A Tuning Parameter Free Nearly Optimal Test of the Autoregressive Unit Root Hypothesis[1]

Morten Ørregaard Nielsen[2]
Cornell University and CREATES

July 13, 2007

[1]I am grateful to James Davidson, Javier Hualde, Tim Vogelsang, and seminar participants at Cornell University, Purdue University, and Queen's University as well as participants at the North American Summer Meeting of the Econometric Society 2006 in Minneapolis and the Summer Econometrics Workshop 2006 in Aarhus for comments and discussion. I also thank the Center for Analytic Economics (CAE) at Cornell University, the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation) at the University of Aarhus, and the Danish Social Sciences Research Council (grant no. FSE 275-05-0220)for research support. Part of this research was carried out while I was visiting the University of Aarhus, their hospitality is gratefully acknowledged.

[2]Please address correspondence to: Morten Ørregaard Nielsen, Department of Economics, Cornell University, 486 Uris Hall, Ithaca, NY 14853, USA; phone: +1 607 255 6338; fax: +1 607 255 2818; email: mon2@cornell.edu

Abstract

This paper presents a family of simple nonparametric tests of the autoregressive unit root hypothesis. The tests are constructed as a ratio of the sample variance of the observed series and that of a fractional partial sum of the series, and the family is indexed by one parameter, d, to determine the order of the (fractional) partial summation. However, there are several important differences between this parameter and the choice of, e.g., lag length in augmented Dickey-Fuller regressions or bandwidth in Phillips-Perron type tests. (i) Each member of the family with $d > 0$ is consistent. (ii) The asymptotic distribution depends on d, and thus reflects the parameter chosen to implement the test. (iii) Since the asymptotic distribution depends on d and the test remains consistent for all $d > 0$, it is possible to locate a member of the family which has the highest (within the family) power against relevant alternatives. The usual Phillips-Perron or Dickey-Fuller type tests, possibly with GLS detrending, have tuning parameters (bandwidth, lag length, etc.), i.e. parameters which change the test statistic but are not reflected in the asymptotic distribution, and thus have none of these three properties. When d is small, the asymptotic local power of the proposed nonparametric test is relatively close to the parametric power envelope, particularly in the case with a linear time-trend. Furthermore, simulations demonstrate that the proposed test has good finite sample properties in the presence of both linear and nonlinear short-run dynamics, and even rivals the (nearly) optimal parametric GLS detrended augmented Dickey-Fuller test with lag length chosen by an information criterion.

JEL Classification: C22.

Keywords: Augmented Dickey-Fuller test, fractional integration, GLS detrending, nonparametric, nuisance parameters, tuning parameters, power envelope, unit root test, variance ratio.

Short Title: Tuning Parameter Free Unit Root Test.

1 Introduction

The problem of testing for an autoregressive unit root is one of the most intensely studied testing problems in time series econometrics over the last two decades; seminal contributions to this literature include Dickey & Fuller (1979, 1981), Phillips (1987*a*), Phillips & Perron (1988), and Elliott, Rothenberg & Stock (1996). For general reviews see, e.g., Stock (1994) or Phillips & Xiao (1998). Remarkably, research on testing for unit roots has been characterized by parallel developments in theoretical and empirical econometrics, and the relevance and importance of this problem to empirical research is undeniable.

Recently, important progress has been made towards constructing unit root tests with better size and power properties. Examples include the point optimal tests and augmented Dickey-Fuller (ADF) tests with GLS detrending of Elliott et al. (1996), and the use of improved data dependent lag selection information criteria as in Ng & Perron (2001). See Haldrup & Jansson (2006) for a review focusing on power properties. The seminal contribution of Elliott et al. (1996) developed a theory of optimal testing in the framework of unit root tests, leading to the construction of parametric power envelopes for such tests, i.e. bounds on the possible power of parametric unit root tests under different conditions allowing for serial correlation, deterministic components, etc.

Nevertheless, all these tests share similar shortcomings. In particular, in the presence of serial correlation nuisance parameters appear in the asymptotic distribution unless the tests are modified to cope with the serial correlation. The ADF type tests, including the ADF-GLS tests of Elliott et al. (1996), are parametric and require the selection of a lag length for the augmentation to handle serial correlation. Similarly, the Phillips-Perron tests of Phillips (1987*a*) and Phillips & Perron (1988), although handling the serial correlation by a nonparametric correction, require the selection of bandwidth and kernel for the estimation of the long-run variance. The performance of the tests depend highly on the choice of lag length or bandwidth parameters, both in terms of finite sample power and size properties (although data dependent lag selection information criteria may improve the tests in this respect, see Ng & Perron (2001)), but also asymptotically since the consistency of the tests requires that the lag length or bandwidth parameters expand at particular rates relative to the sample size.[1] Furthermore, the asymptotic distributions of these test statistics are independent of the lag length, bandwidth, or kernel employed to construct the tests, and thus do not reflect the particular choice of these parameters. That is, the tests are characterized by parameters (lag length, bandwidth,

[1]For example, Agiakloglou & Newbold (1996) study the trade-off between size and power in Dickey-Fuller tests when data-dependent rules are used for the choice of lag order, and Leybourne & Newbold (1999*b*, 1999*b*) examine the behavior (e.g. with respect to the nuisance parameter issue) of both Dickey-Fuller and Phillips-Perron tests under the null and alternative hypotheses.

etc.) which change the value of the test statistics but are not reflected in the corresponding asymptotic distributions, and hence, in particular, not reflected in the critical values for the test statistics – such parameters are referred to as *tuning parameters*.

Existing unit root tests that are free of tuning parameters include the variable addition test of Park & Choi (1988), see also Park (1990), and the nonparametric test of Breitung (2002). The test of Breitung (2002) is a generalization of the KPSS unit root test of Shin & Schmidt (1992), who note that the calculation of their $\hat{\eta}_{\tau}(0)$ test may be done "without the necessity to choose a rule for determining l [bandwidth parameter]." Thus, Shin & Schmidt (1992) explicitly recognized, although only in passing, the importance and usefulness of tuning parameter free tests of the unit root hypothesis. Breitung (2002) demonstrated by simulations the superiority of his test relative to the variable addition test of Park & Choi (1988), so the below comparisons to existing tuning parameter free tests focus on the nonparametric Breitung (2002) test.

This paper presents a family of simple nonparametric tests of the autoregressive unit root hypothesis, which are free of tuning parameters and avoid many of the issues related to nuisance parameters while maintaining highly competitive power properties relative to parametric tests. The nonparametric tests are constructed as a ratio of the sample variance of the observed series and that of a fractional partial sum of the series. Recently, long memory and fractional integration has been attracting increasing attention from both theoretical and empirical researchers in economics and finance, see e.g. Baillie (1996) or Robinson (2003) for reviews. In this paper, fractional integration techniques are exploited to construct a family of tests for an autoregressive unit root.[2]

The proposed procedure is nonparametric and does not rely on the specification of a particular data generating process or model. This feature in particular distinguishes the approach from the well known fully parametric testing approaches, e.g. the ADF test. Of course, this aspect is a consequence of the nonparametric nature of the variance ratio test statistic, and is naturally of great importance in practical applications where the specification of the short-run dynamics is always a matter of some ambiguity and concern, since misspecified short-run dynamics leads to inconsistent estimation of the remainder of the model and hence to erroneous inferences on the order of integration. There is also no need to specify a bandwidth and kernel as in the Phillips-Perron type approach. The family of tests is indexed by one parameter, d, to determine the order of the fractional partial summation. However, there are several important differences between this parameter and the choice of tuning parameters in ADF regressions (lag length) or Phillips-Perron type tests (bandwidth and kernel). First of all, for any member of

[2] In the fractional integration literature, tests of the unit root hypothesis against alternatives of fractional integration have been developed which admit standard asymptotics, see e.g. Robinson (1994) and Tanaka (1999). This paper excludes such alternatives since the unit root hypothesis is nested within the class of autoregressive alternatives.

the family with $d > 0$, the nonparametric test is consistent. Secondly, the asymptotic distribution depends on d, and thus reflects the parameter chosen to implement the test. Thirdly, and consequently, since the asymptotic distribution depends on d and the test remains consistent for all $d > 0$, it is possible to locate a member of the family which is "tailored" to maximize the power against relevant alternatives. The usual ADF/ADF-GLS or Phillips-Perron type tests have none of these three properties. Instead, they are characterized by tuning parameters.

When d is small, the asymptotic local power of the proposed nonparametric test is relatively close to, but naturally below, the parametric power envelope of Elliott et al. (1996). In particular, in the case with a linear time trend the asymptotic local power of the nonparametric variance ratio test with a low value of d is close to the parametric power envelope.

To document the finite sample properties of the methods proposed in this paper a simulation study is conducted. The simulations demonstrate that the nonparametric variance ratio test compares favorably to the (nearly) optimal ADF-GLS test of Elliott et al. (1996) when the two are compared on an even footing by applying the MAIC lag augmentation selection rule of Ng & Perron (2001). In particular, it appears that the nonparametric variance ratio test is useful and that the finite sample size-adjusted power of the nonparametric variance ratio test is similar to, if not higher than, that of the ADF-GLS test of Elliott et al. (1996) in sample sizes that are relevant for empirical research. Thus, even though the ADF-GLS test has superior asymptotic local power properties, the need to select a tuning parameter (lag augmentation) and estimate nuisance parameters (serial correlation) reduces the power of the Dickey-Fuller type tests in more realistic settings.

The remainder of the paper is laid out as follows. In the next section the nonparametric approach is introduced and the variance ratio family of tests is presented along with the asymptotic distribution theory. Section 3 develops the relevant local asymptotic power analysis and introduces a GLS detrended version of the tests. In section 4 simulation evidence is presented to document the finite sample properties of the nonparametric test. Both sections 3 and 4 include comparisons to parametric power envelopes and (nearly) efficient parametric tests. Section 5 offers some concluding remarks. All proofs are gathered in the appendix.

2 The Nonparametric Variance Ratio Test

Suppose the observed univariate time series $\{y_t\}_{t=1}^T$ is generated by the AR(1) model

$$y_t = \phi y_{t-1} + u_t, \quad t = 0, 1, ..., \tag{1}$$

where $y_0 = 0$ and u_t is unobserved short-run dynamics to be defined precisely later.[3] The unit root testing problem is the test of the null hypothesis

$$H_0 : \phi = 1 \text{ vs. } H_1 : |\phi| < 1. \tag{2}$$

Consider, under the null hypothesis, the behavior of the observed time series $\{y_t\}_{t=1}^{T}$ generated according to (1) with $\phi = 1$ and also its fractional partial sum,

$$\tilde{y}_t = \Delta_+^{-d} y_t, \quad d > 0, t = 0, 1, ..., \tag{3}$$

where we have used the definition

$$\Delta_+^{-d} x_t = (1 - L)_+^{-d} x_t = \sum_{j=0}^{t-1} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} x_{t-j} = \sum_{j=0}^{t-1} \pi_j(d) x_{t-j}$$

so that only values corresponding to a positive time index enters the fractional difference/summation expression. This is denoted by the subscript on the difference operator, i.e. Δ_+, which is a truncated version of the binomial expansion in the lag operator L ($Lx_t = x_{t-1}$).

