NBER TECHNICAL WORKING PAPER SERIES

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

Graham Elliott

James H. Stock

Technical Working Paper No. 122

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

The authors thank N.H. Chan, Tony Lancaster, and Mark Watson for helpful discussions and comments on an earlier draft. This research was supported in part by the National Science Foundation (grants no. SES-89-10601 and SES-91-22463). This paper is part of NBER's research program in Monetary Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #122
June 1992

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

ABSTRACT

It is well known that 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. This paper examines two conventional approaches to this problem, finds them unsatisfactory, and proposes a new procedure.

The two conventional approaches -- simply to ignore unit root problems or to use unit root pretests to determine the critical values for second-stage inference -- both often induce substantial size distortions. In the case of unit root pretests, this arises because type I and II pretest errors produce incorrect second-stage critical values, and because, in many empirically plausible situations, the first stage test (the unit root test) and the second stage test (the exclusion restriction test) are dependent. Monte Carlo simulations reveal size distortions even if the regressor is stationary but has a large autoregressive root, a case that might arise for example in a regression of excess stock returns against the dividend yield.

In the proposed alternative procedure, the second-stage test is conditional on a first-stage "unit root" statistic developed in Stock (1992); the second-stage critical values vary continuously with the value of the first-stage statistic. The procedure is shown to have the correct size asymptotically and to have good local asymptotic power against Granger-causality alternatives.

Graham Elliott
Department of Economics
Harvard University
Cambridge, MA 02138

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

1. Introduction

The asymptotic theory of inference in multivariate time series models when regressors have one or more unit roots is well understood (Chan and Wei (1988), Park and Phillips (1988), Sims, Stock and Watson (1990)). 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 tests of the number of lags to include in an equation, are typically unaffected by the presence of unit roots in the system. However, the null distribution of statistics testing certain quantitites 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 paper 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 which has recently received considerable attention in 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 (1989) and Campbell (1991) and, for a review of this literature, Fama (1991). 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 (1991) and in Fama's (1991) review, the importance of this regression arises from its apparently strong evidence against the "random walk" theory of stock prices. As several authors note, 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. While 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 (1981); see Mankiw and Shapiro (1985) and Stock and West (1988) 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 paper 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}. One conventional approach is to model x_t as stationary and to proceed with inference using standard Gaussian critical values. Not surprisingly, if in fact the largest root of the regressor process is near one, this approach can result in tests with sizes that far exceed levels: results in Section 2 indicate that sizes of one-tailed 5% tests can exceed 30% for cases that might reasonably be found in practice. A second approach is to pretest to see whether the regressor has a unit root 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. This approach, however, suffers from two difficulties: inferential errors in the first stage (type I and type II errors) will result in second-stage tests not having the correct size, and the first-stage and second-stage test statistics will, in many empirically plausible situations, fail to be independent. These difficulties conspire to induce large size distortions in the two-stage procedure. For example, as shown in Section 2, when a Dickey-Fuller (1979) t-statistic is used to pretest for a unit root in x_t, a two-sided second stage test with nominal level 10% can have size exceeding 30% for sample sizes encountered in econometric practice.

The second purpose of the paper is to propose an alternative approach to this problem, referred to as the Bayesian mixture approximation. In this approach, inference on the second stage test statistic is explicitly conditioned on a statistic ϕ_T that is informative about the order of integration of

x_t. This conditional distribution is computed as the 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 (it is asymptotically similar). Second, the (second stage) test has good power against local Granger causality alternatives. Specifically, if d=0 then the local asymptotic power of the proposed test is the same as the likelihood ratio test which imposes d=0 *a-priori*. If d=1, the proposed test can be compared to 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 (1991a) and his commenters) and instead entails defining priors over the point hypotheses "I(1)" and "I(0)". 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 paper 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 described above. 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 roots, respectively the local-to-I(1) and local-to-I(0) cases; these results are then use 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 discussed, and a Monte Carlo experiment reported, in Section 6. Section 7 concludes.

2. The model and problems with conventional second-stage inference techniques

A. The model.

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

$$x_t = \mu_x + \alpha(L)x_{t-1} + \beta(L)y_{t-1} + \eta_{1t} \tag{2.1a}$$

$$y_t = \mu_y + \gamma x_{t-1} + \eta_{2t} \tag{2.1b}$$

where L is the lag operator, where $\eta_t = (\eta_{1t}, \eta_{2t})'$ is a martingale difference sequence with $E(\eta_t\eta_t' | \eta_{t-1}, \eta_{t-2}, \ldots) = \Sigma_\eta$ 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 $\gamma = 0$.

The specification (2.1b) ignores the possibility of multiple lags of x_t, or of lagged y_t, being useful in prediction 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 (1988), Park and Phillips (1988) and Sims, Stock and Watson (1990) that, if additional lags of x_t are included in this regression, then Wald tests on these additional lags will have conventional χ^2 asymptotic distributions whether x_t is I(0) or I(1); moreover, in the I(1) case this Wald statistic is asymptotically independent of the Wald statistic testing the levels effect of x_t (γ=0). Thus only inference concerning this levels effect is affected by the order of integration of x_t.

It is useful to reparameterize (2.1) to isolate the largest root of $(1-\alpha(L)L)$. 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 $\Delta = (1-\rho L)$. By additionally rewriting the effect of y_{t-1} on x_t in terms of deviations from its mean under the null, we have,

(2.2a) $$\tilde{\Delta}x_t = \tilde{\mu}_x + \tilde{\alpha}(L)\tilde{\Delta}x_{t-1} + \beta(L)(y_{t-1}-\mu_y) + \eta_{1t}$$

(2.2b) $$y_t = \mu_y + \gamma x_{t-1} + \eta_{2t}$$

In terms of (2.2), the I(1) hypothesis is that $\rho = 1$ and the I(0) hypothesis is that $|\rho| < 1$. It is further assumed that, if $\rho = 1$, $\tilde{\mu}_x = 0$, while if $|\rho| < 1$, $\tilde{\mu}_x$ is unrestricted. Define Ω to be 2π times the spectral density matrix of $(\tilde{\Delta}x_t, \eta_{2t})$ at frequency zero, and let $\delta = \Omega_{21}/(\Omega_{11}\Omega_{22})^{\frac{1}{2}}$. Note that Ω is identified and is consistently estimable under either the I(1) or I(0) hypothesis, the latter being true even if there are multiple largest real roots or no nonzero real roots under I(0).

All regressions studied below include a constant but no time trend, and the non-regression based ϕ_T statistic involves demeaning x_t. In the second regression, this accords with conventional practice. In first-stage inference on the order of integration of x_t, this implicitly assumes that the I(1) model of interest is difference stationary with no drift. The modification for first-stage inference when x_t is detrended rather than demeaned is straightforward and is briefly described at the appropriate points during the exposition of the theory.

B. Size distortions introduced using Gaussian critical values or unit root pretests.

Two conventional approaches to inference in this problem are: (1) to use a standard normal approximation to the distribution of t_γ regardless of any information about the degree of persistence in x_t, and (2) 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 a simple special case of (2.2a) in which there is no feedback, x_t follows an AR(1) process, and $\mu_1=\mu_0=0$. Specifically, let

(2.3a) $$x_t = \rho x_{t-1} + \eta_{1t}$$

(2.3b) $$y_t = \gamma x_{t-1} + \eta_{2t}$$

where η_t satisfies the conditions stated following (2.2). In this simple model, $\Omega = \Sigma$ and the (frequentist) asymptotic distribution of t_γ is determined by ρ and, if $\rho=1$, by δ.

