INFERENCE IN TIME SERIES REGRESSION WHEN THE ORDER OF INTEGRATION OF A REGRESSOR IS UNKNOWN

GRAHAM ELLIOTT AND JAMES H. STOCK
Harvard University

The distribution of statistics testing restrictions on the coefficients in time series regressions can depend on the order of integration of the regressors. In practice, the order of integration is rarely known. We examine two conventional approaches to this problem — simply to ignore unit root problems or to use unit root pretests to determine the critical values for second-stage inference — and show that both exhibit substantial size distortions in empirically plausible situations. We then propose an alternative approach in which the second-stage critical values depend continuously on a first-stage statistic that is informative about the order of integration of the regressor. This procedure has the correct size asymptotically and good local asymptotic power.

1. INTRODUCTION

The asymptotic theory of classical inference in multivariate time series models when regressors have one or more unit roots is well understood (Chan and Wei [7], Park and Phillips [16], Phillips [19], Phillips and Durlauf [24], Sims, Stock, and Watson [29]). This theory has been developed under the assumption that the number and location of unit roots in the system is known a priori. Many inferences, such as inferences on the number of lags to include in a system, are typically unaffected by the presence of unit roots in the system. However, the null distribution of statistics testing certain quantities of economic interest, such as long-run effects of one variable on another, can depend on whether the regressor has a unit root. This poses difficulties in applied work in which it is rarely known whether a series actually has a unit root. This in turn can lead researchers either to ignore the problems that arise if a regressor is integrated or to use pretests (tests for unit roots or cointegration) to check if the regressors are integrated or cointegrated.

This article studies inference in a special case of this general problem, in which there is a single lagged regressor x_{t-1} which is suspected, but not

known, to have a unit root, that is, in which the researcher is unsure whether the regressor is integrated of order 0 or 1 (is $I(0)$ or $I(1)$, respectively). One motivating example is an empirical relation that has recently received considerable attention is the finance literature in which the lagged dividend yield (the dividend–price ratio) appears to be useful in predicting excess stock returns; see, for example, Fama and French [10], Campbell [3] and, for a review of this literature, Fama [9]. A typical regression in this literature is monthly (or longer) excess returns on a portfolio of stocks against a constant, lags of excess returns, and lags of the log dividend yield for the portfolio; the finding is that the lagged level of the log dividend yield enters as a significant predictor of excess returns. As emphasized in Campbell [3] and in Fama's [9] review, the importance of this regression arises from its apparently strong evidence against the "random walk" theory of stock prices. As several econometricians have noted, although finance theories typically predict that the dividend yield will be $I(0)$ even though stock prices are $I(1)$, for actual portfolios the dividend yield is only slowly mean reverting and the evidence that it does not have a unit root is weak. Although our primary motivation is this dividend yield regression, this regression is similar to regressions in the empirical consumption literature in which the growth rate of consumption is regressed against the lagged level of labor income or its logarithm (Flavin [11]; see Mankiw and Shapiro [13] and Stock and West [32] for discussions of the unit root issues in this context). It is also closely related to money–income causality regressions, although unlike the stock return and consumption examples, in the money–income case the null hypothesis (that money is not a useful predictor of income) does not imply that the dependent variable (income growth) is $I(0)$.

The purpose of this article is twofold. The first is to examine difficulties with conventional approaches to inference in this regression in light of the lack of asymptotic similarity of the t-test of the significance of x_{t-1}. If the largest autoregressive root of the regressor is nearly one, then ignoring unit root issues altogether can result in Granger causality tests with asymptotic sizes that far exceed their nominal levels. One reaction to this problem is to pretest for a unit root in the regressor and, depending on whether a unit root is rejected or not, respectively to adopt Gaussian or nonstandard critical values for inference in the second-stage regression. We show in Section 2, however, that this two-stage approach also can produce large distortions in the size of the second-stage test. For example, when a Dickey–Fuller [8] t-statistic is used to pretest for a unit root in x_t, a two-sided second-stage test with nominal level of 10% can have a size that exceeds 30% for sample sizes encountered in econometric practice.

The second purpose of the article is to propose an alternative approach to this problem, referred to as the Bayesian mixture approximation. In the proposed approach, critical values for the second-stage test depend on a statistic ϕ_T that is informative about the order of integration of x_t. The condi-

tional distribution of the test statistic given ϕ_T is computed as a mixture of two asymptotic conditional distributions for the test statistic, conditional on x_t being $I(0)$ or $I(1)$, with the mixture probabilities being given by posterior probabilities that x_t is $I(0)$ or $I(1)$, respectively, $p(I(0)\,|\,\phi_T)$ and $p(I(1)\,|\,\phi_T)$. The proposed procedure has three desirable features. First, under the null that the regressor has no predictive content, asymptotically the mixture distribution will provide the correct critical values, that is, for any fixed $I(0)$ or $I(1)$ model, the test has the correct size. Second, the (second-stage) test has good power against local Granger causality alternatives. Specifically, if the order of integration d is zero, then the local asymptotic power of the proposed test is the same as the likelihood ratio test that imposes $d = 0$ a priori. If $d = 1$, the proposed test can be compared with the test based on the same regression, but using the correct $I(1)$ unconditional (on ϕ_T) critical values; although neither test dominates the other, the proposed test has higher power against most local alternatives. Finally, the proposed procedure avoids the difficulties of defining priors over parametric representations of x_t within the $I(0)$ or $I(1)$ classes (e.g., see the debate between Phillips [21] and his discussants) and instead entails defining priors over the point hypotheses "$I(1)$" and "$I(0)$." Because priors are not specified over the entire parameter space, this is not a fully Bayesian treatment of inference in the second-stage regression. From a classical perspective the proposed procedure can be seen as a device for approximating the distribution of the t-statistic when the researcher has no a priori information on whether x_t is $I(0)$ or $I(1)$. From this perspective, the argument in favor of the approximation is that it provides an asymptotically similar test with desirable power properties, thereby circumventing the pitfalls of unit root pretesting in this application. The prior on $I(1)$ can be considered a tuning parameter to be chosen by the researcher, for example, based on a Monte-Carlo study of the effect of the choice of prior on size or power in leading finite-sample models.[1]

The outline of the article is as follows. Section 2 sets out the model under investigation and documents the size distortions introduced by the two conventional second-stage test procedures just described. The Bayesian mixture approximation, the construction of the posterior probabilities $p(I(0)\,|\,\phi_T)$ and $p(I(1)\,|\,\phi_T)$, and the asymptotic properties of the Bayesian mixture approximation are given in Section 3. Section 4 presents results on the posterior probabilities when x_t has large, but not unit autoregressive or moving average (MA) roots, respectively, the local-to-$I(1)$ and local-to-$I(0)$ cases; these results are then used to examine the performance of the Bayesian mixture approximation when x_t is local-to-$I(1)$. The asymptotic power of this procedure against local Granger causality alternatives is studied in Section 5. Numerical issues are reported, and a Monte-Carlo experiment is discussed in Section 6. Section 7 concludes.

Throughout the paper, it is assumed that a constant is included in the second-stage regression and that x_t is a driftless process. The results for

Granger-causality tests are developed for the special case of demeaned data, that is, when the only deterministic regressor in the second-stage regression is a constant. However, to facilitate extensions to higher-order detrending, the theory of the first-stage Bayesian classifier is developed in Sections 3.1 and 4.1 for general polynomial detrending.

2. THE MODEL AND PROBLEMS WITH CONVENTIONAL SECOND-STAGE INFERENCE TECHNIQUES

2.1. The Model

The data are assumed to be generated by the bivariate autoregressive system,

$$x_t = \mu_x + \alpha(L)x_{t-1} + \beta(L)y_{t-1} + \eta_{1t} \tag{1a}$$

$$y_t = \mu_y + \gamma x_{t-1} + \eta_{2t}, \tag{1b}$$

where L is the lag operator, $\eta_t = (\eta_{1t}, \eta_{2t})'$ is a martingale difference sequence with $E(\eta_t \eta_t' \mid \eta_{t-1}, \eta_{t-2}, \ldots) = \Sigma$ and with $\sup_t E\eta_{it}^4 < \infty$, $i = 1, 2$, and $\alpha(L)$ and $\beta(L)$ have finite orders. If x_t is $I(1)$, then one of the roots of $(1 - \alpha(L)L)$ equals one and the remaining roots are assumed to be fixed and greater than one in modulus. If x_t is $I(0)$, then all the roots of $(1 - \alpha(L)L)$ are fixed and exceed one in modulus. The null hypothesis to be tested is that $\gamma = 0$. It is assumed that x_t has no drift in its univariate representation, that is, $\tilde{\mu}_x = 0$, where $\tilde{\mu}_x = \mu_x + \beta(1)\mu_y$.

The specification in (1b) ignores the possibility of multiple lags of x_t, or of lagged y_t, being useful in predicting y_t given x_{t-1}. Our reason for focusing on this restricted system is that the conceptual difficulties are associated with estimating the levels effect of the possibly integrated regressor x_t. It follows from results in Chan and Wei [7], Park and Phillips [16], and Sims, Stock, and Watson [29] that, if additional lags of x_t are included in this regression, then Wald tests on these additional lags will have conventional chi-square asymptotic distributions whether x_t is $I(0)$ or $I(1)$; moreover, if x_t is $I(1)$, then the test on additional lags is asymptotically independent of the test of the levels effect of x_t. Thus, only inference concerning the levels effect is affected by the order of integration of x_t.

The local-to-$I(1)$ model studied below nests the largest root of $1 - \alpha(L)L$ as being in a $1/T$ neighborhood of 1. We therefore reparameterize (1) to isolate this largest root. Factor $(1 - \alpha(L)L)$ as $(1 - \rho L)(1 - \tilde{\alpha}(L)L)$, where ρ is the largest real root of $(1 - \alpha(L)L)$ and let $\tilde{\Delta} = (1 - \rho L)$. By rewriting y_{t-1} in (1a) as deviations from its mean under the null, we have

$$\tilde{\Delta}x_t = \tilde{\mu}_x + \tilde{\alpha}(L)\tilde{\Delta}x_{t-1} + \beta(L)(y_{t-1} - \mu_y) + \eta_{1t} \tag{2a}$$

$$y_t = \mu_y + \gamma x_{t-1} + \eta_{2t}. \tag{2b}$$

In terms of (2), the $I(1)$ hypothesis is that $\rho = 1$ and the $I(0)$ hypothesis is that $|\rho| < 1$. Define Ω to be 2π times the spectral density matrix of $(\tilde{\Delta}x_t, \eta_{2t})$ at frequency zero, and let $\delta = \Omega_{21}/(\Omega_{11}\Omega_{22})^{1/2}$.

2.2. Size Distortions Introduced by Using Gaussian Critical Values or Unit Root Pretests

Two conventional approaches to inference in this problem are either to use a standard normal approximation to the distribution of t_γ, regardless of any information about the degree of persistence in x_t, or to use conventional unit root pretests to determine second-stage critical values. This subsection examines the consequences of these approaches. To simplify exposition, the problems with these two procedures are illustrated in the special case of (2a) in which there is no feedback and x_t follows an AR(1) process. Specifically, let

$$x_t = \mu_x + \rho x_{t-1} + \eta_{1t} \tag{3a}$$

$$y_t = \mu_y + \gamma x_{t-1} + \eta_{2t}, \tag{3b}$$

where η_t satisfies the conditions stated following (1). The assumption $\tilde{\mu}_x = 0$ here simplifies to $\mu_x = 0$. In this simple model, $\Omega = \Sigma$ so $\delta = \text{corr}(\eta_{1t}, \eta_{2t})$ and the asymptotic distribution of t_γ is determined by ρ and, if ρ is local to one, by δ.

Under the null hypothesis that $\gamma = 0$, it can be shown that the de-meaned Dickey–Fuller t-statistic testing $\rho = 1(t_{DF}^\mu)$ and the Granger causality t-statistic testing $\gamma = 0$, t_γ, are related by the expression,

$$t_\gamma = \delta T(1-\rho)\left(\frac{T^{-2}\sum_{t=1}^{T}(x_{t-1}^\mu)^2}{\hat{\Sigma}_{11}}\right)^{1/2} + \delta t_{DF}^\mu + (1-\delta^2)^{1/2}z_T + o_p(1) \tag{4}$$

where $z_T = \{\sum_{t=1}^{T} x_{t-1}^\mu(\eta_{2t} - \text{Proj}(\eta_{2t}|\eta_{1t}))\}/\{(\sum_{t=1}^{T}(x_{t-1}^\mu)^2)\tilde{\Sigma}_{22}\}^{1/2}$, $\hat{\Sigma}_{11} = T^{-1}\sum_{t=1}^{T}\hat{\eta}_{1t}^2$, $\tilde{\Sigma}_{22} = \Sigma_{22} - \Sigma_{11}^{-1}\Sigma_{12}^2$, $x_t^\mu = x_t - \bar{x}$, and $\text{Proj}(\eta_{2t}|\eta_{1t}) = \Sigma_{21}\Sigma_{11}^{-1}\eta_{1t}$. If $\rho = 1$, $(t_{DF}^\mu, z_T) \Rightarrow (\hat{\tau}^\mu, z)$, where $\hat{\tau}^\mu$ denotes the asymptotic representation of the demeaned Dickey–Fuller t-statistic, z is a standard normal random variable, $\hat{\tau}^\mu$ and z are independent, and $\Rightarrow$ denotes weak convergence of random elements of $D[0,1]$. If $\rho = 1$ and $\delta = \pm 1$, then $t_\gamma = \delta t_{DF}^\mu$; if $\rho = 1$ and $\delta = 0$, then asymptotically t_γ is normally distributed and is independent of t_{DF}^μ; and if $\rho = 1$ and $0 < |\delta| < 1$, then t_γ is asymptotically distributed as a linear combination of independent $\hat{\tau}^\mu$ and z statistics.