It is well known that under regularity conditions on u_t, a functional central limit theorem is obtained for y_t and a similar (fractional) functional central limit theorem is obtained for $\tilde{y}_t$, i.e.

$$T^{-1/2} y_{[Ts]} \Rightarrow \sigma_y W_0(s), \quad 0 \le s \le 1, \tag{4}$$

$$T^{-1/2-d} \tilde{y}_{[Ts]} \Rightarrow \sigma_y W_d(s), \quad 0 \le s \le 1, \tag{5}$$

as $T \to \infty$ for some σ_y to be specified later. Here, $[\bullet]$ denotes the integer part of the argument, "$\Rightarrow$" means weak convergence in $D[0,1]$ endowed with the Skorohod J_1 topology, and W_d is the type II fractional standard Brownian motion of order d ($> -1/2$), see e.g. Marinucci & Robinson (1999, 2000). The fractional standard Brownian motion of type II can be defined as a Holmgren-Riemann-Liouville fractional integral,

$$W_d(r) = 0, \text{ a.s.}, r = 0, \tag{6}$$

$$W_d(r) = \frac{1}{\Gamma(d+1)} \int_0^r (r-s)^d \, dW_0(s), \; r > 0. \tag{7}$$

Note that with this definition W_0 is the standard Brownian motion.

[3] The initial condition can be replaced by other well-known conditions that yield the same functional central limit theorem (4). Note that, if it is known that y_0 is likely to be small then, in the fully parametric setup, this knowledge will generate more discriminatory power for the unit root problem by applying the ADF-GLS tests of Elliott et al. (1996), see Müller & Elliott (2003). In that sense, the zero initial condition poses the greatest challenge for the proposed nonparametric test when compared to the ADF-GLS tests in simulations below.

It follows that the rescaled sample variances of y_t and $\tilde{y}_t$ satisfy

$$T^{-2}\sum_{t=1}^{T} y_t^2 \Rightarrow \sigma_y^2 \int_0^1 W_0(s)^2\, ds \tag{8}$$

$$T^{-2(1+d)}\sum_{t=1}^{T} \tilde{y}_t^2 \Rightarrow \sigma_y^2 \int_0^1 W_d(s)^2\, ds \tag{9}$$

as $T \to \infty$, under the unit root null hypothesis (2). Thus, by forming the variance ratio,

$$\rho(d) = T^{2d}\frac{\sum_{t=1}^{T} y_t^2}{\sum_{t=1}^{T} \tilde{y}_t^2}, \tag{10}$$

the nuisance parameter σ_y^2 is eliminated from the limiting distribution and there is no need to estimate serial correlation parameters. The statistic $\rho(d)$ in (10) defines the family of variance ratio statistics indexed by the fractional partial summation parameter, d.[4]

The statistic (10) generalizes the idea of Shin & Schmidt (1992), Breitung (2002), and Taylor (2005) who used the ratio of the sample variance of y_t and that of the partial sum of y_t to eliminate the nuisance parameter σ_y^2 and avoid estimation of serial correlation parameters in testing for a unit root.[5] The same idea was applied by Vogelsang (1998*a*, 1998*b*) to test for structural breaks, see also the simulation evidence by Harvey, Leybourne & Newbold (2004).

In recent work, Müller (2007*a*, 2007*b*) demonstrates some desirable properties of variance ratio-type unit root test statistics such as (10), which are not necessarily shared by other statistics that have to estimate the long-run variance σ_y^2. In particular, tests based on variance ratio-type statistics are shown to be able to consistently discriminate between the unit root null and the stationary alternative.

To adjust for a non-zero mean and possibly deterministic time trend in the observed time series y_t, suppose $\{y_t\}_{t=1}^{T}$ is generated according to

$$y_t = \alpha'\delta_t + z_t, \quad t = 0, 1, ..., \tag{11}$$

where z_t is unobserved and generated as y_t in (1). Here, $\delta_t = 0$ when there are no deterministic terms, $\delta_t = 1$ when there is a non-zero mean, and $\delta_t = (1, t)'$ when there is correction for

[4] Note that the ratio in (10) has the appearance of a likelihood ratio test statistic constructed based on the implicit null and alternative (point) hypotheses that y_t is *i.i.d.* and that $\tilde{y}_t$ is *i.i.d.*, respectively. In other words, this interpretation would correspond to a null hypothesis that the observed data is $I(0)$ against the alternative that the data is $I(-d)$, that is (fractionally) overdifferenced, and since this does not appear particularly useful we do not consider this interpretation further.

[5] A similar type of variance ratio test, based on the ratio of the sample variance of Δy_t and $y_t - y_{t-k}$ for some $k > 1$, was used by, inter alia, Lo & MacKinlay (1988) and Miller & Newbold (1995) to test the unit root hypothesis in ARIMA models.

a deterministic linear time trend. Thus, the family of variance ratio statistics corrected for deterministic terms is defined as in (10) but with the residuals $\hat{y}_t = y_t - \hat{\alpha}'\delta_t$ replacing the observed time series y_t. For now, $\hat{y}_t$ are ordinary least squares residuals, which are sufficient to generate a consistent test, and in the next section GLS detrending is considered in the spirit of Elliott et al. (1996) which will in fact increase the power of the test, at least against linear alternatives and for an important range of d values.

The following assumption on u_t in (1) is sufficient for the weak convergence in (4) and (5).

Assumption 1 *The unobserved errors u_t are generated by the linear process*

$$u_t = \psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad t = 0, 1, ...,$$

where $\sum_{j=0}^{\infty} j^{1/2} |\psi_j| < \infty$, $\psi(1) = \sum_{j=0}^{\infty} \psi_j \neq 0$, $\psi_0 = 1$, and the ε_t are i.i.d. with $E\varepsilon_t = 0$, $E\varepsilon_t^2 = \sigma_\varepsilon^2 > 0$, $E|\varepsilon_t|^q < \infty$ for some $q > 2$.

Under the null hypothesis that $\phi = 1$ and under Assumption 1 on u_t the convergence in (5) holds with $\sigma_y^2 = \psi(1)^2 \sigma_\varepsilon^2$, see e.g. Akonom & Gourieroux (1987) and Marinucci & Robinson (2000). The conditions on the coefficients ψ_j are similar to those employed by Phillips & Solo (1992) and can in fact be relaxed slightly, e.g. Marinucci & Robinson (2000) and Davidson & de Jong (2000), but are mild enough to cover all stationary and invertible ARMA models.

Note that under the null hypothesis, $\tilde{y}_t$ is integrated of order $1+d > 1$ and the relevant moment condition in that case is $q > 2$, c.f. Marinucci & Robinson (2000). Under the alternative, $\tilde{y}_t$ is integrated of order d thus requiring a different moment condition below.

When the convergence (5) is established, the limiting distribution of the variance ratio statistic $\rho(d)$ is easily derived and is presented in the following theorem.

Theorem 1 *Let y_t be defined by (1) and (11), $\rho(d)$ by (10) with the residuals $\hat{y}_t$ replacing y_t in (3) and (10), and let $j = 0$ when $\delta_t = 0$, $j = 1$ when $\delta_t = 1$, and $j = 2$ when $\delta_t = (1, t)'$. Under the null hypothesis (2), Assumption 1 on u_t, and for $d > 0$,*

$$\rho(d) \Rightarrow U_j(d) = \frac{\int_0^1 B_j(s)^2 ds}{\int_0^1 \tilde{B}_{j,d}(s)^2 ds}, \quad j = 0, 1, 2,$$

as $T \to \infty$, where

$$B_j(s) = W_0(s), \quad j = 0,$$

and the demeaned respectively demeaned and detrended standard Brownian motions are defined as

$$B_j(s) = W_0(s) - \left(\int_0^1 W_0(r) D(r)' dr \right) \left(\int_0^1 D(r) D(r)' dr \right)^{-1} D(s), \quad j = 1, 2,$$

where $D(s) = 1$ *when* $\delta_t = 1$ *(*$j = 1$*),* $D(s) = (1, s)'$ *when* $\delta_t = (1, t)'$ *(*$j = 2$*), and also* $\tilde{B}_{0,d}(s) = W_d(s)$ *and*

$$\tilde{B}_{j,d}(s) = W_d(s) - \left(\int_0^1 W_0(r) D(r)' dr\right) \left(\int_0^1 D(r) D(r)' dr\right)^{-1} \int_0^s \frac{(s-r)^{d-1}}{\Gamma(d)} D(r) dr, \quad j = 1, 2.$$

Note that in this theorem and below, the weak convergence is pointwise in d. Also note that, by the substitution $u = s - r$,

$$\begin{aligned} \int_0^s \frac{(s-r)^{d-1}}{\Gamma(d)} dr &= \frac{s^d}{d\Gamma(d)} = \frac{s^d}{\Gamma(d+1)}, \\ \int_0^s \frac{(s-r)^{d-1}}{\Gamma(d)} r dr &= \frac{s^{d+1}}{d(d+1)\Gamma(d)} = \frac{s^{d+1}}{\Gamma(d+2)}, \end{aligned}$$

so that the term $\int_0^s \frac{(s-r)^{d-1}}{\Gamma(d)} D(r) dr$ appearing in the definition of $\tilde{B}_{j,d}(s)$ for $j = 1$ and $j = 2$ corresponds to fractional powers of s. Furthermore, $\tilde{B}_{j,d}(s)$ $(j = 1, 2)$ is a fractional Brownian motion less the trend correction term from an L_2 regression of a standard (non-fractional) Brownian motion on a fractional polynomial trend of order s^d.

The asymptotic distribution $U_j(d)$ of $\rho(d)$ given in Theorem 1 depends only on the choice of deterministic terms (j) and the parameter d, i.e. the order of (fractional) partial summation indexing the family of tests. Hence, the asymptotic distribution can easily be simulated to obtain quantiles for any member of the family characterized by the value of the parameter d. Quantiles of $U_j(d)$ for several members of the family, i.e. several values of the parameter d, are presented in Table 1.