Under the null hypothesis that $\gamma=0$, the distributions of the demeaned Dickey-Fuller t-statistic testing $\rho=1$ (t^{μ}_{DF}) and the Granger causality t-statistic testing $\gamma=0$ (with a constant in the regression) are related by the expression,

$$t_{\gamma} = \delta T(1-\rho)(T^{-2}\textstyle\sum_{t=1}^{T}(x^{\mu}_{t-1})^2/\hat{\Sigma}_{11})^{1/2} + \delta t^{\mu}_{DF} + (1-\delta^2)^{1/2}z_T + o_p(1) \qquad (2.4)$$

where $z_T = (\sum_{t=1}^{T}x^{\mu}_{t-1}(\eta_{2t}-\text{Proj}(\eta_{2t}|\eta_{1t}))/((\sum_{t=1}^{T}(x^{\mu}_{t-1})^2)\hat{\Sigma}_{22})^{1/2}$, $\hat{\Sigma}_{11} = T^{-1}\sum_{t=1}^{T}\hat{\eta}^2_{1t}$, $\hat{\Sigma}_{22} = \Sigma_{22} - \Sigma_{11}^{-1}\Sigma_{12}^2$, $x^{\mu}_t = x_t - \bar{x}$, and $\text{Proj}(\eta_{2t}|\eta_{1t}) = \Sigma_{21}\Sigma_{11}^{-1}\eta_{1t}$. If $\rho=1$, (t^{μ}_{DF}, 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 $\delta=\pm1$, then $t_{\gamma}=\delta t^{\mu}_{DF}$; if $\rho=1$ and $\delta=0$, then asymptotically t_{γ} is normally distributed and is independent of t^{μ}_{DF}; 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 using (2.4) 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 (1983), Cavanagh (1985), Chan and Wei (1987), Chan (1988), Phillips (1987), and Nabeya and Tanaka (1990) (see Nabeya and Tanaka (1990) for a recent review of theoretical results). If $\rho=1+c/T$, then $T^{-1/2}(x_{\lfloor T\cdot\rfloor}-\bar{x}) \Rightarrow \Sigma_{11}^{1/2}B^{\mu}_c(\cdot)$, where $B^{\mu}_c(s)=B_c(s)-\int_0^1 B_c(r)dr$, where $B_c(s)$ is a diffusion process satisfying $dB_c(s) = cB(s) + dW(s)$ (e.g. Phillips (1987)), and where $\lfloor\cdot\rfloor$ denotes the greatest lesser integer function. Under this local-to-I(1) nesting for x_t,

$$(t_{\gamma}, t^{\mu}_{DF}) \Rightarrow (\delta\{\textstyle\int_0^1 B^{\mu}_c(s)dW\}/\{\int_0^1 B^{\mu}_c(s)^2ds\}^{1/2} + (1-\delta^2)^{1/2}z,$$
$$c\{\textstyle\int_0^1 B^{\mu}_c(s)^2ds\}^{1/2} + \{\int_0^1 B^{\mu}_c(s)dW\}/\{\int_0^1 B^{\mu}_c(s)^2ds\}^{1/2}), \qquad (2.5)$$

where z is asymptotically independent of the functionals of B^{μ}_c. 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 (2.4) and (2.5) permit analyzing the size properties of the two naive 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 importantly, the local-to-unity result (2.5) indicates that if ρ is large, $\delta\neq 0$, and the sample size is moderate, 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 second-stage rejection rates when data are generated according to (2.3) with $\gamma=0$, t_γ is computed by regressing y_t onto (1, x_{t-1}), and tests on t_γ are performed using the standard Gaussian 5% and 95% critical values. 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 using a one-sided test. If the unit root null is rejected for $\rho<1$, then the I(0) standard normal distribution is used, while if the unit root null is not rejected, then the I(1) distribution obtained from (2.5) with c=0 is used. In the context of (1), a natural unit root 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 $\rho=1$ using t^{μ}_{DF} at the α_1 level (with critical value $c_{DF;\alpha_1}$) leads to using 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 if ρ is large and $\delta\neq 0$, and is present in two-sided as well as one-sided tests. Monte Carlo evidence on sizes obtained using this sequential testing

Table 1

Size of t-tests of γ=0 with standard Gaussian critical values

One-sided tests with nominal level 5%

	$Pr(t_\gamma<-1.645)$, ρ =						$Pr(t_\gamma>1.645)$, ρ =					
	.6	.8	.9	.95	.975	1	.6	.8	.9	.95	.975	1
δ=-.9												
T=50	0.02	0.01	0.01	0.00	0.00	0.00	0.10	0.13	0.16	0.22	0.28	0.38
T=100	0.03	0.02	0.01	0.00	0.00	0.00	0.08	0.10	0.13	0.17	0.23	0.39
δ=-.5												
T=50	0.04	0.03	0.02	0.02	0.01	0.01	0.07	0.09	0.11	0.13	0.14	0.18
T=100	0.03	0.03	0.03	0.02	0.01	0.01	0.06	0.08	0.09	0.12	0.12	0.19
δ=0												
T=50	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.05	0.05	0.05	0.06	0.06
T=100	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.05	0.05	0.05

Notes: The entries are rejection rates when t_γ is compared to ± 1.645. The pseudo-data were generated according to (2.3) 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 γ=0 in a regression of y_t onto $(1, x_t)$. Based on 5000 Monte Carlo replications.

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 (1979) 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 δ=0 there is no size distortion introduced by the pretest. However, when $\delta \neq 0$, the size distortions can be large, with rejection rates exceeding 30% for large values of ρ even with T=100.

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 conditions on the relative nominal levels and sizes listed above ($\alpha_2 < \alpha_1$ and $-c_{DF;\alpha_1} > c_{z;\alpha_2}$) eventually would not hold. The drawback of this device is that, without further refinement, it gives no guide to the actual first-stage critical values to use in samples of the size typically found in empirical work, so that 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 probabilites 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 paper, however, we focus on a specific class of ϕ_T statistics developed in Stock (1992). Before turning to the proposed Bayesian mixture procedure, we first briefly review the construction and properties of these statistics.

Table 2

Size of t-tests of γ=0 with critical values selected by pretesting using a Dickey-Fuller t-statistic

Equal-tailed two-sided tests with nominal level 10%

	ρ = .6	.8	.9	.95	.975	1.0
δ=-.9						
T=50	0.13	0.32	0.33	0.31	0.26	0.22
T=100	0.11	0.12	0.30	0.34	0.28	0.23
δ=-.5						
T=50	0.10	0.16	0.17	0.17	0.16	0.16
T=100	0.11	0.11	0.18	0.17	0.16	0.15
δ=0						
T=50	0.12	0.11	0.10	0.12	0.10	0.11
T=100	0.10	0.11	0.09	0.10	0.10	0.11

Notes: Entries are rejection rates in the second stage test of γ=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 (2.3) with i.i.d. $N(0,\Sigma)$ errors. Based on 2000 Monte Carlo replications.

A. 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 in Stock (1992), which 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 (1991).) For future reference, the construction and asymptotic properties of ϕ_T are now briefly summarized. Write x_t as,

(3.1) $$x_t = d_t + u_t$$

where d_t is deterministic and u_t is purely 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 $\gamma_x(j)=\text{cov}(x_t,x_{t-j})$ for a second order stationary process x_t, and let $W(\cdot)$ denote a standard Brownian motion process restricted to the unit interval. The two hypothesis are respectively defined by:

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

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

The statistic V_T statistic is constructed 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 $k(\cdot)$ 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 $\ell^{-1}\sum_{u=1}^{\ell}k(u/\ell)\geq \underline{k}$ for all $\ell\geq 1$, where $\underline{k}>0$. It is assumed that the sequence 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 $\delta_t = \hat{d}_t - d_t$. Let $\|z_t\| = T^{-1}\sum_{t=1}^{T}z_t^2$ for a time series x_t, $D_{0T}(\lambda) = T^{-1/2}\sum_{s=1}^{[T\lambda]}\delta_s$, and $D_{1T}(\lambda) = T^{-1/2}\delta_{[T\lambda]}$. The estimated trend is assumed to satisfy the following conditions:

Detrending Conditions

A. If u_t is I(0), then:

(i) $(U_{0T}, D_{0T}) \Rightarrow \omega_0(W, D_0)$ where $D_0 \in C[0,1]$

(ii) $\lambda_T^2 \|\delta_t\| \to 0$.