The distribution of t_γ can be obtained by using (4) when ρ is nearly one, in the sense that $\rho = 1 + c/T$ where c is a constant. This local-to-unity nesting has been studied extensively by Bobkoski [2], Cavanagh [4], Chan and Wei [6], Chan [5], Nabeya and Tanaka [15], Perron [17], and Phillips [20], (see Nabeya and Tanaka [15] for a recent review of theoretical results).

If $\rho = 1 + c/T$, then $T^{-1/2}(x_{[T\cdot]} - \bar{x}) \Rightarrow \Sigma_{11}^{1/2} B_c^{\mu}(\bullet)$, where $B_c^{\mu}(s) = B_c(s) - \int_0^1 B_c(r)\,dr$, where $B_c(s)$ is a diffusion process that satisfies $dB_c(s) = cB_c(s)ds + dW(s)$ (e.g., Phillips [20]), where $[\bullet]$ denotes the greatest lesser integer function, W is a standard Brownian motion on the unit interval, and $B_c(0) = 0$. Under this local-to-$I(1)$ nesting for x_t,

$$(t_\gamma, t_{DF}^\mu) \Rightarrow \left(\delta \frac{\left\{ \int_0^1 B_c^\mu(s)\,dW(s) \right\}}{\left\{ \int_0^1 B_c^\mu(s)^2\,ds \right\}^{1/2}} + (1 - \delta^2)^{1/2} z, \right.$$

$$\left. c \left(\int_0^1 B_c^\mu(s)^2\,ds \right)^{1/2} + \frac{\left\{ \int_0^1 B_c^\mu(s)\,dW(s) \right\}}{\left\{ \int_0^1 B_c^\mu(s)^2\,ds \right\}^{1/2}} \right), \tag{5}$$

where z is asymptotically independent of the functionals of B_c^μ. Thus, when ρ is local to one and $\gamma = 0$, the qualitative results concerning the distribution of t_γ are similar to the $\rho = 1$ case: asymptotically, when $\delta = 0$, t_γ is normally distributed independently of t_{DF}, but for nonzero δ, t_γ has a nonstandard distribution and in general t_γ and t_{DF} are dependent.

The representations in (4) and (5) permit the analysis of the sizes of the two conventional approaches to inference in this problem. First, consider the case in which standard Gaussian critical values are used to evaluate the significance of t_γ. If $\delta = 0$ or if ρ is fixed and less than one, then t_γ has an asymptotic $N(0,1)$ distribution and this inference is justified. However, if $\rho = 1$ and $\delta \neq 0$, the distribution is nonstandard. Equally important, the local-to-unity result in (5) indicates that if ρ is large and $\delta \neq 0$, then the distribution of t_γ will be nonstandard and the normal distribution will provide a poor approximation.

Table 1 presents evidence on the magnitude of these effects, specifically rejection rates of the t-test of the null hypothesis $\gamma = 0$ when data are generated according to (3) with $\gamma = 0$, t_γ is computed by regressing y_t onto $(1, x_{t-1})$, and rejection occurs when t_γ falls outside the standard Gaussian 5% and 95% critical values. (Because the distribution of t_γ is symmetric in δ, Table 1 and subsequent tables present only results for $\delta \leq 0$.) As the theory predicts, there are no appreciable size distortions when $\delta = 0$, even if ρ is large. However, for nonzero δ, the size distortions can be substantial. For example, when $\delta = -.9$, $\rho = .95$, and $T = 50$, the rejection rates are under 1% in the left tail and 22% in the right tail.

A second approach to inference on γ is to pretest for a unit root in x_t by using a one-sided test. If the unit root null is rejected, then the $I(0)$ standard normal distribution is used, whereas if the unit root null is not rejected, then the $I(1)$ distribution obtained from (5) with $c = 0$ is used. A natural unit root

TABLE 1. Size of t-tests of $\gamma = 0$ with standard Gaussian critical values: One-sided tests with nominal level 5% [a]

δ	T	\multicolumn{6}{c}{$\Pr(t_\gamma < -1.645)$, $\rho =$}	\multicolumn{6}{c}{$\Pr(t_\gamma > 1.645)$, $\rho =$}
		.6	.8	.9	.95	.975	1	.6	.8	.9	.95	.975	1
−.9	50	.02	.01	.01	.00	.00	.00	.10	.13	.16	.22	.28	.38
	100	.03	.02	.01	.00	.00	.00	.08	.10	.13	.17	.23	.39
−.5	50	.04	.03	.02	.02	.01	.01	.07	.09	.11	.13	.14	.18
	100	.03	.03	.03	.02	.01	.01	.06	.08	.09	.12	.12	.19
0	50	.05	.06	.06	.06	.05	.06	.06	.05	.05	.05	.06	.06
	100	.06	.05	.05	.05	.05	.05	.05	.06	.06	.05	.05	.05

[a]The entries are rejection rates when t_γ is compared with ± 1.645. The pseudodata were generated according to (3) with i.i.d. $N(0,\Sigma)$ errors, with $\Sigma_{11} = \Sigma_{22} = 1$ and $\Sigma_{12} = \delta$. t_γ is the t-statistic testing $\gamma = 0$ in a regression of y_t onto $(1, x_{t-1})$. Based on 5,000 Monte-Carlo replications.

pretest is the demeaned Dickey–Fuller t-statistic, t^μ_{DF}. The difficulty with this two-stage procedure arises when ρ is one or local-to-one and $\delta \neq 0$, so that t_γ and t^μ_{DF} are asymptotically dependent; then inference on t_γ, conditional on t^μ_{DF}, differs from unconditional inference. Consider the extreme case $(\rho, \delta) = (1, -1)$, so that $t_\gamma = -t^\mu_{DF}$. Then one-sided (left-tail) failure to reject at the α_1 level in the first stage ensures one-sided (right-tail) acceptance at the α_2 level in the second stage for any $\alpha_2 < \alpha_1$. First-stage rejection of $\rho = 1$ by using t^μ_{DF} at the α_1 level (with critical value $c_{DF;\alpha_1}$) leads to using standard normal critical values. As long as $-c_{DF;\alpha_1} > c_{z;\alpha_2}$ (typically true because of the skewness of the $\hat{\tau}^\mu$ asymptotic distribution), first-stage rejection implies a second-stage rejection with probability one. Thus, the asymptotic size of a second-stage right-tailed test of nominal level α_2 is in fact α_1, as long as $\alpha_2 < \alpha_1$ and $-c_{DF;\alpha_1} > c_{z;\alpha_2}$.

This size distortion is found more generally when ρ is large and $\delta \neq 0$, and it is present in two-sided as well as one-sided tests. Monte-Carlo evidence on sizes obtained by using this sequential testing procedure for various values of ρ, δ, and T are summarized in Table 2. The first-stage test is a 20% one-sided Dickey–Fuller [8] t-test for a unit root with a constant and no lags of Δx_t in the regression; the second-stage test is equal-tailed with nominal size 10%, that is, the 5% and 95% quantiles of the $I(0)$ or $I(1)$ distribution are used, depending on the outcome of the first-stage Dickey–Fuller test. As the theory predicts, when $\delta = 0$, there is no size distortion introduced by the pretest. (When $\delta = 0$, t_γ has an asymptotic normal distribution in both the $I(0)$ and $I(1)$ cases, so the same critical values (± 1.645) are used whether or not t^μ_{DF} rejects.) However, when $\delta \neq 0$, the $I(0)$ and $I(1)$ asymptotic distribu-

TABLE 2. Size of t-tests of $\gamma = 0$ with critical values selected by pretesting with a Dickey–Fuller t-statistic: Equal-tailed two-sided tests with nominal level 10%[a]

δ	T	.6	.8	.9	.95	.975	1.0
				ρ			
−.9	50	.17	.29	.30	.26	.24	.22
	100	.11	.15	.30	.32	.28	.23
−.5	50	.13	.16	.16	.17	.15	.14
	100	.10	.12	.16	.17	.16	.15
0	50	.11	.11	.10	.10	.10	.10
	100	.10	.10	.10	.11	.10	.10

[a]Entries are rejection rates in the second-stage test of $\gamma = 0$ when critical values are chosen either from the standard Gaussian distribution or from the appropriate $I(1)$ asymptotic distribution, given δ, depending on the outcome of a preliminary Dickey–Fuller test for a unit root in x_t (from a regression of Δx_t on $(1, x_{t-1})$). The first-stage test is one-sided against the stationary alternative at the 20% level; the second-stage test is two-sided based on the nominal 5% and 95% percentiles of the marginal $I(0)$ or $I(1)$ distributions of t_γ. The data were generated according to (3) with i.i.d. $N(0, \Sigma)$ errors. Based on 20,000 Monte-Carlo replications.

tions of t_γ differ, and the size distortions can be large, with rejection rates exceeding 30% for large values of $|\rho|$ even with $T = 100$.[2]

A source of the size distortions in Table 2 is that the critical values in the first stage are fixed, so the Dickey–Fuller test does not consistently select the true order of integration. A theoretical solution to this problem is to consider sequences of unit root pretests in which the critical values are indexed to the sample size such that the type I and type II error rates simultaneously tend to a limit of zero. Thus, asymptotically, the correct null distribution for t_γ would be selected with probability one, and the second-stage test would have the correct size, at least for ρ fixed. But this device, although theoretically attractive, has little practical value: without further refinement, it fails to specify the first-stage critical values to use in samples of the size typically found in empirical work, so the difficulties outlined in the preceding paragraphs would remain an accurate description of the pitfalls facing applied researchers.

3. THE BAYESIAN MIXTURE APPROXIMATION

Our proposed approach to this problem is to consider the distribution of t_γ conditional on a statistic ϕ_T that is informative about the root ρ or (less parametrically) about whether x_t is $I(0)$ or $I(1)$. Because of the asymptotically different distributions under the $I(1)$ and $I(0)$ cases, it is useful to treat

the order of integration d as a dichotomous unknown parameter. Instead of performing a pretest on this unknown parameter, a Bayesian procedure is used to construct posterior probabilities for d given ϕ_T. This approach can be developed for general ϕ_T as long as these posterior probabilities can be computed. In this article, however, we focus on a specific class of ϕ_T statistics developed by Stock [30]. Before turning to the proposed Bayesian mixture procedure, we briefly review the construction and properties of these statistics. The theory for ϕ_T is set out for general detrending, although only the results for the demeaned case are used in the subsequent analysis of inference on γ.

3.1. Construction of Posterior Probabilities for $I(1)$ and $I(0)$

The construction of the proposed approximation relies on a class of statistics $\phi_T = \phi(V_T)$, introduced by Stock [30], that permit computing the posterior probability that x_t is $I(1)$ or $I(0)$ when prior probabilities are placed solely on the point hypotheses $I(1)$ and $I(0)$, that is, without reference to a specific prior distribution over parametric representations of the x_t process. (For an alternative approach see Phillips and Ploberger [25].) Write x_t as

$$x_t = d_t + u_t, \tag{6}$$

where d_t is deterministic and u_t is stochastic. In the general case, the $I(0)$ and $I(1)$ hypotheses are taken to refer to properties of partial sums of u_t. Let $U_{0T}(\lambda) = T^{-1/2} \sum_{s=1}^{[T\lambda]} u_s$ and $U_{1T}(\lambda) = T^{-1/2} u_{[T\lambda]}$, let $\gamma_x(j) = \mathrm{cov}(x_t, x_{t-j})$ for a second-order stationary process x_t. Throughout it is assumed that initial conditions have finite variance. The $I(0)$ and $I(1)$ models are respectively defined by:

$$I(0): U_{0T} \Rightarrow \omega_0 W, \quad \text{where } \omega_0^2 = \sum_{j=-\infty}^{\infty} \gamma_u(j), \quad 0 < \omega_0^2 < \infty, \tag{7}$$

$$I(1): U_{1T} \Rightarrow \omega_1 W, \quad \text{where } \omega_1^2 = \sum_{j=-\infty}^{\infty} \gamma_{\Delta u}(j), \quad 0 < \omega_1^2 < \infty. \tag{8}$$

The V_T statistic is constructed by using detrended data x_t^d. Let $x_t^d = x_t - \hat{d}_t$, where $\hat{d}_t$ is an estimator of d_t. Let $V_T(\lambda) = \hat{\omega}^{-1} T^{-1/2} \sum_{s=1}^{[T\lambda]} x_s^d$, where $\hat{\omega}^2 = \sum_{m=-\ell_T}^{\ell_T} k(m/\ell_T) \hat{\gamma}_{xd}(|m|)$ and $\hat{\gamma}_{xd}(m) = T^{-1} \sum_{t=m+1}^{T} x_t^d x_{t-m}^d$. The kernel $k(\bullet)$ is assumed to satisfy $k(w) = 0$ for $|w| \geq 1$; $k(w) = k(-w)$; $0 < k(w) \leq 1$ for $|w| < 1$; $k(0) = 1$; and $\ell^{-1} \sum_{u=1}^{\ell} k(u/\ell) \geq \kappa$ for all $\ell > 1$, where $\kappa > 0$. It is assumed that the sequence of lag truncation parameters, ℓ_T, satisfies $\ell_T^2 \ln T/T \to 0$, $\ell_T \to \infty$. Let $N_T = T/\sum_{m=-\ell_T}^{\ell_T} k(m/\ell_T)$. Let the trend estimation error be $\tilde{d}_t = \hat{d}_t - d_t$. Let $\|z_t\| = T^{-1} \sum_{t=1}^{T} z_t^2$ for a time series z_t, $D_{0T}(\lambda) = T^{-1/2} \sum_{s=1}^{[T\lambda]} \tilde{d}_s$, and $D_{1T}(\lambda) = T^{-1/2} \tilde{d}_{[T\lambda]}$. The estimated trend is assumed to satisfy the following conditions.