Table 1 about here

A very important property of the variance ratio statistic (10) and its asymptotic distribution in Theorem 1 is that there is no need to specify or estimate any particular parametric or nonparametric model for the short-run dynamics in $\psi(L)$. Thus, the statistic is asymptotically invariant to any short-run dynamics in the data generating process for y_t. As a result, any hypothesis test based on a member of the family of variance ratio statistics will share this useful property.

Thus, consider using $\rho(d)$ as a test of the unit root hypothesis, i.e. of the null hypothesis (2), where large values of $\rho(d)$ are associated with rejection of H_0. Under the alternative, $\tilde{y}_t$ is a fractionally integrated process of order d, and a strengthening of Assumption 1 is needed.

Assumption 2 *Assumption 1 is satisfied with* $q > 2$ *when* $d < 1/2$ *and* $q > \max(2, 2/(2d-1))$ *when* $d > 1/2$.

Under Assumption 2 more moments are required (higher q) compared to Assumption 1. The moment condition $E|\varepsilon_t|^q < \infty$ for some $q > \max(2, 2/(2d-1))$ in Assumption 2 is needed when analyzing the behavior of $\rho(d)$ under the alternative hypothesis, and in particular to obtain (5) when $\tilde{y}_t$ is integrated of order $d \in (1/2, 1)$ (when $d \geq 1$ the moment condition is $q > 2$). Thus, the innovations ε_t are assumed to be identically and independently distributed, and less dependence (smaller d) requires the existence of more moments (higher q) for ε_t when $d \in (1/2, 1)$. When $d < 1/2$, $\tilde{y}_t$ is (asymptotically) stationary and the relevant moment condition is $q > 2$.

The rejection region of the test and the alternatives against which it is consistent are given in the following theorem.

Theorem 2 *Let the assumptions of Theorem 1 be satisfied, with Assumption 2 replacing Assumption 1. Then the test that rejects H_0 in (2) when $\rho(d) > CV_{j,\gamma}(d)$, where $CV_{j,\gamma}(d)$ is found from*

$$\Pr(U_j(d) > CV_{j,\gamma}(d)) = \gamma, \tag{12}$$

has asymptotic size γ and is consistent against the alternative H_1 in (2).

Note that, although the parameter d indexing the family of tests is specified by the econometrician, it is not a tuning parameter in the sense described in the introduction above. This is because the choice of d is reflected in the limiting distribution of the variance ratio statistic, unlike the tuning parameters, e.g. lag length and bandwidth parameters, in the usual Dickey-Fuller or Phillips-Perron unit root tests. Thus, it may be possible to locate a member of the family of tests which is tailored in such a way that power is maximized against relevant alternatives. Indeed, this is considered in the following asymptotic local power analysis, where results are provided which recommend $d = 0.1$, c.f. Theorem 3. See also the simulations in section 4 below. Another typical choice could be $d = 1$, i.e. partial summation, based on computational simplicity, which would in fact lead to (the inverse of) the statistic suggested by Breitung (2002) to test for a unit root against nonlinear alternatives, see also Taylor (2005) for seasonal unit root tests. An important reason to choose $d < 1/2$ rather than $d \in (1/2, 1)$ is that the moment condition $q > 2/(2d-1)$ is very strict when d is greater than, but close to, $1/2$, whereas when d is less then $1/2$ the moment condition is simply $q > 2$.

The variance ratio statistic (10) is related to many well known statistics such as the KPSS statistic of Kwiatkowski, Phillips, Schmidt & Shin (1992) and Shin & Schmidt (1992), and also earlier statistics such as the Durbin-Watson statistic, to mention just a few. Indeed, variance ratio type statistics have a very long tradition in time series analysis. However, there is a fundamental difference between those statistics and the variance ratio statistic in (10). The former statistics are mostly based on the ratio of the sample variance of y_t and that of Δy_t

(corresponding to $d = -1$ in the present setup) and then the σ_y^2 that appears in the limiting distribution is divided out by employing some form of long-run variance estimator. On the other hand, the statistic (10) is the ratio of the sample variance of y_t and that of the (fractional) partial sum of y_t, which implies that σ_y^2 cancels from the limiting distribution and there is no need to estimate serial correlation parameters or the long-run variance. In the next section, the asymptotic local power is analyzed and the subsequent section investigates the finite sample properties, where in both sections comparisons are made to the GLS detrended ADF tests of Elliott et al. (1996).

3 Asymptotic Local Power Analysis

In this section, the asymptotic local power of the autoregressive unit root test described in Theorem 2 is analyzed to guide the choice of the parameter d. Since d is the only parameter indexing the family of tests and the only parameter needed to calculate the variance ratio test statistic (10), and is also the only parameter in the asymptotic distribution, it is of interest to examine the power function for a range of values of d. In particular, one might ask if there is a member of the family with maximum (within the family) power against relevant alternatives, i.e. if there is a power maximizing value of d. This value could then be chosen by the researcher in order to "tailor" the test to obtain high power, i.e. to select the member of the test family with the best power properties.

Instead of attempting to calculate the exact power function of the test as a function of d, the power is described qualitatively using local-to-unity asymptotics. To obtain non-degenerate power under the alternative, consider the well-known sequence of local alternatives where $\{y_t\}_{t=1}^T$ is generated according to

$$y_t = \phi_T y_{t-1} + u_t, \quad \phi_T = 1 - c/T, \tag{13}$$

i.e. near-integrated alternatives with some $c \geq 0$, c.f. Chan & Wei (1987) and Phillips (1987*b*). For any fixed T, y_t is stationary (the alternative) provided T is large enough that $c/T \in (0, 2)$. On the other hand, y_t is nonstationary (the null hypothesis) in the limit since $\phi_T \to 1$ when $T \to \infty$. Thus, the model (13) provides alternatives local to $\phi = 1$. The next two subsections first consider the asymptotic local power of the above family of variance ratio tests, and subsequently introduce a GLS detrended version to be compared to the GLS detrended ADF test of Elliott et al. (1996).

3.1 Asymptotic Local Power of the Variance Ratio Test

The following theorem presents the distribution of the variance ratio statistic under the near-integrated local alternatives.

Theorem 3 *Let the assumptions of Theorem 1 be satisfied except (13) replaces (1) in the definition of y_t (or z_t if $\delta_t \neq 0$). Then, as $T \to \infty$,*

$$\rho(d) \Rightarrow U_{j,NI}(c,d) = \frac{\int_0^1 J_{j,c}(s)^2 \, ds}{\int_0^1 \tilde{J}_{j,c,d}(s)^2 \, ds}, \quad j = 0,1,2,$$

where $J_{0,c}(s)$ is the Ornstein-Uhlenbeck process,

$$J_{0,c}(s) = W(s) - c\int_0^s e^{-(s-r)c} W(r) dr, \quad J_{0,c}(0) = 0,$$

$J_{1,c}(s), J_{2,c}(s)$ are the demeaned respectively demeaned and detrended Ornstein-Uhlenbeck processes,

$$J_{j,c}(s) = J_{0,c}(s) - \left(\int_0^1 J_{0,c}(r) D(r)' dr\right)\left(\int_0^1 D(r) D(r)' dr\right)^{-1} D(s), \quad j = 1,2,$$

and

$$\begin{aligned} \tilde{J}_{0,c,d}(s) &= W_d(s) - c\int_0^s e^{-(s-r)c} W_d(r) \, dr, \\ \tilde{J}_{j,c,d}(s) &= \tilde{J}_{0,c,d}(s) - \left(\int_0^1 J_{0,c}(r) D(r)' dr\right)\left(\int_0^1 D(r) D(r)' dr\right)^{-1} \int_0^s \frac{(s-r)^{d-1}}{\Gamma(d)} D(r) \, dr, \quad j = 1,2. \end{aligned}$$

Theorem 3 describes the distribution of $\rho(d)$ under the sequence of near-integrated local alternatives (13). It follows that the asymptotic local power of any member of the family of variance ratio tests of (2) can be described in terms of $U_{j,NI}(c,d)$ whose distribution is a continuous function of the local noncentrality parameter $c \geq 0$ and the index $d > 0$. Note that $U_{j,NI}(0,d) = U_j(d)$, $j = 0,1,2$. Also note that the process $\tilde{J}_{0,c,d}(s)$ appearing in Theorem 3 is a fractional version of the well known Ornstein-Uhlenbeck process $J_{0,c}(s)$, see e.g. Brockwell & Marquardt (2005) or Buchmann & Klüppelberg (2006).

The local asymptotic power function of any member of the family of variance ratio tests can thus be calculated as

$$P\left(U_{j,NI}(c,d) > CV_{j,\gamma}(d)\right),$$

where $CV_{j,\gamma}(d)$ is defined in Theorem 2.

Figure 1 about here

Figure 1 displays simulated asymptotic local power curves for several members of the variance ratio test family (with $\gamma = 0.05$) as a function of the local noncentrality parameter, $c \geq 0$. The simulated power functions are based on 20,000 Monte Carlo replications of (13) with $T = 500$, u_t *i.i.d.* standard normal, and either no deterministic terms (Panel A), constant mean (Panel B), or linear trend (Panel C). In each of the graphs, the power curves are drawn for $d \in \{0.1, 0.25, 0.5, 1.0\}$. The critical values are simulated for each d based on 20,000 Monte Carlo replications and sample size 500.

From Figure 1 it is clear that the asymptotic local power of the variance ratio test is monotonic in d, and that $d = 0.1$ appears to be the "power maximizing" choice among those power functions depicted, in the sense that it has uniformly (in c) higher power relative to $d = 0.25$, $d = 0.5$, and $d = 1.0$. It should be noted that other choices of d conform to the monotonicity apparent in Figure 1 although the gain in power from choosing an even smaller value of d is minor. Furthermore, it also seems unwise to choose d too small, since then d acts as if it depends (inversely) on the sample size which may distort the size properties of the test (and result in poor size properties in finite samples). Obviously, if $d = 0$ the test statistic degenerates. Thus, Figure 1 suggests that $d = 0.1$ provides a good choice of the parameter d indexing the family of tests, in the sense that local asymptotic power is better uniformly in c relative to higher values of d.[6] Note, however, that the local asymptotic power gains relative to higher values of d are smaller when allowing non-zero mean (and possibly trend) correction as in Panels B and C. In section 4 below, further support of the test with $d = 0.1$ relative to the one with $d = 1$ is presented based on simulation evidence.