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) $\|\Delta\delta_t\| = O_p(1)$.

The properties of specific functionals of V_T under the I(1) and I(0) models for several detrending processes have variously been studied by Kwiatkowski, Phillips and Schmidt (1990), Phillips (1991b), and Perron (1991). These results are extended to the general detrending conditions A and B in Stock (1992, Theorem 1), where it is shown that if y_t is I(0), then $V_T \Rightarrow W_0^d$, while if y_t is I(1) then $N_T^{-1/2}V_T \Rightarrow V_1^d$, where $V_1^d(\lambda) = \int_0^\lambda W_1^d(s)ds / \{\int_0^1 W_1^d(s)^2 ds\}^{1/2}$, 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 focus here on the case of constants included in the regression so that a mean is subracted 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 $\phi(\cdot)$ be a functional such that (i) ϕ is a continuous mapping from $D[0,1] \to \mathbb{R}^1$; (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 $\phi_T \Rightarrow \phi(W_0^d)$, while 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,

(3.4a) $p(I(0)|\phi_T) = f_0(\phi_T)\pi_0/p(\phi_T)$

(3.4b) $p(I(1)|\phi_T) = f_1(\phi_T|\ln N_T)\pi_1/p(\phi_T)$

A consequence of theorems 1 and 2 in Stock (1992) is that the posteriors asymptotically converge to zero or one: if $0<\pi_0, \pi_1<1$, then if x_t is I(0), then $p(I(0)|\phi_T) \xrightarrow{p} 1$ and $p(I(1)|\phi_T) \xrightarrow{p} 0$, while if x_t is I(1), then $p(I(0)|\phi_T) \xrightarrow{p} 0$ and $p(I(1)|\phi_T) \xrightarrow{p} 1$.

B. Bayes mixture approximation to the distribution of t_γ.

The consistency of the posteriors in (3.4) suggests using them to construct the asymptotic approximation to the distribution of t_γ conditional on ϕ_T, specifically,

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

As is made precise in theorem 1 below, under (2.2) 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 (2.2), δ, and so are readily computed for general systems.

The formulation (3.5) 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 using (3.5) has the property that the correct (i.e., I(0) or I(1)) conditional distribution of t_γ is used asymptotically.

The second motivation comes from recognizing (3.5) as a mixture of the I(0) and I(1) distributions with probility weights given by the posteriors $p(I(d)|\phi_T)$. In finite samples, these posteriors will in general be strictly positive and less than one; 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 (3.5) 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),

$$p(\gamma|S_\rho, S_\gamma) = \int_\rho \int_\theta p(S_\rho, S_\gamma|\gamma, \rho, \theta)p(\gamma, \rho, \theta)d\theta d\rho / p(S_\rho, S_\gamma). \tag{3.6}$$

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, $p(S_\gamma|S_\rho, \gamma, \rho, \theta) = p(S_\gamma|S_\rho, \gamma, d=1, \theta)1(\rho=1) + p(S_\gamma|S_\rho, \gamma, d=0, \theta)1(|\rho|<1)$; (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, $p(S_\rho|\gamma, \theta, \rho) = p(S_\rho|d=1)1(\rho=1) + p(S_\rho|d=0)1(|\rho|<1)$; and (iii) the priors on γ are flat and $p(\gamma, \rho, \theta)$ satisfies $p(\gamma, \rho, \theta) \propto p(\theta)p(d=1)\delta^*(\rho-1) + p(\theta)p(\rho)1(|\rho|<1)$, where $\int_\theta p(\theta)d\theta=1$, $\delta^*(\cdot)$ is the Dirac delta function, and $\int_{|\rho|<1} p(\rho)d\rho = p(d=0)$ where $0 \le p(d=0) \le 1$ and $p(d=0)+p(d=1)=1$. (An implication of theorem 1 below is that assumptions (i) and (ii) are satisfied asymptotically for $(S_\rho, S_\gamma) = (\phi_T, t_\gamma)$.) With these assumptions (3.6) simplifies to,

$$\begin{aligned} p(\gamma|S_\rho, S_\gamma) \propto & \{\int_\theta p(S_\gamma|S_\rho, \gamma, d=0, \theta)p(\theta)d\theta\}p(d=0|S_\rho) \\ & + \{\int_\theta p(S_\gamma|S_\rho, \gamma, d=1, \theta)p(\theta)d\theta\}p(d=1|S_\rho). \end{aligned} \tag{3.7}$$

Except for the integration over θ, if $(S_\rho, S_\gamma) = (\phi_T, t_\gamma)$, then the right-hand-side of (3.7), evaluated at $\gamma=0$, is the same as the right-hand side of (3.5) evaluated under the $\gamma=0$ null. This leads to the motivation of (3.5) as a large-sample approximation to the posterior (3.7). For the system (2.2), the dependence of the asymptotic approximation of $p(S_\gamma|S_\rho, \gamma, d=0, \theta)$ on θ is limited to the single parameter δ. While 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 (3.7) and (3.5) would be more compelling were (S_ρ, S_γ) sufficient for (ρ, γ), which (ϕ_T, t_γ) are not. For tractability in the general model (2.2) (in particular, to permit computation of the posterior distributions on I(1) and I(0)), however, we restrict attention to the statistics (ϕ_T, t_γ).

The value of (3.5) 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 (2.2) and let t_γ be the t-statistic testing $\gamma=0$ in (2.2) (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) $p(I(0)|\phi_T) \xrightarrow{p} 1$ and $p(I(1)|\phi_T) \xrightarrow{p} 0$.

(b) If x_t is I(1) then

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

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

All proofs are given in the Appendix.

Although the results in theorem 1 are presented for x_t generated according to (2.2a), in fact they apply more generally to x_t satisfying either the I(0) or I(1) conditions (3.2) and (3.3), of which the parametric model (2.2a) 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 the 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 using the limiting representation in theorem 1(b). Despite the presence of nuisance parameters in (2.2), 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. Because the I(1) distribution is continuous in δ and δ is consistently estimable, in practice inference can proceed by replacing δ by its estimated value.

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 concerning the resulting distributions when 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 peformance of the Bayesian mixture approximation (3.5) is studied when x_t is local to I(1).

A. First-stage posterior probabilities under local-to-I(0) and local-to-I(1) models.

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,

$$d_t = z_t'\beta \tag{4.1}$$

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 (4.1) the detrended data are $x_t^d = x_t - z_t'\hat{\beta} = u_t - \delta_t$, where $\delta_t = z_t'(\sum_{t=1}^T z_t z_t')^{-1}\sum_{t=1}^T z_t u_t$.

The local-to-I(0) model considered 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

$$x_t = d_t + u_t,\ u_t = u_{0T} + H_T u_{1T} \tag{4.2}$$

where u_{0T} and u_{1T} are respectively I(0) and I(1) as defined in (3.2) and (3.3), and where $H_T=h/T$, where h is a constant. This representation has a natural interpretation as a 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 (4.2) 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 the standard model in which the largest autoregressive root ρ_T is nested as converging to one at rate T:

$$y_t = d_t + u_t,\ u_t = \rho_T u_{t-1} + v_t,\ \text{where } \rho_T = 1+c/T \tag{4.3}$$

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, $U_{1T}(\cdot) = T^{-\frac{1}{2}}u_{1[T\cdot]}$ converges to a diffusion process, that is, $U_{1T} \Rightarrow \omega_1 B_c$, where $B_c(s)$ satisifies $dB_c(s) = cB(s) + dW(s)$.

Theorem 2 summarizes the behavior of V_T under these local processes with polynomial time trends of the form (4.1), 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 here between the limiting representations of these two processes, that is, in the I(0) case $U_{0T} \Rightarrow \omega_0 W_0$, while 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 (4.1) and detrending be by OLS.

