NBER TECHNICAL WORKING PAPER SERIES

DECIDING BETWEEN I(1) AND I(0)

James H. Stock

Technical Working Paper No. 121

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1992

The author thanks T.W. Anderson, Clive Granger, Lars Hansen, Herman van Dijk, Emanual Parzen, Mark Watson, 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, 4 and 6). Graham Elliott provided excellent research assistance. This research was supported in part by the National Science Foundation (grant no. SES-89-10601) and by the Sloan Foundation. This paper is part of NBER's research program in Monetary Economics. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #121
June 1992

DECIDING BETWEEN I(l) AND I(O)

ABSTRACT

This paper proposes 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)) for general I(0) and I(1) processes. These procedures entail the evaluation of the asymptotic likelihoods of certain statistics under the I(0) and I(1) hypotheses. These likelihoods do not depend on nuisance parameters describing short-run dynamics and diverge asymptotically, so their ratio provides a consistent basis for classifying a process as I(1) or I(0). Bayesian inference can be performed by placing prior mass only on the point hypotheses "I(0)" and "I(1)" without needing to specify parametric priors within the classes of I(0) and I(1) processes; the result is posterior odds ratios for the I(0) and I(1) hypotheses.

These procedures are developed for general polynomial and piecewise linear detrending. When applied to the Nelson-Plosser data with linear detrending, they largely support the original Nelson-Plosser inferences. With piecewise-linear detrending these data are typically uninformative, producing Bayes factors that are close to one.

James H. Stock
Kennedy School of Government
Harvard University
Cambridge, MA 02138
and NBER

1. Introduction

It is often useful to have a sense of whether a univariate economic time series is better modeled as being integrated of order one (is I(1)) or of order zero (I(0)). For example, this distinction can be important in the subsequent application of asymptotic theory to obtain approximate distributions of estimators and test statistics. Alternatively, different economic theories can have different implications for the trending 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 (1989), Sims (1988), Sims and Uhlig (1988), Sowell (1991), and Phillips (1991a), and the discussions of Phillips (1991a) in the October-December 1991 issue of the *Journal of Applied Econometrics*.[1] However, this work has the important practical drawback of relying 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 some authors have described these priors as "flat", 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) that avoids the problem of making explicit parametric assumptions about priors within the

I(1) or I(0) class under general assumptions on deterministic trends. These procedures are based on a scaled cumulative sum process of the detrended data, V_T. The process V_T has the following properties: (i) $N_{iT}^{-\frac{1}{2}} V_T$ has a (classical) asymptotic distribution that depends on no unknown nuisance parameters; and (ii) for I(0), $N_{iT} = 1$, for I(1), $N_{iT} = N_T \to \infty$. 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 under the I(1) and under the I(0) hypothesis, 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. Of course, consistent I(0)/I(1) classification procedures already exist, at least in theory; such a rule can be constructed simply by using a consistent test which is similar under either the I(0) or I(1) null, for example the augmented Dickey-Fuller (1979) t-statistic, and critical values which depend on the sample size and which tend to infinity at a suitable rate. The principal advantage of the classification scheme proposed here is that it incorporates point priors in an explicitly Bayesian framework and thereby produces interpretable posterior odds ratios.

The statistics based on the cumulative sum V_T are closely related to statistics studied in three related literatures: cusum statistics that are used to test for structural breaks (Gardner (1969), Brown, Durbin and Evans (1975), and MacNeill (1978); see Perron (1991) for a recent review); LM tests for a random walk drift under the null of a stationary time series model (and Nabeya and Tanaka (1988); see Harvey (1989) for a review of the earlier literature); and recently developed tests of the I(0) null against the I(1) alternative, sometimes cast as tests of the null of a unit moving average root in the first difference of the series (Park (1990), Park and Choi (1988), Bierens (1989), Saikkonen and Luukkonen (1990), and Kwiatkowski, Phillips and Schmidt (1990), and the references therein).

The procedures proposed here are developed for two trend specifications. 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) and Perron (1991)). Perron (1989) and Rappoport and Reichlin (1989) suggested that an alternative to the unit root model for many aggregate economic time series is that 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 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 considered therefore is the broken trend model in which the break date is treated as unknown.

In a paper closely related to this one, Phillips and Ploberger (1991) have also recently proposed the use of Bayesian posterior odds ratios for constructing 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

A. General Results

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

(2.1) $$y_t = d_t + u_t$$

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) follows recent convention: a purely stochastic process is said to be I(d) if the process formed from the partial sums of its d-th difference, scaled by $T^{-½}$, obeys a functional central limit theorem and converges to a constant times a standard Brownian motion. Let $U_{0T}(\lambda)=T^{-½}\sum_{s=1}^{[T\lambda]}u_s$ and $U_{1T}(\lambda)=T^{-½}u_{[T\lambda]}$, where $[\cdot]$ denotes the greatest lesser integer function, and let $\gamma_x(j)=cov(x_t,x_{t-j})$ for a second order stationary process x_t. Also let "=>" 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:

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

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

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 assume 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). In stating $\gamma_u(j)=cov(u_t,u_{t-j})$ (I(0) case) or $\gamma_{\Delta u}(j)=cov(\Delta u_t,\Delta u_{t-j})$ (I(1) case), we are further assuming that I(0) variables are second order stationary.

The basis of the proposed statistics is 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,

$$(2.4) \qquad V_T(\lambda) = \hat{\omega}^{-1} T^{-\frac{1}{2}} \sum_{s=1}^{[T\lambda]} y_s^d$$

where

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

where ℓ_T is an increasing sequence of integers and $k(\cdot)$ is a kernel satisfying: $k(x)=0$ for $|x|\geq 1$; $k(x)=k(-x)$; $0<k(x)\leq 1$ for $|x|<1$; $k(0)=1$; and $\ell^{-1}\sum_{u=1}^{\ell} k(u/\ell)\geq \underline{k}$ for all $\ell>1$, where $\underline{k}>0$. 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 (2.4), 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, $D_{0T}(\lambda) = T^{-\frac{1}{2}}\sum_{s=1}^{[T\lambda]}\delta_s$, and $D_{1T}(\lambda) = T^{-\frac{1}{2}}\delta_{[T\lambda]}$. 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) $\ell_T^2 \|\delta_t\| \xrightarrow{p} 0$.

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 $Y_{1T}^d(\lambda) = T^{-1/2} y_{[T\lambda]} = U_{1T}(\lambda) - D_{1T}(\lambda)$; condition B(i) implies that, if u_t is I(1), then $Y_{1T}^d(\cdot) \Rightarrow \omega_0 W_1^d(\cdot)$, where $W_1^d(\cdot) = W(\cdot) - D_1(\cdot)$.

Under these conditions, because $\ell_T \to \infty$, 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, $\|\Delta\delta_t\| \xrightarrow{p} 0$ 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 and Schmidt (1990) (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 generalizes their results to general trends and provides limiting representations for the statistic $V_T \in D[0,1]$. Let $N_T = T/\sum_{m=-\ell_T}^{\ell_T} k(m/\ell_T)$.

Theorem 1. Suppose $\ell_T^2 \ln T/T \to 0$, $\ell_T \to \infty$, and assumptions 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 $V_1^d(\lambda) = \int_0^\lambda W_1^d(s)ds / \{\int_0^1 W_1^d(s)^2 ds\}^{1/2}$.

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).

B. 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 $C[0,1] \rightarrow \mathfrak{R}^1$; (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

$$(2.5) \qquad B_T = f_1(\phi_T - \ln N_T)/f_0(\phi_T).$$

It is readily seen that (2.5) 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 $\phi_T = \phi(V_T) \Rightarrow \phi(W_0^d) = O_p(1)$, so $f_0(\phi_T) = O_p(1)$ but $f_1(\phi_T - \ln N_t) \xrightarrow{p} 0$; thus $B_T \xrightarrow{p} 0$ and I(0) is chosen with probability one. On the other hand, if u_t is I(1), then $\phi_T - \ln N_T \Rightarrow \phi(V_1^d) = O_p(1)$, but $f_0(\phi_T) \xrightarrow{p} 0$; thus $1/B_T \xrightarrow{p} 0$ 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,

$\phi(N_T^{-\frac{1}{2}}V_T)$ 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 and Choi (1988), Saikkonen and Luukkonen (1990), and Kwiatkowski, Phillips and Schmidt (1991).

The likelihood ratio (2.5) 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 (2.5),

$$(2.6) \qquad P_T = (\pi_1/\pi_0)B_T.$$

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:

$$(2.7a) \qquad \phi_1(g) = \ln\{\textstyle\int_0^1 g(s)^2 ds\}$$

$$(2.7b) \qquad \phi_2(g) = \ln\{(\sup_{s\in(0,1)} g(s) - \inf_{s\in(0,1)} g(s))^2\}$$

$$(2.7c) \qquad \phi_3(g) = \ln\{\textstyle\sum_{j=1}^J |\int_0^1 g(s)e^{-i2\pi js}ds|^2\}$$

The statistic ϕ_{1T} and close variants have been studied in several related literatures. In terms of the original data, $\phi_{1T} = \phi_1(V_T) = \ln\{\hat{\omega}^{-2}T^{-1}\sum_{t=1}^T(T^{-\frac{1}{2}}\sum_{s=1}^t y_s^d)^2\}$. One motivation for using ϕ_{1T} comes from recognizing that, with no trend or detrending, $\tilde{\phi}_{1T} = T^{-1}\sum_{t=1}^T(T^{-\frac{1}{2}}\sum_{s=1}^t u_s)^2/\hat{\gamma}_u(0)$ (which is appropriate if $d_t = 0$ and u_t is serially uncorrelated) is the Sargan-Bhargava (1983) 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 Gaussian random walk. If the rejection region is the right tail, $\tilde{\phi}_{1T}$ accordingly can be interpreted as testing for 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 Saikkonen and Luukkonen (1990) (to test for a unit MA root). Saikkonen and Luukkonen (1990) proposed a generalization of $\tilde{\phi}_{1T}$ test I(0) vs. I(1) with ARMA errors, although their generalization differs slightly from that examined here. Kwiatkowski, Phillips and Schmidt (1990) proposed the generalization $\exp(\phi_{1T})$ as a test of the general I(0) null against the I(1) alternative. Gardner (1969) and MacNeill (1978) studied $\tilde{\phi}_{1T}$ as a test for a broken time trend; Perron (1991) extended their statistic to general error terms and proposed $\exp(\phi_{1T})$. Both Kwiatkowski, Phillips and Schmidt (1990) and Perron (1991) derived asymptotic representations for $\exp(\phi_{1T})$ under the general I(0) and I(1) hypotheses.

The statistic ϕ_{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 (2.7c) and applied it to financial time series data.

The statistic ϕ_{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, ϕ_{2T} 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.

A. Polynomial detrending by OLS.

Consider the polynomial time trend,

$$d_t = z_t'\beta \tag{3.1}$$

where $z_t=(1, t, t^2, \ldots, t^q)$, 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 (3.1) the detrended data are $y_t^d = y_t - z_t'\hat{\beta} = u_t - \delta_t$, where $\delta_t = z_t' M_T^{-1} \sum_{t=1}^T z_t u_t$.

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

Theorem 2. If d_t is given by (3.1) 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 $D_0(\lambda) = \Phi' M^{-1}\nu(\lambda)$, where Φ, M, and ν are respectively (q+1)×1, (q+1)×(q+1), and (q+1)×1, and $\Phi_i = W(1) - (i-1)\int_0^1 s^{i-2}W(s)ds$, $M_{ij} = 1/(i+j-1)$, and $\nu_i(\lambda) = \lambda^i/i$.

(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 $\Psi = \int_0^1 \xi(s) W(s) ds$.

