$. Using previously defined expressions and results, we have $\int_0^1 \zeta_T(\lambda)^2\mathrm{d}\lambda = \int_0^1 \{\hat{\theta}'_T \xi_T(\lambda,\hat{\tau}) - e_T(\lambda,\hat{\tau})\}^2 \mathrm{d}\lambda \Rightarrow \omega_0^2 \int_0^1 \{\theta^{*\prime}\xi(\lambda,\tau_0) - k^*\mathbf{1}(\lambda > \tau_0)\}^2\mathrm{d}\lambda = \omega_0^2 \Phi' \{\int_0^1 \xi^\dagger(s,\tau_0)\xi^\dagger(s,\tau_0)' \times \mathrm{d}s\}\Phi$, where $\xi^\dagger(s,\tau) = (-\mathbf{1}(s>\tau), \xi(s,\tau)')'$, which is the desired result.

(b)(i) Let $F_{1T}(\lambda,\tau) = T^{-1/2}\delta_{[T\lambda]}([T\tau])$, so that $D_{1T}(\lambda) = F_{1T}(\lambda,\hat{\tau})$. The strategy of the proof is first to obtain a limiting representation for the process $F_{1T}(\cdot,\cdot)$, which will be continuous in its two arguments, next to obtain a limiting representation for $\hat{\tau}$, and then to use these two results and the continuous mapping theorem to obtain the desired limiting representation for $D_{1T}(\cdot)$.

Using terms defined in the proof of part (a), write F_{1T} as

$$F_{1T}(\lambda,\tau) = \xi_T(\lambda,\tau)' T^{-1}\hat{\theta}_T(\tau) + T^{-1} e_T(\lambda,\tau).$$

It was previously shown that $\xi_T \to \xi$. Next consider $T^{-1}e_T(\lambda,\tau)$. From (A.5),

$$\begin{aligned}
&|T^{-1}e_T(t/T, k/T)| \\
&= |T^{-1/2}\gamma_T(t-\min(k,k_0))|\mathbf{1}(\min(k,k_0) \leqslant t < \max(k,k_0)) \\
&\quad + |T^{-1/2}\gamma_T(k-k_0)|\mathbf{1}(t \geqslant \max(k,k_0)) \\
&\leqslant |T^{-1/2}\gamma_T(k-k_0)| \leqslant |T^{1/2}\gamma_T|,
\end{aligned}$$

where the two inequalities are uniform in t and k and the second follows from $|k-k_0| \leqslant T$. By assumption, $T^{1/2}\gamma_T \to 0$; thus $T^{-1}e_T(\cdot,\cdot) \to 0$.

Next consider $T^{-1}\hat{\theta}_T(\tau)$. Now $T^{-1}\hat{\theta}_T(\tau) = T^{-1/2}\Upsilon_T(\hat{\theta}([T\tau]) - \theta) = M_T(\tau)^{-1}\{N_T(\tau) + \Psi_T(\tau)\}$, where $N_T(\tau) = T^{-3/2}\sum_{t=1}^T \Upsilon_T^{-1} z_t([T\tau])(z_t([T\tau_0]) - z_t([T\tau]))'\theta$, $\Psi_T(\tau) = T^{-3/2}\sum_{t=1}^T \Upsilon_T^{-1} z_t([T\tau])u_t$, and $M_T(\tau) = T^{-1}\Upsilon_T^{-1} \times \sum_{t=1}^T z_t([T\tau])z_t([T\tau])'\Upsilon_T^{-1}$. Direct calculation reveals that $M_T(\cdot) \to M(\cdot)$, where $M(\cdot) = \int_0^1 \xi(s,\tau)\xi(s,\tau)'\mathrm{d}s$ as defined in the statement of Theorem 3. N_T can be rewritten $N_T(\tau) = \int_0^1 \xi_T(s,\tau)(T^{-1}e_T(s,\tau))\mathrm{d}s$; the results $\xi_T \to \xi$ and $T^{-1}e_T \to^p 0$ imply that $N_T \to^p 0$. Under the I(1) assumption, the remaining term, $\Psi_T(\tau)$, has the limit $\Psi_T(\cdot) = \int_0^1 \xi_T(s,\cdot)U_{1T}(s)\mathrm{d}s \Rightarrow \omega_1 \int_0^1 \xi(s,\cdot)W(s)\mathrm{d}s \equiv \omega_1\Psi(\cdot)$. Combining these various expressions, we have $T^{-1}\hat{\theta}_T(\cdot) \Rightarrow \omega_1 M(\cdot)^{-1}\Psi(\cdot)$, so $F_{1T}(\cdot,\cdot) \Rightarrow \omega_1 F_1(\cdot,\cdot)$, where $F_1(\lambda,\tau) = \xi(\lambda,\tau)'M(\tau)^{-1}\Psi(\tau)$.