Finally, Figure 1 clearly demonstrates that significant power gains can be achieved by considering non-integer values of d (< 1). Comparing the $d = 1$ curve with the other curves, it is seen that $d = 1$ provides the lowest power in all the panels of Figure 1. In other words, the variance ratio test with $d = 1$, which was suggested by Breitung (2002) for testing the unit root hypothesis against nonlinear models, can be vastly improved upon (at least against near-integrated alternatives) by admitting non-integer values of $d < 1$.

3.2 GLS Detrending and Comparison to ADF-GLS Tests

Now consider applying GLS detrending to correct for deterministic terms instead of the simple OLS detrending above. Thus, for any generic series $\{x_t\}_{t=1}^T$ and some constant $\bar{c}$ define $x_{\bar{c},1} = x_1$ and $x_{\bar{c},t} = x_t - (1 - \bar{c}/T)\, x_{t-1}, t = 2, ..., T$. With this definition the observed GLS detrended time series, denoted $\{\hat{y}_{\bar{c},t}\}_{t=1}^T$, is given by

$$\hat{y}_{\bar{c},t} = y_t - \tilde{\alpha}'\delta_t, \tag{14}$$

where

$$\tilde{\alpha} = \arg\min_{\alpha} \sum_{t=1}^{T} \left(y_{\bar{c},t} - \alpha'\delta_{\bar{c},t}\right)^2 .$$

[6] This analysis is based on local power arguments. An alternative, which is not explored here, is to consider power against fixed alternatives by examining the rate of divergence under the alternative. However, this is known to be possibly misleading, e.g. this would make one believe that the F-test in the standard regression model is superior to the t-test because it has a faster rate of divergence under the alternative.

The use of GLS detrended time series for the ADF test was proposed by Elliott et al. (1996) who in particular suggest $\bar{c} = 7$ and $\bar{c} = 13.5$ for $\delta_t = 1$ and $\delta_t = (1, t)'$, respectively, resulting in the ADF-GLS test. These values of $\bar{c}$ correspond to the (local) point alternatives against which the local asymptotic power for significance level 5% equals one-half. With respect to the choice of lag augmentation, i.e. tuning parameter, for the ADF-GLS tests, Ng & Perron (2001) show that the tests have both good size and power properties when employing a modified version of the well known Akaike information criterion, which is applied in the simulations below. In the asymptotic comparisons, the lag augmentation is (unrealistically, of course) assumed to be chosen correctly and optimally.

Consider constructing the variance ratio test based on the GLS detrended time series (14). That is,

$$\rho(\bar{c}, d) = T^{2d} \frac{\sum_{t=1}^{T} \hat{y}_{\bar{c},t}^2}{\sum_{t=1}^{T} \tilde{y}_{\bar{c},t}^2}, \tag{15}$$

where $\tilde{y}_{\bar{c},t} = \Delta_+^{-d} \hat{y}_{\bar{c},t}$ as in (3). The distribution of the GLS detrended variance ratio test (15) under the sequence of local alternatives (13) depends on the stochastic processes

$$\begin{aligned} V_{\bar{c},c}(s) &= J_{0,c}(s) - b_1 s, \\ \tilde{V}_{\bar{c},c,d}(s) &= \tilde{J}_{0,c,d}(s) - b_1 \frac{s^{d+1}}{\Gamma(d+2)}, \\ b_1 &= \frac{(1+\bar{c})}{1+\bar{c}+\bar{c}^2/3} J_{0,c}(1) + \frac{\bar{c}^2}{1+\bar{c}+\bar{c}^2/3} \int_0^1 r J_{0,c}(r)\, dr, \end{aligned}$$

and is presented in the next theorem.

Theorem 4 *Let the assumptions of Theorem 3 be satisfied except y_t is GLS detrended as in (14) and the variance ratio statistic is given by (15). Then, as $T \to \infty$,*

$$\rho(\bar{c}, d) \Rightarrow U_{j,GLS}(\bar{c}, c, d), \quad j = 1, 2,$$

where

$$\begin{aligned} U_{1,GLS}(\bar{c}, c, d) &= U_{0,NI}(c, d), \\ U_{2,GLS}(\bar{c}, c, d) &= \frac{\int_0^1 V_{\bar{c},c}(s)^2\, ds}{\int_0^1 \tilde{V}_{\bar{c},c,d}(s)^2\, ds}, \end{aligned}$$

and $U_{0,NI}(c, d)$ and $J_{0,c}(s)$ are defined in Theorem 3.

To implement the GLS detrending procedure for the variance ratio test, a recommendation regarding the choice of local detrending parameter $\bar{c}$ is needed.

Table 2 about here

Following Elliott et al. (1996), the values of $\bar{c} = c$ that attain asymptotic local power equal to one-half at 5% significance level are presented in Panel A of Table 2 for $\delta_t = 1$ and $\delta_t = (1, t)'$ and several values of d. These values of $\bar{c} = c$ are those for which the power envelope type function $\Pr\left(U_{j,GLS}\left(\bar{c}, \bar{c}, d\right) > CV_{j,0.05}\left(\bar{c}, d\right)\right)$ is equal to one-half at the 5% significance level, where $CV_{j,\gamma}\left(\bar{c}, d\right)$ satisfies $\Pr\left(U_{j,GLS}\left(\bar{c}, 0, d\right) > CV_{j,\gamma}\left(\bar{c}, d\right)\right) = \gamma$.

Critical values of the variance ratio test for the particular $\bar{c}$ and d values presented in Panel A of Table 2 are presented in Panel B of Table 2 for significance levels $\gamma = 1\%, 5\%$, and 10%. Note that the table presents the critical values for $j = 2$ only, since the $j = 1$ case has the same asymptotic null distribution and hence the same critical values as $j = 0$ in Table 1.

Figure 2 about here

In Figure 2 the asymptotic local power functions of the GLS detrended variance ratio tests with $d = 0.001, 0.1, 1$ are presented for the no deterministics case (Panel A), the case with a constant mean (Panel B), and the case with a linear trend (Panel C). The asymptotic local power functions are simulated based on 20,000 Monte Carlo replications with $T = 500$. Also included are the local power functions of the Dickey-Fuller and GLS detrended Dickey-Fuller tests. The local power functions of the latter are indistinguishable from the parametric power envelope, c.f. Elliott et al. (1996).

The main focus is the comparison between the Dickey-Fuller tests, the GLS detrended Dickey-Fuller tests, and the GLS detrended variance ratio test with $d = 0.1$, since the latter was preferred to tests with higher values of d in the asymptotic local power analysis in Figure 1 above. The test with $d = 0.001$ is included to examine how close the asymptotic local power curve of the nonparametric variance ratio test can be pushed towards the parametric power envelope, essentially given by the asymptotic local power function of the GLS detrended Dickey-Fuller tests.

Note that, as observed by Breitung & Taylor (2003), the Breitung (2002) test does not benefit from GLS detrending – on the contrary – whereas the variance ratio test based on fractional partial summation (e.g. $d = 0.1$) does benefit significantly from GLS detrending, in terms of asymptotic local power. In both the case with mean correction (Panel B) and the case with trend correction (Panel C), the asymptotic local power of the variance ratio test with $d = 0.1$ is approximately the same as that of the Dickey-Fuller test. When GLS detrending is employed in the construction of the variance ratio test the power is increased, and in particular the power of the GLS detrended variance ratio test with $d = 0.1$ is significantly higher than that of the Dickey-Fuller test although still below that of the GLS detrended Dickey-Fuller test. In all three panels of Figure 2, the GLS detrended variance ratio tests conform to the same monotonicity in d as in Figure 1.

One method to measure and compare the asymptotic local power of the GLS detrended variance ratio test with that of the ADF-GLS tests (whose asymptotic local power essentially coincides with the parametric power envelope) is to calculate the Pitman asymptotic relative efficiency (ARE) of the $d = 0.1$ and $d = 1$ tests relative to the ADF-GLS test. In the framework of asymptotic local power, this is done by comparing the values of c at which the tests obtain a specified power, e.g. power one-half following Elliott et al. (1996). The interpretation is that if the Pitman ARE of test A relative to test B is 1.25, then 25% more observations would be needed to obtain asymptotic local power of one-half using test A instead of test B. In the constant mean case, using 5% tests, the Pitman ARE of the VR-GLS tests with $d = 0.1$ and $d = 1$ relative to the ADF-GLS test are 1.34 and 2.97. In the linear trend case the corresponding AREs are 1.12 and 2.07. Thus, in the linear trend case only 12% more observations would be required for the VR-GLS test with $d = 0.1$ than for the ADF-GLS test to achieve asymptotic local power of one-half.

It is clear from the above asymptotic analysis that the nearly optimal ADF-GLS test is more powerful in a local asymptotic sense than the nonparametric GLS detrended variance ratio test with $d = 0.1$. However, these considerations are assuming that the tuning parameter, i.e. lag length, in the ADF-GLS test is chosen optimally, even though in any applied situation the correct/optimal lag length is unknown. It is well known that in more realistic scenarios where the lag length is unknown and must be chosen/estimated from the data, using e.g. an information criterion, and the serial correlation nuisance parameters must be estimated, the properties of the Dickey-Fuller type tests may deteriorate relative to the above "perfect knowledge" case. Indeed, they may be inferior to the nonparametric variance ratio test, which does not require the selection of any tuning parameters. This possibility is examined using simulations in the next section.