(a) If y_t is local to I(0) as specified by (4.2), then $V_T \Rightarrow W^d_{0,r}$, where $W^d_{0,r}(\lambda) = W^d_0(\lambda) + r\int_0^\lambda W^d_1(s)ds$, where $r = h\omega_1/\omega_0$, $W^d_0(\lambda) = W_0(\lambda) - \nu(\lambda)'M^{-1}\Phi$, $W^d_1(\lambda) = W_1(\lambda) - \xi(\lambda)'M^{-1}\Psi$; M, ν, and ξ are respectively (q+1)×(q+1), (q+1)×1, and (q+1)×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 y_t *is is local to I(1) as specified by (4.3), then* $N_T^{-\frac{1}{2}}V_T \Rightarrow V_c^d$, *where* $V_c^d(\lambda) = \int_0^\lambda B_c^d(s)ds/[\int_0^1 B_c^d(r)^2dr]^{\frac{1}{2}}$, *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 alternatives. First consider the local-to-I(0) alternative. For functionals ϕ discussed in Section 3A, $\phi(V_T) \Rightarrow \phi(W_{0,r}^d(\lambda)) = O_p(1)$, 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 to note that this asymptotic misclassification of these 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 $100(1-\alpha)$ 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.[2] 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, using a sequence of critical values $c_{\alpha,T}$ indexed to the sample size such that $c_{\alpha,T} \to \infty$ but that $c_{\alpha,T}/\ln N_T \to 0$. 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.

B. 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 (3.2a) be nested as $\rho = 1+c/T$, where c is a constant, and let $\tilde{\mu}_x$ in (3.2a) be given by the sequence $\tilde{\mu}_{x,T} = -c(1-\tilde{\sigma}(1))\tilde{\mu}_x/T$ for general nonzero $\tilde{\mu}_x$. Asymptotic representations for the various statistics are given in the next theorem.

Theorem 3. Suppose that (x_t, y_t) are generated according to (2.2) with $\rho = 1+c/T$ and that the null hypothesis H_0: $\gamma = 0$ is true. Then:

(a) $(t_\gamma, \phi_T^{-1}lnN_T) \Rightarrow (\delta\int_0^1 B_c^\mu(s)dW/(\int_0^1 B_c^\mu(s)^2ds)^{1/2} + (1-\delta^2)^{1/2}z, \phi(V_{1,c}^\mu))$, where $V_{1,c}^\mu(\lambda) = \int_0^\lambda B_c^\mu(s)ds/(\int_0^1 B_c^\mu(s)ds)^{1/2}$, 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 $dB_c(s) = cB_c(s) + dW(s)$.

(b) $p\lim(I(1)/\phi_T) \to 1$ and $p\lim(I(0)/\phi_T) \to 0$.

The result (a) implies that 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 (3.5) 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 now 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 using the Bayesian mixture approximation (3.5). As a simplification, the local power is analyzed for a special case of (2.1) 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,

$$x_t = \rho x_{t-1} + \beta y_{t-1} + \eta_{1t} \tag{5.1a}$$

$$y_t = \gamma_T x_{t-1} + \eta_{2t}. \tag{5.1b}$$

Although the constants equal zero in (5.1), the statistics are computed in the demeaned case for comparability to the previous results.

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,

$$\text{If } |\rho| < 1: \quad \gamma_T = g/T^{1/4} \tag{5.2}$$

$$\text{If } |\rho| = 1: \quad \gamma_T = g/T. \tag{5.3}$$

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 (5.1).

(a) If $|\rho|<1$ and γ_T is given by (5.2), then

(i) $t_\gamma \Rightarrow z + \{\gamma_x(0)/\Sigma_{22}\}^{1/4} g$, where z is a standard normal random variable.

(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 (5.3), then

(i) $(t_\gamma, \phi_T | lnNT) \Rightarrow (g\{\Omega_{11}\int_0^1 B_c^\mu(s)^2 ds/\Sigma_{22}\}^{1/4} + \delta\{\int_0^1 B_c^\mu(s)dW\}/\{\int_0^1 B_c^\mu(s)^2 ds\}^{1/4} + (1-\delta^2)^{1/4} z,$ $\phi(V_{1,c}^\mu))$, where $V_{1,c}^\mu(\lambda) = \int_0^\lambda B_c^\mu(s)ds/\{\int_0^1 B_c^\mu(s)ds\}^{1/4}$, $\Omega_{11} = \Sigma_{11}+2\beta\Sigma_{12}+\beta^2\Sigma_{22}$, 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 $dB_c(s) = cB_c(s) + dW(s)$ with $c=\beta g$.

(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.

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 implies 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

$$(5.4) \qquad P[\text{Reject } H_0\colon \gamma=0 \mid \gamma_T = g/T] = E\{\Phi((c_{\ell,\alpha/2}(\hat{\phi}) - g\theta_c - \delta\tilde{\tau}_c^{\mu})/(1-\delta^2)^{1/2}) + \Phi((-c_{u,\alpha/2}(\hat{\phi}) + g\theta_c + \delta\tilde{\tau}_c^{\mu})/(1-\delta^2)^{1/2})\},$$

where $\hat{\phi} = \phi(V^{\mu}_{1,c})$, $\theta_c = \{\Omega_{11}\int_0^1 B_c^{\mu}(s)^2 ds/\Sigma_{22}\}^{1/2}$, $\tilde{\tau}_c^{\mu} = \{\int_0^1 B_c^{\mu}(s)dW\}/\{\int_0^1 B_c^{\mu}(s)^2 ds\}^{1/2}$, and $c_{\ell,\alpha/2}(\phi)$ and $c_{u,\alpha/2}(\phi)$ are respectively the lower and upper $\alpha/2$ quantiles of the conditional distribution $p(t_\gamma|\phi_T=\phi, d=1)$. The difference between (5.4) 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 (5.4) 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 ρ=1 is known *a-priori* – are plotted in figure 1a for δ = -.5 and in figure 1b for δ = -.9, in both cases for β=0 and $\Sigma_{11}=\Sigma_{22}$ (so Ω_{11}/Σ_{22}=1). Neither test dominates the other, although for alternatives with moderate power the proposed test has higher power than the test based on the unconditional d=1 critical values. If ρ=1 is known *a-priori*, then the neither test based on t_γ is optimal relative to the system likelihood ratio test which imposes the ρ=1 restriction. The power function of the system likelihood ratio test, computed for β=0, is also plotted in figure 1 (the short-dashed curve)[3]. For δ=-.5 the power loss of the proposed procedure is moderate relative to the system likelihood ratio test; for δ=-.9 it is substantial. In summary, if ρ=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 (2.2b), the proposed procedure often outperforms the test based on the marginal I(1) null distribution of t_γ.

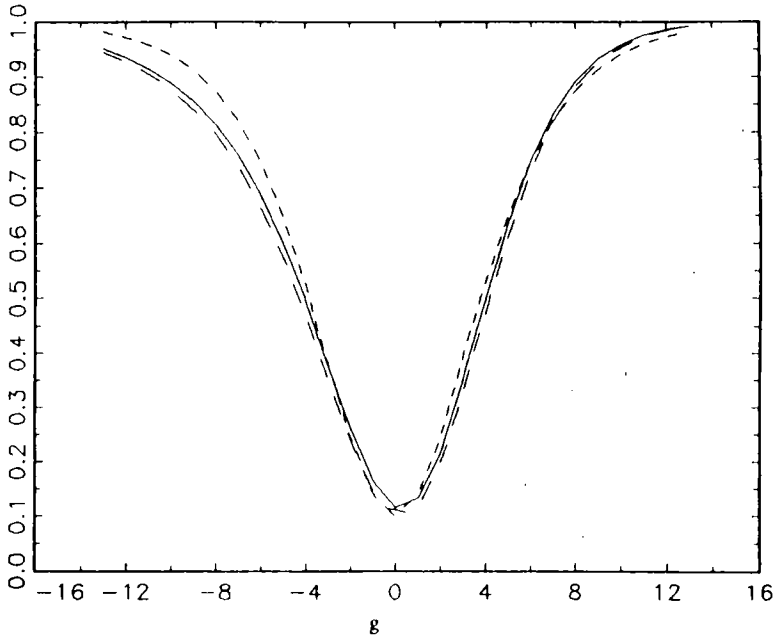
6. Numerical issues and Monte Carlo results