(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 result stated in Theorem 1 (that is, $\ell_T^2 \ln T/T \to 0$ and $T\|\delta_t\| \Rightarrow \omega_0^2 \Phi' M^{-1} \Phi$ implies that $\ell_T^2\|\delta_t\| \xrightarrow{p} 0$ as desired). Part b(ii) implies $\|\Delta\delta_t\| \xrightarrow{p} 0$, which verifies condition B(ii).

In the case of a linear time trend, an alternative estimator for the slope coefficient is the "first difference" estimator, $(y_T - y_1)/T$. This detrending procedure is used by Bhargava (1986) in his construction of a locally most powerful invariant test of the unit root null against the stationary alternative, under the maintained assumption of i.i.d. Gaussian errors and a linear time trend component. As Watson (1992) has pointed out, if u_t is I(1) then this results in a trend estimator that has a standard limiting representation, but if u_t is I(0), the large-sample distribution of the estimated trend depends on the exact marginal distributions of u_T and u_1, which in general will not be Gaussian. Thus detrending procedures which use this estimator, such as Bhargava's (1986), will producing limiting representations for V_T in the I(0) case that depend on the nuisance parameters describing the marginal distribution of u_t, making such procedures impractical for our purposes.

B. Piecewise-linear ("broken-trend") detrending

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

$$d_t(k_0) = \alpha + \beta t + \gamma_T(t-k_0)\mathbf{1}(t>k_0) = z_t(k_0)'\theta \tag{3.2}$$

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. Following Picard (1985), the sequence of coefficients on the trend-break term, γ_T, is assumed to be local to zero and to satisfy, (3.3) $$T^{½}|\gamma_T| \to 0, \; T^{3/2}|\gamma_T| \to \infty.$$

The estimated trend is

(3.4) $$\hat{d}_t(\hat{k}) = z_t(\hat{k})'\hat{\theta}(\hat{k})$$

where

(3.5) $$\hat{\theta}(k) = (\sum_{t=1}^{T} z_t(k)z_t(k)')^{-1}\sum_{t=1}^{T} z_t(k)y_t$$

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

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

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

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

Theorem 3. Let d_t be given by (3.2), where γ_T satisfies (3.3), and let $d_t(\hat{k})$ be given by (3.4) - (3.6).

(a) If y_t is I(0), then:

(i) $(U_{0T}, D_{0T}) \Rightarrow \omega_0(W, D_0)$ where $D_0(\lambda) = \nu(\lambda,\tau_0)'\Phi(\tau_0)$, where $\nu(\lambda,\tau) = (\lambda, ½\lambda^2, ½(\lambda-\tau)^2 1(\lambda>\tau), -(\lambda-\tau)1(\lambda>\tau))'$ and the 4×1 random vector $\Phi(\tau_0)$ is a functional of W that is distributed $N(0, \Omega(\tau_0)^{-1})$, where $\Omega_{11} = 1-\tau$, $\Omega_{12} = -(1-\tau)$, $\Omega_{13} = -½(1-\tau^2)$, $\Omega_{14} = -½(1-\tau)^2$, $\Omega_{22} = 1$, $\Omega_{23} = ½$, $\Omega_{24} = ½(1-\tau)^2$, $\Omega_{33} = 1/3$, $\Omega_{34} = (2-3\tau+\tau^3)/6$, and $\Omega_{44} = (1-\tau)^3/3$.

(ii) $T\|\delta_t\| \Rightarrow \omega_0^2\Phi(\tau_0)'\{\int_0^1 \xi^\dagger(s,\tau_0)\xi^\dagger(s,\tau_0)'ds\}\Phi(\tau_0)$ where $\xi^\dagger(s,\tau) =$

$(\xi(s,\tau)', (s-\tau)1(s>\tau))'$ and where $\xi(s,\tau) = (1, s, (s-\tau)1(s>\tau))'$.

(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 $\Psi(\tau) = \int_0^1\xi(s,\tau)W(s)ds$, $M(\tau)$ and $\xi(\lambda,\tau)$ are defined in part a(ii) of this theorem, and τ^* has the distribution, $\mathrm{argmin}_{\tau\in[\tau_{min},\tau_{max}]}\int_0^1\{W(s)-F_1(s,\tau)\}^2ds$.

(ii) Let $\eta_T(\lambda,\tau) = \delta_{[T\lambda]}([T\tau]) - \delta_{[T\lambda]-1}([T\tau])$. Then $\eta_T(\cdot,\cdot) \overset{p}{\to} 0$.

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 $\ell_T^2/T \to 0$, 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 (3.2), 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 (3.2), let $d_t(k)$ be given by (3.4) - (3.6), and suppose that the true value of γ_T is 0.

(a) If y_t is I(0), then:

(i) $(U_{0T}, D_{0T}) \Rightarrow \omega_0(W, D_0)$ where $D_0(\lambda) = \bar{\nu}(\lambda,\tau^\dagger)'\Phi(\tau^\dagger)$, where $\Phi(\tau) = \int_0^1\xi(s,\tau)dW(s)$, $\bar{\nu}(\lambda,\tau) = (\lambda, \tfrac{1}{2}\lambda_2, \tfrac{1}{2}(\lambda-\tau)^2 1(\lambda>\tau))'$, and $\tau^\dagger$ has the distribution $\mathrm{argmax}_{\tau\in[\tau_{min}, \tau_{max}]}\Phi(\tau)'M(\tau)^{-1}\Phi(\tau)$, where $\xi(\lambda,\tau)$ and $M(\tau)$ are defined in theorem 3(a).

(ii) $T\|\delta_t\| \Rightarrow \Phi(\tau^\dagger)'M(\tau^\dagger)^{-1}\Phi(\tau^\dagger)$.

(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 (2.5) will depend on γ and, if $\gamma \neq 0$, on τ_0. Our proposed solution is to estimate $f_0(x)$ by $f_0(x;\hat{\gamma}/\hat{\sigma}_u,\hat{\tau})$, where $f_0(x;\hat{\gamma}/\hat{\sigma}_u,\hat{\tau})$ is the density of $W_0^d(\cdot;\hat{\gamma}/\hat{\sigma},\hat{\tau})$, which in turn 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 $d_t = (\hat{\gamma}/\hat{\sigma}_u)(t-[T\hat{\tau}])1(t>[T\hat{\tau}])$, where $\hat{\gamma}$, $\hat{\sigma}_u$ and $\hat{\tau}$ are the OLS estimates of γ, σ_u and τ.

Assuming that u_t is is fact I(0), this procedure is justified in two steps, first for γ_T local to zero, next for γ_T=0. First suppose that γ is local to zero as in theorem 3; to be concrete, let γ_T=b/T, where b is a constant, a sequence that satisfies (3.3). Under this local nesting $\hat{b} \xrightarrow{p} b$, $\hat{\sigma} \xrightarrow{p} \sigma$, and $\hat{\tau} \xrightarrow{p} \tau_0$ (Picard (1985), Bai (1991); see the proof to theorem 3), $W_1^d(\lambda;\gamma,\tau)$ is continuous in τ, and W_1^d does not depend explicitly on b beyond the maintained assumption that b≠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 $W_1^d(\cdot;\hat{\gamma}/\hat{\sigma}_u,\hat{\tau})$. Next suppose that γ=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 → 0 and moreover $\hat{b} \xrightarrow{p} 0$.[3] It follows that $W_1^d(\cdot;\hat{\gamma}/\hat{\sigma}_u,\hat{\tau})$ has the limiting distribution in theorem 4, so that for b=0 or b≠0 this procedure yields a (pointwise) 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) 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 γ_T=b/T (where b might be zero), then the distribution of $\phi(W_0^d(\cdot;\hat{\gamma}/\hat{\sigma}_u,\hat{\tau}))$ converges to the distribution resulting from Theorem 3(a), with the random variable $\tau^\dagger$ (defined in theorem 3(b)) replacing τ_0. This follows from the fact that, if u_t is I(0), $T\hat{\gamma}/\hat{\sigma}_u = O_p(1)$ 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 (3.3) and therefore will satisfy the conditions of theorem 3(a). Although $f_0(x;\hat{\gamma}/\hat{\sigma}_u,\hat{\tau})$ for fixed x will asymptotically be a random variable in this case (because $\hat{\tau} \Rightarrow \tau^\dagger$), the posterior odds and Bayes ratios will still yield consistent decision procedures.[4]

4. Numerical Issues and Finite Sample Performance

A. 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 P_T were evaluated numerically using a kernel density estimator of the likelihood. The approach was first to produce a matrix of discretized 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 $u_t = \epsilon_t$, 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 $\Delta u_t = \epsilon_t$, ϵ_t i.i.d. N(0,1), T = 100, and $\phi(N_T^{-\frac{1}{2}}V_T)$ was computed (with ℓ_T=0) and saved. Both cases entailed 8000 Monte Carlo replications. Given a realization $\hat{\phi}_T$, the densities $f_0(\phi_T)$ and $f_1(\phi_T)$ 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 estimated by OLS, first with γ=0 and second with γ=0.5 and τ_0=0.5. In this final case, the pseudo-data were generated using $y_t = \gamma(t-[T\tau_0])1(t>[T\tau_0]) + u_t$ 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 γ=0 and γ_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 subsection C below. The Monte Carlo and empirical work with the trend-break specifications all used τ_{min}=.15 and τ_{max}=.85.

The densities f_0 and f_1 for ϕ_1 (that is, the densities of $\phi(W_0^d)$ and $\phi_1(V_1^d)$) and the corresponding cdf's are plotted in figures 1 - 5 for the five trend specifications. In each 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 substantially less skewed and more bell-shaped the greater is the amount of detrending. The densities for the two detrending cases differ little, suggesting that as a practical matter the numerical imprecision introduced by the dependence of the distribution on (γ, λ) will have little practical importance.

B. Large-sample Approximate Error Rates

These densities can be used to compute large-sample approximations to the error rates for this procedure, or equivalently for the probability of correctly classifying a series. For a posterior odds ratio of $P_T(\phi_T)$, if the series is truly I(0), the probability of correctly classifying the procedure is $\Pr[P_T<1|u_t \text{ is } I(0)] = \int 1(P_T(\phi_T)<1)dF_{0T}(\phi_T)$, where $F_{0T}(\phi_T)$ is the cdf of ϕ_T. Because $\phi_T \Rightarrow \phi(W_0^d)$, this can be approximated by $\int 1(\pi_1 f_1(\phi - \ln N_T) < \pi_0 f_0(\phi))f_0(\phi)d\phi$. Similarly, $\Pr[P_T>1|u_t \text{ is } I(1)]$ can be approximated by $\int 1(\pi_1 f_1(\phi) > \pi_0 f_0(\phi+\ln N_T))f_1(\phi)d\phi$. These two probabilities are respectively large-sample approximations to the correct classification rates under the I(0) and I(1) hypothesis.

These large-sample correct classification rates for ϕ_{1T} are plotted as a function of N_T in figures 6 (demeaned case) and figure 7 (detrended case) for a prior odds ratio of 1. These figures suggest four conclusions. First, in both cases the probability of correctly classifying the series initially declines with N_T. The mechanical explanation for this is that the I(1) distribution is to the left of the I(0) distribution, but as N_T increases the asymptotic approximation to the finite-sample distribution of ϕ_T shifts to the rate (at the rate $\ln N_T$), so that for $\ln N_T$ small these distributions overlap and likelihood ratios are large. Second, for all values of N_T the probability of correctly classifying an I(1) process is high, exceeding 50% in both the demeaned and the detrended cases. Although this probability drops to 35% if the process is I(0), it quickly rises above 75% for $\ln N_T \geq 3$. Third, the correct classification rates for the demeaned case lie above those for the detrended case for $N_T > 2$. Thus detrending inhibits the ability of this procedure to distinguish between the two classes of processes. Fourth, high correct classification rates are obtained for moderate values of N_T. If the task were to distinguish an i.i.d. process from a random walk, then the MacNeill/Sargan-Bhargava statistic $\tilde{\phi}_{1T}$ would be appropriate (that is, $\hat{\omega}^2 = \hat{\gamma}_{yd}(0)$). In this case, with T=55, $\ln N_T = 4$ and the correct classification rate exceeds 95% in both the demeaned and linearly detrended cases. In general, however, ℓ_T will be nonzero. For example, if $\ell_T = 5$, T=100, and a Bartlett kernel is used, then $\ln N_T = 2.4$ and the correct classification rates are, for the detrended case, approximately .65 (I(0)) and .50 (I(1)). This suggests limitations on the ability of the procedure to distinguish between processes with substantial short-run dependence.