4 Finite Sample Performance

In this section simulation evidence is provided to evaluate the finite sample performance of the proposed nonparametric variance ratio test compared to the (nearly) optimal parametric GLS detrended ADF test of Elliott et al. (1996). The time series y_t is simulated according to the autoregressive model

$$y_t = \phi y_{t-1} + u_t, \quad t = 1, ..., T, \tag{16}$$

where $y_0 = 0$. Several different linear and nonlinear generating mechanisms are considered for u_t, in particular,

$$\text{AR} \quad : \quad u_t = a u_{t-1} + \varepsilon_t, \quad t = 1, ..., T, \tag{17}$$

$$\text{MA} \quad : \quad u_t = \varepsilon_t + a \varepsilon_{t-1}, \quad t = 1, ..., T, \tag{18}$$

$$\text{GARCH} \quad : \quad u_t = h_t^{1/2} \varepsilon_t, \quad h_t = 1 + a h_{t-1} + (0.95 - a) \, u_{t-1}^2, \quad t = 1, ..., T, \tag{19}$$

$$\text{Bilin} \quad : \quad u_t = a \varepsilon_{t-1} u_{t-1} + \varepsilon_t, \quad t = 1, ..., T, \tag{20}$$

$$\text{VCM} \quad : \quad u_t = \alpha_t u_{t-1} + \varepsilon_t, \quad \alpha_t = a \cos(2\pi t / T), \quad t = 1, ..., T, \tag{21}$$

$$\text{TAR} \quad : \quad u_t = \begin{cases} a u_{t-1} + \varepsilon_t & \text{if } |u_{t-1}| < 2, \\ -a u_{t-1} + \varepsilon_t & \text{if } |u_{t-1}| \geq 2, \end{cases} \quad t = 1, ..., T. \tag{22}$$

The models (17) and (18) are the traditional autoregressive (AR) and moving average (MA) models of order one for u_t with coefficient a. In (19), u_t is serially uncorrelated but exhibits time-varying variance, GARCH, of order (1,1). The parameterization is such that the sum of the two GARCH parameters (here denoted a and $(0.95 - a)$) equals 0.95 reflecting typical empirical values. Model (20) is the bilinear (Bilin) model with parameter a, (21) is a variable coefficient model (VCM) where the autoregressive coefficient is cyclical with amplitude determined by the parameter a, and (22) is the threshold autoregressive (TAR) model with parameter equal to a or $-a$ determined by the threshold condition. Finally, y_t is also simulated from the model

$$\text{STUR} : y_t = \alpha_t y_{t-1} + \varepsilon_t, \quad \alpha_t = a + (\phi - a) \, \alpha_{t-1} + 0.05 \eta_t, \quad t = 1, ..., T, \tag{23}$$

where $y_0 = 0$ and $E(\alpha_t) = a / (1 - \phi + a)$, which is a variant of the stochastic unit root model considered by, e.g., McCabe & Tremayne (1995) and Granger & Swanson (1997). Note that some of the nonlinear models considered here induce trends in y_t, see e.g. Granger & Anderson (1978), so only the case with correction for a linear trend is considered for the nonlinear models.

In all models, ε_t and η_t are *i.i.d.* standard normal and independent. The sample sizes considered are $T = 100$ and $T = 500$, and $20,000$ Monte Carlo replications are used in the simulations. Throughout, a 5% nominal significance level is employed. All calculations were made in Ox 3.4, see Doornik (2001).

In all the simulations, comparisons are made to the well known ADF test and to the ADF-GLS test of Elliott et al. (1996). To make the tests comparable, the lag augmentations (say k) in the ADF and ADF-GLS regressions are chosen using the data dependent modified Akaike information criterion (MAIC) of Ng & Perron (2001) who show that this criterion "dominates all other criteria from both theoretical and numerical perspectives." In particular, the lag augmentation was chosen to optimize the MAIC with $k_{\min} = 0$ and $k_{\max} = [12(T/100)^{1/4}]$ as in Ng & Perron (2001). Note that, in the simulations, this upper bound binds rarely for the small sample size ($T = 100$) and never for the larger sample size ($T = 500$). Also note that the

ADF-GLS test favors small initial conditions, see Müller & Elliott (2003), so in that sense the zero initial condition poses the greatest challenge for the proposed nonparametric test when compared to the ADF-GLS test.

Tables 3 and 4 about here

Tables 3 and 4 present the simulated size and size-adjusted rejection frequencies with $T = 100$ for the constant mean and the linear trend cases, respectively, under the simple autoregressive and moving average models (17) and (18). The results are reported for the variance ratio statistic with $d = 0.1$ (denoted $\rho(0.1)$), the corresponding GLS detrended variance ratio statistic (denoted $\rho(\bar{c}, 0.1)$), the Breitung (2002) test (BT), and the ADF and ADF-GLS tests using the MAIC to select lag augmentation. For each statistic, entries in the rows marked $\phi = 1.00$ are the rejection frequencies under the unit root null hypothesis, i.e. the size of the test, and all other entries are size-adjusted finite sample rejection frequencies.[7] In both tables, the BT test appears dominated by either $\rho(\bar{c}, 0.1)$ or ADF-GLS or both, in terms of size as well as size-adjusted power, so the focus will be on the comparison of the $\rho(\bar{c}, 0.1)$ and ADF-GLS tests.

The results of Table 3 for the constant mean case show that the variance ratio test has some size distortion in the presence of a negative moving average or autoregressive root. On the other hand, the ADF and ADF-GLS tests handle the size issue very well, and have sizes very close to the nominal level for all the models considered in this table. With respect to the (size-adjusted) finite sample power of the tests, the variance ratio tests, and especially the GLS detrended version, are clearly superior to the Dickey-Fuller type tests. Thus, the size control of the ADF-GLS tests, which results from the application of the MAIC lag selection criterion, comes at the price of a great decrease in power, at least for this sample size, $T = 100$. Specifically, for $\phi = 0.90$, the finite sample rejection frequencies of the GLS detrended variance ratio test are very similar to the ones of the ADF-GLS test, but for $\phi < 0.9$ the GLS detrended variance ratio test clearly outperforms the ADF-GLS test. For the latter range of ϕ-values, the finite sample power of the GLS detrended variance ratio test is mostly around 20% higher than that of the ADF-GLS test.

In Table 4, presenting the results for the linear trend case, it is clear that the close proximity of the asymptotic local power of the $\rho(\bar{c}, 0.1)$ test in this case to the parametric power envelope carries over to the simulation results. In particular, the results from Table 3 are reinforced here: the size distortion remains in the presence of negative autoregressive or moving average roots, but the finite sample power of the GLS detrended variance ratio test is now higher than that of

[7]The unadjusted rejection frequencies are not shown here for reasons of space, but are available from the author upon request.

the ADF-GLS test throughout the table. In particular, when there is a positive autoregressive root, the $\rho(\bar{c}, 0.1)$ test is, in some cases, more than twice as powerful as the ADF-GLS test. The results presented in Tables 3 and 4 thus demonstrate that, at the cost of some size distortion in specific cases, significant power gains may be obtained for relevant sample sizes by considering the proposed variance ratio test.

Tables 5 and 6 about here

In Tables 5 and 6, laid out as the previous two tables, the simulated size and size-adjusted rejection frequencies for sample size $T = 500$ are reported under the same models as in Tables 3 and 4. The results for the constant mean case in Table 5 show that the size distortion evident in the smaller sample size has vanished, and all the tests now have reasonable size properties. The power of the ADF-GLS test has increased dramatically relative to the variance ratio test in this larger sample size, presumably due to better lag augmentation selection, and now dominates that of the variance ratio tests for small deviations from the null. In particular, the ADF-GLS test now has somewhat higher finite sample power against $\phi = 0.98$, whereas for $\phi = 0.96$ the GLS detrended variance ratio test has nearly the same power as the ADF-GLS test. For $\phi < 0.96$, both the $\rho(\bar{c}, 0.1)$ and ADF-GLS tests reject in very nearly all replications.

The results in Table 6 for the linear trend case with $T = 500$ show that the variance ratio test remains very competitive in the presence of a linear trend, even with the larger sample size and therefore better lag augmentation selection by the ADF-GLS test. The variance ratio tests are much less over-sized when $T = 500$ compared to $T = 100$ in Table 4, and size distortion is now only significant for the largest negative moving average root. With respect to finite sample power, the rejection frequencies of the GLS detrended variance ratio test and the ADF-GLS tests are extremely similar for all the alternatives shown in the table.

Tables 7 and 8 about here

Tables 7 and 8 present simulation results for the models (19)-(23). Since some of the nonlinear models induce trends in the observed time series, only the linear trend case is considered here. The results for the smaller sample size, $T = 100$, in Table 7 emphasize the usefulness of the nonparametric variance ratio test. All tests have excellent size properties, but overall the variance ratio test is clearly superior in terms of finite sample power. The results for the variance ratio test under the GARCH model (19) are similar to the results for the *i.i.d.* error case in Table 4, whereas the ADF-GLS test has somewhat reduced power in the presence of GARCH. In the bilinear model (20), GLS detrending appears to have no effect on power. Under this model, the finite sample power of the variance ratio tests (with or without GLS detrending) is similar to that of the Dickey-Fuller type tests for $\phi \geq 0.8$, but for moderate to large deviations from the null the variance ratio tests are clearly superior, i.e. when $\phi < 0.8$. For the TAR model (22), the situation is very similar to that for the GARCH model or the *i.i.d.* error case, where the finite sample powers of the $\rho(\bar{c}, 0.1)$ and ADF-GLS tests are similar for $\phi = 0.9$ but the GLS detrended variance ratio test is superior for $\phi < 0.9$. Finally, in the VCM and STUR models (21) and (23), the GLS detrended variance ratio test has higher finite sample power than the ADF and ADF-GLS tests for all alternatives considered. In the STUR model, GLS detrending also appears to have no significant effect on the power of the variance ratio test whereas the ADF-GLS test has lower power than the ADF test.

For the larger sample size, $T = 500$, the results in Table 8 confirm the previous results for the GARCH, VCM, and TAR models: that the GLS detrended variance ratio test and the ADF-GLS test both have excellent size properties, that the two tests have similar size-adjusted power for small deviations from the null hypothesis, and that the GLS detrended variance ratio test has slightly higher power when moving further away from the null. For the bilinear model, size is again well controlled by all the tests, but the GLS detrended tests (both $\rho(\bar{c}, 0.1)$ and ADF-GLS) have very low power. For the STUR model, all tests (including the ADF and ADF-GLS tests) exhibit severe size distortions when $a = 0.1$ but not when $a = 0.5$. In terms of power, the variance ratio test is clearly superior when $a = 0.1$ and the tests are similar for $a = 0.5$.

In general, it is apparent from the simulations that the nonparametric variance ratio test is useful and that non-trivial power gains may be obtained relative to the ADF-GLS test of Elliott et al. (1996) in sample sizes that are relevant for empirical research. Thus, even though the ADF-GLS test has superior asymptotic local power properties, as documented in section 3 above, the need to select a tuning parameter (lag augmentation) and estimate nuisance parameters (serial correlation) reduces the power of Dickey-Fuller type tests in more realistic settings. Although the power loss of the ADF-GLS test relative to the power envelope is somewhat alleviated in larger samples, where the MAIC comes closer to selecting the optimal (but unknown to the researcher) lag augmentation, it remains an issue and the variance ratio test is still able to achieve similar, if not higher, power without the need to select any tuning parameters.

5 Concluding Remarks

The family of nonparametric variance ratio tests of the unit root hypothesis presented here has the property that the tests are free of tuning parameters. That is, there are no parameters involved in calculating the test which are not reflected in the asymptotic distribution. The tests are constructed as a ratio of the sample variance of the observed series and that of a fractional partial sum of the series, possibly applying GLS detrending to handle deterministic terms, and the family is thus indexed by the parameter d which determines the order of the fractional partial summation. However, unlike the choice of tuning parameters, e.g., lag length in augmented Dickey-Fuller regressions or bandwidth in Phillips-Perron type tests, each member of the family with $d > 0$ is consistent and its asymptotic distribution depends on d, thus reflecting the parameter chosen to implement the test. Consequently, using local power asymptotics, the member of the family which has the highest (within the family) power against relevant (near-integrated) alternatives was derived, and in particular, it was shown that when d is small the asymptotic local power of the proposed nonparametric test is relatively close to the parametric power envelope, especially in the case with a linear time trend.