A. *The specific ϕ_T statistic and numerical issues*

The specific ϕ_T statistic used here is one of the statistics studied in Stock (1992):

$$\phi(V_T) = \ln\{\int_0^1 V_T(s)^2 ds\} - \ln\{\tilde{\omega}^{-1}T^{-1}\sum_{t=1}^{T}(x_t^d)^2\}. \tag{6.1}$$

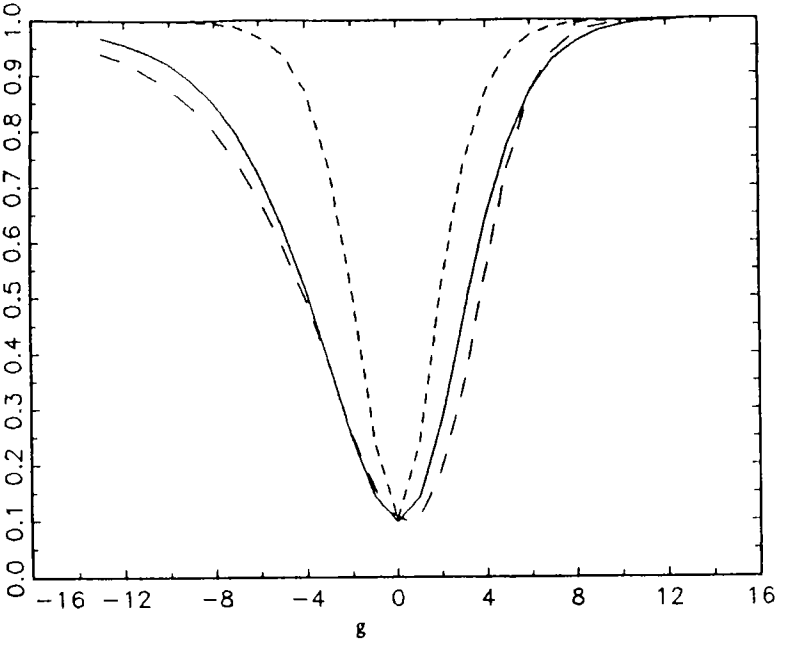
With some modifications, this statistic appears in different literatures variously as a test for random coefficients (Nabeya and Tanaka (1988)), as a test of the null of a unit MA root against a MA root less than one (Saikkonen and Luukkonen (1990)), as test of the I(0) hypothesis against the I(1) alternative (Kwiatkowski, Phillips and Schmidt (1990)), and as a test for breaks in deterministic trend components in an I(0) time series (Perron (1991)), and it is also related to the Sargan-Bhargava (1983) test of the unit root null; see Stock (1992) and Perron (1991) for references. The primary motivation for using this statistic here is that, of the three ϕ functionals studied in the Monte Carlo in Stock (1992), (6.1) appeared to have the highest rates of correctly classifying a series as I(0) or I(1) in the demeaned case of interest here.



(A) $\delta = -0.5$

Figure 1. Asymptotic power of tests of $\gamma = 0$ against the local alternative $\gamma_T = g/T$ when x_t is I(1), for $\Sigma_{11} = \Sigma_{22}$ and β=0 for (A) $\delta = -0.5$ and (B) $\delta = -0.9$.

Key: solid line: Bayesian mixture approximation test
long-dashed line: test based on t_γ using the I(1) marginal null distribution of t_γ
short-dashed line: system likelihood ratio test imposing ρ=1.



(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), for $\Sigma_{11} = \Sigma_{22}$ and $\beta=0$ for (A) $\delta = -0.5$ and (B) $\delta = -0.9$.

Key: solid line: Bayesian mixture approximation test
long-dashed line: test based on t_γ using the I(1) marginal null distribution of t_γ
short-dashed line: system likelihood ratio test imposing $\rho=1$.

The evaluation of the conditional distribution (3.5) requires computing the posterior probabilities $p(I(0)|\phi_T)$ and $p(I(1)|\phi_T)$ and computing the conditional distributions $p(t_\gamma|\phi_T, d=0)$ and $p(t_\gamma|\phi_T, d=1)$. 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 using a Parzen kernel with lag truncation parameter $\ell_T = \min(\ell_T, \ell_{max}(T/100)^{.49})$, where ℓ_T is Andrews' (1991) 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 computed here, ℓ_{max} was set to 10. Given priors Π_0 and Π_1, the posteriors are computed using kernel density estimates of the asymptotic distributions f_0 and f_1 for a given functional ϕ. For additional discussion, see Stock (1992).

The conditional distributions of $p(t_\gamma|\phi_T,I(0))$ and $p(t_\gamma|\phi_T,I(1))$ are computed 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 ϕ_T-lnN_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 is computed using a nearest neighbor algorithm implemented using 16,000 Monte Carlo replications of (ϕ_T-lnN_T, t_γ), 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.[4]

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 the 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 (3.5).

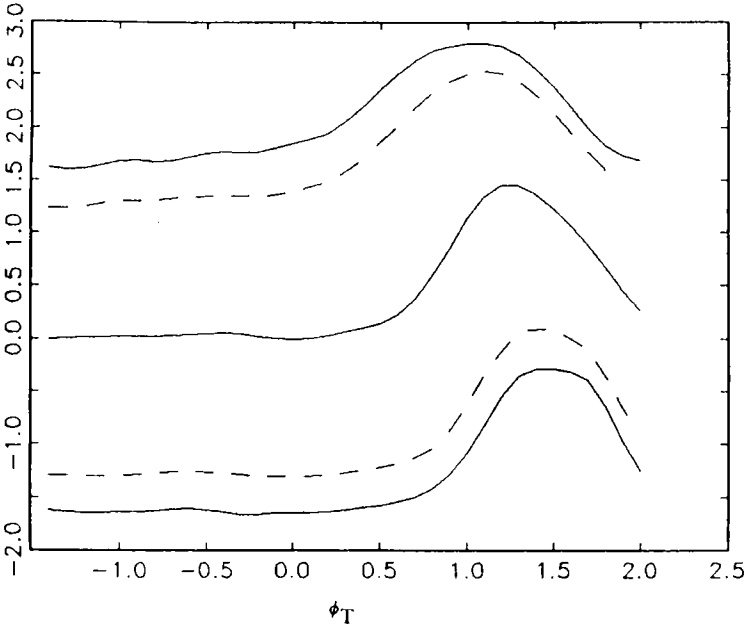


Figure 2.

Quantiles of $p(t_\gamma|\phi_T=\phi)$ as a function of ϕ based on the Bayesian mixture approximation for T=100 and δ = -0.9. Quantiles are upper 5%, 10% (dotted line), median, and lower 10% (dotted line) and 5%. The first-stage statistic ϕ_T was computed using ℓ_T = 5 with a Parzen kernel.

B. Monte Carlo Results.

The Monte Carlo experiment studies the model examined in Section 2B for which the two naive procedures were found to work poorly. Specifically, the data were generated according to (2.3) 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. Where the procedure has difficulty is for ρ large: the mean I(1) posterior initially increases with T rather than decreases for ρ = .90, .95, and .975, although in each case for T sufficiently large 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 using the Bayes mixture approximation conditional critical values. For T=100, the largest discrepency is the size of .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 preliminary 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, either by ignoring unit root problems altogether or by pretesting using a unit root test, can lead to substantial over-rejections 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 in fact x_t is I(0), and the numerical results of Section 5 indicate the proposed procedure has better 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

Table 3

Mean posterior probabilities $p(I(1)|\phi_T)$ for first-stage inference

	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

Notes: The x_t pseuo-random data were computed according to (2.3a) with the indicated value of ρ. The ϕ_T functional used is given in (6.1). The process V_T was computed using demeaned data as described in Section 3A. The Parzen kernel wasused 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' (1991) automatic lag truncation estimator (details are given in Section 6A). For a given value of ϕ_T, the posterior probabilities were computed using a kernel density algorithm (based on 16,000 pseudo-random realizations of the limiting Brownian motion functional) described in Stock (1992). Prior probabilities are $p(I(0))=p(I(1))=0.5$. Based on 5000 Monte Carlo replications.