These results are based on the large-sample approximations to the distributions and do not take into account the differences between the finite sample and asymptotic distributions, which are likely to be particularly important for series with substantial short-run dependence. The following Monte Carlo analysis investigates this finite-sample behavior.

C. Finite-sample Performance: Monte Carlo Results

The Monte Carlo experiment examines the ability of this procedure to classify various Gaussian ARMA(1,1) processes as I(1) or I(0). 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 $\ell_T=\min(\hat{\lambda}_T, \ell_{T,max})$, where $\hat{\ell}_T$ is Andrews' (1991) automatic bandwidth selector. Because $\hat{\ell}_T$ is unbounded in the I(1) case it was truncated at $\ell_{T,max} = [10(T/100)^2]$, where the rate is taken from Andrews (1991) and satisfies the condition of theorem 1, and where 10 was picked arbitrarily. The prior odds ratio π_1/π_0 is the relative prior weight on the I(1) and I(0) hypotheses, respectively.

The Monte Carlo results – the 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 Tables 1-3 for, respectively, the ϕ_{1T}, ϕ_{2T}, and ϕ_{3T} statistics. These results suggest five observations.

First, the asymptotic approximate classification rates summarized in Figures 6 and 7 provide an approximate guide to the finite-sample performance, even for T=50, for the i.i.d. and pure random walk models. In the i.i.d. case, the true value of ℓ_T is one for which N_T=T; for T=50 in the demeaned case, the Monte Carlo error rate is 2%, comparable to the approximate error rate of 1% from figure 6. For the demeaned ϕ_{1T} statistic with T=50 in the random walk case, the median value of N_T is 9.6; from figure 6, this corresponds to an error rate of .38, comparable to the Monte Carlo error rate of .30 from Table 1. For the other parameterizations, however, the rates in figure 6 and Table 1 differ, reflecting differences between the finite sample distributions of the cumulative sums and their large-sample approximations as martingales or integrated martingales.

Second, the prior odds can have a substantial effect on the classification rates for moderate sample sizes. For example, for T=100, ρ=1, and θ=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 58%.

Third, these results permit some preliminary comparisons across estimators. One way to make such comparisons is to consider the performance of the classifiers, standardized so that their error rate is constant for a certain model; this is effectively the same as comparing size-adjusted power. For the leading case of linear detrending with T=100, comparison of the π_1/π_0=1 results for ϕ_{1T} with the π_1/π_0=.5 results for ϕ_{2T} indicates higher correct classification rates for ϕ_{2T} than ϕ_{1T} for the AR models with large roots and comparable rates for the IMA models, holding the random walk classification rate constant. A similar comparison of the π_1/π_0=.25 results for ϕ_{2T} with the π_1/π_0=1 results for ϕ_{3T} (T=100, linear detrending) suggests that ϕ_{2T} outperforms ϕ_{3T} for the I(0) AR(1) models. The results for ϕ_{3T} suggest an additional difficulty with interpreting this statistic with linear detrending, the large incorrect classification rate of the random walk with even prior odds (41% for T=100). In contrast, the error rates for ϕ_{2T} in the same case is 12% for a pure random walk and 6% for the i.i.d. process. For linear detrending and typical macroeconomic sample sizes, this initial evidence suggests that, of the three tests, ϕ_{2T} has the best performance. The relative performance of the estimators is similar under broken-trend detrending (in particular compare the T=200 results across estimators), except that the error rate of ϕ_{1T} for the pure random walk, T=100 is much higher than for the other statistics, making posterior odds based on ϕ_{1T} difficult to interpret.

Fourth, for each estimator, increasing the extent of detrending through the first four cases (panels A-D) reduces the discriminatory power of the statistics. For example, for π_1=π_0 and T=100 for ϕ_{2T}, the random walk error rates for the demeaned, detrended, and broken trend-detrended (γ=0 case) are comparable, respectively .13, .12, and .09, but the I(0) correct classification rates for the ρ=.9, θ=0 case drop sharply, respectively being .47, .27, and .17. Moreover, the IMA error rates increase with the nature of the detrending, rising from .36 to .57

to .73 for ρ=1, θ= -.875 for the three 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 suggests that in practice these statistic might not be very informative when broken trend detrending is used.

Fifth, the results indicate that the misclassification rate for models near the I(1)/I(0) boundary increases as the sample size increases. To be precise, consider models that have roots local to one, in the sense that ρ = 1+c/T, where c is a constant. For demeaned ϕ_{2T} with π_1=π_0 and with c = -10, for example, the correct classification rate drops from 57% for T=50 to 47% for T=100 to 39% for T=200. The same pattern is present for the other statistics. For c as negative as -20, these I(0) models that are close to the I(1) boundary are misclassified with a probability that increases with the sample size.

5. Empirical Results

This section presents I(1)/I(0) posterior odds ratios for Nelson and Plosser's (1982) annual data on 14 aggregate economic time series for the United States for linear detrending and for broken-trend detrending. The results with linear detrending and even prior odds are reported in Table 4. With linear detrending, the Monte Carlo study found a high misclassification rate for the ϕ_{3T} statistic in the random walk case relative both to the i.i.d. case and to the other statistics, so table 4 gives results only for the ϕ_{1T} and ϕ_{2T} statistics.

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 the inferences commonly drawn from Nelson and Plosser's (1982) results for the Dickey-Fuller 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). However, the Monte

Carlo results indicate that the ϕ_{2T} statistic has greater discriminatory power than the ϕ_{1T} statistic, suggesting that greater weight should be placed on the ϕ_{2T} results. Moreover, because the unemployment rate is bounded below and above, it is arguably more appropriate to demean than to linearly detrend the data. For the demeaned case, the posterior odds ratios (even prior odds) are .44 and .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, (.687, 1.030), and barely contains 1, while the two ϕ_T posterior odds ratios exceed .8, providing only weak evidence in favor of the I(0) model.

From a Bayesian perspective, the posterior odds ratios in the final two columns of Table 4 provide information about the relative certainty of the I(1) and I(0) hypotheses. 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, even that based on the ϕ_2 statistic. 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 presented in Table 5. Because of its relatively large error rate found in the Monte Carlo experiment under the random walk model with even priors, the ϕ_{1T} statistic is not considered. The Monte Carlo analysis suggested better performance of ϕ_{2T} than ϕ_{3T}, so we place greater weight on the ϕ_{2T} results. As discussed in Section 3, in the I(0) case the asymptotic distribution depends on the break parameters γ and τ

(if $\gamma \neq 0$), so the ϕ_{2T} and ϕ_{3T} likelihood ratios are evaluated under two cases, $\gamma=0$ and ($\gamma=50/T$, $\tau=.5$). In addition, the ϕ_{2T} statistic is evaluated for estimated γ, τ, where the limiting distribution under I(0) was approximated by the distribution of $\phi_2(W_0^d(\cdot;\hat{\gamma}/\hat{\sigma}_u,\hat{\tau}))$ as described in Section 3. The likelihood ratio statistic for the estimated γ, τ case was computed as described in Section 4(a), 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 the statistic used or 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 .8 to 1.3, 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 strongly favor the I(1) model.[6] 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 $f_0(\cdot;\hat{\gamma}/\hat{\sigma}_u,\hat{\tau})$ 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 scaled estimates of b, $T\hat{\gamma}/\hat{\sigma}_u$, are small and always less than 50. The one series for which this is not the case 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.

6. Conclusions

The empirical results for the long U.S. economic time series data suggests three main conclusions. First, 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 being the money supply, for which the posterior odds slightly favor I(0). Second, 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. Third, when the series are detrended using piecewise linear trends, the evidence in these data is much weaker, with Bayes ratios often in the range .8 - 1.3. This final point accords with the Monte Carlo experiment, which found that the difficulty of ascertaining whether the stochastic component of the series is I(0) or I(1) increases as the severity of detrending increases from demeaning to linear detrending to broken-trend detrending. In addition, although the presence of nuisance parameters in the I(0) distribution in the trend-break model poses a potential practical difficulty, in the empirical application we found that the results were largely insensitive to the specific method used to estimate the I(0) distribution with the exception of one series (the bond yield), which seemed to be well described by neither the I(0) nor the I(1) models.

At the level of econometric theory, these results suggest several areas for further work. Primary among these is the desirability of constructing optimal classifiers among the set considered here, that is, of constructing optimal ϕ_T statistics. 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 increase with 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 and related problems are left for future research.

Appendix A

Proof of Theorem 1

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 $K_T(m) = k(m/\ell_T)/\sum_{j=-\ell_T}^{\ell_T} k(j/\ell_T)$.

Proof of Theorem 1.

(a) By Assumption A(i),

$$T^{-\frac{1}{2}}\sum_{t=1}^{[T\lambda]} y_t^d = T^{-\frac{1}{2}}\sum_{t=1}^{[T\lambda]}(u_t - \delta_t)$$

$$\text{(A.1)} \qquad = U_{0T}(\lambda) - D_{0T}(\lambda) \Rightarrow \omega_0(W(\lambda) - D_0(\lambda)) = \omega_0 W_0^d(\lambda).$$

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

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

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

(b) Write $N_T^{-\frac{1}{2}}V_T(\lambda) = N_T^{-\frac{1}{2}}\hat{\omega}^{-1}T^{-\frac{1}{2}}\sum_{s=1}^{[T\lambda]} y_s^d = B_T^{-\frac{1}{2}}A_T(\lambda)$, where $A_T(\lambda) = T^{-3/2}\sum_{s=1}^{[T\lambda]} y_s^d$ and $B_T = T^{-1}\sum_{m=-\ell_T}^{\ell_T} K_T(m)\hat{\gamma}_{yd}(|m|)$. By Assumption B(i), $A_T(\cdot) \Rightarrow \omega_1\int_{s=0}^{\cdot} W_1^d(s)ds$.

In the case of no detrending, it was shown by Phillips (1991b, appendix) that if u_t is I(1) then $B_T \Rightarrow \omega_1^2\int_0^1 W_1(s)^2 ds$. This result was extended to linear trends (OLS detrending) by Kwiatkowski, Phillips and Schmidt (1990) 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 $\Xi_T = T^{-2}\sum_{t=1}^T (y_t^d)^2 - B_T \xrightarrow{p} 0$, so that by Assumption B(i) and the continuous mapping theorem, $B_T \Rightarrow \omega_1^2\int_0^1 W_1^d(s)^2 ds$. 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 $B_T = T^{-1}\sum_{m=-\ell_T}^{\ell_T} K_T(m)\hat{\gamma}_{y^d}(|m|)$. If y_t is I(1), $\ell_T^2 \ln T/T \to 0$, and assumption B holds, then $\Xi_T \xrightarrow{p} 0$.

Proof. Use $\sum_{m=-\ell_T}^{\ell_T} K_T(m) = 1$ and $K_T(m) = K_T(-m)$ to write $\Psi_T = 2T^{-1}\sum_{m=1}^{\ell_T} K_T(m)(\hat{\gamma}_y(0) - \hat{\gamma}_y(m))$. For m≥1,

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

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

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

say. These two terms are handled in turn.