The next step of the proof is to obtain a limiting representation for $\hat{\tau}$. By definition, $\hat{\tau}$ solves (13), which can be rewritten as the problem of minimizing $S_T(\tau)$ over $\tau_{\min} \leqslant \tau \leqslant \tau_{\max}$, where $S_T(\tau) = T^{-2}\sum_{t=1}^T \hat{u}_t([T\tau])^2$. Let $\delta_t(k) = \hat{d}_t(k) - d_t(k_0)$, where $\hat{d}_t(k) = \hat{\theta}(k)'z_t(k)$. Then $S_T(\tau) = T^{-2}\sum_{t=1}^T\{y_t - \hat{d}_t([T\tau])\}^2 = T^{-2}\sum_{t=1}^T\{u_t - \delta_t([T\tau])\}^2 = T^{-1}\sum_{t=1}^T\{U_{1T}(t/T) - F_{1T}(t/T,[T\tau]/T)\}^2 = \int_{s=0}^1\{U_{1T}(s) - F_{1T}(s,[T\tau]/T)\}^2\mathrm{d}s$. It follows that $S_T(\cdot) \Rightarrow S(\cdot)$, where $S(\tau) = \omega_1^2\int_{s=0}^1\{W(s) - F_1(s,\tau)\}^2\mathrm{d}s$. Thus $\hat{\tau} \Rightarrow \tau^*$, where τ^* has the distribution

$\operatorname{argmin}_{\tau\in[\tau_{\min},\tau_{\max}]}\int_0^1\{W(s)-F_1(s,\tau)\}^2\mathrm{d}s$. Because F_1 is continuous in τ, it follows that $D_{1T}(\cdot)=F_{1T}(\cdot,\hat\tau)\Rightarrow\omega_1F_1(\cdot,\tau^*)\equiv\omega_1D_1(\cdot)$.

(b)(ii) By direct calculation,

$$\begin{aligned}
&|\eta_T(t/T,k/T)|\\
&=|\delta_t(k)-\delta_{t-1}(k)|\\
&=|(\hat\theta(k)-\theta)'\{z_t(k)-z_{t-1}(k)\}+\theta'\{z_t(k)-z_t(k_0)-z_{t-1}(k)+z_{t-1}(k_0)\}|\\
&\leqslant|\hat\beta(k)-\beta|+|\hat\gamma_T(k)-\gamma_T|\mathbf{1}(t>k)+|\gamma_T|\mathbf{1}(\min(k,k_0)\leqslant t<\max(k,k_0)).
\end{aligned}\tag{A.6}$$

In the proof of part (b)(i) it was shown that $T^{-1}\hat\theta_T(\cdot)\Rightarrow M(\cdot)^{-1}\Psi(\cdot)$, where $T^{-1}\hat\theta_T(\tau)=T^{-1/2}\Upsilon_T(\hat\theta([T\tau])-\theta)=\{T^{-1/2}(\hat\alpha([T\tau])-\alpha),T^{1/2}(\hat\beta([T\tau])-\beta),T^{1/2}(\hat\gamma_T([T\tau])-\gamma_T)\}$. Thus, in particular, $\sup_\tau|\hat\beta([T\tau])-\beta|\to^p0$ and $\sup_\tau|\hat\gamma_T([T\tau])-\gamma_T|\to^p0$, so the first two terms in (A.6) converge to zero uniformly in $\lambda=t/T$, $\tau=k/T$. In addition, $\gamma_T\to0$ by assumption, so the final term in (A.6) vanishes. Thus $\sup_{\lambda,\tau}|\eta_T(\lambda,\tau)|\to^p0$, so $\eta_T(\cdot,\cdot)\to^p0$ as desired. □

Proof of Theorem 4

The proofs of parts (a) and (b) are, respectively, modifications of the proofs of Theorem 3(a) and 3(b), and notation and expressions refer to those proofs.