Table 4

Size of t-tests of $\gamma=0$ with critical values computed using the Bayesian mixture approximation (3.5)

Nominal level 10%

δ	T	ρ = 0.0	.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

Notes: The data were generated according to (2.3). 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 6A. Based on 5000 Monte Carlo replications.

one. This is somewhat surprising, since the theoretical results of Section 4 indicate that the size of the second stage test will be incorrect, even asymptotically, when x_t is local to I(1); for the ϕ statistic and models studied here, these initial results suggest that this might not pose an important problem in practice.

The work presented here is preliminary in several regards. 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. Finally, the extension to including additional lags of x_t or y_t as regressors in the second stage regression, while conceptually straightforward, is of considerable practical importance since such lags are typically included in empirical practice. These and related problems are areas of ongoing research.

Appendix A

Proofs of Theorems

Proof of Theorem 1.

(a) First write x_t in the form (3.1). By assumption, $(1-\alpha(L)L)$ is invertible, so under H_0, $x_t = (1-\alpha(1))^{-1}\bar{\mu}_x + (1-\alpha(L)L)^{-1}(\eta_{1t}+\beta(L)\eta_{2t-1}) = \bar{\mu}_x + v_t$, 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 a N(0,1) marginal distribution. To show the asymptotic independence of t_γ and ϕ_T, note that (2.1) is a special case of the system analyzed by Chan and Wei (1988), so that their theorem 2.2 implies $(T^{-\frac{1}{2}}\sum_{s=1}^{[T\cdot]}\eta_s, T^{-\frac{1}{2}}\sum_{s=1}^{[T\cdot]}x_{s-1}\eta_{2s}) \Rightarrow (\Sigma^{\frac{1}{2}}B(\cdot), \zeta)$, 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 $\hat{\omega}^2$, $\hat{\Sigma}_{22}$, $T^{-1}\sum_{t=1}^{T}x_{t-1}^2$, and from writing $V_T^{\mu}(\cdot)$ in terms of the partial sum process $T^{-\frac{1}{2}}\sum_{s=1}^{[T\cdot]}\eta_s$.

(b) Write x_t as $x_t = x_0 + v_t$, where x_0 is a fixed 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 (3.3), so $\phi(V_T) \Rightarrow \phi(V_0^{\mu})$ (Stock (1992, theorems 1 and 2); this also follows, with minor modifications, from Kwiatkowski, Phillips and Schmidt (1990), Perron (1991), and Phillips (1991b)).

The joint representation of (ϕ_T, t_γ) is obtained using now-conventional frequency zero projection arguments. Note that $T^{-\frac{1}{2}}v_{[T\cdot]} = T^{-\frac{1}{2}}\sum_{s=1}^{[T\cdot]}\zeta_s + o_p(1)$, 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 $T^{-\frac{1}{2}}\sum_{s=1}^{[T\cdot]}\zeta_s \Rightarrow \Omega_{11}^{\frac{1}{2}}W(\cdot)$. In addition, let $\tilde{\eta}_{2t} = \eta_{2t} - P(\eta_{2t}|\zeta_t) = \eta_{2t} - \Omega_{11}^{-1}\Omega_{21}\zeta_t$, so that $(\zeta_t, \tilde{\eta}_{2t})$ are a martingale difference sequence and $E(\zeta_t\tilde{\eta}_{2t})=0$. With this notation, t_γ can be written,

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

Now $(T^{-\frac{1}{2}}v_{\lfloor T\cdot\rfloor}, T^{-\frac{1}{2}}\sum_{s=1}^{\lfloor T\cdot\rfloor}\eta_{2s}) \Rightarrow \Omega_{11}^{\frac{1}{2}}W(\cdot), \Sigma_{22}^{\frac{1}{2}}G(\cdot))$, where $\Sigma_{22} = \Omega_{22} - \Omega_{11}^{-1}\Omega_{21}^2$ and $W(\cdot)$ and $G(\cdot)$ are independent standard Brownian motions. In addition, $\Sigma_{22} \overset{p}{\to} \Sigma_{22} = \Omega_{22}$, so

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

The argument in Phillips (1991c) or Saikkonen (1991) imply that the second term in (A.1) is independent of the first term and has a N(0,1) marginal distribution, which leads to the representation given in the statement of the theorem. □

Proof of Theorem 2.

Throughout, let Υ_T=diag(1, T, . . . , 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×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 $U_T(\lambda) = T^{-\frac{1}{2}}\sum_{t=1}^{\lfloor T\lambda\rfloor}u_t$, and let $D_T(\lambda)=T^{-\frac{1}{2}}\sum_{s=1}^{\lfloor T\lambda\rfloor}\delta_s$. The desired result will follow if it can be shown that (i) (U_T, D_T) satisfies detrending condition A(i), and (ii) $\hat{\omega}^2 \overset{p}{\to} \omega_0$. These are shown in turn.

(i) By direct calculation, $U_T = U_{0T} + TH_T U_{1T}$, where $U_{1T}(\lambda)=T^{-1}\sum_{t=1}^{\lfloor T\lambda\rfloor}U_{1T}(t/T)$. 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 $D_T(\lambda)$ $= \nu_T(\lambda)'M_T^{-1}\Gamma_T$, where $\nu_T(\lambda) = T^{-1}\sum_{s=1}^{\lfloor T\lambda\rfloor}\Upsilon_T^{-1}z_s$ and $\Gamma_T = T^{-\frac{1}{2}}\sum_{t=1}^T\Upsilon_T^{-1}z_t u_t = \Phi_T + h\Psi_T$, where $\Phi_T = T^{-\frac{1}{2}}\sum_{t=1}^T\Upsilon_T^{-1}z_t u_{0t}$ 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 $\xi_T(\lambda) = \Upsilon_T^{-1}z_{\lfloor T\lambda\rfloor}$. From theorem 2 of Stock (1992), $(\Phi_T, \Psi_T, \nu_T, M_T) \Rightarrow (\omega_0\Phi, \omega_1\Psi, \nu, M)$, where Φ is (q+1)×1 with $\Phi_i = W_0(1) - (i-1)\int_0^1 s^{i-2}W_0(s)ds$, $i = 1, \ldots, q+1$, $\Psi = \int_0^1\xi(s)W_1(s)ds$ where $\xi_i(\lambda) = \lambda^{i-1}$, and ν is (q+1)×1 with $\nu_i(\lambda) = \lambda^i/i$, $i = 1, \ldots, q+1$. Thus $\Gamma_T \Rightarrow \omega_0(\Phi + r\Psi)$ and $D_T(\cdot)$ $\Rightarrow \omega_0 D(\cdot)$, where $D(\lambda) = \nu(\lambda)'M^{-1}(\Phi + r\Psi)$. This verifies detrending condition A(i). Let $X_{0T}^d(\lambda) = T^{-\frac{1}{2}}\sum_{s=1}^{\lfloor T\lambda\rfloor}y_s^d = U_T(\lambda) - D_T(\lambda)$, 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 $X_{0T}^d(\cdot) \Rightarrow \omega_0 W_{0,r}^d(\cdot)$, where $W_{0,r}^d(\lambda) = W_0(\lambda) + r\int_0^\lambda W_1(s)ds - \nu(\lambda)'M^{-1}(\Phi + r\Psi) = W_0^d(\lambda) - r\int_0^\lambda W_1^d(s)ds$, where the second equality uses $\nu(\lambda) = \int_0^\lambda \xi(s)ds$.