Detrending Condition A. If u_t is $I(0)$, then

(i) $(U_{0T}, D_{0T}) \Rightarrow \omega_0(W, D_0)$, where $D_0 \in C[0,1]$, and
(ii) $\ell_T^2 \|\tilde{d}_t\| \overset{p}{\to} 0$.

Detrending Condition B. If u_t is $I(1)$, then

(i) $(U_{1T}, D_{1T}) \Rightarrow \omega_1(W, D_1)$, where $D_1 \in C[0,1]$, and
(ii) $\|\Delta \tilde{d}_t\| = O_p(1)$.

The properties of specific functionals of V_T under the $I(1)$ and $I(0)$ models for several detrending processes have been studied by Kwiatkowski et al. [12], Phillips [22], and Perron [18]. These results are extended to the general Detrending Conditions A and B in Stock [30, Theorem 1], where it is shown that if x_t is $I(0)$, then $V_T \Rightarrow W_0^d$, whereas if x_t is $I(1)$, then $N_T^{-1/2} V_T \Rightarrow V_1^d$, where $V_1^d(\lambda) = \int_0^\lambda W_1^d(s) \, ds / \{\int_0^1 W_1^d(s)^2 \, ds\}^{1/2}$, where $W_0^d(s) = W(s) - D_0(s)$ and $W_1^d(s) = W(s) - D_1(s)$.

The general detrending conditions are satisfied by polynomial detrending by ordinary least squares (OLS) and by piecewise linear ("broken-trend") detrending. We will focus on the case of constants included in the regression so that a mean is subtracted from the data. Accordingly, denote the demeaned processes W_0^d and W_1^d by W_0^μ and W_1^μ, respectively, where $W_0^\mu(s) = W(s) - sW(1)$ and $W_1^\mu(s) = W(s) - \int_0^1 W(r) \, dr$.

These results permit computing the posteriors that x_t is $I(1)$ or $I(0)$. Let $\phi(\bullet)$ be a functional such that (i) ϕ is a continuous mapping from $D[0,1] \to \mathfrak{R}^1$; (ii) $\phi(ag) = \phi(g) + 2 \ln a$, where a is a scalar and $g \in D[0,1]$; 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 $\phi_T = \phi(V_T)$; under these conditions, if x_t is $I(0)$, then $\phi_T \Rightarrow \phi(W_0^d)$, whereas if x_t is $I(1)$, $\phi_T - \ln N_T \Rightarrow \phi(V_1^d)$. The posterior probability that the series is $I(d)$, given the statistic ϕ_T, is $p(I(d) | \phi_T) = p(\phi_T | I(d)) \Pi_d / p(\phi_T)$, where $\Pi_d = p(I(d))$ is the prior probability that the process is $I(d)$, $d = 0,1$. In large samples, $p(\phi_T | I(0))$ and $p(\phi_T | I(1))$, respectively, can be approximated by $f_0(\phi_T)$ and $f_1(\phi_T - \ln N_T)$.

With these asymptotic approximations, the posterior probabilities can be computed as

$$p(I(0) | \phi_T) = \frac{f_0(\phi_T) \Pi_0}{p(\phi_T)}, \tag{9a}$$

$$p(I(1) | \phi_T) = \frac{f_1(\phi_T - \ln N_T) \Pi_1}{p(\phi_T)}. \tag{9b}$$

A consequence of Theorems 1 and 2 in Stock [30] is that the posteriors asymptotically converge to zero or one: if $0 < \Pi_0, \Pi_1 < 1$ and if x_t is $I(0)$, then $p(I(0) | \phi_T) \overset{p}{\to} 1$ and $p(I(1) | \phi_T) \overset{p}{\to} 0$, whereas if x_t is $I(1)$, then $p(I(0) | \phi_T) \overset{p}{\to} 0$ and $p(I(1) | \phi_T) \overset{p}{\to} 1$.

3.2. Bayes Mixture Approximation to the Distribution of t_γ

The consistency of the posteriors in (9) suggests their use for constructing the asymptotic approximation to the distribution of t_γ conditional on ϕ_T, specifically,

$$p(t_\gamma | \phi_T) = p(t_\gamma | \phi_T, d = 0)p(I(0) | \phi_T) + p(t_\gamma | \phi_T, d = 1)p(I(1) | \phi_T).$$

$$(10)$$

As is made precise in Theorem 1, under (2), with $\gamma = 0$, the conditional distributions $p(t_\gamma | \phi_T, d = 0)$ and $p(t_\gamma | \phi_T, d = 1)$ asymptotically depend on only one nuisance parameter in the system in (2), δ, and so are readily computed.

The formulation in (10) has three parallel motivations. The first comes from its asymptotic properties. Because the posterior probabilities are consistent for zero or one, depending on d, $p(t_\gamma | \phi_T)$ constructed by using (10) has the property that, for fixed ρ, the correct distribution for t_γ is used asymptotically.

The second motivation comes from recognizing (10) as a mixture of the $I(0)$ and $I(1)$ distributions with weights given by the posteriors $p(I(d) | \phi_T)$; the greater the posterior weight on $d = 1$ for a realization of ϕ_T, say, the greater the weight given to the $d = 1$ conditional distribution.

The third motivation comes from drawing an analogy between (10) and the Bayesian posterior distribution for γ. Suppose that we had available statistics S_ρ and S_γ, where S_ρ is informative for ρ and (S_ρ, S_γ) are informative for γ. Let θ denote the vector of nuisance parameters so that (γ, ρ, θ) comprise the complete parameter vector. From a Bayesian perspective, one might be interested in the posterior distribution of γ given (S_ρ, S_γ),

$$p(\gamma | S_\rho, S_\gamma) = \int_\rho \int_\theta p(S_\rho, S_\gamma | \gamma, \rho, \theta)p(\gamma, \rho, \theta) \, d\theta \, d\rho / p(S_\rho, S_\gamma). \qquad (11)$$

Next make three assumptions: (i) the dependence of $p(S_\gamma | S_\rho, \gamma, \rho, \theta)$ on ρ reduces to whether $\rho = 1$ or $|\rho| < 1$, specifically, $p(S_\gamma | S_\rho, \gamma, \rho, \theta) = p(S_\gamma | S_\rho, \gamma, d = 1, \theta)\mathbf{1}(\rho = 1) + p(S_\gamma | S_\rho, \gamma, d = 0, \theta)\mathbf{1}(|\rho| < 1)$; (ii) $p(S_\rho | \gamma, \rho, \theta)$ does not depend on θ or γ, and depends on ρ only through $\rho = 1$ or $|\rho| < 1$, specifically, $p(S_\rho | \gamma, \theta, \rho) = p(S_\rho | d = 1)\mathbf{1}(\rho = 1) + p(S_\rho | d = 0)\mathbf{1}(|\rho| < 1)$; and (iii) the priors on γ are flat and $p(\gamma, \rho, \theta)$ satisfies $p(\gamma, \rho, \theta) \propto p(\theta)p(d = 1)\delta^*(\rho - 1) + p(\theta)p(\rho)\mathbf{1}(|\rho| < 1)$, where $\int_\theta p(\theta) \, d\theta = 1$, $\delta^*(\bullet)$ is the Dirac delta function, and $\int_{|\rho| < 1} p(\rho) \, d\rho = p(d = 0)$, where $0 \leq p(d = 0) \leq 1$ and $p(d = 0) + p(d = 1) = 1$. (An implication of Theorem 1 is that assumptions (i) and (ii) are satisfied asymptotically for $(S_\rho, S_\gamma) = (\phi_T, t_\gamma)$.) With these assumptions, (11) simplifies to

$$p(\gamma \mid S_\rho, S_\gamma) \propto \left\{ \int_\theta p(S_\gamma \mid S_\rho, \gamma, d = 0, \theta) p(\theta) \, d\theta \right\} p(d = 0 \mid S_\rho)$$

$$+ \left\{ \int_\theta p(S_\gamma \mid S_\rho, \gamma, d = 1, \theta) p(\theta) \, d\theta \right\} p(d = 1 \mid S_\rho). \tag{12}$$

Except for the integration over θ, if $(S_\rho, S_\gamma) = (\phi_T, t_\gamma)$, then the right-hand side of (12), evaluated at $\gamma = 0$, is the same as the right-hand side of (10) evaluated under the $\gamma = 0$ null. This leads to the motivation of (10) as a large-sample approximation to the posterior in (12). For the system in (2), the dependence of the asymptotic approximation of $p(S_\gamma \mid S_\rho, \gamma, d = 0, \theta)$ on θ is limited to the single parameter δ. Although in principle one could integrate over a prior on this parameter, in general prior beliefs about δ are likely to be weak, and in any event δ is consistently estimable, so in keeping with previous appeals to first-order asymptotic approximations, we treat δ as known. The analogy between (12) and (10) would be more compelling were (S_ρ, S_γ) sufficient for (ρ, γ), which (ϕ_T, t_γ) are not. For tractability, however, we restrict attention to the statistics (ϕ_T, t_γ).

The value of (10) is that it provides an approximation to the conditional distribution of t_γ, which is readily computable, depends on only one nuisance parameter δ, and asymptotically delivers the correct null distribution of t_γ whether x_t is $I(0)$ or $I(1)$ as determined by the fixed parameter ρ. These properties are implied by the following theorem.

THEOREM 1. *Let (x_t, y_t) be generated according to (2) and let t_γ be the t-statistic that tests $\gamma = 0$ in (2) (with a constant included in the regression). Suppose that $\gamma = 0$.*

    (a) If x_t is $I(0)$, then

        (i) $(t_\gamma, \phi_T) \Rightarrow (z_0, \phi(W_0^\mu))$, where z_0 is a standard normal random variable distributed independently of $\phi(W_0^\mu)$; and

        (ii) $p(I(0) \mid \phi_T) \overset{P}{\to} 1$ and $p(I(1) \mid \phi_T) \overset{P}{\to} 0$.

    (b) If x_t is $I(1)$, then

        (i) $(t_\gamma, \phi_T - \ln N_T) \Rightarrow \left(\delta \int_0^1 W^\mu(s) \, dW(s) / \left\{ \int_0^1 W^\mu(s)^2 \, ds \right\}^{1/2} + (1 - \delta^2)^{1/2} z_1, \right.$ $\left. \phi(V_1^\mu) \right)$, where $V_1^\mu(\lambda) = \int_0^\lambda W_1^\mu(s) \, ds / \left\{ \int_0^1 W_1^\mu(s) \, ds \right\}^{1/2}$ and z_1 is a standard normal random variable distributed independently of W^μ; and

        (ii) $p(I(0) \mid \phi_T) \overset{P}{\to} 0$ and $p(I(1) \mid \phi_T) \overset{P}{\to} 1$.

Proof. All proofs are given in the Appendix.

Although the results in Theorem 1 are presented for x_t generated according to (2a), in fact they apply more generally to x_t satisfying either the $I(0)$ or $I(1)$ conditions in (7) and (8), of which the parametric model (2a) is a special case. In addition, the results are readily extended to more general deterministic terms than the constant considered here. For example, if x_t

is linearly detrended by OLS, then Theorem 1 holds, except that W^μ, W_0^μ, and W_1^μ are replaced by their linearly detrended counterparts.

These results provide a straightforward mechanism for computing asymptotic approximations to the conditional distributions $p(t_\gamma | \phi_T, d = 0)$ and $p(t_\gamma | \phi_T, d = 1)$. In the $I(0)$ case, t_γ and ϕ_T are asymptotically independent, so this conditional distribution is simply a standard normal. In the $I(1)$ case, the limiting conditional distribution is nonstandard but can be computed as $p(t_\gamma, \phi_T - \ln N_T | d = 1)/p(\phi_T - \ln N_T | d = 1)$, where the joint distribution is computed by using the limiting representation in Theorem 1(b). Despite the presence of nuisance parameters in (2), only the long-term correlation δ between $(1 - \rho L)x_t$ and η_{2t} enters the asymptotic distributions of (t_γ, ϕ_T), and then only in the $I(1)$ case.

The dependence of the joint limiting distribution of (t_γ, ϕ_T) on δ means that in practice δ must be estimated to implement the Bayes mixture approximation in (10). However, this joint distribution is continuous in δ and, moreover, δ is a function of the spectral density matrix at frequency zero, Ω, which in turn is consistently estimable (see, e.g., Andrews [1]). For the first-order asymptotic treatment here, we therefore treat δ as known.