(i) A_{1T}. Note that $|T^{-1}\sum_{t=m+1}^T y_t^d \Delta_m y_t^d| \le \{T^{-1}\sum_{t=1}^T (y_t^d)^2\}^{1/2}\{T^{-1}\sum_{t=m+1}^T (\Delta_m y_t^d)^2\}^{1/2}$.
Using the two inequalities $x^{1/2} < 1+x$ for all x>≥0 and, for $\ell_T > 1$, $K_T(m) \le \ell_T^{-1}\underline{k}^{-1}$ for all T,m, we have

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

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

Now $D_{1T} \Rightarrow \omega_1^2\int_0^1 W_1^d(s)^2 ds$ by Assumption B(i). Also, $D_{2T} = 1/\underline{k}\ln T \to 0$. Thus $A_{1T} \xrightarrow{p} 0$ if $D_{3T} \xrightarrow{p} 0$. Also, $D_{3T} = (\ln T/\underline{k}\ell_T T)\sum_{m=1}^{\ell_T}\|\Delta_m y_t^d\| \le (\ln T/\underline{k}\ell_T T)\sum_{m=1}^{\ell_T}\{\|\Delta_m u_t\| + \|\Delta_m \delta_t\|\}^{\frac{1}{2}}$. Also, $\|\Delta_m u_t\| = \|\sum_{j=0}^{m-1}\Delta u_{t-j}\| \le m^2\|\Delta u_t\|$ and $\|\Delta_m \delta_t\| \le m^2\|\Delta\delta_t\|$. Thus

$$
\begin{aligned}
D_{3T} &\le (\ln T/\underline{k}\ell_T T)\sum_{m=1}^{\ell_T}\{m^2\|\Delta u_t\| + m^2\|\Delta\delta_t\|\}^{\frac{1}{2}} \\
&\le (c/\underline{k})(\ell_T^2\ln T/T)\{\|\Delta u_t\| + \|\Delta\delta_t\|\}^{\frac{1}{2}} \qquad \text{(A.2)}
\end{aligned}
$$

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

(ii) A_{2T}.
$$
\begin{aligned}
A_{2T} &= 2T^{-1}\sum_{m=1}^{\ell_T}K_T(m)T^{-1}\sum_{t=1}^{m}(y_t^d)^2 \\
&\le 2T^{-1}\sum_{m=1}^{\ell_T}K_T(m)(T^{-1}\sum_{t=1}^{\ell_T}(y_t^d)^2) \\
&\le T^{-2}\sum_{t=1}^{\ell_T}(y_t^d)^2 = \int_0^{\ell_T/T}\{U_{1T}(\lambda) - D_{1T}(\lambda)\}^2 d\lambda \xrightarrow{p} 0
\end{aligned}
$$

by assumption B(i) and because $\ell_T/T \to 0$. □

Proof of Theorem 2.

Throughout, let T_T=diag(1, T, . . . , T^q) and let $M_T = T^{-1}T_T^{-1}\sum_{t=1}^{T} z_t z_t' T_T^{-1}$. The nonstochastic q×q 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 $D_{0T}(\lambda) = T^{-\frac{1}{2}}\sum_{s=1}^{[T\lambda]}\delta_s = \phi_T(\lambda)'M_T^{-1}\Phi_T$, where $\phi_T(\lambda) = T^{-1}\sum_{s=1}^{[T\lambda]}T_T^{-1}z_s$ and $\Phi_T = T^{-\frac{1}{2}}\sum_{t=1}^{T}T_T^{-1}z_t u_t$. The (q+1)×1-dimensional process $\phi_T(\lambda)$ is nonstochastic and has the limit, $\phi_T(\lambda) \to \phi(\lambda)$, where the i-th element is $\phi_i(\lambda) = \lambda^i/i$. Under the I(0) assumption (2.2), the random (q+1)-vector Φ_T has the limit, $\Phi_T \Rightarrow \omega_0\Phi$, where $\Phi_1 = W(1)$ and $\Phi_i = W(1)-(i-1)\int_0^1 s^{i-2}W(s)ds$ for i=2, . . . , q+1. Thus $(U_{0T}, D_{0T}(\cdot)) \Rightarrow \omega_0(W(\cdot), D_0(\cdot))$ where $D_0(\lambda) = \Phi'M^{-1}\phi(\lambda)$, which verifies condition A(i).

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

(b)(i) $D_{1T}(\lambda) = T^{-\frac{1}{2}}\delta_{[T\lambda]} = \xi_T(\lambda)'M_T^{-1}\Psi_T$, where $\xi_T(\lambda) = T_T^{-1}z_{[T\lambda]}$ and $\Psi_T = T^{-3/2}\sum_{t=1}^{T}T_T^{-1}z_t u_t = \int_0^1\xi_T(s)U_{1T}(s)ds$. The nonstochastic (q+1)×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 (2.2), and the continuous mapping theorem, $\Psi_T \Rightarrow \omega_1\Psi$, where $\Psi = \int_0^1\xi(s)W(s)ds$. Thus $D_{1T}(\cdot) \Rightarrow \omega_1\Psi'M^{-1}\xi(\cdot) = \omega_1 D_1(\cdot)$.

(ii) Write $\Delta\delta_t = (\hat\beta-\beta)'\Delta z_t = (\hat\beta-\beta)'Rz_t^-$, where R is q×q and lower triangular with $R_{1i} = 0$, i = 1, . . . , q and $R_{ij} = \binom{j}{i-1}$, j = 2, . . . , q, i = 2, . . . , j, and where $z_{tj}^- = t^{j-1} - (t-1)^{j-1}$. Also define $S_T(\lambda) = \{T^{\frac{1}{2}}(\delta_{[T\lambda]}-\delta_{[T\lambda]-1})\}^2$. Then $S_T(t/T) = (T^{\frac{1}{2}}\Delta\delta_t)^2 = \beta_T^{*\prime}M_T^-(t/T)\beta_T^*$ where $M_T^-(t/T) = T^{-2}T_T^{-1}z_t^-z_t^{-\prime}T_T^{-1}$ and $\beta_T^* = T^{-\frac{1}{2}}T_T(\hat\beta-\beta)$. 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 $\xi_T^-(\lambda) = TT_T^{-1}z_{[T\lambda]}^-$; 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 $\|T^{\frac{1}{2}}\Delta\delta_t\| = \int_0^1 S_T(\lambda)d\lambda \Rightarrow \omega_1^2\Psi'M^{-1}M^\dagger M^{-1}\Psi$, where $M^\dagger = \int_0^1\xi(\lambda)\xi(\lambda)'d\lambda$. Direct evaluation of $M^\dagger$ shows that $M_{11}^\dagger = M_{1i}^\dagger = 0$ and $M_{ij}^\dagger = (i-1)(j-1)/(i+j-3)$, i, j≥2. □

Proof of Theorem 3.

(a) The proof uses Proposition 1 in Bai (1991). Under the conditions of Theorem 3(a), Bai (1991) shows that $(T^{\frac{1}{2}}\gamma_T(\hat{k}-k), T^{\frac{1}{2}}\Upsilon_T(\hat{\theta}-\theta)) => \omega_0\Phi$, where Υ_T = diag(1, T, T) and $\Phi = (k^{*}\ \theta^{*})'$ is distributed N(0, $\Omega(\tau_0)^{-1}$), where $\Omega(\tau)$ is given in the statement of Theorem 3(a). Under this nesting, $\hat{k}-k \neq O_p(1)$, but $\hat{\tau} = \hat{k}/T$ is consistent for τ_0: because $T^{\frac{1}{2}}\gamma_T(\hat{k}-k_0) => \omega_0 k^* = O_p(1)$, $T^{3/2}\gamma_T(\hat{\tau}-\tau_0) => \omega_0 k^*$, but $T^{3/2}\gamma_T \to \infty$ by assumption, so $\hat{\tau} \xrightarrow{p} \tau_0$.

(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,

$$\text{(A.3)} \qquad D_{0T}(\lambda,\hat{\tau}) = T^{-\frac{1}{2}}\sum_{s=1}^{[T\lambda]}\delta_s(\hat{k}) = \tilde{\nu}_T(\lambda,\hat{\tau})'\vartheta_T(\hat{\tau}) + \rho_T(\lambda,\hat{\tau})$$

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

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

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

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

$\vartheta_T(\hat{\tau})$. From Bai (1991), $\vartheta_T(\hat{\tau}) => \theta^*$ as defined previously.

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

$$\text{(A.5)} \quad e_T(t/T,k/T) = T^{\frac{1}{2}}\gamma_T \text{sign}(k_0-k)(t-\min(k,k_0))1(\min(k,k_0)\le t<\max(k,k_0)) - T^{\frac{1}{2}}\gamma_T(k-k_0)1(t\ge\max(k,k_0)).$$

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

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

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

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

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

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

where the two inequalities are uniform in t and k and the second follows from $|k-k_0|\le T$. 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 $T^{-1}\hat{\theta}_T(\tau) = T^{-\frac{1}{2}}\Upsilon_T(\hat{\theta}([T\tau])-\theta) = M_T(\tau)^{-1}\{N_T(\tau) + \Psi_T(\tau)\}$, where $M_T(\tau)$ is defined in the proof of part (a) and has the limit $M_T(\cdot) \to M(\cdot)$, where $N_T(\tau) = T^{-3/2}\sum_{t=1}^{T}\Upsilon_T^{-1}z_t([T\tau])(z_t([T\tau_0])-z_t([T\tau]))'\theta$, and where $\Psi_T(\tau) = T^{-3/2}\sum_{t=1}^{T}\Upsilon_T^{-1}z_t([T\tau])u_t$. N_T can be rewritten, $N_T(\tau) = \int_0^1\xi_T(s,\tau)(T^{-1}e_T(s,\tau))ds$; the results $\xi_T \to \xi$ and $T^{-1}e_T \overset{p}{\to} 0$ imply that $N_T \overset{p}{\to} 0$. Under the I(1) assumption, the remaining term, $\Psi_T(\tau)$, has the limit, $\Psi_T(\cdot) = \int_0^1\xi_T(s,\cdot)U_{1T}(s)ds \Rightarrow \omega_1\int_0^1\xi(s,\cdot)W(s)ds \equiv \Psi(\cdot)$. Combining these various expressions, we have $T^{-1}\hat{\theta}_T(\cdot) \Rightarrow M(\cdot)^{-1}\Psi(\cdot)$, so $F_{1T}(\cdot,\cdot) \Rightarrow \omega_1F_1(\cdot,\cdot)$, 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 (3.6), which can be rewritten as the problem of minimizing $S_T(\tau)$ over $\tau_{min}\le\tau\le\tau_{max}$, 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 $S_T(\tau) = T^{-2}\sum_{t=1}^{T}\{y_t - \hat{d}_t([T\tau])\}^2 = T^{-2}\sum_{t=1}^{T}\{u_t - \delta_t([T\tau])\}^2 = T^{-1}\sum_{t=1}^{T}\{U_{1T}(t/T) - F_{1T}(t/T, [T\tau]/T)\}^2$ $= \int_{s=0}^{1}\{U_{1T}(s) - F_{1T}(s,[T\tau]/T)\}^2ds$. It follows that $S_T(\cdot) \Rightarrow S(\cdot)$, where $S(\tau) = \omega_1^2\int_{s=0}^{1}\{W(s) - F_1(s,\tau)\}^2ds$. Thus $\hat{\tau} \Rightarrow \tau^*$, where τ^* has has the distribution, $\text{argmin}_{\tau\in[\tau_{min},\tau_{max}]}\int_0^1\{W(s)-F_1(s,\tau)\}^2ds$. Because F_1 is continuous in τ, it follows that $D_{1T}(\cdot) = F_{1T}(\cdot,\hat{\tau}) \Rightarrow \omega_1F_1(\cdot,\tau^*) \equiv \omega_1D_1(\cdot)$.