Furthermore, simulation evidence demonstrates that the proposed test has good finite sample properties in the presence of both linear and nonlinear short-run dynamics. Indeed, the finite sample size-adjusted power of the nonparametric tuning parameter free variance ratio test is similar, and sometimes superior, to that of the GLS detrended ADF test when the two are compared on an even footing by applying the MAIC to select the lag augmentation of the Dickey-Fuller regressions.

Appendix: Proofs

Proof of Theorem 1. In the case with no deterministic terms, the result follows immediately by the Continuous Mapping Theorem since Assumption 1 implies (4) and (5). In the presence of deterministic terms, recall that $\hat{y}_t = z_t - (\hat{\alpha} - \alpha)' \delta_t$, where z_t is generated as y_t in (1). From (4), the convergence

$$y_T(s) = T^{-1/2} z_{[Ts]} \Rightarrow \sigma_y W_0(s)$$

holds. Now define $N(T) = diag(1, T^{-1})$ and write

$$T^{-1/2} (\hat{\alpha} - \alpha)' \delta_{[Ts]} = \left(T^{-1} \sum_{s=1}^{T} T^{-1/2} z_s \delta_s' N(T) \right) \left(T^{-1} \sum_{s=1}^{T} N(T) \delta_s \delta_s' N(T) \right)^{-1} N(T) \delta_{[Ts]},$$

where

$$\begin{aligned} T^{-1/2} (\hat{\alpha} - \alpha)' N(T)^{-1} &= \left(T^{-1} \sum_{s=1}^{T} T^{-1/2} z_s \delta_s' N(T) \right) \left(T^{-1} \sum_{s=1}^{T} N(T) \delta_s \delta_s' N(T) \right)^{-1} \quad (24) \\ &= \left(T^{-1} \sum_{s=1}^{T} T^{-1/2} z_s D(s/T) \right) \left(T^{-1} \sum_{s=1}^{T} D(s/T) D(s/T)' \right)^{-1} \\ &\Rightarrow \sigma_y \left(\int_0^1 W_0(s) D(s)' ds \right) \left(\int_0^1 D(s) D(s)' ds \right)^{-1} \end{aligned}$$

by application of (4) and the Continuous Mapping Theorem, and

$$N(T)\delta_{[Ts]} = D([Ts]/T) \to D(s) \text{ as } T \to \infty. \tag{25}$$

It thus follows that

$$\hat{y}_T(s) = T^{-1/2}\hat{y}_{[Ts]} \Rightarrow \sigma_y B_j(s), \quad j = 0, 1, 2. \tag{26}$$

Next, for $\tilde{y}_T(s) = T^{-d}\Delta_+^{-d}\hat{y}_T(s) = T^{-1/2-d}\sum_{j=0}^{[Ts]-1}\pi_j(d)\,\hat{y}_{[Ts]-j} = T^{-1/2-d}\sum_{j=1}^{[Ts]}\pi_{[Ts]-j}(d)\,\hat{y}_j$, where $\hat{y}_t = z_t - (\hat{\alpha}-\alpha)'\delta_t$ and $\pi_j(d) = \Gamma(j+d)/(\Gamma(d)\Gamma(j+1))$, the convergence

$$\tilde{y}_T(s) = T^{-1/2-d}\sum_{j=1}^{[Ts]}\pi_{[Ts]-j}(d)\,z_j \Rightarrow \sigma_y W_d(s)$$

holds from (5). For the remaining term,

$$T^{-1/2-d}\sum_{j=1}^{[Ts]}\pi_{[Ts]-j}(d)\,(\hat{\alpha}-\alpha)'\delta_j = \left(T^{-1/2}(\hat{\alpha}-\alpha)'N(T)^{-1}\right)\left(T^{-d}\sum_{j=1}^{[Ts]}\pi_{[Ts]-j}(d)\,N(T)\delta_j\right),$$

the first term converges by (24) and the last term is deterministic and satisfies the convergence

$$\begin{aligned} T^{-d}\sum_{j=1}^{[Ts]}\pi_{[Ts]-j}(d)\,N(T)\delta_j &= T^{-d}\sum_{j=1}^{[Ts]}\pi_{[Ts]-j}(d)\,D(j/T) \\ &= T^{-d}\sum_{j=1}^{[Ts]}\frac{([Ts]-j)^{d-1}}{\Gamma(d)}D(j/T) + o(1) \\ &= T^{-1}\sum_{j=1}^{[Ts]}\frac{\left(\frac{[Ts]}{T}-\frac{j}{T}\right)^{d-1}}{\Gamma(d)}D(j/T) + o(1) \\ &\to \int_0^s \frac{(s-r)^{d-1}}{\Gamma(d)}D(r)\,dr \text{ as } T \to \infty. \end{aligned} \tag{27}$$

Hence, it follows that

$$\tilde{y}_T(s) \Rightarrow \sigma_y \tilde{B}_{j,d}(s), \quad j = 0, 1, 2,$$

which proves the desired result. ■

Proof of Theorem 2. The test has asymptotic size γ by Theorem 1 and the definition of $CV_{j,\gamma}(d)$. Consistency is proved in the case $\delta_t = 0$; the remaining cases follow similarly. Under the alternative hypothesis H_1 in (2) and Assumption 2,

$$T^{-1}\sum_{t=1}^{T} y_t^2 \xrightarrow{P} \omega_y^2 = Ey_t^2,$$

$$T^{-2d}\sum_{t=1}^{T}\tilde{y}_t^2 \Rightarrow \sigma_y^2\int_0^1 W_{d-1}(s)^2\,ds, \quad d > 1/2,$$

where the first line holds by the law of large numbers for stationary ergodic time series since $\omega_y^2 = Ey_t^2 < \infty$, e.g. White (1984, p. 42), and the second line holds by (5) and the Continuous Mapping Theorem. If $d < 1/2$, define the stationary time series $\breve{y}_t = \Delta^{-d} y_t = \sum_{j=0}^{\infty} \pi_j(d) y_{t-j}$ (with no truncation), for which the law of large numbers for stationary ergodic time series implies that

$$T^{-1} \sum_{t=1}^{T} \breve{y}_t^2 \xrightarrow{P} \breve{\omega}_y^2 = E\breve{y}_t^2, \quad d < 1/2,$$

since $\breve{\omega}_y^2 = E\breve{y}_t^2 < \infty$ when $d < 1/2$. Under H$_1$, y_t can be written as $y_t = \sum_{j=0}^{\infty} \tau_j \varepsilon_{t-j}$, where the τ_j are functions of ϕ and $\psi_j, j = 0, 1, \ldots$. Thus, $\breve{y}_t$ and $\tilde{y}_t$ can be written as $\breve{y}_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}$ and $\tilde{y}_t = \sum_{j=0}^{\infty} \tilde{\theta}_{t,j} \varepsilon_{t-j}$, where the coefficients $\theta_j = \sum_{k=0}^{j} \pi_k(d) \tau_{j-k}$ and $\tilde{\theta}_{t,j} = \sum_{k=0}^{\min(j,t-1)} \pi_k(d) \tau_{j-k}$ such that $\theta_j = \tilde{\theta}_{t,j}$ for $j \le t-1$. The coefficients satisfy $\theta_j \le C j^{d-1}$ (because $\pi_j(d)$ satisfies the same inequality) and $\tilde{\theta}_{t,j} \le C t^{d-1}$. The process $\tilde{y}_t = \sum_{j=0}^{\infty} \tilde{\theta}_{t,j} \varepsilon_{t-j}$ is asymptotically stationary in the sense that

$$\begin{aligned} Var\left(\tilde{y}_t - \breve{y}_t\right) &= \sigma_\varepsilon^2 \sum_{j=t}^{\infty} (\theta_j - \tilde{\theta}_{t,j})^2 \\ &\le C \sum_{j=t}^{\infty} j^{2d-2} \\ &\le C t^{2d-1} \to 0 \end{aligned}$$

as $t \to \infty$. It therefore follows that also $T^{-1} \sum_{t=1}^{T} \tilde{y}_t^2 \xrightarrow{P} E\breve{y}_t^2$ if $T^{-1} \sum_{t=1}^{T} \left(\tilde{y}_t - \breve{y}_t\right)^2 \xrightarrow{P} 0$. But this is a consequence of

$$\begin{aligned} E\left| T^{-1} \sum_{t=1}^{T} \left(\tilde{y}_t - \breve{y}_t\right)^2 \right| &= T^{-1} \sum_{t=1}^{T} Var\left(\tilde{y}_t - \breve{y}_t\right) \\ &\le C T^{2d-1} \to 0. \end{aligned}$$

Hence,

$$\begin{aligned} T^{-1} \rho\left(d\right) &\Rightarrow \frac{\omega_y^2}{\sigma_y^2} \left(\int_0^1 W_{d-1}\left(s\right)^2 ds \right)^{-1}, \quad d > 1/2, \\ T^{-2d} \rho\left(d\right) &\xrightarrow{P} \frac{\omega_y^2}{\breve{\omega}_y^2}, \quad d < 1/2, \end{aligned}$$

i.e. $\rho\left(d\right) \in O_P\left(T^{\min(1,2d)}\right)$ under H$_1$ and thus diverges in probability to $+\infty$ when $d > 0$ (note that $\rho\left(d\right) > 0$ by construction). Consistency against the alternative H$_1$ follows. ■

Proof of Theorem 3. Recall that $\hat{y}_t = z_t - (\hat{\alpha} - \alpha)' \delta_t$, where z_t is generated by (13). Under assumptions implied by Assumption 1, Chan & Wei (1987) and Phillips (1987*b*) proved that

$$T^{-1/2} z_{[Ts]} \Rightarrow \sigma_y J_{0,c}\left(s\right),$$

where $J_{0,c}(s) = W(s) - c\int_0^s e^{-(s-r)c} W(r)dr$, $J_{0,c}(0) = 0$, is the Ornstein-Uhlenbeck process which is sometimes also written as $J_{0,c}(s) = \int_0^s e^{-c(s-r)} dW_0(r)$. As in (24) it follows that

$$T^{-1/2}(\hat{\alpha} - \alpha)' N(T)^{-1} \Rightarrow \sigma_y \left(\int_0^1 J_{0,c}(r) D(r)' dr \right) \left(\int_0^1 D(r) D(r)' dr \right)^{-1}, \tag{28}$$

which combined with (25) implies that

$$\hat{y}_T(s) = T^{-1/2} \hat{y}_{[Ts]} \Rightarrow \sigma_y J_{j,c}(s), \tag{29}$$

where $J_{j,c}(s)$ is the demeaned ($j = 1$) respectively demeaned and detrended ($j = 2$) Ornstein-Uhlenbeck processes defined in Theorem 3.