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

$$
\begin{aligned}
|\hat{\omega}^2 - \hat{\omega}_0^2| &\le |\hat{\omega}^2 - \tilde{\omega}_0^2| + |\tilde{\omega}^2 - \hat{\omega}_0^2| \\
&= \Big|\sum_{m=-\ell_T}^{\ell_T} k(m/\ell_T)T^{-1}\sum_{t=|m|+1}^{T}\{H_T^2 u_{1t}u_{1t-|m|} + H_T(u_{1t}u_{0,t-|m|} + u_{0t}u_{1,t-|m|})\}\Big| \\
&\quad + \Big|\sum_{m=-\ell_T}^{\ell_T} k(m/\ell_T)T^{-1}\sum_{t=|m|+1}^{T}\{\delta_t\delta_{t-|m|} + (\delta_t u_{t-|m|} + u_t\delta_{t-|m|})\}\Big| \\
&\le (2\ell_T+1)\{TH_T^2\|T^{-\frac{1}{4}}u_{1t}\| + T^{\frac{1}{4}}H_T\|T^{-\frac{1}{4}}u_{1t}\|^{\frac{1}{4}}\|u_{0t}\|^{\frac{1}{4}}) \\
&\quad + (2\ell_T+1)\{2\|u_t\|^{\frac{1}{4}}\|\delta_t\|^{\frac{1}{4}} + \|\delta_t\|\}
\end{aligned}
\tag{A.2}
$$

Because $\|T^{-\frac{1}{4}}u_{1t}\| \Rightarrow \int_0^1 W_1(s)^2 ds$, and $\|u_{0t}\| \xrightarrow{p} \gamma_{u_0}(0)$, the first term in (A.2) $\xrightarrow{p} 0$ if $(2\ell_T+1)TH_T^2 \to 0$ and $(2\ell_T+1)T^{\frac{1}{4}}H_T \to 0$. These in turn follow from the rate condition $\ell_T^2 \ln T/T \to 0$. Because $\|u_t\| - \|u_{0t}\| \le H_T^2\|u_{1t}\| + 2H_T\|u_{0t}\|^{\frac{1}{4}}\|u_{1t}\|^{\frac{1}{4}} \xrightarrow{p} 0$, $\|u_t\| \xrightarrow{p} \gamma_{u_0}(0)$ so the second term in (A.2) $\xrightarrow{p} 0$ if $\ell_T^2\|\delta_t\| \xrightarrow{p} 0$, which is condition A(ii).

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

(b) It is first demonstrated that conditions B(i) and B(ii) hold for u_t given by (4.3). Concerning condition B(i), write $D_{1T} = \xi_T(\lambda)'M_T^{-1}\Psi_T$, 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 $D_{1T}(\cdot) \Rightarrow D_1(\cdot)$, where $D_1(\lambda) = \xi(\lambda)'M^{-1}\Psi$, which verifies B(i). Let $X_{1T}^d(\lambda) = T^{-\frac{1}{4}}y_{\lfloor T\lambda\rfloor} = U_{1T}(\lambda) - D_{1T}(\lambda)$, so that $V_T(\lambda) = \hat{\omega}^{-1}\int_0^\lambda X_{1T}^d(s)ds$. It follows that $X_T^d(\cdot) \Rightarrow \omega_1 B_c^d(\cdot)$, 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 $\|T^{-\frac{1}{2}}\Delta\delta_t\| = \int_0^1 S_T(\lambda)d\lambda$, where $S_T(\lambda) = \psi_T' M_T^{-1}\xi_T(\lambda)\xi_T(\lambda)'M_T^{-1}\psi_T$, where $\xi_T(\lambda) = TT_T^{-1}\Delta z_{[T\lambda]}$. 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 $\|\Delta\delta_t\| \xrightarrow{p} 0$, verifying B(ii).

Next write $N_T^{-\frac{1}{2}}V_T(\lambda) = B_T^{-\frac{1}{2}}A_T(\lambda)$, where $A_T(\lambda) = T^{-3/2}\sum_{s=1}^{[T\lambda]}y_s^d$ and $B_T = T^{-1}\sum_{m=-\ell_T}^{\ell_T}k_T(m)\hat{\gamma}_{yd}(|m|)/\sum_{m=-\ell_T}^{\ell_T}k(m/\ell_T)$. 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 $B_T - T^{-2}\sum_{t=1}^T(y_t^d)^2 \xrightarrow{p} 0$. This was shown in lemma A.1 of Stock (1992) when u_t is I(1). The only part of the proof of that lemma which requires modification for the local-to-I(1) case is demonstrating that the term $D_{3T} \xrightarrow{p} 0$. From (A.1) in Stock (1992), $D_{3T} \xrightarrow{p} 0$ if $\|\Delta\delta_t\| \xrightarrow{p} 0$ (just shown) and if $\|\Delta u_t\| = O_p(1)$. This final relation is now shown. Because $(1-\rho_T L)u_t = v_t$,

$$\begin{aligned}\|\Delta u_t\| - \|v_t\| &= \|v_t + (\rho_T - 1)u_{t-1}\| - \|v_t\| \\ &\le 2|\rho_T - 1|\|u_t\|^{\frac{1}{2}}\|v_t\|^{\frac{1}{2}} + (\rho_T - 1)^2\|u_t\|.\end{aligned}$$

Now $(\rho_T - 1)^2\|u_t\| = c^2T^{-1}\int_0^1 U_{1T}(s)^2 ds \xrightarrow{p} 0$ 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 holds, completing the proof. □

Proof of Theorem 3.

(a) Under the stated assumptions, algebraic manipulations permit rewriting x_t as $x_t = \bar{\mu} + v_t$ 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 $T^{-\frac{1}{2}}v_{[T\cdot]} = T^{-\frac{1}{2}}\sum_{s=1}^{[T\cdot]}\rho^{[T\cdot]-s}(1-\tilde{\alpha}(L)L)^{-1}(\eta_{1t} + \beta(L)\eta_{2,t-1}) = T^{-\frac{1}{2}}\sum_{s=1}^{[T\cdot]}\rho^{[T\cdot]-s}\zeta_s + o_p(1)$, 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 $(T^{-\frac{1}{2}}\sum_{s=1}^{[T\cdot]}\rho^{[T\cdot]-s}\zeta_s, T^{-\frac{1}{2}}\sum_{s=1}^{[T\cdot]}\tilde{\eta}_{2s}) \Rightarrow (\Omega_{11}^{\frac{1}{2}}B_c(\cdot), \Sigma_{22}^{\frac{1}{2}}G(\cdot))$, 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,

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

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 (4.3), 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 (5.3) and (5.1b) into (5.1a) to obtain, $(1-\rho L-\beta\gamma_T L^2)x_{t-1}^2 = v_t$, 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 (3.2) and in particular $T^{-1}\sum_{t=1}^T(x_{t-1}^\mu)^2 \xrightarrow{p} \gamma_x(0)$. Thus, under the local nesting (5.2), $t_\gamma = (\sum_{t=1}^T x_{t-1}^\mu(\gamma_T x_{t-1}^\mu+\eta_{2t}))/$
$(\sum_{t=1}^T(x_{t-1}^\mu)^2\Sigma_{22})^{1/2} = g(T^{-1}\sum_{t=1}^T(x_{t-1}^\mu)^2/\Sigma_{22})^{1/2} + (\sum_{t=1}^T x_{t-1}^\mu\eta_{2t})/(\sum_{t=1}^T(x_{t-1}^\mu)^2\Sigma_{22})^{1/2}$
$\Rightarrow g(\gamma_x(0)/\Sigma_{22})^{1/2} + z$, the stated result.

(ii) Because x_t is I(0) in the sense (3.2) for T sufficiently large, the results of Stock (1992) apply and, for the reasons discussed in Section 3A, the posteriors are consistent as stated.