(ii). By direct calculation,

$$
\begin{aligned}
|\eta_T(t/T,k/T)| &= |\delta_{[T\lambda]}([T\tau]) - \delta_{[T\lambda]-1}([T\tau])| \\
&= |(\hat{\theta}(k)-\theta)'\{z_t(k)-z_{t-1}(k)\} - \theta'\{z_t(k)-z_t(k_0)-z_{t-1}(k)+z_{t-1}(k_0)\}| \\
&\le |\hat{\beta}(k)-\beta| + |\hat{\gamma}_T(k)-\gamma_T|1(t>k)| + |\gamma_T|1(\min(k,k_0)\le t<\max(k,k_0))
\end{aligned} \tag{A.6}
$$

In the proof of part (b)(i) it was shown that $T^{-1}\hat{\theta}_T(\cdot) \Rightarrow M(\cdot)^{-1}\Psi(\cdot)$, where $T^{-1}\hat{\theta}_T(\tau) = T^{-\frac{1}{2}}\Upsilon_T(\hat{\theta}([T\tau])-\theta) = \{T^{\frac{1}{2}}(\hat{\alpha}([T\tau])-\alpha), T^{-\frac{1}{2}}(\hat{\beta}([T\tau])-\beta), T^{-\frac{1}{2}}(\hat{\gamma}_T([T\tau])-\gamma_T)\}$. Thus in particular $\sup_\tau|\hat{\beta}([T\tau])-\beta| \overset{p}{\to} 0$ and $\sup_\tau|\hat{\gamma}_T([T\tau])-\gamma_T| \overset{p}{\to} 0$, so the first two terms in (A.6) converge to zero uniformly in λ=t/T, τ=k/T. In addition, $\gamma_T \to 0$ by assumption, so the final term in (A.6) vanishes. Thus $\sup_{\lambda,\tau}|\eta_T(\lambda,\tau)| \overset{p}{\to} 0$, so $\eta_T(\cdot,\cdot) \overset{p}{\to} 0$ 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 $D_{0T}(\lambda,\tau) = \tilde{\nu}_T(\lambda,\tau)'\theta_T(\tau)$. As in the proof of theorem 3, $\tilde{\nu}_T(\cdot,\cdot) \to \tilde{\nu}(\cdot,\cdot)$. Because $\gamma_T = 0$, $\theta_T^* = M_T(\tau)^{-1}\Phi_T(\tau)$, where $\Phi_T(\tau) = T^{-\frac{1}{2}}\Gamma_T^{-1}\sum_{t=1}^{T} z_t([T\tau])u_t$. It follows from the FCLT that $\Phi_T(\cdot) \Rightarrow \omega_0\Phi(\cdot)$ as defined in the statement of theorem 4; thus $\theta_T^*(\cdot) \Rightarrow \theta^*(\cdot)$, where $\theta^*(\tau) = \omega_0 M(\tau)^{-1}\Phi(\tau)$, from which it follows that $D_{0T}(\cdot,\cdot) \Rightarrow \omega_0 D_0(\cdot,\cdot)$, where $D_0(\lambda,\tau) = \tilde{\nu}(\lambda,\tau)'\theta^*(\tau)$.

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 $\gamma_T = 0$, $H_T(\cdot) \Rightarrow H(\cdot) = \theta^*(\cdot)'M(\cdot)\theta^*(\cdot) = \Phi(\cdot)'M(\cdot)^{-1}\Phi(\cdot)$. By the continuity of the distribution of the argmax, the limiting representation for $\hat{\tau}$ as $\text{argmax}_{\tau\in[\tau_{\min},\,\tau_{\max}]}H(\tau)$ follows.

(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 $N_T = 0$ identically. In particular, the key result that $T^{-1}\theta_T^*(\cdot) \Rightarrow M(\cdot)^{-1}\Psi(\cdot)$ still holds.

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

Footnotes

1. Since the original draft of this paper was written, three additional closely related papers have appeared, by Kwiatkowski, Phillips and Schmidt (1990), Phillips and Ploberger (1991), and Perron (1991); these are discussed below.

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 motivated by computational experience, which indicates that the the additive property enhances the numerical stability of the techniques described in Section 4.

3. From Bai (1991), 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 $\hat{b} \overset{p}{\rightarrow} b$ for general b. From theorem 3(a) and its proof, for $b \neq 0$, $D_{0T}(\lambda) = \tilde{\phi}(\lambda,\hat{\tau})'\Phi(\hat{\tau}) + bh(\lambda)\tau^* + o_p(1)$ uniformly in λ, where τ^* is an $O_p(1)$ random variable that does not depend on λ. It follows that the limit as $b \rightarrow 0$ 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 $\max_{\tau\in[\tau_{min}, \tau_{max}]}\Phi(\tau)'M(\tau)^{-1}\Phi(\tau) + bQ(\tau)$, where $Q(\tau) = O_p(1)$ uniformly in a $T^{-1/2}$ neighborhood of τ_0. As $b \rightarrow 0$, the objective function converges to the objective function in theorem 4(a), suggesting the continuity of the distribution of $\hat{\tau}$ as $b \rightarrow 0$.

4. Because the limit distributions in theorems 3 and 4 do not depend on γ, except that γ is respectively local to or equal to zero, there is some ambiguity in scaling γ in the estimator of f_0. Although $\hat{\gamma}/\hat{\sigma}_u$ is used here, an alternative would be to use $\hat{\gamma}/\hat{\omega}$. For reasons paralleling those just given, both estimators have qualitatively similar asymptotic properties. Which procedure works better in finite samples is a subject for future investigation.

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 $\kappa = .1$. Figures 1-7 were computed using a Gaussian kernel with the same bandwidth. With one exception discussed in Section 5, 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 κ=.02 - .20.

6. It is curious that the evidence that velocity is I(1) is strong but the evidence concerning

money and income is ambiguous. This suggests handling these series in a multiviarate setting in which inference is performed over the number of unit roots in a system with logarithms of nominal money and income.

References

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858

Bai, J. (1991), "Estimation of Structural Change in Econometric Models," manuscript, Department of Economics, University of California, Berkeley.

Banerjee, A., Lumsdaine, R.L., and Stock, J.H. (1992), "Recursive and Sequential Tests of the Unit Root and Trend Break Hypotheses: Theory and International Evidence," Working Paper no. 3510, National Bureau of Economic Research, Cambridge MA; forthcoming, *Journal of Business and Economic Statistics.*

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.

Brown, R.L., Durbin, J. and Evans, J.M. (1975), "Techniques for Testing the Constancy of Regression Relationships over Time with Comments," *Journal of the Royal Statistical Society*, Series 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*, no. 2, 221-254.

Dickey, D.A. and Fuller, W.A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* 74, no. 366, 427-431.

Durbin, J. and G.S. Watson (1950), "Testing for Serial Correlation in Least Squares Regression I," *Biometrika 37*, 409-428.

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

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.* New York: Academic Press.

Harvey, A.C. (1989), *Forecasting, Structural Models and the Kalman Filter*, Cambridge: Cambridge University Press.

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, and P. Schmidt (1990), "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?," manuscript, Yale University.

Lo, A.W. (1991), "Long-Term Memory in Stock Market Prices," *Econometrica 59*, no. 5, 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," *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 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 Plosser, C.I. (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*, Dordrecht: Kluwer, 7-28

Park, J. (1990), "Testing for Unit Roots and Cointegration by Variable Addition," in G.F. Rhodes and T.B. Fomby (eds.) *Advances in Econometrics.*

Park J. and C. Choi (1988), "A New Approach to Testing for a Unit Root," CAE Working Paper No. 88-23, Cornell University.

Park, J.Y. and P.C.B. Phillips (1988), "Statistical Inference in Regressions with Integrated Processes: Part I," *Econometric Theorey 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.

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, UK.

Phillips, P.C.B. and W. Ploberger (1991), "Time Series Modeling with a Bayesian Frame of Reference: I. Concepts and Illustrations," manuscript, Yale University.

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 (1990), "Testing for Moving Average Unit Root," manuscript, University of Helsinki.

Sargan, J.D. and A. Bhargava (1983), "Testing for Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk," *Econometrica*, 51, 153-174.

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, no. 6, 1591-1600

Sowell, F. (1991), "On DeJong and Whiteman's Bayesian Inference for the Unit Root Model," *Journal of Monetary Economics 28*, no. 2, 255-264.

Stock, J.H. (1991), "Confidence Intervals for the Largest Autoregressive Root in U.S. Economic Time Series," *Journal of Monetary Economics* 28, no. 3, 435-460.

Watson, M. (1992), "A Note on Estimating Deterministic Trends in the Presence of Serially Correlated Errors," manuscript, Northwestern University.

Zivot, E. and Andrews, D.W.K. (1992), "Further Evidence on the Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," manuscript, Cowles Foundation, Yale University; forthcoming, *Journal of Business and Economic Statistics.*

Table 1
Monte Carlo Results: I(0) Classification Rates for the ϕ_{1T} Statistic

Model: $(1-\rho L)x_t = (1+\theta L)\epsilon_t$, ϵ_t i.i.d. N(0,1)

T	π_1/π_0	$\theta = 0, \rho =$ 0	.6	.8	.9	.95	.975	1.0	$\rho = 1.0, \theta =$ -.875	-.75	-.5
A. No Detrending											
50	1.00	0.97	0.84	0.68	0.49	0.36	0.27	0.19	0.23	0.20	0.18
50	0.50	0.98	0.92	0.82	0.62	0.47	0.37	0.26	0.29	0.26	0.26
50	0.25	0.99	0.97	0.90	0.75	0.59	0.49	0.34	0.35	0.36	0.32
50	0.10	1.00	0.99	0.96	0.87	0.72	0.63	0.52	0.53	0.51	0.52
100	1.00	0.97	0.85	0.79	0.57	0.41	0.28	0.14	0.15	0.12	0.13
100	0.50	0.99	0.93	0.89	0.73	0.56	0.40	0.21	0.20	0.20	0.19
100	0.25	0.99	0.98	0.94	0.82	0.64	0.48	0.28	0.25	0.25	0.26
100	0.10	1.00	1.00	0.97	0.92	0.76	0.59	0.35	0.34	0.32	0.32
200	1.00	0.99	0.92	0.83	0.67	0.53	0.32	0.10	0.08	0.09	0.10
200	0.50	0.99	0.96	0.92	0.81	0.67	0.45	0.15	0.12	0.12	0.16
200	0.25	0.99	0.98	0.96	0.88	0.72	0.52	0.19	0.15	0.15	0.18
200	0.10	1.00	0.99	1.00	0.92	0.82	0.60	0.24	0.20	0.22	0.25
B. Demeaned											
50	1.00	0.93	0.49	0.39	0.28	0.25	0.27	0.30	0.65	0.38	0.32
50	0.50	0.97	0.78	0.72	0.63	0.63	0.61	0.66	0.75	0.60	0.67
50	0.25	0.98	1.00	1.00	1.00	1.00	1.00	1.00	0.94	0.91	0.99
50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00
100	1.00	0.98	0.71	0.56	0.40	0.25	0.20	0.12	0.38	0.17	0.11
100	0.50	0.99	0.84	0.74	0.56	0.37	0.33	0.24	0.45	0.27	0.21
100	0.25	0.99	0.95	0.87	0.76	0.65	0.59	0.53	0.56	0.49	0.46
100	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.76	0.92	1.00
200	1.00	0.99	0.87	0.75	0.58	0.35	0.21	0.09	0.13	0.09	0.08
200	0.50	0.99	0.94	0.85	0.69	0.47	0.28	0.14	0.17	0.14	0.14
200	0.25	1.00	0.98	0.92	0.79	0.56	0.38	0.19	0.23	0.18	0.18
200	0.10	1.00	0.99	0.96	0.90	0.66	0.50	0.26	0.28	0.29	0.24
C. Linear trend/OLS detrending											
50	1.00	0.87	0.37	0.35	0.36	0.39	0.45	0.46	0.68	0.55	0.45
50	0.50	0.94	1.00	1.00	1.00	1.00	1.00	1.00	0.86	0.78	0.96
50	0.25	0.98	1.00	1.00	1.00	1.00	1.00	1.00	0.93	0.90	1.00
50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.97	1.00
100	1.00	0.95	0.62	0.42	0.30	0.21	0.21	0.20	0.59	0.29	0.20
100	0.50	0.96	0.89	0.78	0.65	0.56	0.58	0.48	0.66	0.45	0.53
100	0.25	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.71	0.71	0.99
100	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.84	0.89	1.00
200	1.00	0.98	0.76	0.67	0.45	0.28	0.15	0.10	0.28	0.10	0.11
200	0.50	0.99	0.88	0.80	0.62	0.43	0.28	0.19	0.34	0.18	0.17
200	0.25	1.00	0.96	0.90	0.74	0.58	0.42	0.28	0.42	0.27	0.28
200	0.10	1.00	0.99	0.97	0.88	0.78	0.64	0.50	0.51	0.43	0.47