(a)(i) In the notation of the proof of Theorem 3(a), because $\gamma_T=0$, $\rho_T(\lambda,\tau)=0$ identically, so $D_{0T}(\lambda,\tau)=\tilde v_T(\lambda,\tau)'\theta_T(\tau)$. As in the proof of Theorem 3, $\tilde v_T(\cdot,\cdot)\to\tilde v(\cdot,\cdot)$. Because $\gamma_T=0$, $\hat\theta_T(\hat\tau)=M_T(\hat\tau)^{-1}\Phi_T(\hat\tau)$, where $\Phi_T(\tau)=T^{-1/2}\Upsilon_T^{-1}\sum_{t=1}^T z_t([T\tau])u_t$. It follows from the FCLT that $\Phi_T(\cdot)\Rightarrow\omega_0\Phi(\cdot)$ as defined in the statement of Theorem 4; thus $\hat\theta_T(\cdot)\Rightarrow\theta^*(\cdot)$, where $\theta^*(\tau)=\omega_0M(\tau)^{-1}\tilde\Phi(\tau)$, from which it follows that $D_{0T}(\cdot,\cdot)\Rightarrow\omega_0D_0(\cdot,\cdot)$, where $D_0(\lambda,\tau)=\tilde v(\lambda,\tau)'\theta^*(\tau)$.

Because $D_0(\lambda,\tau)$ is continuous in τ, $D_{0T}(\cdot,\hat\tau)\Rightarrow D_0(\cdot,\tau^\dagger)$, where $\tau^\dagger$ is the limiting representation for $\hat\tau$ (obtained jointly with the other expressions comprising D_0). Because $\|u_t\|$ does not depend on τ, the solution to the problem $\min_{\tau\in[\tau_{\min},\tau_{\max}]}\|\hat u_t([T\tau])\|$ is equivalent to the solution to the problem $\max_{\tau\in[\tau_{\min},\tau_{\max}]}H_T(\tau)$, where $H_T(\tau)=T(\|u_t\|-\|\hat u_t([T\tau])\|)$, where $\hat u_t(k)=y_t-\hat d_t(k)$. A standard calculation reveals that, when $\gamma_T=0$, $H_T(\cdot)\Rightarrow H(\cdot)=\theta^*(\cdot)'M(\cdot)\theta^*(\cdot)=\tilde\Phi(\cdot)'M(\cdot)^{-1}\tilde\Phi(\cdot)$. By the continuity of the distribution of the argmax, the limiting representation for $\hat\tau$ as $\operatorname{argmax}_{\tau\in[\tau_{\min},\tau_{\max}]}H(\tau)$ follows.

(a)(ii) By direct calculation, $T\|\delta_t([T\tau])\|=H_T(\tau)$, so $T\|\delta_t\|=H_T(\hat\tau)\Rightarrow H(\tau^\dagger)$, the desired result.

(b)(i) The proof of Theorem 3b(i) applies directly, with the simplifications that $e_T = 0$ and $N_T = 0$.

In particular, the key result that $T^{-1}\theta_T^*(\cdot) \Rightarrow M(\cdot)^{-1}\Psi(\cdot)$ still holds.

(b)(ii) This follows from the proof of Theorem 3b(ii), using $T^{-1}\theta_T^*(\cdot) \Rightarrow M(\cdot)^{-1}\Psi(\cdot)$. □

References

Andrews, D.W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, 817–858.

Bai, J., 1992, Econometric estimation of structural change, Ph.D. dissertation (Department of Economics, University of California, Berkeley, CA).

Banerjee, A., R.L. Lumsdaine, and J.H. Stock, 1992, Recursive and sequential tests of the unit root and trend break hypotheses: Theory and international evidence, Journal of Business and Economic Statistics 10, 271–288.

Bhargava, A., 1986, On the theory of testing for unit roots in observed time series, Review of Economic Studies 53, No. 174, 369–384.

Bierens, Herman J., 1989, Testing stationarity against the unit root hypothesis, SERIE research memorandum 1989-81 (Vrije Universiteit Amsterdam, Amsterdam).

Bierens, Herman J. and S. Guo, 1993, Testing stationarity and trend stationarity against the unit root hypothesis, Econometric Reviews 12, 1–32.

Brown, R.L., J. Durbin, and J.M. Evans, 1975, Techniques for testing the constancy of regression relationships over time with comments, Journal of the Royal Statistical Society B 37, 149–192.

Christiano, L.C. and M. Eichenbaum, 1990, Unit roots in GNP: Do we know and do we care?, Carnegie–Rochester Conference Series on Public Policy 32, 7–62.

DeJong, D.N and C.H. Whiteman, 1991, Reconsidering 'Trends and random walks in macroeconomic time series', Journal of Monetary Economics 28, 221–254.

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, no. 366, 427–431.

Durbin, J. and G.S. Watson, 1950, Testing for serial correlation in least squares regression I, Biometrika 37, 409–428.