4. PERFORMANCE UNDER LOCAL-TO-$I(1)$ AND LOCAL-TO-$I(0)$ MODELS

One might suspect that the first-order asymptotic results of Section 3, which hinge on whether ρ is equal to or less than one, might provide poor approximations when x_t is $I(0)$ but ρ is large or, alternatively, when x_t is $I(1)$ with a large moving average root. This section provides some theoretical results for the case that x_t is local to $I(1)$ ($I(0)$ with a large autoregressive root) or, alternatively, is local to $I(0)$ ($I(1)$ with a large moving average root). This is done by first examining the properties of the ϕ_T-based posteriors and decision rules when x_t is local to either $I(1)$ or $I(0)$. Next, the performance of the Bayesian mixture approximation (10) is studied when x_t is local to $I(1)$.

4.1. First-Stage Posterior Probabilities Under Local-to-$I(0)$ and Local-to-$I(1)$ Models (General Detrending)

The results of this subsection are developed for general polynomial trends with OLS detrending; this contains the demeaning procedure considered in Section 3 as a special case. The trend component d_t is given by

$$d_t = z_t' \beta, \tag{13}$$

where $z_t = (1, t, t^2, \ldots, t^q)$, where the unknown parameters β are estimated by regressing x_t onto z_t to obtain the OLS estimator $\hat{\beta}$ of β. Thus, $q = 0$ corresponds to subtracting from x_t its sample mean and $q = 1$ corresponds to

linear detrending by OLS. For general q, under (13) the detrended data are $x_t^d = x_t - z_t'\hat{\beta} = u_t - \tilde{d}_t$, where $\tilde{d}_t = z_t'(\sum_{t=1}^T z_t z_t')^{-1} \sum_{t=1}^T z_t u_t$.

The local-to-$I(0)$ model considered here combines the $I(0)$ model with the $I(1)$ model, with a weight on the $I(1)$ component that vanishes at rate T. Specifically, let

$$x_t = d_t + u_t, \qquad u_t = u_{0t} + H_T u_{1t}, \tag{14}$$

where u_{0t} and u_{1t} are, respectively, $I(0)$ and $I(1)$ as defined in (7) and (8) and where $H_T = h/T$, where h is a constant. This representation has a natural interpretation as an unobserved components time series model in which the $I(1)$ component is small relative to the $I(0)$ component. Because of this analogy to unobserved components models, the two components here are taken to be independent, although the results below are readily extended to the case of a nonzero cross-spectrum between Δu_{1t} and u_{0t}.

The local components model can be rewritten as a moving average model in first differences where the largest MA root approaches one at the rate T. In the special case that Δu_{1t} and u_{0t} are serially uncorrelated, the stochastic element in the local-to-$I(0)$ model in (14) has the MA(1) representation $\Delta u_t = \eta_t - \theta_T \eta_{t-1}$, where $\theta_T = 1 - (h\gamma_{u_0}(0)/\gamma_{\Delta u_1}(0))/T + O(T^{-2})$.

The local-to-$I(1)$ model is given by

$$x_t = d_t + u_t, \qquad u_t = \rho_T u_{t-1} + v_t, \quad \text{where } \rho_T = 1 + c/T, \tag{15}$$

where c is a constant and v_t is $I(0)$ with spectral density at frequency zero equal to $2\pi\omega_1^2$ (say). Under this local-to-$I(1)$ specification, $U_{1T}(\bullet) = T^{-1/2} u_{1[T\bullet]}$ converges to a diffusion process, that is, $U_{1T} \Rightarrow \omega_1 B_c$, where $B_c(s)$ satisfies $dB_c(s) = cB_c(s)\,ds + dW(s)$ with $B_c(0) = 0$.

Theorem 2 summarizes the behavior of V_T under these local processes with polynomial time trends of the form in (13) detrended by OLS. Because the local-to-$I(0)$ specification is a combination of both $I(0)$ and $I(1)$ processes, we make the distinction between the limiting representations of these two processes, that is, in the $I(0)$ case $U_{0T} \Rightarrow \omega_0 W_0$, whereas in the $I(1)$ case $U_{1T} \Rightarrow \omega_1 W_1$, where W_0 and W_1 are independent standard Brownian motions.

THEOREM 2. *Let d_t be given by* (13) *and detrending by OLS.*

(a) *if x_t is local to $I(0)$ as specified by* (14), *then $V_T \Rightarrow W_{0,r}^d$, where $W_{0,r}^d(\lambda) = W_0^d(\lambda) + r\int_0^\lambda W_1^d(s)\,ds$, where $r = h\omega_1/\omega_0$, $W_0^d(\lambda) = W_0(\lambda) - \nu(\lambda)'M^{-1}\Phi$, $W_1^d(\lambda) = W_1(\lambda) - \xi(\lambda)'M^{-1}\Psi$; M, ν, and ξ are, respectively, $(q+1) \times (q+1)$, $(q+1) \times 1$, and $(q+1) \times 1$, with elements $M_{ij} = 1/(i+j-1)$, $\nu_i(\lambda) = \lambda^i/i$, and $\xi_i(\lambda) = \lambda^{i-1}$, and Φ, Ψ are $(q+1) \times 1$ with $\Phi_i = W_0(1) - (i-1)\int_0^1 s^{i-2} W_0(s)\,ds$, $i = 1, \ldots, q+1$ and $\Psi = \int_0^1 \xi(s) W_1(s)\,ds$.*

(b) *If x_t is local to $I(1)$ as specified by* (15), *then $N_T^{-1/2} V_T \Rightarrow V_c^d$, where $V_c^d(\lambda) = \int_0^\lambda B_c^d(s)\,ds/\{\int_0^1 B_c^d(r)^2\,dr\}^{1/2}$, where $B_c^d(\lambda) = B_c(\lambda) - \xi(\lambda)'M^{-1}\int_0^1 \xi(s) B_c(s)\,ds$.*

This theorem permits studying the posteriors under the local models. Consider the local-to-$I(0)$ model. For functionals ϕ discussed in Section 3A,

$\phi(V_T) \Rightarrow \phi(W_{0,r}^d) = O_p(1)$, so $f_0(\phi(V_T)) = O_p(1)$ but $f_1(\phi(V_T) - \ln N_T) \xrightarrow{p} 0$. For priors $0 < \Pi_0, \Pi_1 < 1$, $p(I(0)|\phi_T) \xrightarrow{p} 1$ and $p(I(1)|\phi_T) \xrightarrow{p} 0$, that is, under the local-to-$I(0)$ alternative, x_t will be classified as $I(0)$ with probability tending to one. Similarly, under the local-to-$I(1)$ alternative, $\phi(V_T) - \ln N_T \Rightarrow \phi(V_c) = O_p(1)$, so that $p(I(0)|\phi_T) \xrightarrow{p} 0$ and $p(I(1)|\phi_T) \xrightarrow{p} 1$ and x_t will be classified as $I(1)$ asymptotically.

It is instructive that this asymptotic misclassification of these local processes contrasts with the behavior of classical hypothesis tests based on ϕ_T. To be concrete, suppose that a one-sided test with asymptotic level α of the $I(0)$ hypothesis is performed by rejecting if $\phi_T > c_\alpha$, where c_α is the upper $100(1 - \alpha)$th percentile of the distribution of $\phi(W_0^d)$. Then the asymptotic probability of rejecting the local-to-$I(0)$ alternative is $\Pr[\phi(W_{0,r}^d) > c_\alpha]$. Because $\phi(W_{0,r}^d) = O_p(1)$ in general, this probability will exceed the level but will be less than one.[3] Classical tests will have nontrivial — indeed, possibly high — power against these local-to-$I(0)$ alternatives, but in large samples the Bayesian decision rules will classify them as $I(0)$ with probability one for any nontrivial choice of priors. The same conclusions apply to the local-to-$I(1)$ model for the same reasons: classical tests will have nontrivial power against this alternative, even though asymptotically the Bayesian decision rules will classify a local-to-$I(1)$ process as $I(1)$ with probability one.

This contrast with classical tests highlights the source of the asymptotic misclassification by this Bayesian procedure. Because the rate of convergence of $\phi(V_T)$ differs by $\ln N_T$ under the null and alternative hypotheses, one could perform classical hypothesis tests of (for example) the $I(0)$ null against the $I(1)$ alternative by using a sequence of critical values $c_{\alpha,T}$ indexed to the sample size such that $c_{\alpha,T} \to \infty$ but that $c_{\alpha,T} - \ln N_T \to -\infty$. If u_t is truly $I(0)$, the test would reject with asymptotic probability 0; but if u_t is truly $I(1)$, it would reject with unit asymptotic probability, so that this too would form a consistent classifier. Because $\phi(V_T) = O_p(1)$ under the local-to-$I(0)$ alternative, this classifier would also reject the local-to-$I(0)$ model with probability zero asymptotically, although it would reject the local-to-$I(1)$ model with asymptotic probability one. The cost of eliminating the type I error is to introduce asymptotic misclassification in a vanishingly small neighborhood of the $I(0)$ and $I(1)$ models.

4.2. The Bayesian Mixture Approximation Under Local-to-I(1) Models

The results of Theorem 2 permit analyzing the distribution of t_γ for x_t generated according to a local-to-$I(1)$ process, so that the largest root of x_t is local-to-$I(1)$. Specifically, let ρ in (2a) be nested as $\rho = 1 + c/T$, where c is a constant, and let $\tilde{\mu}_x$ in (2a) be given by the sequence $\tilde{\mu}_{x,T} = -c(1 - \tilde{\alpha}(1))\bar{\mu}_x/T$ where $\bar{\mu}_x$ is a constant. Asymptotic representations for the various statistics are given in the next theorem.

THEOREM 3. *Suppose that* (x_t, y_t) *are generated according to* (2) *with* $\rho = 1 + c/T$ *and that the null hypothesis* $H_0 : \gamma = 0$ *is true. Then*:

(a) $(t_\gamma, \phi_T - \ln N_T) \Rightarrow \left(\delta \int_0^1 B_c^\mu(s) \, dW(s) / \{ \int_0^1 B_c^\mu(s)^2 \, ds \}^{1/2} + (1 - \delta^2)^{1/2} z, \phi(V_{1,c}^\mu) \right)$,
where $V_{1,c}^\mu(\lambda) = \int_0^\lambda B_c^\mu(s) \, ds / \{ \int_0^1 B_c^\mu(s)^2 \, ds \}^{1/2}$, z *is a standard normal random variable distributed independently of* (B_c^μ, W), *and* $B_c^\mu = B_c - \int_0^1 B_c(s) \, ds$,
where $B_c(s)$ *satisfies* $dB_c(s) = cB_c(s) \, ds + dW(s)$.

(b) $p(I(1) \mid \phi_T) \overset{p}{\to} 1$ *and* $p(I(0) \mid \phi_T) \overset{p}{\to} 0$.

The result (a) implies that the asymptotic joint distribution of (t_γ, ϕ_T), and therefore the distribution of t_γ given ϕ_T, is different when $c \neq 0$ than in the unit root case $c = 0$ (given in Theorem 1(b)). The result (b) implies that the local-to-$I(1)$ process will be misclassified as $I(1)$ with probability one, so that the mixture distribution in (10) will asymptotically place all weight on the $I(1)$ conditional distribution. Taken together, these two results imply that, when ρ is local to one, the Bayesian mixture approximation will yield the incorrect asymptotic distribution. The magnitude of the resulting size distortions in local-to-$I(1)$ models is investigated numerically in the Monte-Carlo analysis in Section 6.

5. POWER OF THE PROPOSED TESTS AGAINST LOCAL ALTERNATIVES

We turn to an investigation of the theoretical power properties of the test of $\gamma = 0$ against a local sequence of $\gamma_T \neq 0$, performed by using the Bayesian mixture approximation in (10). As a simplification, the local power is analyzed for a special case of (1) in which only the first lag of x_t and y_t enter the equation for x_t. That is, (x_t, y_t) are assumed to be generated by

$$x_t = \mu_x + \rho x_{t-1} + \beta y_{t-1} + \eta_{1t} \tag{16a}$$

$$y_t = \mu_y + \gamma_T x_{t-1} + \eta_{2t}, \tag{16b}$$

where $\mu_x + \beta \mu_y = 0$. Although there are no deterministic terms in (16), the statistics are computed by including a constant in both (16a) and (16b) to maintain comparability to the results in Section 2.

Because of the different orders in probability of x_t, the local alternatives in the $I(0)$ and $I(1)$ cases differ: for some constant g,

If $|\rho| < 1$: $\gamma_T = g/T^{1/2}$ \hfill (17)

If $|\rho| = 1$: $\gamma_T = g/T$. \hfill (18)

The asymptotic representations of the relevant statistics are summarized in the next theorem.

THEOREM 4. *Suppose that* (x_t, y_t) *are generated according to* (16).

(a) *If* $|\rho| < 1$ *and* γ_T *is given by* (17), *then*

(i) $t_\gamma \Rightarrow z + \{\gamma_x(0)/\Sigma_{22}\}^{1/2}g$, *where* z *is a standard normal random variable*; *and*

(ii) $p(I(0)\,|\,\phi_T) \xrightarrow{p} 1$ *and* $p(I(1)\,|\,\phi_T) \xrightarrow{p} 0$.