(b)(i) When $\rho=1$, $(1-\rho L-\beta\gamma_T L^2)$ can be factored, $(1-r_1L)(1-r_2L)$, where $r_1 = 1 + \upsilon_T$ and $r_2 = \upsilon_T$, where $\upsilon_T = \frac{1}{2}((1+4\beta\gamma_T)^{1/2}-1)$. Let $\xi_{T,t} = (1-\upsilon_T L)^{-1}(\eta_{1t} + \beta\eta_{2t-1})$. Then, with $x_0 = 0$, $T^{-1/2}x_{\lfloor T\cdot\rfloor} = T^{-1/2}\sum_{s=1}^{\lfloor T\cdot\rfloor}(1+\upsilon_T)^{\lfloor T\cdot\rfloor-s}\xi_{T,s}$. Now $\upsilon_T = \beta g/T + O(T^{-2})$, so $T^{-1/2}\sum_{s=1}^{\lfloor T\cdot\rfloor}(1+\upsilon_T)^{\lfloor T\cdot\rfloor-s}\xi_{T,t} -$
$T^{-1/2}\sum_{s=1}^{\lfloor T\cdot\rfloor}\exp\{\beta g(\lfloor T\cdot\rfloor-s)/T\}\zeta_t \xrightarrow{p} 0$, where $\zeta_t = \eta_{1t}+\beta\eta_{2t}$. Thus $T^{-1/2}x_{\lfloor T\cdot\rfloor} \Rightarrow \Sigma_{11}^{1/2}B_c(\cdot)$, where B_c is defined in the statement of the theorem, and where W is defined by $T^{-1/2}\sum_{s=1}^{\lfloor T\cdot\rfloor}\zeta_s \Rightarrow \Sigma_{11}^{1/2}W(\cdot)$.

To obtain the representation for t_γ, note that Ω converges to the covariance matrix of (ζ_T, η_{2t}) and write $t_\gamma = g(T^{-2}\sum_{t=1}^T(x_{t-1}^\mu)^2/\Sigma_{22})^{1/2} + (T^{-1}\sum_{t=1}^T x_{t-1}^\mu\eta_{2t})/(T^{-2}\sum_{t=1}^T(x_{t-1}^\mu)^2\Sigma_{22})^{1/2}$.

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(b). Combining this result for the first term with (A.3), we have,

$$(A.4) \qquad t_\gamma \Rightarrow g\{\Omega_{11}\int_0^1 B_c^{\mu}(s)^2 ds/\Sigma_{22}\}^{1/2} + \delta\int B_c^{\mu}(ds)dW(s)/(\int B_c^{\mu}(s)^2 ds)^{1/2} + (1-\delta^2)^{1/2}\int B_c^{\mu}(s)dG(s)/(\int B_c^{\mu}(s)^2 ds)^{1/2}.$$

As in theorem 3, G is distributed independently of (B_c^{μ}, W), so the final term in (A.4) is 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 that x_t is local-to-I(1), so the results of theorem 2(b) apply and $p(I(1)|\phi_T) \xrightarrow{p} 1$ and $p(I(0)|\phi_T) \xrightarrow{p} 0$. □

Footnotes

1. The results in this paper complement those in Toda and Phillips (1991), who consider the problem of sequential inference when some of the variables are cointegrated. Toda and Phillips' (1991) theory maintains that the series are cointegrated and studies 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. 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 $1-2\Phi(\{\chi^2_{1;\alpha}/(1+r^2/3)\}^{1/4})$, where $\Phi(\cdot)$ is the standard normal cdf. This power function has its minimum at r=0 and is monotone increasing to 1 as r increases.

3. This power function is derived as follows. The Gaussian MLE for γ in (5.1) with $\rho=1$ and $\beta=0$ can be obtained using the triangularized system, $\Delta x_t = \eta_{1t}$ and $y_t = \mu_y + \gamma x_{t-1} + d\Delta x_t + \tilde{\eta}_{2t}$, where $d = \Sigma_{21}\Sigma_{11}^{-1}$, so $d\Delta x_t = \text{Proj}(\eta_{2t}|\eta_{1t})$ and $\tilde{\eta}_{2t} = \eta_{2t} - \text{Proj}(\eta_{2t}|\eta_{1t})$; see Phillips (1991c), Saikkonen (1991), and Stock and Watson (1989, Section 2). Then t_γ is asymptotically N(0,1) and the local asymptotic power against $\gamma_T = g/T$ is $E\{\Phi(-c_{1-\alpha} - g(1-\delta^2)^{1/4}\theta_0) + \Phi(-c_{1-\alpha} + g(1-\delta^2)^{1/4}\theta_0)\}$ where $\theta_0 = \theta_c$ evaluated at c=0 as defined following (5.4) and $c_{1-\alpha}$ is the $1-\alpha$ Gaussian critical value. All power functions in figure 1 were computed by Monte Carlo integration of the relevant expression (e.g. (5.4)) over $(\delta, \theta_c, \tilde{r}_c^{\mu})$ for c=0 with 5000 Monte Carlo replications.

4. Specifically, 16,000 pseudo-random realizations from the limiting joint distribution were

computed using pseudo-data generated from (2.2) with ρ=1, $\beta(L) = 0$, $\tilde{\alpha}(L) = 0$, $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = \delta$, and ℓ_T=L. 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 (3.5) 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 for $\phi_T=\phi$ were computed using the resulting Monte Carlo empirical mixture cdf 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

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

Bobkoski, M.J. (1983), *Hypothesis testing in nonstationary time series*, unpublished Ph.D. thesis, Department of Statistics, University of Wisconsin.

Campbell, J.Y. (1991), "A Variance Decomposition for Stock Returns," *The Economic Journal* 101, 157-179.

Cavanagh, C. (1985), "Roots local to unity," manuscript, Department of Economics, Harvard University.

Chan, N.H., (1988), "On The Parameter Inference for Nearly Nonstationary Time Series," *Journal of the American Statistical Association* 83, no. 403, 857-862.

Chan, N.H. and C.Z. Wei, (1987), "Asymptotic Inference For Nearly Nonstationary AR(1) Processes," *Annals of Statistics* 15, 1050-1063.

Chan, N.H. and Wei, C.Z. (1988), "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *The Annals of Statistics* 16, no. 1, 367-401.

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.

Fama, E.F. (1991), "Efficient Capital Markets II," *Journal of Finance 46*, no. 5, 1575-1617.

Fama, E.F. and K.R. French (1989), "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics 25*, 23-49.

Flavin, Marjorie A. (1981), "The Adjustment of Consumption to Changing Expectations about Future Income," *Journal of Political Economy 89*, 974-1009.

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.

Mankiw, N. G. and M.D. Shapiro (1985), "Trends, Random Walks and Tests of the Permanent Income Hypothesis," *Journal of Monetary Economics 16* 165-174.

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.

Nabeya, S. and K. Tanaka, (1990), "A general approach to the limiting distribution for estimators in time series regression with nonstable autoregressive errors," *Econometrica* 58, 145-163.

Park, J.Y. and P.C.B. Phillips (1988), "Statistical Inference in Regressions with Integrated Processes: Part I," *Econometric Theory 4*, 468-497.

Perron, P. (1991), "A Test for Changes in a Polynomial Trend Function for a Dynamic Time Series," manuscript, Princeton University.

Phillips, P.C.B. (1987), "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika* 74, 535-547.

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. (1991c), "Optimal Inference In Cointegrated Systems," *Econometrica*, 59, 283-306.

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

Saikkonen, P. (1991), "Aymptotically Efficient Estimation of Cointegrating Regressions, " *Econometric Theory* 7, no. 1, 1-21.

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., J.H. Stock, and M.W. Watson (1990), "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica* 58, 113-144.

Stock, J.H. (1992), "Deciding Between I(0) and I(1)," manuscript, Kennedy School of Government, Harvard University.

Stock, J.H. and M.W. Watson (1989), "A Simple MLE of Cointegrating Vectors in General Integrated Systems," NBER Technical Working Paper no. 83.

Stock, J.H., and K.D. West (1988), "Integrated Regressors and Tests of the Permanent Income Hypothesis," *Journal of Monetary Economics 21*, no. 1, 85-96.

Toda, H.Y. and Phillips, P.C.B. (1991), "Vector Autoregression and Causality: A Theoretical Overview and Simulation Study," Cowles Foundation Working Paper no. 1001.