Elliott, G., T.J. Rothenberg, and J.H. Stock, 1992, Efficient tests for an autoregressive unit root, NBER technical working paper 130.

Ethier, S.N. and T.G. Kurtz, 1986, Markov processes: Characterization and convergence (Wiley, New York, NY).

Gardner, L.A., 1969, On Detecting changes in the mean of normal variates, Annals of Mathematical Statistics 40, 116–126.

Hall, P. and C.C. Heyde, 1980, Martingale limit theory and its applications (Academic Press, New York, NY).

Herrndorf, N.A., 1984, A functional central limit theorem for weakly dependent sequences of random variables, Annals of Probability 12, 141–153.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, Journal of Econometrics 54, 159–178.

Lo, A.W., 1991, Long-term memory in stock market prices, Econometrica 59, 1279–1314.

MacNeill, I.B., 1978, Properties of sequences of partial sums of polyomial regression residuals with applications to tests for change of regression at unknown times, Annals of Statistics 6, 422–433.

Mandelbrot, B.B., 1975, Limit theorems on the self-normalized range for weakly and strongly dependent processes, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 31, 271–285.

Mandelbrot, B.B. and J.W. Van Ness, 1968, Fractional Brownian motions, fractional noise and applications, SIAM Review 10, 422–437.

Nabeya, S and K. Tanaka, 1988, Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative, Annals of Statistics 16, 218–235.

Nelson, C.R. and C.I. Plosser, 1982, Trends and random walks in macro-economic time series: Some evidence and implications, Journal of Monetary Economics 10, 139–162.

Ouliaris, S., J.Y. Park, and P.C.B. Phillips, 1989, Testing for a unit root in the presence of a maintained trend, in: B. Raj, ed., Advances in econometrics and modelling (Kluwer, Dordrecht, 7–28.

Park, J., 1990, Testing for unit roots and cointegration by variable addition, in: T.B. Fomby and G.F. Rhodes, eds., Advances in econometrics (JAI Press, Greenwich, CT).

Park, J. and C. Choi, 1988, A new approach to testing for a unit root, CAE working paper 88-23 (Cornell University, Ithaca, NY).

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

Perron, P., 1989, The great crash, the oil price shock and the unit root hypothesis, Econometrica 57, 1361–1401.

Perron, P., 1991, A test for changes in a polynomial trend function for a dynamic time series, Manuscript (Princeton University, Princeton, NJ).

Phillips, P.C.B., 1991a, To criticize the critics: An objective Bayesian analysis of stochastic trends, Journal of Applied Econometrics 6, 333–364.

Phillips, P.C.B., 1991b, Spectral regression for cointegrated time series, in: W. Barnett, J. Powell, and G. Tauchen, eds., Nonparametric and semiparametric methods in econometrics and statistics (Cambridge University Press, Cambridge).

Phillips, P.C.B. and W. Ploberger, 1991, Time series modeling with a Bayesian frame of reference: I. Concepts and illustrations, Manuscript (Yale University, New Haven, CT).

Picard, D., 1985, Testing and estimating change-points in time series, Advances in Applied Probability 176, 841–867.

Rappoport, P. and L. Reichlin, 1989, Segmented trends and non-stationary time series, Economic Journal 99, 168–177.

Saikkonen, P. and R. Luukkonen, 1993, Testing for moving average unit root in autoregressive integrated moving average models, Journal of the American Statistical Association 88, 596–601.

Sargan, J.D. and A. Bhargava, 1983, Testing residuals from least squares regression for being generated by the Gaussian random walk, Econometrica 51, 153–174.

Shively, T.S., 1988, An exact text for a stochastic coefficient in a time series regression model, Journal of Time Series Analysis 9, 81–88.

Sims, C.A., 1988, Bayesian skepticism on unit root econometrics, Journal of Economic Dynamics and Control 12, 463–474.

Sims, C.A. and H. Uhlig, 1991, Understanding unit rooters: A helicopter tour, Econometrica 59, 1591–1600.

Sowell, F., 1991, On DeJong and Whiteman's Bayesian inference for the unit root model, Journal of Monetary Economics 28, 255–264.

Stock, J.H., 1991, Confidence intervals for the largest autoregressive root in U.S. economic time series, Journal of Monetary Economics 28, 435–460.

Stock, J.H., 1992, Deciding between I(1) and I(0), NBER technical working paper 121.

Zivot, E. and D.W.K. Andrews, 1992, Further evidence on the great crash, the oil price shock, and the unit root hypothesis, Journal of Business and Economic Statistics 10, 251–270.