As in the proof of Theorem 1, define $\tilde{y}_T(s) = T^{-d} \Delta_+^{-d} \hat{y}_T(s) = T^{-1/2-d} \sum_{j=1}^{[Ts]} \pi_{[Ts]-j}(d) \hat{y}_j$. First suppose there are no deterministic terms ($j = 0$), in which case $\hat{y}_t = y_t$. Since $e^{-c/T} = 1 - c/T + O(T^{-2})$ it follows that $y_t = \sum_{j=1}^t e^{-(t-j)c/T} u_j$ (where a negligible remainder term has been left out), and using summation by parts the representation

$$y_t = \sum_{j=1}^t e^{-(t-j)c/T} u_j = \sum_{j=1}^t u_j + \sum_{j=1}^{t-1} \left(e^{-(t-j)c/T} - e^{-(t-j-1)c/T} \right) \sum_{k=1}^j u_k$$

is obtained. By the Mean Value Theorem, for $0 \le x \le 1$,

$$\begin{aligned} e^{-(t-j-1)c/T} &= e^{-(t-j)c/T} + \frac{c}{T} e^{-(t-j)c/T} + \frac{1}{2}\left(\frac{c}{T}\right)^2 e^{-(t-j-x)c/T} \\ &= e^{-(t-j)c/T} + \frac{c}{T} e^{-(t-j)c/T} \left(1 + O\left(T^{-1}\right)\right), \end{aligned}$$

which implies that

$$y_t = \sum_{j=1}^t u_j - \frac{c}{T} \sum_{j=1}^{t-1} e^{-(t-j)c/T} \sum_{k=1}^j u_k \left(1 + O_P\left(T^{-1}\right)\right).$$

The $O_P(T^{-1})$ term is uniform in t and is therefore ignored in the following.

Now, still in the case with no deterministic terms, $\tilde{y}_T(s) = T^{-1/2-d} \sum_{j=1}^{[Ts]} \pi_{[Ts]-j}(d) y_j$ is

$$\begin{aligned} \tilde{y}_T(s) &= T^{-1/2-d} \sum_{j=1}^{[Ts]} \pi_{[Ts]-j}(d) \sum_{k=1}^j u_k - T^{-1/2-d} \sum_{j=2}^{[Ts]} \pi_{[Ts]-j}(d) \frac{c}{T} \sum_{k=1}^{j-1} e^{-(j-k)c/T} \sum_{l=1}^k u_l \\ &= \tilde{y}_{1T}(s) - \tilde{y}_{2T}(s), \end{aligned}$$

where $\tilde{y}_{1T}(s) = T^{-1/2-d} \sum_{j=1}^{[Ts]} \pi_{[Ts]-j}(d) \sum_{k=1}^j u_k \Rightarrow \sigma_y W_d(s)$ by (5). By interchanging the order of the summations the second term can be rearranged as

$$\begin{aligned}
\sum_{j=2}^{t}\pi_{t-j}(d)\frac{c}{T}\sum_{k=1}^{j-1}e^{-(j-k)c/T}\sum_{l=1}^{k}u_l &= \frac{c}{T}\sum_{k=0}^{t-2}e^{-(k+1)c/T}\sum_{j=2}^{t-k}\pi_{t-k-j}(d)\sum_{k=1}^{j-1}u_k \\
&= \frac{c}{T}\sum_{i=2}^{t}e^{-(t-i+1)c/T}\sum_{j=2}^{i}\pi_{i-j}(d)\sum_{k=1}^{j-1}u_k \\
&= \frac{c}{T}\sum_{k=1}^{t-1}e^{-(t-k)c/T}\sum_{j=2}^{k+1}\pi_{k+1-j}(d)\sum_{k=1}^{j-1}u_k \\
&= \frac{c}{T}\sum_{k=1}^{t-1}e^{-(t-k)c/T}\sum_{j=1}^{k}\pi_{k-j}(d)\sum_{k=1}^{j}u_k,
\end{aligned}$$

and thus $\tilde{y}_{2T}(s)$ is

$$\begin{aligned}
\tilde{y}_{2T}(s) &= \frac{c}{T}\sum_{k=1}^{[Ts]-1}e^{-([Ts]-k)c/T}\tilde{y}_{1T}(k/T) \\
&= c\sum_{k=1}^{[Ts]-1}e^{-([Ts]-k)c/T}\int_{k/T}^{(k+1)/T}\tilde{y}_{1T}(r)dr \\
&= c\sum_{k=1}^{[Ts]-1}\int_{k/T}^{(k+1)/T}e^{-([Ts]-[Tr])c/T}\tilde{y}_{1T}(r)dr \\
&= c\int_{1/T}^{[Ts]/T}e^{-([Ts]-[Tr])c/T}\tilde{y}_{1T}(r)dr \\
&= c\int_{0}^{s}e^{-(s-r)c}\tilde{y}_{1T}(r)dr + R_T(s).
\end{aligned}$$

By application of (5) and the Continuous Mapping Theorem (since the functional $\int_0^s e^{-(s-r)c}f(r)dr$ is a continuous mapping from $D[0,1]$ to $D[0,1]$) it follows that

$$c\int_0^s e^{-(s-r)c}\tilde{y}_{1T}(r)dr \Rightarrow \sigma_y c\int_0^s e^{-(s-r)c}W_d(r)\,dr.$$

Thus, it only remains to show that the approximation error $R_T(s)$ is negligible uniformly in $s\in[0,1]$. Write $R_T(s)$ as

$$\begin{aligned}
R_T(s) &= c\int_0^{1/T}e^{-([Ts]-[Tr])c/T}\tilde{y}_{1T}(r)dr \\
&\quad +c\int_{[Ts]/T}^{s}e^{-([Ts]-[Tr])c/T}\tilde{y}_{1T}(r)dr \\
&\quad +c\int_0^s\left(e^{-([Ts]-[Tr])c/T}-e^{-(s-r)c}\right)\tilde{y}_{1T}(r)dr.
\end{aligned}$$

It is easily seen that $\int_0^{1/T} e^{-([Ts]-[Tr])c/T}\tilde{y}_{1T}(r)dr = 0$ because $\tilde{y}_{1T}(r) = 0$ for $r < 1/T$. For the next term we have that

$$\begin{aligned}
\sup_{0\leq s\leq 1} c\int_{[Ts]/T}^{s} e^{-([Ts]-[Tr])c/T}\tilde{y}_{1T}(r)dr &= \sup_{0\leq s\leq 1} C\int_{[Ts]/T}^{s} \tilde{y}_{1T}(r)dr \\
&\leq \sup_{0\leq s\leq 1} C\left(\frac{Ts-[Ts]}{T}\right)\left|\tilde{y}_{1T}\left(\frac{[Ts]}{T}\right)\right| \\
&\leq \frac{C}{T}\sup_{0\leq s\leq 1}\left|\tilde{y}_{1T}(s)\right|,
\end{aligned}$$

which is $O_P\left(T^{-1}\right)$ since $\sup_{0\leq s\leq 1}\left|\tilde{y}_{1T}(s)\right| \Rightarrow \sigma_y \sup_{0\leq s\leq 1}\left|W_d(s)\right|$ by (5) and the Continuous Mapping Theorem. The last term of $R_T(s)$ is bounded by

$$\begin{aligned}
& \sup_{0\leq s\leq 1} c\int_0^s \left(e^{-([Ts]-[Tr])c/T} - e^{-(s-r)c}\right)\tilde{y}_{1T}(r)dr \\
\leq\ & \sup_{0\leq s\leq 1} c\int_0^s \left(e^{-([Ts]/T-[Tr]/T)c} - e^{-([Ts]/T-r)c}\right)\tilde{y}_{1T}(r)dr \\
& + \sup_{0\leq s\leq 1} c\int_0^s \left(e^{-([Ts]/T-r)c} - e^{-(s-r)c}\right)\tilde{y}_{1T}(r)dr \\
\leq\ & \sup_{0\leq r\leq 1} C\left|\left(e^{([Tr]/T)c} - e^{rc}\right)\tilde{y}_{1T}(r)\right| + \sup_{0\leq r\leq 1}\frac{C}{T}e^{rc}\left|\tilde{y}_{1T}(r)\right| \\
\leq\ & \frac{C}{T}\sup_{0\leq r\leq 1}\left|\tilde{y}_{1T}(r)\right|,
\end{aligned}$$

which is $O_P(T^{-1})$ by (5) and the Continuous Mapping Theorem. Hence $\sup_{0\leq s\leq 1} R_T(s) \xrightarrow{P} 0$.

Finally, the result for the fractional partial sum of the detrended process, i.e. $\tilde{y}_T(s) = T^{-1/2-d}\sum_{j=1}^{[Ts]}\pi_{[Ts]-j}(d)\,\hat{y}_j$, can easily be proven. Following the steps in the proof of Theorem 1 above and using $\hat{y}_t = z_t - (\hat{\alpha}-\alpha)'\,\delta_t$, where z_t is generated by (13), it is seen that

$$T^{-1/2-d}\sum_{j=1}^{[Ts]}\pi_{[Ts]-j}(d)\,z_j \Rightarrow \sigma_y\tilde{J}_{0,c,d}(s),$$

which combined with (27) and (28) yields

$$\tilde{y}_T(s) \Rightarrow \sigma_y\tilde{J}_{0,c,d}(s) - \sigma_y\left(\int_0^1 J_{0,c}(r)\,D(r)'\,dr\right)\left(\int_0^1 D(r)\,D(r)'\,dr\right)^{-1}\int_0^s \frac{(s-r)^{d-1}}{\Gamma(d)}D(r)\,dr$$

in the same way as above. ■

Proof of Theorem 4. From Elliott et al. (1996, pp. 834-835) it follows that $T^{-1/2}\hat{y}_{\bar{c},[Ts]} \Rightarrow \sigma_y J_{0,c}(s)$ if $j=1$ and $T^{-1/2}\hat{y}_{\bar{c},[Ts]} \Rightarrow \sigma_y V_{\bar{c},c}(s)$ if $j=2$. It therefore also follows immediately that $T^{-1/2-d}\tilde{y}_{\bar{c},[Ts]} \Rightarrow \sigma_y\tilde{J}_{0,c,d}(s)$ for $j=1$ when $\tilde{y}_{\bar{c},t}$ is based on $\hat{y}_{\bar{c},t}$.