(b) *If* $\rho = 1$ *and* γ_T *is given by* (18), *then*

(i) $(t_\gamma, \phi_T - \ln N_T) \Rightarrow \left(g\{\Omega_{11}\int_0^1 B_c^\mu(s)^2\,ds/\Sigma_{22}\}^{1/2} + \delta\{\int_0^1 B_c^\mu(s)\,dW(s)\}/\{\int_0^1 B_c^\mu(s)^2\,ds\}^{1/2} + (1 - \delta^2)^{1/2}z, \phi(V_{1,c}^\mu)\right)$, *where* $V_{1,c}^\mu(\lambda) = \int_0^\lambda B_c^\mu(s)\,ds/\{\int_0^1 B_c^\mu(s)\,ds\}^{1/2}$, $\Omega_{11} = \Sigma_{11} + 2\beta\Sigma_{12} + \beta^2\Sigma_{22}$, z *is a standard normal random variable distributed independently of* (B_c^μ, W) *and* $B_c^\mu = B_c(s) - \int_0^1 B_c(s)\,ds$, *where* $B_c(s)$ *satisfies* $dB_c(s) = cB_c(s)\,ds + dW(s)$ *with* $c = \beta g$; *and*

(ii) $p(I(1)\,|\,\phi_T) \xrightarrow{p} 1$ *and* $p(I(0)\,|\,\phi_T) \xrightarrow{p} 0$.

The result in the $|\rho| < 1$ case is conventional. Because the posterior probabilities asymptotically place all mass on the $I(0)$ distribution, the local asymptotic power function for a two-sided level α test with Gaussian critical values $\pm c_{\alpha/2}$ is $\Phi(-c_{\alpha/2} - g\{\gamma_x(0)/\Sigma_{22}\}^{1/2}) + \Phi(-c_{\alpha/2} + g\{\gamma_x(0)/\Sigma_{22}\}^{1/2})$, where Φ is the cumulative normal distribution, the same local power function that would arise were it known a priori that $|\rho| < 1$, so that Gaussian critical values would be used at the outset for the t_γ test. In this sense, if $|\rho| < 1$, the use of the Bayesian mixture distribution results in no asymptotic loss in power of the second-stage test regardless of the priors Π_0 and Π_1.

The results for the $\rho = 1$ case are more unusual and arise from the maintained possibility of feedback from y_t to x_t, which results in x_t being local to $I(1)$. This implies that the mixture distribution asymptotically places all weight on the $I(1)$ conditional distribution. In addition, in this case y_t is local-to-$I(0)$. The results of Section 4 imply that a posterior odds ratio based on ϕ_T would, with high probability, classify y_t as $I(0)$. Thus finite-sample evidence based on the ϕ_T classifier that y_t is $I(0)$ does not imply that $\gamma = 0$ but merely that γ is not too large.

For $|\delta| < 1$ in the $\rho = 1$ case, the local asymptotic power function of the two-sided equal-tailed test of $\gamma = 0$ is given by

$$P[\text{Reject } H_0 : \gamma = 0\,|\,\gamma_T = g/T]$$

$$= E\,\Phi\left[\frac{\{c_{\ell,1/2\alpha}(\tilde\phi) - g\Theta_c - \delta\tilde\tau_c^\mu\}}{(1 - \delta^2)^{1/2}}\right] + E\,\Phi\left[\frac{\{-c_{u,1/2\alpha}(\tilde\phi) + g\Theta_c + \delta\tilde\tau_c^\mu\}}{(1 - \delta^2)^{1/2}}\right],$$

(19)

where $\tilde\phi = \phi(V_{1,c}^\mu)$, $\Theta_c = \{\Omega_{11}\int_0^1 B_c^\mu(s)^2\,ds/\Sigma_{22}\}^{1/2}$, $\tilde\tau_c^\mu = \{\int_0^1 B_c^\mu(s)\,dW(s)\}/\{\int_0^1 B_c^\mu(s)^2\,ds\}^{1/2}$, and $c_{\ell,\alpha/2}(\phi)$ and $c_{u,\alpha/2}(\phi)$ are, respectively, the lower and upper $\frac{1}{2}\alpha$th quantiles of the conditional distribution $p(t_\gamma\,|\,\phi_T = \phi, d = 1)$. The difference between (19) and the local power function of the test based on t_γ when it is known a priori that x_t is $I(1)$ is that the critical values of the

latter test do not depend on ϕ_T, that is, the critical values in (19) are replaced by the constant critical values $\bar{c}_{\ell,\alpha/2}$ and $\bar{c}_{u,\alpha/2}$ taken from the *marginal* $I(1)$ asymptotic distribution of t_γ. For $\delta = 0$, these two tests have the same critical values and thus the same local asymptotic power, but for $\delta \neq 0$ their power functions will differ and must be compared numerically.

The local asymptotic power functions of these two tests—the proposed conditional test and the test based on t_γ when $\rho = 1$ is known a priori—are plotted in Figure 1a for $\delta = -.5$ and in Figure 1b for $\delta = -.9$, in both cases for $\beta = 0$ and $\Sigma_{11} = \Sigma_{22}$ (so $\Omega_{11}/\Sigma_{22} = 1$). Neither test dominates the other, although for most values of g the proposed test has higher power than the test based on the unconditional $d = 1$ critical values. If $\rho = 1$ is known a priori, then neither test based on t_γ is optimal relative to the system likelihood ratio test, which imposes the $\rho = 1$ restriction. The power function of the system likelihood ratio test, computed for $\beta = 0$, is also plotted in Figure 1 (the short-dashed curve).[4] For $\delta = -.5$, the power loss of the proposed procedure is moderate relative to the system likelihood ratio test; for $\delta = -.9$, it is substantial. In summary, if $\rho = 1$ and $|\delta|$ is large, substantial power is lost by not imposing this restriction a priori; but if this restriction is not imposed and tests are based on the single equation (2b), the proposed procedure often outperforms the test based on the marginal $I(1)$ null distribution of t_γ.

6. NUMERICAL ISSUES AND MONTE-CARLO RESULTS

6.1. The Specific ϕ_T Statistic and Numerical Issues

The specific ϕ_T statistic used here is one of the statistics studied in Stock [30]:

$$\phi(V_T) = \ln\left\{ \int_0^1 V_T(s)^2 \, ds \right\} = \ln\left\{ \hat{\omega}^{-1} T^{-1} \sum_{t=1}^{T} (x_t^d)^2 \right\}. \tag{20}$$

With some modifications, this statistic appears in different literatures variously as a test for random coefficients (Nabeya and Tanaka [14]), as a test of the null of a unit MA root against an MA root less than one (Saikkonen and Luukkonen [27]), as test of the $I(0)$ hypothesis against the $I(1)$ alternative (Kwiatkowski et al. [12]), and as a test for breaks in deterministic trend components in an $I(0)$ time series (Perron [18]), and it is also related to the Sargan–Bhargava [28] test of the unit root null. See Stock [30] and Perron [18] for further discussion.

All computations reported here place equal weight on the $I(1)$ and $I(0)$ hypotheses, that is, $\Pi_0 = \Pi_1 = 0.5$. The spectral density of x_t at frequency zero, ω^2, used to construct V_T, was estimated by using a Parzen kernel with lag truncation parameter $\ell_T = \min(\hat{\ell}_T, \ell_{\max}(T/100)^{.49})$, where $\hat{\ell}_T$ is Andrews's [1] data-dependent estimated lag length for the AR(1) model for the Parzen kernel. (Note that ℓ_T satisfies the rate conditions for the asymptotic representations of V_T.) For the results reported here, $\ell_{\max}$ was set to 10. Given
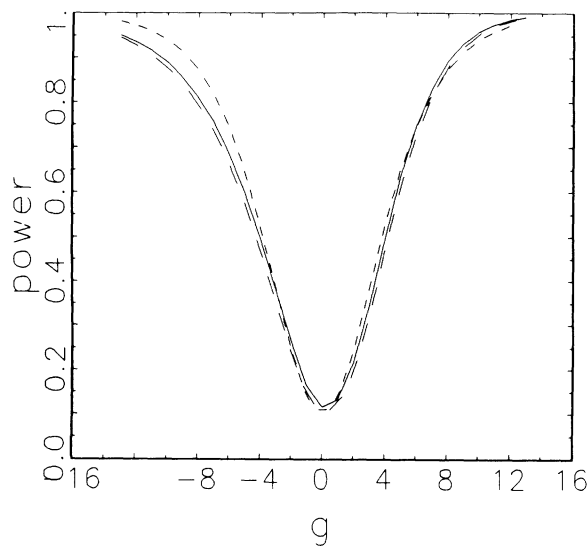
(A) $\delta = -0.5$
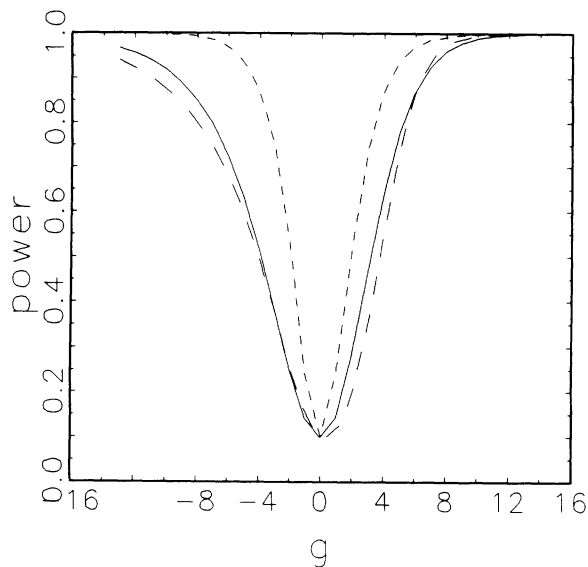
(B) $\delta = -0.9$

FIGURE 1. Asymptotic power of tests of $\gamma = 0$ against the local alternative $\gamma_T = g/T$ when x_T is $I(1)$, $\Sigma_{11} = \Sigma_{22}$, and $\beta = 0$ for (A) $\delta = -.5$ and (B) $\delta = -.9$. Solid line: Bayesian mixture approximation test; long-dashed line: test based on t_γ by using the $I(1)$ marginal null distribution of t_γ; short-dashed line: system likelihood ratio test that imposes $\rho = 1$.

priors Π_0 and Π_1, the posteriors $p(I(0)|\phi_T)$ and $p(I(1)|\phi_T)$ were computed by using kernel density estimates of the asymptotic distributions f_0 and f_1 for a given functional ϕ. For additional discussion, see Stock [30].

The conditional distributions of $p(t_\gamma|\phi_T, I(0))$ and $p(t_\gamma|\phi_T, I(1))$ were computed by using the limiting representations in Theorem 1. In the $I(0)$ case, the limiting conditional distribution is simply $N(0,1)$. In the $I(1)$ case, t_γ and $\phi_T - \ln N_T$ are asymptotically dependent random variables for $\delta \neq 0$. The conditional distribution $p(t_\gamma|\phi_T = \phi, d = 1)$ for a value of ϕ was computed by using a nearest-neighbor algorithm that was implemented by using 16,000 Monte-Carlo replications of $(\phi_T - \ln N_T, t_\gamma)$, generated under the $I(1)$ model with $T = 400$. The mixture distribution was computed by drawing randomly from the independent $N(0,1)$ $I(0)$ distribution and this nearest-neighbor estimate of the $I(1)$ conditional distribution.[5]

The quantiles of the mixture distribution $p(t_\gamma|\phi_T = \phi)$ for the case $\Pi_0 = \Pi_1 = 0.5$, $\delta = -0.9$, and $N_T = 26.65$ (corresponding to $\ell_T = 5$ and $T = 100$ for the Parzen window) are plotted in Figure 2 as a function of ϕ. For low ϕ, the posteriors place most weight on $I(0)$, and the critical values are close to the $N(0,1)$ critical values. As ϕ increases, more weight is placed on the $I(1)$ distribution, and the critical values shift up sharply.

As discussed in Section 3, the mixture distribution depends on one nuisance parameter, the long-run correlation δ. In practice, δ is unknown and would need to be estimated. As noted in Section 3, however, δ can be estimated consistently whether x_t is $I(0)$ or $I(1)$. In Monte-Carlo analysis, we therefore adopt the expedient of treating δ as known. An extension for future research is to study the effect of estimating δ on the finite sample performance of (10).

6.2. Monte-Carlo Results

The Monte-Carlo experiment studies the model examined in Section 2.2 for which the two naive procedures were found to work poorly. Specifically, the data were generated according to (3) with $\Sigma_{11} = \Sigma_{22} = 1$ and $\Sigma_{12} = \delta$. The performance of the statistic ϕ_T is examined in Table 3, which reports the mean posterior probabilities from the first-stage estimation. These results support several of the theoretical predictions in the preceding sections. For $\rho \leq 8$, the posteriors tend to zero monotonically as T increases, so that the series are being correctly classified as $I(0)$. For $\rho = 1$, the mean $I(1)$ posterior increases to .899 for $T = 400$, again correctly classifying the series. The procedure has difficulty for large ρ when the mean $I(1)$ posterior initially increases with T rather than decreases ($\rho = .90$, $.95$, and $.975$), although in each case for sufficiently large T the posteriors eventually decline.