Table 1, continued

T	π_1/π_0	$\theta = 0, \rho =$ 0	.6	.8	.9	.95	.975	1.0	$\rho = 1.0, \theta =$ -.875	-.75	-.5
D. Broken trend/OLS detrending, $\gamma=0$											
50	1.00	0.74	0.94	0.98	0.98	0.98	0.98	0.99	0.73	0.63	0.75
50	0.50	0.92	0.99	1.00	1.00	1.00	1.00	1.00	0.91	0.84	0.91
50	0.25	0.97	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.95	0.98
50	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.97	0.99
100	1.00	0.89	0.69	0.63	0.63	0.57	0.53	0.53	0.73	0.41	0.49
100	0.50	0.94	0.99	1.00	1.00	1.00	1.00	1.00	0.83	0.58	0.96
100	0.25	0.97	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.75	0.98
100	0.10	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.87	1.00
200	1.00	0.95	0.60	0.48	0.35	0.21	0.16	0.17	0.52	0.15	0.16
200	0.50	0.97	0.85	0.81	0.66	0.49	0.42	0.44	0.59	0.29	0.43
200	0.25	0.98	0.95	0.90	0.85	0.73	0.64	0.64	0.66	0.42	0.63
200	0.10	0.99	0.99	0.98	0.97	0.92	0.84	0.81	0.73	0.59	0.85
E. Broken trend/OLS detrending, $\gamma=50/T$, $\lambda=.5$											
50	1.00	0.77	0.95	0.96	0.96	0.97	0.97	0.98	0.70	0.66	0.83
50	0.50	0.96	0.99	1.00	1.00	0.99	1.00	1.00	0.92	0.90	0.95
50	0.25	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.97	0.97
50	0.10	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00
100	1.00	0.93	0.90	0.91	0.93	0.96	0.96	0.96	0.75	0.45	0.82
100	0.50	0.98	1.00	1.00	1.00	1.00	1.00	1.00	0.86	0.68	0.96
100	0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.83	0.99
100	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.95	1.00
200	1.00	0.99	0.61	0.51	0.32	0.18	0.14	0.11	0.52	0.11	0.14
200	0.50	0.99	0.90	0.87	0.70	0.55	0.44	0.36	0.60	0.22	0.44
200	0.25	1.00	0.98	0.96	0.89	0.78	0.69	0.59	0.68	0.42	0.72
200	0.10	1.00	1.00	1.00	0.99	0.95	0.91	0.83	0.78	0.67	0.93

Notes: 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 ϕ_T statistics are defined in (2.7) in the text. The spectral density was estimated using the Parzen kernel with bandwidth truncation parameter ℓ_T was estimated using Andrews' (1991) automatic procedure, truncated at $10(T/100)^{.2}$, as discussed in the text. For the model in panel E the coefficient γ on the trend-break is set to accord with the local nesting (3.3). Based on 500 Monte Carlo replications for each entry.

Table 2

Monte Carlo Results: I(0) Classification Rates for the ϕ_{2T} statistic

Model: $(1-\rho L)x_t = (1+\theta L)\epsilon_t$, ϵ_t i.i.d. N(0,1)

		$\theta = 0$, $\rho =$							$\rho = 1.0$, $\theta =$		
T	π_1/π_0	0	.6	.8	.9	.95	.975	1.0	-.875	-.75	-.5
A. No Detrending											
50	1.00	0.88	0.73	0.53	0.40	0.34	0.29	0.23	0.21	0.23	0.21
50	0.50	0.96	0.82	0.65	0.50	0.46	0.38	0.31	0.30	0.32	0.32
50	0.25	0.99	0.91	0.79	0.63	0.58	0.50	0.44	0.40	0.44	0.43
50	0.10	1.00	0.99	0.94	0.89	0.81	0.82	0.76	0.71	0.76	0.73
100	1.00	0.91	0.80	0.69	0.52	0.40	0.32	0.23	0.17	0.19	0.17
100	0.50	0.97	0.88	0.81	0.64	0.51	0.40	0.29	0.24	0.27	0.25
100	0.25	0.99	0.94	0.88	0.73	0.61	0.48	0.39	0.30	0.35	0.31
100	0.10	0.99	0.98	0.94	0.86	0.74	0.61	0.52	0.42	0.46	0.44
200	1.00	0.97	0.86	0.77	0.62	0.50	0.38	0.15	0.15	0.18	0.16
200	0.50	0.98	0.96	0.88	0.80	0.65	0.48	0.25	0.21	0.26	0.24
200	0.25	0.99	0.97	0.92	0.84	0.72	0.55	0.30	0.24	0.29	0.30
200	0.10	0.99	0.99	0.97	0.92	0.80	0.65	0.41	0.32	0.38	0.41
B. Demeaned											
50	1.00	0.93	0.72	0.57	0.37	0.35	0.26	0.18	0.49	0.26	0.21
50	0.50	0.97	0.86	0.75	0.56	0.49	0.39	0.30	0.61	0.37	0.30
50	0.25	0.99	0.94	0.85	0.68	0.61	0.50	0.40	0.74	0.50	0.42
50	0.10	1.00	0.97	0.93	0.80	0.76	0.64	0.54	0.86	0.63	0.57
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
C. Linear trend/OLS detrending											
50	1.00	0.87	0.45	0.28	0.17	0.16	0.16	0.15	0.69	0.48	0.19
50	0.50	0.95	0.74	0.57	0.40	0.41	0.37	0.38	0.81	0.60	0.38
50	0.25	0.98	0.91	0.82	0.71	0.72	0.68	0.68	0.90	0.73	0.68
50	0.10	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.90	0.99
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

Table 2, continued

T	π_1/π_0	$\theta = 0, \rho =$ 0	.6	.8	.9	.95	.975	1.0	$\rho = 1.0, \theta =$ -.875	-.75	-.5
D. Broken trend/OLS detrending, γ=0											
50	1.00	0.72	0.62	0.59	0.54	0.53	0.52	0.53	0.66	0.52	0.34
50	0.50	0.86	0.96	0.99	1.00	1.00	1.00	1.00	0.84	0.76	0.80
50	0.25	0.94	0.99	1.00	1.00	1.00	1.00	1.00	0.93	0.88	0.91
50	0.10	0.98	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.95	0.98
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
E. Broken trend/OLS detrending, γ=50/T, λ=.5											
50	1.00	0.70	0.91	0.98	0.99	0.99	0.99	0.98	0.62	0.49	0.59
50	0.50	0.88	0.99	1.00	1.00	1.00	1.00	1.00	0.81	0.77	0.84
50	0.25	0.97	1.00	1.00	1.00	1.00	1.00	1.00	0.94	0.93	0.94
50	0.10	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.97	0.98
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

Notes: See the notes to table 1.

Table 3
Monte Carlo Results: I(0) Classification Rates for the ϕ_{3T} statistic

Model: $(1-\rho L)x_t = (1+\theta L)\epsilon_t$, ϵ_t i.i.d. N(0,1)

		$\theta = 0, \rho =$							$\rho = 1.0, \theta =$		
T	π_1/π_0	0	.6	.8	.9	.95	.975	1.0	-.875	-.75	-.5
A. No Detrending											
50	1.00	0.95	0.78	0.64	0.57	0.49	0.35	0.28	0.52	0.38	0.30
50	0.50	0.99	1.00	0.98	0.95	0.94	0.95	0.94	0.98	0.96	0.95
50	0.25	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00
50	0.10	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00
100	1.00	0.89	0.58	0.50	0.40	0.33	0.37	0.32	0.48	0.41	0.33
100	0.50	0.93	0.85	0.79	0.64	0.58	0.60	0.58	0.71	0.67	0.65
100	0.25	0.97	0.96	0.93	0.83	0.74	0.75	0.78	0.84	0.84	0.81
100	0.10	0.99	1.00	0.99	0.97	0.96	0.96	0.96	0.91	0.96	0.98
200	1.00	0.86	0.41	0.29	0.21	0.17	0.17	0.16	0.29	0.20	0.21
200	0.50	0.89	0.64	0.51	0.38	0.30	0.27	0.27	0.40	0.33	0.30
200	0.25	0.93	0.83	0.72	0.55	0.44	0.40	0.40	0.51	0.45	0.41
200	0.10	0.96	0.90	0.81	0.66	0.57	0.50	0.50	0.60	0.54	0.55
B. Demeaned											
50	1.00	0.99	0.89	0.80	0.71	0.65	0.62	0.52	0.85	0.69	0.60
50	0.50	1.00	0.95	0.93	0.87	0.83	0.80	0.78	0.96	0.88	0.84
50	0.25	1.00	0.97	0.97	0.97	0.95	0.95	0.97	1.00	0.99	0.98
50	0.10	1.00	0.98	0.98	0.99	0.99	0.98	0.99	1.00	1.00	0.99
100	1.00	0.95	0.78	0.67	0.51	0.39	0.39	0.26	0.43	0.34	0.26
100	0.50	0.98	0.89	0.82	0.65	0.53	0.50	0.38	0.49	0.43	0.37
100	0.25	0.98	0.95	0.89	0.76	0.65	0.60	0.46	0.57	0.50	0.43
100	0.10	1.00	0.98	0.95	0.85	0.75	0.68	0.58	0.67	0.59	0.54
200	1.00	0.82	0.58	0.41	0.29	0.22	0.16	0.11	0.13	0.10	0.10
200	0.50	0.88	0.69	0.55	0.40	0.31	0.22	0.16	0.18	0.15	0.13
200	0.25	0.91	0.82	0.68	0.52	0.41	0.28	0.18	0.22	0.20	0.17
200	0.10	0.96	0.89	0.77	0.61	0.48	0.37	0.22	0.25	0.25	0.22
C. Linear trend/OLS detrending											
50	1.00	0.95	0.71	0.66	0.63	0.58	0.58	0.59	0.85	0.82	0.59
50	0.50	0.98	0.86	0.82	0.81	0.77	0.78	0.76	0.96	0.93	0.80
50	0.25	0.99	0.89	0.87	0.84	0.81	0.82	0.80	0.98	0.96	0.82
50	0.10	1.00	0.94	0.91	0.88	0.86	0.87	0.83	0.99	0.98	0.88
100	1.00	0.90	0.75	0.60	0.54	0.47	0.42	0.41	0.68	0.49	0.45
100	0.50	0.97	0.92	0.79	0.73	0.65	0.62	0.60	0.82	0.64	0.62
100	0.25	0.98	0.97	0.89	0.83	0.78	0.79	0.76	0.87	0.76	0.75
100	0.10	0.99	1.00	0.99	0.98	0.97	0.97	0.96	0.94	0.88	0.94
200	1.00	0.69	0.31	0.31	0.22	0.16	0.17	0.22	0.24	0.21	0.21
200	0.50	0.80	0.61	0.56	0.41	0.33	0.32	0.36	0.33	0.32	0.32
200	0.25	0.86	0.75	0.68	0.51	0.42	0.41	0.44	0.41	0.43	0.40
200	0.10	0.92	0.87	0.81	0.64	0.54	0.49	0.54	0.53	0.50	0.50