Table 4 summarizes size results for the equal-tailed 10%-level test of $\gamma = 0$ for various values of δ, ρ, and T by using the Bayes mixture approximation conditional critical values. For $T = 100$, the largest discrepancy is the size of
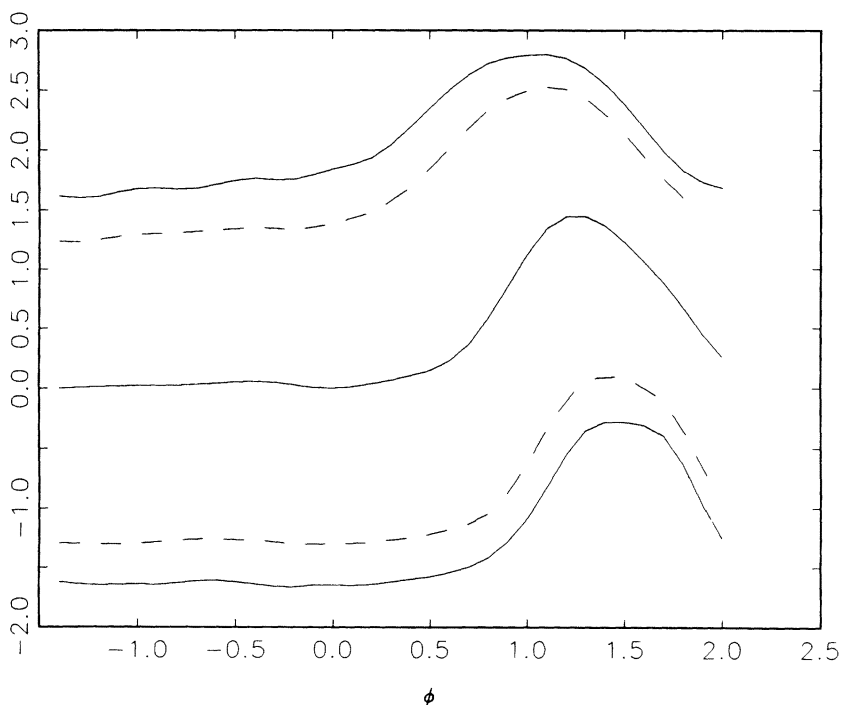
FIGURE 2. Quantiles of $p(t_\gamma | \phi_T = \phi)$ as a function of ϕ based on the Bayesian mixture approximation for $T = 100$ and $\delta = -.9$. Quantiles are 5% (solid line), 10% (dashed line), median, 90% (dashed line), and 95% (solid line). The first-stage statistic ϕ_T was computed by using $\ell_T = 5$ with a Parzen kernel.

.168 in the $\delta = -.9$ case with $\rho = 1$. Overall, the distortions are much less than for either of the conventional approaches analyzed in Tables 1 and 2.

7. CONCLUSIONS

These results suggest several conclusions. First, the Monte-Carlo evidence in Section 2 indicates that conventional approaches to the choice of critical values for second-stage inference, by either ignoring unit root problems altogether or pretesting with a unit root test, can lead to substantial overrejections when the null $\gamma = 0$ is true.

Second, the proposed Bayesian mixture approach has the desirable property that it asymptotically selects the correct $I(0)$ or $I(1)$ distribution for fixed $I(0)$ or $I(1)$ models. The use of the first-stage statistic ϕ_T has no effect on the local asymptotic power of the second-stage test if x_t is $I(0)$, and the numerical results of Section 5 indicate that the proposed procedure has bet-

TABLE 3. Mean posterior probabilities $p(I(1)|\phi_T)$ for first-stage inference[a]

				ρ			
T	0	.6	.8	.9	.95	.975	1.0
50	.089	.437	.471	.490	.489	.486	.488
100	.027	.334	.405	.514	.588	.629	.655
200	.006	.193	.262	.404	.576	.694	.815
400	.001	.102	.146	.261	.471	.675	.899

[a]The x_t pseudorandom data were computed according to (3a) with the indicated value of ρ. The ϕ_T functional used is given in (20). The process V_T was computed by using demeaned data as described in Section 3.1. The Parzen kernel was used to estimate the spectral density with lag truncation parameter $\ell_T = \min[\hat{\ell}_T, 10(T/100)^{.49}]$, where $\hat{\ell}_T$ is Andrews's [1] automatic lag truncation estimator (details are given in Section 6.1). For a given value of ϕ_T, the posterior probabilities were computed by using a kernel density algorithm (based on 16,000 pseudorandom realizations of the limiting Brownian motion functional) described by Stock [30]. Prior probabilities are $p(I(0)) = p(I(1)) = 0.5$. Based on 5,000 Monte-Carlo replications.

ter power against most alternatives than the test of $\gamma = 0$ based on critical values from the marginal $I(1)$ distribution of t_γ.

Third, the Monte-Carlo evidence is encouraging and suggests good size properties for two-sided tests based on these procedures for a wide range of values of ρ, including ρ close to, but less than, one. This is somewhat surprising because our theoretical results show that the size of the second-stage test will be incorrect, even asymptotically, when x_t is local to $I(1)$. For the ϕ_T statistic and models studied here, these initial results suggest that this might not pose an important problem in practice.

The results presented here suggest several directions for future research. Throughout, we have treated the important nuisance parameter δ as known. Although δ can be estimated consistently, in finite samples its estimation presumably will affect the performance of the proposed procedure and this remains to be investigated. The Monte-Carlo analysis has focused on a single ϕ functional and uses only flat priors; the use of a different functional or informative priors might improve the finite sample performance. In addition, the performance of the second-stage statistic should be investigated for a wider range of x_t processes than the AR(1) specifications considered here. More difficult is the extension to include additional potentially integrated and cointegrated variables in the second-stage regression. This increases the Bayesian selection procedure nontrivially and introduces additional nuisance parameters to the problem. The second of these difficulties has been examined by Toda and Phillips [33]. Finally, the extension to include additional lags of x_t or y_t as regressors in the second-stage regression, although conceptually straightforward, is of considerable practical importance because such

TABLE 4. Size of t-tests of $\gamma = 0$ with critical values computed by using the Bayesian mixture approximation (10): Nominal level 10% [a]

					ρ			
δ	T	0	.6	.8	.9	.95	.975	1.0
−0.9	50	.077	.057	.054	.057	.060	.070	.105
	100	.093	.071	.072	.069	.068	.083	.168
	200	.089	.075	.079	.093	.098	.094	.141
	400	.102	.089	.091	.105	.135	.167	.160
−0.5	50	.101	.082	.085	.096	.101	.095	.119
	100	.108	.087	.095	.088	.093	.095	.121
	200	.093	.083	.094	.099	.096	.094	.110
	400	.101	.098	.098	.110	.119	.122	.116
0.0	50	.106	.114	.110	.109	.122	.111	.115
	100	.100	.105	.106	.106	.103	.108	.102
	200	.101	.099	.107	.106	.097	.103	.103
	400	.098	.105	.097	.103	.101	.105	.110

[a]The data were generated according to (3). The first-stage ϕ_T statistic was computed as described in the notes to Table 3. The critical values for t_γ (the t-statistic on x_{t-1} in the regression of y_t onto $(1, x_{t-1})$) were computed as described in Section 6.1. Based on 5,000 Monte-Carlo replications.

lags are typically included in empirical practice. These and related problems are areas of ongoing research.

NOTES

1. The results in this paper complement those in Toda and Phillips [33], who considered the problem of sequential inference when some of the variables are cointegrated. Toda and Phillips's [33] theory maintains that the series are cointegrated, and they studied sequences of Wald tests of the null of no Granger causality in a vector autoregression, and inference is always χ^2. Their problem differs from the one considered here: our second-stage regression always examines a levels effect of the regressor, and the issue is whether standard or nonstandard distributions should be used to evaluate the significance of the estimated coefficient.

2. Similar results are obtained for a 10% pretest and a 10% equal-tailed second-stage test. For example, for $T = 100$ and $\delta = -.9$, if $\rho = .9$, the rejection rate is .34, whereas if $\rho = .95$, it is .27.

3. The nonstandard distribution of V_T under the null and the alternative make it difficult to make general statements about the power function of such a test against the local alternative without resorting to numerical calculations. An illustrative case, however, is for the statistic $\phi(V_T) = \ln(V_T(1)^2)$ in the case of no detrending. Under the $I(0)$ null, this statistic has representation $\ln(W(1)^2)$, which has a critical value of $\ln(\chi^2_{1;\alpha})$, where $\chi^2_{1;\alpha}$ is the α-level χ^2_1 critical value. Under the local alternative, $\phi(V_T) \Rightarrow \ln\{(W_0(1) + r\int_0^1 W_1(s)\,ds)^2\}$, which is distributed as the logarithm of $(1 + r^2/3)$ times a standard χ^2_1. The power of this test of $I(0)$ against the local-to-$I(0)$ alternative is therefore $\Pr[\chi^2_1 > \chi^2_{1;\alpha}/(1 + r^2/3)]$. This power function has its minimum at $r = 0$ and is monotone, increasing to one as r increases.

4. This power function is derived as follows. The Gaussian MLE for γ in (16) with $\rho = 1$, $\beta = 0$, and an intercept in (16b) can be obtained by using the triangularized system, $\Delta x_t = \eta_{1t}$ and $y_t = \mu_y + \gamma x_{t-1} + \kappa \Delta x_t + \tilde{\eta}_{2t}$, where $\kappa = \Sigma_{21}\Sigma_{11}^{-1}$, so $\kappa \Delta x_t = \text{Proj}(\eta_{2t}|\eta_{1t})$ and $\tilde{\eta}_{2t} = \eta_{2t} - \text{Proj}(\eta_{2t}|\eta_{1t})$; see Phillips [23], Saikkonen [26], and Stock and Watson [31]). Then t_γ is asymptotically $N(0,1)$ and the local asymptotic power against $\gamma_T = g/T$ is $E\{\Phi(-c_{\alpha/2} - g(1-\delta^2)^{-1/2}\Theta_0) + \Phi(-c_{\alpha/2} + g(1-\delta^2)^{-1/2}\Theta_0)\}$, where $\Theta_0 = \Theta_c$ evaluated at $c = 0$ as defined following (19), and $c_{\alpha/2}$ is the $\frac{1}{2}\alpha$ standard normal critical value. The power functions in Figure 1 were computed by Monte-Carlo integration of the relevant expression (e.g., (19)) over $(\bar{\phi}, \theta_c, \tilde{\tau}_c^\mu)$ for $c = 0$ with 5,000 Monte-Carlo replications.

5. Specifically, 16,000 pseudorandom realizations from the limiting joint distribution were computed by using pseudodata that was generated from (2) with $\rho = 1$, $\beta(L) = 0$, $\tilde{\alpha}(L) = 0$, $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = \delta$, and $\ell_T = 1$. With the exception of δ, from Theorem 1(b) the choice of parameter values does not affect the asymptotic distribution of $(\phi_T - \ln N_T, t_\gamma)$. These can be interpreted as discretized approximations (discretized to 400 equispaced points) to the limiting Brownian motion functionals in Theorem 1(b). For a given value ϕ, the mixture conditional distribution (10) was estimated by drawing K realizations, $p(I(0)|\phi_T = \phi)K$ from a standard normal distribution and $p(I(1)|\phi_T = \phi)K$ from the conditional distribution $p(t_\gamma|\phi_T = \phi, d = 1)$; these latter $p(I(1)|\phi_T)K$ draws were computed as those t_γ for which the associated ϕ_T were the $p(I(1)|\phi_t)K$ nearest neighbors of ϕ. Critical values of t_γ given $\phi_T = \phi$ were computed by using the resulting Monte-Carlo mixture c.d.f. for a grid of ϕ (grid size .05), which was then smoothed by Gaussian kernel regression to reduce Monte-Carlo error. Critical values for arbitrary ϕ were computed by linear interpolation of the resulting table. These tables were constructed for various values of δ. The results reported here use $K = 750$.

REFERENCES

1. Andrews, D.W.K. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59 (1991): 817–858.
2. Bobkoski, M.J. *Hypothesis Testing in Nonstationary Time Series.* Ph.D. Dissertation, University of Wisconsin, 1983.
3. Campbell, J.Y. A variance decomposition for stock returns. *Economic Journal* 101 (1991): 157–179.
4. Cavanagh, C. *Roots Local to Unity.* Harvard University, 1985.
5. Chan, N.H. On the parameter inference for nearly nonstationary time series. *Journal of the American Statistical Association* 83 (1988): 857–862.
6. Chan, N.H. & C.Z. Wei. Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15 (1987): 1050–1063.
7. Chan, N.H. & C.Z. Wei. Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16 (1988): 367–401.
8. Dickey, D.A. & W.A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74 (1979): 427–431.
9. Fama, E.F. Efficient capital markets II. *Journal of Finance* 46 (1991): 1575–1617.
10. Fama, E.F. & K.R. French. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25 (1989): 23–49.
11. Flavin, M.A. The adjustment of consumption to changing expectations about future income. *Journal of Political Economy* 89 (1981): 974–1009.
12. Kwiatkowski, D., P.C.B. Phillips, P. Schmidt & Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54 (1992): 159–178.
13. Mankiw, N.G. & M.D. Shapiro. Trends, random walks and tests of the permanent income hypothesis. *Journal of Monetary Economics* 16 (1985): 165–174.