Table 3, continued

T	π_1/π_0	$\theta = 0, \rho =$ 0	.6	.8	.9	.95	.975	1.0	$\rho = 1.0, \theta =$ -.875	-.75	-.5
D. Broken trend/OLS detrending, γ=0											
50	1.00	0.96	0.79	0.73	0.66	0.67	0.62	0.68	0.94	0.83	0.72
50	0.50	0.99	0.95	0.94	0.92	0.91	0.91	0.92	0.99	0.95	0.94
50	0.25	1.00	0.97	0.95	0.95	0.94	0.94	0.96	1.00	0.98	0.98
50	0.10	1.00	0.98	0.97	0.96	0.97	0.96	0.98	1.00	1.00	0.99
100	1.00	0.88	0.57	0.44	0.38	0.33	0.29	0.25	0.72	0.42	0.32
100	0.50	0.95	0.83	0.72	0.67	0.55	0.52	0.50	0.79	0.57	0.51
100	0.25	0.97	0.94	0.86	0.82	0.72	0.68	0.65	0.83	0.66	0.67
100	0.10	0.99	1.00	0.97	0.95	0.91	0.86	0.89	0.89	0.76	0.89
200	1.00	0.72	0.23	0.14	0.08	0.09	0.06	0.07	0.24	0.06	0.07
200	0.50	0.79	0.48	0.30	0.20	0.19	0.15	0.12	0.30	0.11	0.13
200	0.25	0.83	0.63	0.43	0.31	0.26	0.20	0.18	0.33	0.17	0.21
200	0.10	0.90	0.76	0.58	0.42	0.34	0.26	0.27	0.42	0.24	0.28
E. Broken trend/OLS detrending, γ=50/T, λ=.5											
50	1.00	0.93	0.73	0.73	0.71	0.73	0.71	0.74	0.91	0.87	0.73
50	0.50	0.99	0.89	0.86	0.87	0.87	0.87	0.89	0.98	0.97	0.93
50	0.25	1.00	0.93	0.91	0.90	0.91	0.90	0.91	1.00	0.99	0.95
50	0.10	1.00	0.97	0.97	0.96	0.95	0.95	0.96	1.00	0.99	0.98
100	1.00	0.90	0.53	0.44	0.38	0.34	0.32	0.31	0.76	0.50	0.33
100	0.50	0.95	0.86	0.84	0.74	0.71	0.66	0.62	0.85	0.67	0.62
100	0.25	0.98	1.00	0.99	0.99	0.98	0.96	0.96	0.92	0.79	0.95
100	0.10	0.99	1.00	0.99	1.00	0.99	1.00	0.99	0.97	0.93	0.98
200	1.00	0.72	0.19	0.17	0.13	0.10	0.09	0.06	0.26	0.13	0.10
200	0.50	0.80	0.41	0.33	0.23	0.20	0.14	0.12	0.34	0.20	0.16
200	0.25	0.86	0.60	0.51	0.39	0.34	0.26	0.20	0.41	0.28	0.26
200	0.10	0.93	0.85	0.77	0.64	0.53	0.44	0.35	0.52	0.41	0.44

Notes: See the notes to table 1.

Table 4

Posterior Odds Ratios for I(1) vs. I(0) for the Nelson-Plosser Data Set:
Linear Trend, OLS detrending

Prior odds ratio = π_1/π_0 = 1.0

Series	T	ADF statistics k	$\hat{\tau}_\tau$	90% conf. interval	Post. odds for: $B(\phi_1)$	$B(\phi_2)$
REAL GNP	62	1	-2.994	(.604, 1.042)	1.44	3.89
NOMINAL GNP	62	1	-2.321	(.757, 1.060)	1.54	4.06
REAL PER CAPITA GNP	62	1	-3.045	(.591, 1.041)	1.35	2.43
INDUSTRIAL PRODUCTION	111	5	-2.529	(.836, 1.031)	5.10	5.88
EMPLOYMENT	81	2	-2.655	(.757, 1.039)	1.62	2.14
UNEMPLOYMENT RATE	81	3	-3.552	(.577, .950)	1.07	0.44
GNP DEFLATOR	82	1	-2.516	(.787, 1.041)	1.05	1.15
CONSUMER PRICES	111	3	-1.972	(.901, 1.037)	7.75	44.64
WAGES	71	2	-2.236	(.800, 1.054)	1.37	2.31
REAL WAGES	71	1	-3.049	(.644, 1.035)	2.07	12.71
MONEY STOCK	82	1	-3.078	(.687, 1.030)	0.89	0.84
VELOCITY	102	0	-1.663	(.929, 1.042)	2.18	38.00
BOND YIELD	71	2	.686	(1.032, 1.075)	2.10	6.38
S&P 500	100	2	-2.122	(.873, 1.039)	4.84	22.67

Notes: T is the total number of observations on each series, including observations used for initial conditions in the augmented Dickey-Fuller (1979) regressions regressions. The columns headed "k" 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, k was taken from Nelson and Plosser (1982). The 90% confidence interval column, taken from Stock (1991), is the confidence interval that results from inverting the ADF t-statistics. The final two columns present the posterior odds ratios (with even prior odds, the Bayes ratios) based on ϕ_{1T} and ϕ_{2T}, where $\hat{\omega}$ was computed using the Parzen kernel with lag length estimated using Andrews' (1991) procedure, truncated as described in the notes to table 1. 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.

Table 5

Posterior Odds Ratios for I(1) vs. I(0) for the Nelson-Plosser Data Set:
Broken Trend, OLS detrending

Prior odds ratio $= \pi_1/\pi_0 = 1.0$

Series	(a) $\gamma=0$		(b) $\gamma=50/T$, $\tau_0=.5$		(c) γ, τ estimated			
	$B(\phi_2)$	$B(\phi_3)$	$B(\phi_2)$	$B(\phi_3)$	$\hat{\gamma}/\hat{\sigma}_u$	$T\hat{\gamma}/\hat{\sigma}_u$	$\hat{k}$	$B(\phi_2)$
REAL GNP	1.01	0.83	0.85	0.81	0.204	12.67	1934	0.90
NOMINAL GNP	1.03	0.64	0.79	0.74	0.188	11.66	1936	0.85
REAL PER CAPITA GNP	1.03	1.29	0.85	1.04	0.178	11.06	1933	0.92
INDUSTRIAL PRODUCTION	3.63	8.43	7.33	3.79	-0.076	-8.42	1901	3.89
EMPLOYMENT	1.46	0.85	1.54	0.91	-0.216	-17.51	1907	1.08
UNEMPLOYMENT RATE	1.39	1.12	1.48	1.04	-0.035	-2.81	1926	1.67
GNP DEFLATOR	1.85	3.97	1.83	2.07	0.089	7.37	1940	2.02
CONSUMER PRICES	1.21	1.53	1.53	1.37	0.219	24.36	1899	1.35
WAGES	1.40	0.73	1.15	0.78	0.120	8.51	1939	1.21
REAL WAGES	0.97	0.62	0.97	0.75	0.209	14.85	1933	1.06
MONEY STOCK	1.47	1.82	1.57	1.89	-0.084	-6.88	1918	1.57
VELOCITY	4.87	8.48	6.41	3.17	0.221	22.52	1944	3.05
BOND YIELD	1.34	7.78	2.00	1.79	0.386	27.42	1955	.01-.09
S&P 500	0.92	0.47	1.11	0.57	0.310	30.98	1945	1.07

Notes: The $B(\phi_1)$ and $B(\phi_2)$ entries are the posterior odds ratios (with even odds, the Bayes ratios) for the indicated statistic, computed using asymptotic I(1) distribution and (a) the $\gamma=0$ asymptotic I(0) distribution; (b) the ($\gamma=50/T$, $\tau=.5$) I(0) asymptotic distribution; and (c) the distribution of $\phi_2(W_1^d(\cdot;\hat{\gamma}/\hat{\sigma}_u,\hat{\tau}))$, computed as described in Section 3 using 4000 Monte Carlo replications. The estimates in columns (c), $\hat{\gamma}/\hat{\sigma}_u$ and $T\hat{\gamma}/\hat{\sigma}_u$ were computed by OLS with $\tau_{min}=.15$ and $\tau_{max}=.85$. The second-to-final column gives the estimated break date $\hat{k}$. 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 .1 to .25. See the notes to Table 4.

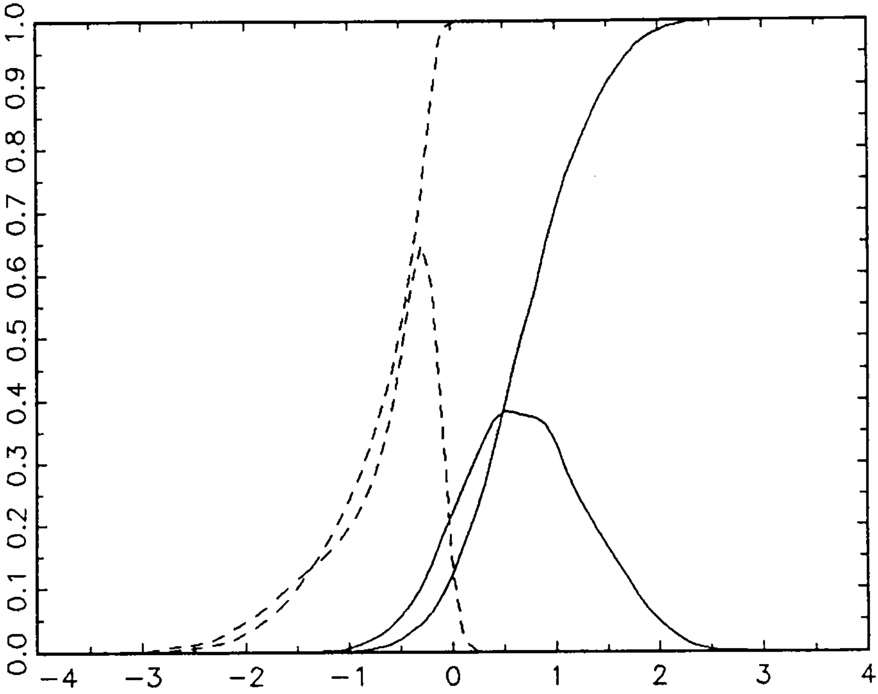

Figure 1

Asymptotic cdf and pdf of ϕ_{1T} under I(0) (solid line) and I(1) (dashed line): no detrending

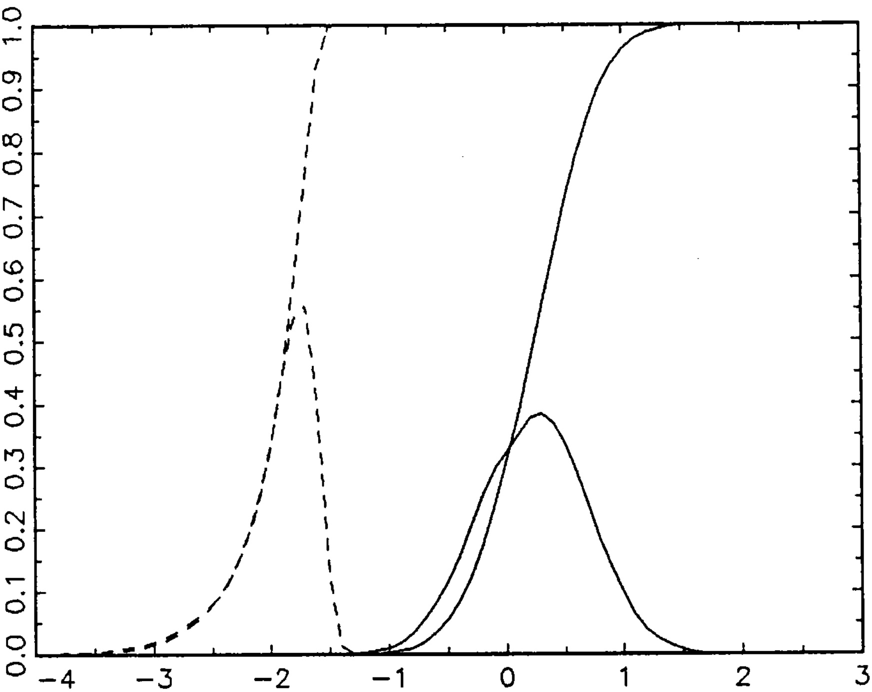

Figure 2

Asymptotic cdf and pdf of ϕ_{1T} under I(0) (solid line) and I(1) (dashed line): demeaned data