14. Nabeya, S. & K. Tanaka. Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Annals of Statistics* 16 (1988): 218-235.
15. Nabeya, S. & K. Tanaka. A general approach to the limiting distribution for estimators in time series regression with nonstable autoregressive errors. *Econometrica* 58 (1990): 145-163.
16. Park, J.Y. & P.C.B. Phillips. Statistical inference in regressions with integrated processes: Part I. *Econometric Theory* 4 (1988): 468-497.
17. Perron, P. The calculation of the limiting distribution of the least-squares estimator in a near-integrated model. *Econometric Theory* 5 (1989): 241-255.
18. Perron, P. *A Test for Changes in a Polynomial Trend Function for a Dynamic Time Series.* Princeton University, 1991.
19. Phillips, P.C.B. Understanding spurious regressions in econometrics. *Journal of Econometrics* 33 (1986): 311-340.
20. Phillips, P.C.B. Toward a unified asymptotic theory for autoregression. *Biometrika* 74 (1987): 535-547.
21. Phillips, P.C.B. To criticize the critics: An objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics* 6 (1991): 333-364.
22. Phillips, P.C.B. Spectral regression for cointegrated time series. In W. Barnett, J. Powell & G. Tauchen (eds), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 413-436. Cambridge University Press, Cambridge, UK. 1991.
23. Phillips, P.C.B. Optimal inference in cointegrated systems. *Econometrica* 59 (1991): 283-306.
24. Phillips, P.C.B. & S.N. Durlauf. Multiple time series regression with integrated processes. *Review of Economic Studies* 53 (1986): 473-496.
25. Phillips, P.C.B. & W. Ploberger. *Time Series Modeling with a Bayesian Frame of Reference: I. Concepts and Illustrations.* Yale University, 1991.
26. Saikkonen, P. Asymptotically efficient estimation of cointegrating regressions. *Econometric Theory* 7 (1991): 1-21.
27. Saikkonen, P. & R. Luukkonen. Testing for moving average unit root in autoregressive integrated moving average models. *Journal of the American Statistical Association* 88 (1993): 596-601.
28. Sargan, J.D. & A. Bhargava. Testing for residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica* 51 (1983): 153-174.
29. Sims, C.A., J.H. Stock & M.W. Watson. Inference in linear time series models with some unit roots. *Econometrica* 58 (1990): 113-144.
30. Stock, J.H. Deciding between $I(1)$ and $I(0)$. *Journal of Econometrics*, forthcoming.
31. Stock, J.H. & M.W. Watson. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61 (1993): 783-820.
32. Stock, J.H. & K.D. West. Integrated regressors and tests of the permanent income hypothesis. *Journal of Monetary Economics* 21 (1988): 85-96.
33. Toda, H.Y. & P.C.B. Phillips. Vector autoregression and causality. *Econometrica* 61 (1993): 1367-1394.

APPENDIX

Proof of Theorem 1. (a) First write x_t in the form of (6). By assumption, $(1 - \alpha(L)L)$ is invertible, so under H_0, $x_t = (1 - \alpha(1))^{-1}\tilde{\mu}_x + (1 - \alpha(L)L)^{-1}(\eta_{1t} + \beta(L)\eta_{2t-1}) = \bar{\mu}_x + v_t$, say. Because the roots of $\beta(L)$ are outside the unit circle and η_t is a martingale difference sequence, standard arguments imply that $V_T^\mu \Rightarrow W_0^\mu$. In

addition, t_γ has an $N(0,1)$ marginal distribution. To show the asymptotic independence of t_γ and ϕ_T, note that (1) is a special case of the system analyzed by Chan and Wei [7], so that their Theorem 2.2 implies $\left(T^{-1/2}\sum_{s=1}^{[T\cdot]}\eta_s, T^{-1/2}\sum_{s=1}^{[T\cdot]} x_{s-1}\eta_{2s}\right) \Rightarrow (\Sigma^{1/2}B(\cdot), \zeta)$, where B is a standard bivariate Brownian motion, ζ is distributed $N(0, \Sigma_{22}\gamma_x(0))$ and where $W(\cdot)$ and ζ are independent. With this result, the asymptotic independence of t_γ and ϕ_T follows by the consistency of $\hat{\omega}^2, \hat{\Sigma}_{22}$ (where $\hat{\Sigma}_{22}$ is the OLS estimator of $\mathrm{var}(\eta_{2t})$), $T^{-1}\sum_{t=1}^T x_{t-1}^2$, and from writing $V_T^\mu(\cdot)$ in terms of the partial sum process $T^{-1/2}\sum_{s=1}^{[T\cdot]}\eta_s$. Part (ii) follows from Stock [30].

(b) Write x_t as $x_t = x_0 + v_t$, where x_0 is an initial condition and $\Delta v_t = (1 - \tilde\alpha(L)L)^{-1}(\eta_{1t} + \beta(L)\eta_{2t-1})$. Now v_t is $I(1)$ as defined in (8), so $\phi(V_T) \Rightarrow \phi(V_0^\mu)$ [30, Theorems 1 and 2]; this also follows, with minor modifications, from Kwiatkowski et al. [12], Perron [18], and Phillips [22].

The joint representation of (ϕ_T, t_γ) is obtained by using now-conventional frequency zero projection arguments. Note that $T^{-1/2}v_{[T\cdot]} = T^{-1/2}\sum_{s=1}^{[T\cdot]}\zeta_s + o_p(1)$, where $\zeta_t = (1 - \tilde\alpha(1))^{-1}(\eta_{1t} + \beta(1)\eta_{2t})$, and that Ω is the variance–covariance matrix of (ζ_t, η_{2t}). Define W to be the limiting standard Brownian motion such that $T^{-1/2}\sum_{s=1}^{[T\cdot]}\zeta_s \Rightarrow \Omega_{11}^{1/2}W(\cdot)$. In addition, decompose η_{2t} as $\eta_{2t} = \mathrm{Proj}(\eta_{2t}\,|\,\zeta_t) + \tilde\eta_{2t} = \Omega_{11}^{-1}\Omega_{21}\zeta_t + \tilde\eta_{2t}$, so that $(\zeta_t, \tilde\eta_{2t})$ are a martingale difference sequence, and $E(\zeta_t\tilde\eta_{2t}) = 0$ and $\mathrm{var}(\tilde\eta_{2t}) \equiv \tilde\Sigma_{22} = \Omega_{22} - \Omega_{11}^{-1}\Omega_{12}^2$. With this notation, t_γ can be written:

$$t_\gamma = \Omega_{11}^{-1}\Omega_{21}\frac{\left(T^{-1}\sum_{t=1}^T x_{t-1}^\mu \zeta_t\right)}{\left(T^{-2}\sum_{t=1}^T (x_{t-1}^\mu)^2\hat\Sigma_{22}\right)^{1/2}} + \frac{\left(T^{-1}\sum_{t=1}^T x_{t-1}^\mu \tilde\eta_{2t}\right)}{\left(T^{-2}\sum_{t=1}^T (x_{t-1}^\mu)^2\hat\Sigma_{22}\right)^{1/2}}.$$

Now, $\left(T^{-1/2}v_{[T\cdot]}, T^{-1/2}\sum_{s=1}^{[T\cdot]}\tilde\eta_{2s}\right) \Rightarrow (\Omega_{11}^{1/2}W(\cdot), \tilde\Sigma_{22}^{1/2}G(\cdot))$, where $W(\cdot)$ and $G(\cdot)$ are independent standard Brownian motions. In addition, $\hat\Sigma_{22} \xrightarrow{P} \Sigma_{22} = \Omega_{22}$, so

$$t_\gamma \Rightarrow \frac{\delta\int W^\mu(s)\,dW(s)}{\left(\int W^\mu(s)^2 ds\right)^{1/2}} + \frac{(1-\delta^2)^{1/2}\int W^\mu(s)\,dG(s)}{\left(\int W^\mu(s)^2 ds\right)^{1/2}}. \tag{A.1}$$

The argument in Phillips [23] or Saikkonen [26] imply that the second term in (A.1) is independent of the first term and has an $N(0,1)$ marginal distribution, which leads to the representation given in the statement of the theorem. Part (ii) follows from Stock [30].   ∎

Proof of Theorem 2. Throughout, let $\Upsilon_T = \mathrm{diag}(1, T, \ldots, T^q)$ and let $M_T = T^{-1}\Upsilon_T^{-1}\sum_{t=1}^T z_t z_t'\Upsilon_T^{-1}$. The nonstochastic $q \times q$ 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 x_t is $I(0)$ or $I(1)$.

(a) Let $U_T(\lambda) = T^{-1/2}\sum_{t=1}^{[T\lambda]} u_t$, and let $D_T(\lambda) = T^{-1/2}\sum_{s=1}^{[T\lambda]}\tilde d_s$. The desired result will follow if it can be shown that (i) (U_T, D_T) satisfies detrending condition A(i), and (ii) $\hat\omega^2 \xrightarrow{P} \omega_0$. These are shown in turn.

(i) By direct calculation, $U_T = U_{0T} + TH_T U_{1T}^\dagger$, where $U_{1T}^\dagger(\lambda) = T^{-1}\sum_{t=1}^{[T\lambda]} U_{1T}(t/T)$. Because $TH_T = h$, $U_T \Rightarrow \omega_0 W_r$, where $W_r(\lambda) = W_0(\lambda) + r\int_0^\lambda W_1(s)\,ds$, where $r = h\omega_1/\omega_0$. Now, $D_T(\lambda) = \nu_T(\lambda)'M_T^{-1}\Upsilon_T$, where $\nu_T(\lambda) = T^{-1}\sum_{s=1}^{[T\lambda]}\Upsilon_T^{-1}z_s$ and

$\Gamma_T = T^{-1/2} \sum_{t=1}^{T} \Upsilon_T^{-1} z_t u_t = \Phi_T + h\Psi_T$, where $\Phi_T = T^{-1/2} \sum_{t=1}^{T} \Upsilon_T^{-1} z_t u_{0t}$ and $\Psi_T = T^{-3/2} \sum_{t=1}^{T} \Upsilon_T^{-1} z_t u_{1t} = \int_0^1 \xi_T(s) U_{1T}(s)\, ds$, where $\xi_T(\lambda) = \Upsilon_T^{-1} z_{[T\lambda]}$. From Stock's [30] Theorem 2 $(\Phi_T, \Psi_T, \nu_T, M_T) \Rightarrow (\omega_0\Phi, \omega_1\Psi, \nu, M)$, where Φ is $(q+1) \times 1$ with $\Phi_i = W_0(1) - (i-1)\int_0^1 s^{i-2} W_0(s)\, ds$, $i = 1, \dots, q+1$, $\Psi = \int_0^1 \xi(s) W_1(s)\, ds$, where $\xi_i(\lambda) = \lambda^{i-1}$, and ν is $(q+1) \times 1$ with $\nu_i(\lambda) = \lambda^i/i$, $i = 1, \dots, q+1$. Thus, $\Gamma_T \Rightarrow \omega_0(\Phi + r\Psi)$ and $D_T(\bullet) \Rightarrow \omega_0 D(\bullet)$, where $D(\lambda) = \nu(\lambda)' M^{-1}(\Phi + r\Psi)$. This verifies Detrending Condition A(i). Let $X_{0T}^d(\lambda) = T^{-1/2} \sum_{s=1}^{[T\lambda]} x_s^d = U_T(\lambda) - D_T(\lambda)$, so that $V_T(\lambda) = \hat{\omega}^{-1} X_{0T}^d(\lambda)$. Define $W_0^d(\lambda) = W_0(\lambda) - \nu(\lambda)' M^{-1}\Phi$ and $W_1^d(s) = W_1(s) - \xi(\lambda)' M^{-1}\Psi$. It follows that $X_{0T}^d(\bullet) \Rightarrow \omega_0 W_{0,r}^d(\bullet)$, where $W_{0,r}^d(\lambda) = W_0(\lambda) + r\int_0^1 W_1(s)\, ds - \nu(\lambda)' M^{-1}(\Phi + r\Psi) = W_0^d(\lambda) - r\int_0^\lambda W_1^d(s)\, ds$, where the second equality uses $\nu(\lambda) = \int_0^\lambda \xi(s)\, ds$.

(ii) Let $\tilde{\omega}^2 = \sum_{m=-\ell_T}^{\ell_T} k(m/\ell_T) T^{-1} \sum_{t=|m|+1}^{T} u_t u_{t-m}$ and $\tilde{\omega}_0^2 = \sum_{m=-\ell_T}^{\ell_T} k(m/\ell_T) T^{-1} \sum_{t=|m|+1}^{T} u_{0t} u_{0t-m}$. Now,

$$|\hat{\omega}^2 - \tilde{\omega}_0^2| \leq |\tilde{\omega}^2 - \tilde{\omega}_0^2| + |\hat{\omega}^2 - \tilde{\omega}^2|$$

$$= \left| \sum_{m=-\ell_T}^{\ell_T} k(m/\ell_T) T^{-1} \sum_{t=|m|+1}^{T} \{ H_T^2 u_{1t} u_{1t-|m|} + H_T(u_{1t} u_{0, t-|m|} \right.$$

$$\left. + u_{0t} u_{1, t-|m|}) \} \right|$$

$$+ \left| \sum_{m=-\ell_T}^{\ell_T} k(m/\ell_T) T^{-1} \sum_{t=|m|+1}^{T} \{ \tilde{d}_t \tilde{d}_{t-|m|} - (\tilde{d}_t u_{t-|m|} + u_t \tilde{d}_{t-|m|}) \} \right|$$

$$\leq (2\ell_T + 1)\{ T H_T^2 \|T^{-1/2} u_{1t}\| + 2T^{1/2} H_T \|T^{-1/2} u_{1t}\|^{1/2} \|u_{0t}\|^{1/2} \}$$

$$+ (2\ell_T + 1)\{ 2\|u_t\|^{1/2} \|\tilde{d}_t\|^{1/2} + \|\tilde{d}_t\| \}. \tag{A.2}$$

Because $\|T^{-1/2} u_{1t}\| \Rightarrow \omega_1^2 \int_0^1 W_1(s)^2\, ds$, and $\|u_{0t}\| \overset{p}{\to} \gamma_{u_0}(0)$, the first term in (A.2) $\overset{p}{\to} 0$ if $(2\ell_T + 1) T H_T^2 \to 0$ and $(2\ell_T + 1) T^{1/2} H_T \to 0$. These in turn follow from the rate condition $\ell_T^2 \ln T/T \to 0$. Because $|\|u_t\| - \|u_{0t}\|| \leq H_T^2 \|u_{1t}\| + 2H_T \|u_{0t}\|^{1/2} \|u_{1t}\|^{1/2} \overset{p}{\to} 0$, $\|u_t\| \overset{p}{\to} \gamma_{u_0}(0)$, the second term in (A.2) $\overset{p}{\to} 0$ if $\ell_T^2 \|\tilde{d}_t\| \overset{p}{\to} 0$, which is Condition A(ii).