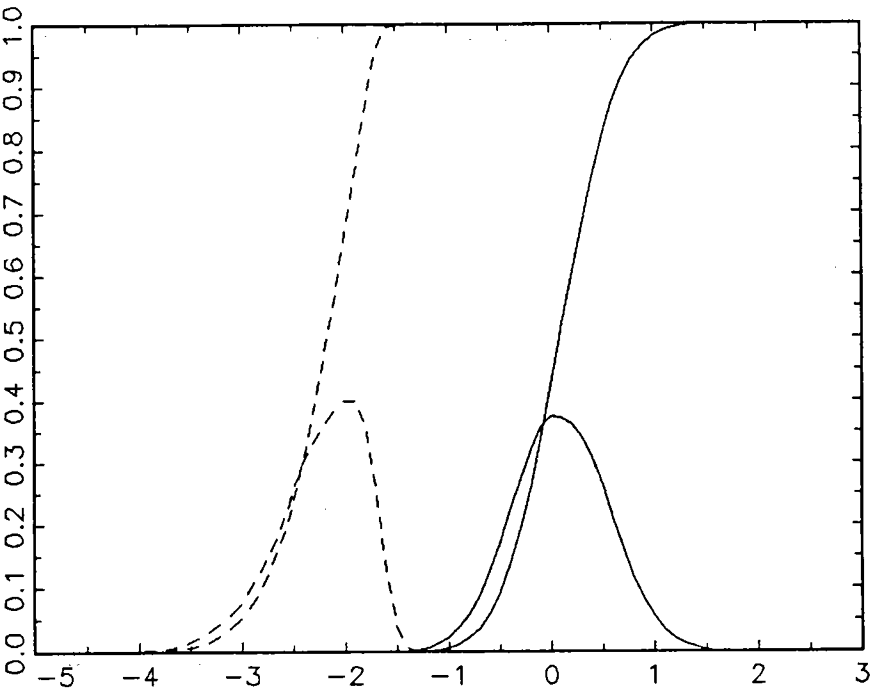

Figure 3

Asymptotic cdf and pdf of ϕ_{1T} under I(0) (solid line) and I(1) (dashed line): linear trend with OLS detrending

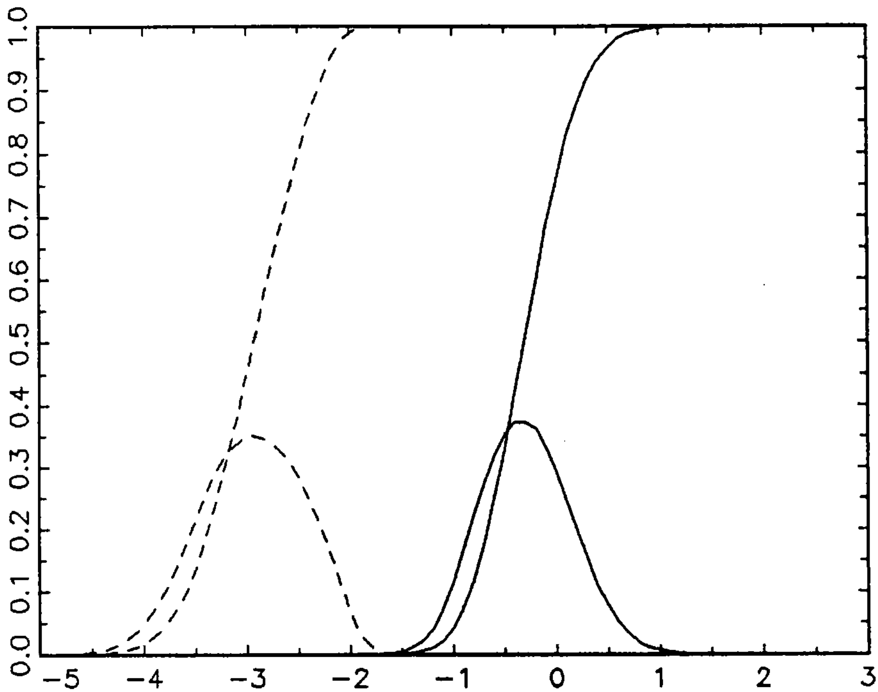

Figure 4

Asymptotic cdf and pdf of ϕ_{1T} under I(0) (solid line) and I(1) (dashed line): broken trend with OLS detrending, $\gamma=0$ case

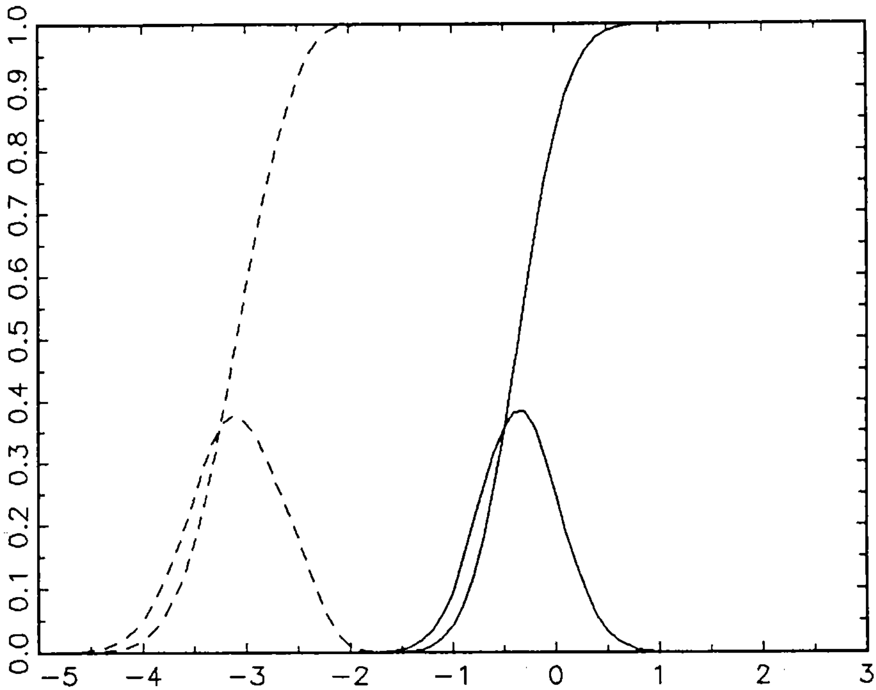

Figure 5

Asymptotic cdf and pdf of ϕ_{1T} under I(0) (solid line) and I(1) (dashed line): broken trend with OLS detrending, γ=50/T, τ_0 = .5 case

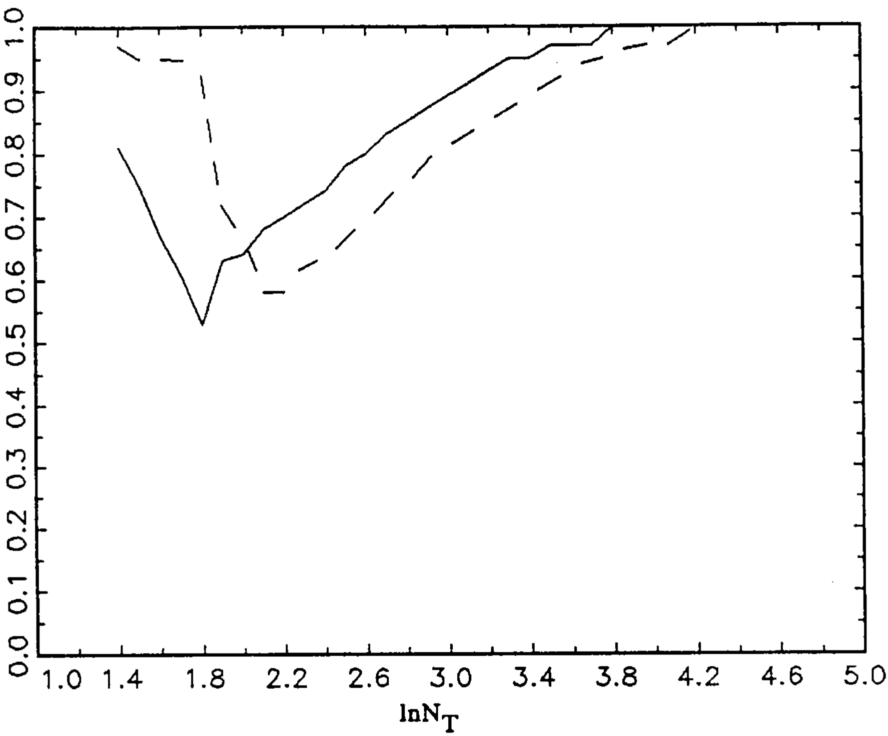

Figure 6

Approximate theoretical correct-classification rates for ϕ_{1T} as a function of $\ln N_T$ under I(0) (solid line) and I(1) (dashed line): demeaned data

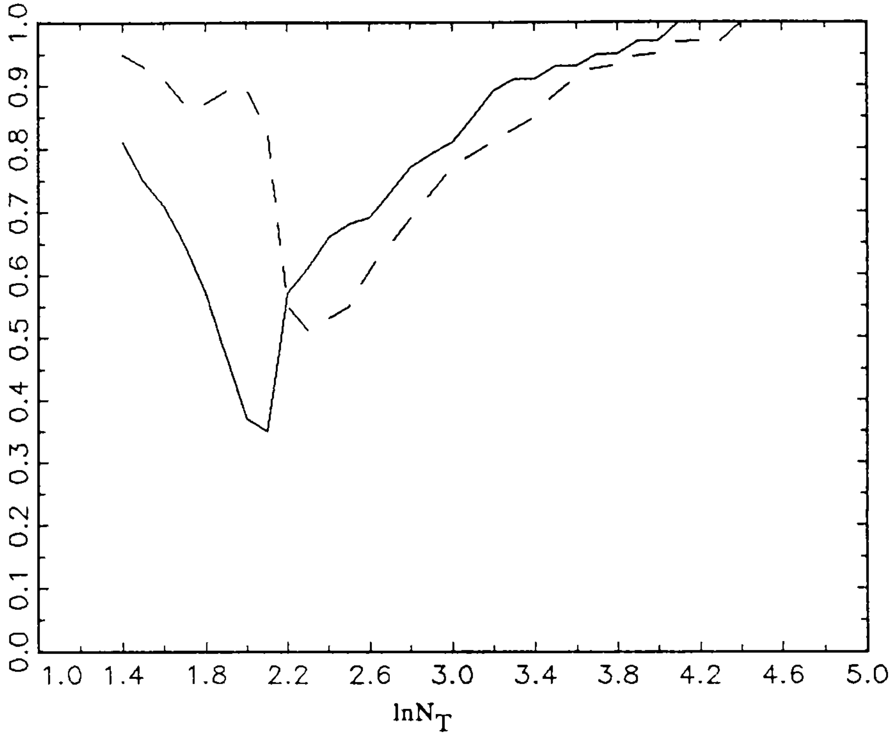

Figure 7

Approximate theoretical correct-classification rates for ϕ_{1T} as a function of $\ln N_T$ under I(0) (solid line) and I(1) (dashed line): linear trend with OLS detrending