To show A(ii), write $T\|\tilde{d}_T\| = \Gamma_T' M_T^{-1} \Gamma_T$, from which it follows that $T\|\tilde{d}_T\| \Rightarrow \omega_0^2 (\Phi + r\Psi)' M^{-1}(\Phi + r\Psi)$. This result and the rate condition $\ell_T^2 \ln T/T \to 0$ together imply that $\ell_T^2 \|\tilde{d}_t\| \overset{p}{\to} 0$, thereby verifying A(ii) so that both terms in (A.2) $\overset{p}{\to} 0$. Because $\tilde{\omega}_0^2 \overset{p}{\to} \omega_0$, it follows that $\hat{\omega}^2 \overset{p}{\to} \omega_0$. Thus $V_T(\lambda) = \hat{\omega}^{-1} X_{0T}^d(\lambda) \Rightarrow W_{0,r}^d(\bullet)$, the desired result.

(b) It is first demonstrated that Conditions B(i) and B(ii) hold for u_t given by (15). Concerning Condition B(i), write $D_{1T}(\lambda) = \xi_T(\lambda)' M_T^{-1} \Psi_T$, where the terms are defined in the proof of part (a). Now, $(\xi_T, M_T, \Psi_T) \Rightarrow (\xi, M, \omega_1\Psi)$, where $\Psi = \int_0^1 \xi(s) B_c(s)\, ds$. Thus, $D_{1T}(\bullet) \Rightarrow \omega_1 D_1(\bullet)$, where $D_1(\lambda) = \xi(\lambda)' M^{-1}\Psi$, which verifies B(i). Let $X_{1T}^d(\lambda) = T^{-1/2} x_{[T\lambda]} = U_{1T}(\lambda) - D_{1T}(\lambda)$, so that $V_T(\lambda) = \hat{\omega}^{-1} \int_0^\lambda X_{1T}^d(s)\, ds$. It follows that $X_{1T}^d(\bullet) \Rightarrow \omega_1 B_c^d(\bullet)$, where $B_c^d(\lambda) = B_c(\lambda) - \xi(\lambda)' M^{-1} \int_0^1 \xi(s) B_c(s)\, ds$.

Condition B(ii) follows from writing $\|T^{1/2}\Delta \tilde{d}_t\| = \int_0^1 S_T(\lambda)\, d\lambda$, where $S_T(\lambda) = \Psi_T' M_T^{-1} \xi_T(\lambda) \xi_T(\lambda)' M_T^{-1} \Psi_T$, where $\xi_T(\lambda) = T \Upsilon_T^{-1} \Delta z_{[T\lambda]}$. Direct calculation shows

that $S_T \Rightarrow S$, where $S(\lambda) = \omega_1^2 \Psi' M^{-1} \xi^-(\lambda) \xi^-(\lambda)' M^{-1} \Psi$, where $\xi^-(\lambda) = (0, 1, 2\lambda, \ldots, q\lambda^{q-1})'$, so $\|\Delta \tilde{d}_t\| \xrightarrow{P} 0$, thereby verifying B(ii).

Next write $N_T^{-1/2} V_T(\lambda) = B_T^{-1/2} A_T(\lambda)$, where $A_T(\lambda) = T^{-3/2} \sum_{s=1}^{[T\lambda]} x_s^d$ and $B_T = T^{-1} \sum_{m=-\ell_T}^{\ell_T} k_T(m/\ell_T) \hat{\gamma}_{xd}(|m|) / \sum_{m=-\ell_T}^{\ell_T} k(m/\ell_T)$. Now $A_T(\lambda) \Rightarrow \omega_1 \int_0^\lambda B_c^d(s)\, ds$. To show that B_T has the desired limit, it remains only to show that $B_T - T^{-2} \sum_{t=1}^{T} (x_t^d)^2 \xrightarrow{P} 0$. This was shown in Stock's [30] Lemma A.1 when u_t is $I(1)$. The only part of the proof of that lemma that requires modification for the local-to-$I(1)$ case is demonstrating that the term D_{3T} (as defined in that proof) $\xrightarrow{P} 0$. From (A.1) in Stock [30], $D_{3T} \xrightarrow{P} 0$ if $\|\Delta \tilde{d}_t\| \xrightarrow{P} 0$ (just shown) and if $\|\Delta u_t\| = O_p(1)$. This final relation is now shown. Because $(1 - \rho_T L) u_t = v_t$,

$$\|\Delta u_t\| - \|v_t\| = \|v_t + (\rho_T - 1) u_{t-1}\| - \|v_t\|$$

$$\leq 2|\rho_T - 1|\, \|u_t\|^{1/2} \|v_t\|^{1/2} + (\rho_T - 1)^2 \|u_t\|.$$

Now $(\rho_T - 1)^2 \|u_t\| = c^2 T^{-1} \int_0^1 U_{1T}(s)^2 ds \xrightarrow{P} 0$ and $\|v_t\| \xrightarrow{P} \gamma_v(0)$ by assumption. Thus, $\|\Delta u_t\| \xrightarrow{P} \gamma_v(0)$, so $D_{3T} \xrightarrow{P} 0$ and Lemma A.1 in Stock [30] holds, thereby completing the proof.   ∎

Proof of Theorem 3. (a) Under the stated assumptions, algebraic manipulations permit rewriting x_t as $x_t = \bar{\mu} + v_t$ plus initial conditions, where $\bar{\mu}$ is constant and v_t satisfies $(1 - \rho L)(1 - \tilde{\alpha}(L)L) v_t = \beta(L)(y_{t-1} - \mu_y) + \eta_{1t}$, where $v_0 = 0$. Thus, $T^{1/2} v_{[T \cdot]} = T^{-1/2} \sum_{s=1}^{[T \cdot]} \rho^{[T \cdot] - s} (1 - \tilde{\alpha}(L)L)^{-1} (\eta_{1t} + \beta(L)\eta_{2t-1}) = T^{-1/2} \sum_{s=1}^{[T \cdot]} \rho^{[T \cdot] - s} \zeta_s + o_p(1)$, where $\zeta_t = (1 - \tilde{\alpha}(1))^{-1} (\eta_{1t} + \beta(1)\eta_{2t})$. Note that (ζ_t, η_{2t}) is a martingale difference sequence with variance–covariance matrix Ω. As in the proof of Theorem 1(b), let $\tilde{\eta}_{2t} = \eta_{2t} - (\Omega_{21}/\Omega_{11}) \zeta_t$. Then $\left(T^{-1/2} \sum_{s=1}^{[T \cdot]} \rho^{[T \cdot] - s} \zeta_s, T^{-1/2} \sum_{s=1}^{[T \cdot]} \tilde{\eta}_{2s} \right) \Rightarrow (\Omega_{11}^{1/2} B_c(\cdot), \tilde{\Sigma}_{22}^{1/2} G(\cdot))$, where $dB_c(s) = cB_c(s)\, ds + dW(s)$ and G is a standard Brownian motion distributed independently of W. The argument leading to (A.1), applied here, yields,

$$t_\gamma \Rightarrow \frac{\delta \int B_c^\mu(s)\, dW(s)}{\left(\int B_c^\mu(s)^2 ds \right)^{1/2}} + \frac{(1 - \delta^2)^{1/2} \int B_c^\mu(s)\, dG(s)}{\left(\int B_c^\mu(s)^2 ds \right)^{1/2}}. \tag{A.3}$$

As in (A.1), the independence of G and (W, B_c, B_c^μ), and the $N(0,1)$ distribution of the second term in (A.3) conditional on B_c^μ yield the representation given in the statement of the theorem. (b) It was shown in the proof of (a) that x_t has the local-to-$I(1)$ representation (15), so the results of Theorem 2(b) apply. The consistency of the posterior probabilities follows according to the remarks following the statement of Theorem 2.   ∎

Proof of Theorem 4. Substitute (16b) into (16a) to obtain $(1 - \rho L - \beta \gamma_T L^2) x_t = v_t$, where $v_t = \eta_{1t} + \beta \eta_{2t-1}$.

(a)(i) For T sufficiently large, if $|\rho| < 1$, then for any finite g the roots of $(1 - \rho L - \beta \gamma_T L^2)$ will both be outside the unit circle, so that x_t is $I(0)$ in the sense of (7) and in particular $T^{-1} \sum_{t=1}^{T} (x_{t-1})^2 \xrightarrow{P} \gamma_x(0)$. Thus, under the local nesting in (17), $t_\gamma = \left\{ \sum_{t=1}^{T} x_{t-1}^\mu (\gamma_T x_{t-1}^\mu + \eta_{2t}) \right\} / \left\{ \sum_{t=1}^{T} (x_{t-1}^\mu)^2 \hat{\Sigma}_{22} \right\}^{1/2} = g \{ T^{-1} \sum_{t=1}^{T} (x_{t-1}^\mu)^2 / \hat{\Sigma}_{22} \}^{1/2} + \left\{ \sum_{t=1}^{T} x_{t-1}^\mu \eta_{2t} \right\} / \left\{ \sum_{t=1}^{T} (x_{t-1}^\mu)^2 \hat{\Sigma}_{22} \right\}^{1/2} \Rightarrow g \{ \gamma_x(0) / \Sigma_{22} \}^{1/2} + z$, the stated result. (ii) Because x_t is $I(0)$ in the sense of (7) for T sufficiently large, the results of Stock

[30] apply and, for the reasons discussed in Section 3.1, the posteriors are consistent as stated.

(b)(i) When $\rho = 1$, $(1 - \rho L - \beta \gamma_T L^2)$ can be factored, $(1 - r_1 L)(1 - r_2 L)$, where $r_1 = 1 + v_T$ and $r_2 = -v_T$, where $v_T = \frac{1}{2}\{(1 + 4\beta\gamma_T)^{1/2} - 1\}$. Let $\xi_{T,t} = (1 - v_T L)^{-1}(\eta_{1t} + \beta\eta_{2t-1})$. Then, with $x_0 = 0$, $T^{-1/2} x_{[T\cdot]} = T^{-1/2} \sum_{s=1}^{[T\cdot]} (1 + v_T)^{[T\cdot]-s} \xi_{T,s}$. Now $v_T = \beta g / T + O(T^{-2})$, so $T^{-1/2} \sum_{s=1}^{[T\cdot]} (1 + v_T)^{[T\cdot]-s} \xi_{T,t} - T^{-1/2} \sum_{s=1}^{[T\cdot]} \exp\{\beta g([T\cdot] - s)/T\}\zeta_t \overset{p}{\to} 0$, where $\zeta_t = \eta_{1t} + \beta\eta_{2t}$. Thus $T^{-1/2} x_{[T\cdot]} \Rightarrow \Sigma_{11}^{1/2} B_c(\cdot)$, where B_c is defined in the statement of the theorem, and where W is defined by $T^{-1/2} \sum_{s=1}^{[T\cdot]} \zeta_s \Rightarrow \Sigma_{11}^{1/2} W(\cdot)$.

To obtain the representation for t_γ, note that Ω is the covariance matrix of (ζ_t, η_{2t}) and write $t_\gamma = g\{T^{-2} \sum_{t=1}^{T} (x_{t-1}^\mu)^2 / \hat{\Sigma}_{22}\}^{1/2} + \{T^{-1} \sum_{t=1}^{T} x_{t-1}^\mu \eta_{2t}\} / \{T^{-2} \sum_{t=1}^{T} (x_{t-1}^\mu)^2 \hat{\Sigma}_{22}\}^{1/2}$. The first term has the limit $g\{\Omega_{11} \int_0^1 B_c^\mu(s)^2 ds / \Sigma_{22}\}^{1/2}$, where $\Omega_{11} = \Sigma_{11} + 2\beta\Sigma_{12} + \beta^2\Sigma_{22}$. The behavior of the second term is the same as in Theorem 3(a). By combining this result for the first term with (A.3), we have

$$t_\gamma \Rightarrow g\left(\frac{\Omega_{11} \int_0^1 B_c^\mu(s)^2 \, ds}{\Sigma_{22}}\right)^{1/2} + \frac{\delta \int B_c^\mu(s) \, dW(s)}{\left(\int B_c^\mu(s)^2 ds\right)^{1/2}} + \frac{(1 - \delta^2)^{1/2} \int B_c^\mu(s) \, dG(s)}{\left(\int B_c^\mu(s)^2 ds\right)^{1/2}}.$$

$$(A.4)$$

As in Theorem 3, G is distributed independently of (B_c^μ, W), so the final term in (A.4) has a standard normal distribution and is independent of the other terms, which yields the representation stated in the theorem. (ii) It was shown in the proof of b(i) that x_t is local-to-$I(1)$, so the results of Theorem 2(b) apply and $p(I(1) | \phi_T) \overset{p}{\to} 1$ and $p(I(0) | \phi_T) \overset{p}{\to} 0$.   ■