It only remains to be shown that $T^{-1/2-d}\tilde{y}_{\bar{c},[Ts]} \Rightarrow \sigma_y \tilde{V}_{\bar{c},c,d}(s)$ when $j=2$ and $\tilde{y}_{\bar{c},t}$ is based on

$$\hat{y}_{\bar{c},t} = z_t - (\tilde{\alpha}_0 - \alpha_0) - (\tilde{\alpha}_1 - \alpha_1)t \tag{30}$$

which is GLS detrended as in (14). From Elliott et al. (1996, p. 835) it is known that $\tilde{\alpha}_0 \in O_P(1)$ and

$$\sqrt{T}(\tilde{\alpha}_1 - \alpha_1) \Rightarrow \sigma_y \frac{(1+\bar{c})}{1+\bar{c}+\bar{c}^2/3} J_{0,c}(1) + \sigma_y \frac{\bar{c}^2}{1+\bar{c}+\bar{c}^2/3}\int_0^1 rJ_{0,c}(r)\,dr = \sigma_y b_1. \tag{31}$$

Following the steps in the proofs of Theorems 1 and 3, write

$$T^{-1/2-d}\tilde{y}_{\bar{c},[Ts]} = T^{-1/2-d}\sum_{j=1}^{[Ts]} \pi_{[Ts]-j}(d)\,\hat{y}_{\bar{c},j}$$

and use (30) to obtain the representation

$$\begin{aligned} T^{-1/2-d}\tilde{y}_{\bar{c},[Ts]} &= T^{-1/2-d}\sum_{j=1}^{[Ts]} \pi_{[Ts]-j}(d)\,z_j \\ &\quad -(\tilde{\alpha}_0 - \alpha_0)T^{-1/2-d}\sum_{j=1}^{[Ts]} \pi_{[Ts]-j}(d) \\ &\quad -\sqrt{T}(\tilde{\alpha}_1 - \alpha_1)T^{-d}\sum_{j=1}^{[Ts]} \pi_{[Ts]-j}(d)\frac{[Ts]}{T}. \end{aligned}$$

For the first term it clearly holds that

$$T^{-1/2-d}\sum_{j=1}^{[Ts]} \pi_{[Ts]-j}(d)\,z_j \Rightarrow \sigma_y \tilde{J}_{0,c,d}(s).$$

Next,

$$\begin{aligned} \sup_{0\le s\le 1} T^{-1/2-d}\sum_{j=1}^{[Ts]} \pi_{[Ts]-j}(d) &\le C\sup_{0\le s\le 1} T^{-1/2-d}\sum_{j=1}^{[Ts]-1} ([Ts]-j)^{d-1} \\ &\le C\sup_{0\le s\le 1} T^{-1/2-d}\sum_{j=1}^{[Ts]-1} j^{d-1} \\ &\le CT^{-1/2}, \end{aligned}$$

and since $\tilde{\alpha}_0 \in O_P(1)$ this implies that the second term is $O_P(T^{-1/2})$. For the third term use (31) and

$$T^{-d}\sum_{j=1}^{[Ts]} \pi_{[Ts]-j}(d)\frac{[Ts]}{T} \to \frac{s^{d+1}}{\Gamma(d+2)},$$

such that $T^{-1/2-d}\tilde{y}_{\bar{c},[Ts]} \Rightarrow \sigma_y \tilde{J}_{0,c,d}(s) - \sigma_y b_1 \frac{s^{d+1}}{\Gamma(d+2)} = \sigma_y \tilde{V}_{\bar{c},c,d}(s)$ for $j=2$. ■

References

Agiakloglou, C. & Newbold, P. (1996), 'The balance between size and power in Dickey-Fuller tests with data-dependent rules for the choice of truncation lag', *Economics Letters* **52**, 229–234.

Akonom, J. & Gourieroux, C. (1987), 'A functional central limit theorem for fractional processes', *Technical Report 8801, CEPREMAP, Paris* .

Baillie, R. T. (1996), 'Long memory processes and fractional integration in econometrics', *Journal of Econometrics* **73**, 5–59.

Breitung, J. (2002), 'Nonparametric tests for unit roots and cointegration', *Journal of Econometrics* **108**, 342–363.

Breitung, J. & Taylor, A. M. R. (2003), 'Corrigendum to "Nonparametric tests for unit roots and cointegration" [J. Econom. 108 (2002) 343–363]', *Journal of Econometrics* **117**, 401–404.

Brockwell, P. J. & Marquardt, T. (2005), 'Lévy-driven and fractionally integrated ARMA processes with continuous time parameter', *Statistica Sinica* **15**, 477–494.

Buchmann, B. & Klüppelberg, C. (2006), 'Fractional integral equations and state space transforms', *Bernoulli* **12**, 431–456.

Chan, N. H. & Wei, C. Z. (1987), 'Asymptotic inference for nearly nonstationary AR(1) processes', *Annals of Statistics* **15**, 1050–1063.

Davidson, J. & de Jong, R. M. (2000), 'The functional central limit theorem and weak convergence to stochastic integrals II: Fractionally integrated processes', *Econometric Theory* **16**, 643–666.

Dickey, D. A. & Fuller, W. A. (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association* **74**, 427–431.

Dickey, D. A. & Fuller, W. A. (1981), 'Likelihood ratio statistics for autoregressive time series with a unit root', *Econometrica* **49**, 1057–1072.

Doornik, J. A. (2001), *Ox: An Object-Oriented Matrix Language*, 4th edn, Timberlake Consultants Press, London.

Elliott, G., Rothenberg, T. J. & Stock, J. H. (1996), 'Efficient tests for an autoregressive unit root', *Econometrica* **64**, 813–836.

Granger, C. W. J. & Anderson, A. P. (1978), *An Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Göttingen.

Granger, C. W. J. & Swanson, N. R. (1997), 'An introduction to stochastic unit-root processes', *Journal of Econometrics* **80**, 35–62.

Haldrup, N. & Jansson, M. (2006), Improving size and power in unit root testing, *in* T. C. Mills & K. Patterson, eds, 'Palgrave Handbook of Econometrics, Vol. I', Palgrave Macmillan, pp. 252–277.

Harvey, D. I., Leybourne, S. J. & Newbold, P. (2004), 'Tests for a break in level when the order of integration is unknown', *Oxford Bulletin of Economics and Statistics* **66**, 133–146.

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

Leybourne, S. J. & Newbold, P. (1999*a*), 'The behaviour of Dickey-Fuller and Phillips-Perron tests under the alternative hypothesis', *Econometrics Journal* **2**, 92–106.

Leybourne, S. J. & Newbold, P. (1999*b*), 'On the size properties of Phillips-Perron tests', *Journal of Time Series Analysis* **20**, 51–61.

Lo, A. W. & MacKinlay, A. C. (1988), 'Stock market prices do not follow random walks: Evidence from a simple specification test', *Review of Financial Studies* **1**, 41–66.

Marinucci, D. & Robinson, P. M. (1999), 'Alternative forms of fractional Brownian motion', *Journal of Statistical Planning and Inference* **80**, 111–122.

Marinucci, D. & Robinson, P. M. (2000), 'Weak convergence of multivariate fractional processes', *Stochastic Processes and their Applications* **86**, 103–120.

McCabe, B. M. P. & Tremayne, A. R. (1995), 'Testing a time series for difference stationarity', *Annals of Statistics* **23**, 1015–1028.

Miller, J. P. & Newbold, P. (1995), 'A generalized variance ratio test of ARIMA(p,1,q) model specification', *Journal of Time Series Analysis* **16**, 403–413.

Müller, U. K. (2007*a*), 'The impossibility of consistent discrimination between I(0) and I(1) processes', *forthcoming in Econometric Theory* .

Müller, U. K. (2007*b*), 'A theory of robust long-run variance estimation', *forthcoming in Journal of Econometrics* .

Müller, U. K. & Elliott, G. (2003), 'Tests for unit roots and the initial condition', *Econometrica* **71**, 1269–1286.

Ng, S. & Perron, P. (2001), 'Lag length selection and the construction of unit root tests with good size and power', *Econometrica* **69**, 1519–1554.

Park, J. Y. (1990), Testing for unit roots and cointegration by variable addition, *in* T. B. Fomby & G. F. Rhodes, eds, 'Advances in Econometrics Volume 8: Co-Integration, Spurious Regressions, and Unit Roots', Elsevier Science, Ltd., Oxford, pp. 107–133.

Park, J. Y. & Choi, B. (1988), 'A new approach to testing for a unit root', *CAE Working paper 88-23, Cornell University* .

Phillips, P. C. B. (1987*a*), 'Time series regression with a unit root', *Econometrica* **55**, 277–301.

Phillips, P. C. B. (1987*b*), 'Towards a unified asymptotic theory for autoregression', *Biometrika* **74**, 535–547.

Phillips, P. C. B. & Perron, P. (1988), 'Testing for a unit root in time series regression', *Biometrika* **75**, 335–346.

Phillips, P. C. B. & Solo, V. (1992), 'Asymptotics for linear processes', *Annals of Statistics* **20**, 971–1001.

Phillips, P. C. B. & Xiao, Z. (1998), 'A primer on unit root testing', *Journal of Economic Surveys* **12**, 423–469.

Robinson, P. M. (1994), 'Efficient tests of nonstationary hypotheses', *Journal of the American Statistical Association* **89**, 1420–1437.

Robinson, P. M. (2003), Long-memory time series, *in* P. M. Robinson, ed., 'Time Series With Long Memory', Oxford University Press, Oxford, pp. 4–32.

Shin, Y. & Schmidt, P. (1992), 'The KPSS stationarity test as a unit root test', *Economics Letters* **38**, 387–392.

Stock, J. H. (1994), Unit roots, structural breaks and trends, *in* R. F. Engle & D. L. McFadden, eds, 'Handbook of Econometrics, Vol. IV', North-Holland, Amsterdam, pp. 2843–2915.

Tanaka, K. (1999), 'The nonstationary fractional unit root', *Econometric Theory* **15**, 549–582.

Taylor, A. M. R. (2005), 'Variance ratio tests of the seasonal unit root hypothesis', *Journal of Econometrics* **124**, 33–54.

Vogelsang, T. J. (1998*a*), 'Testing for a shift in mean without having to estimate serial-correlation parameters', *Journal of Business and Economic Statistics* **16**, 73–80.

Vogelsang, T. J. (1998*b*), 'Trend function hypothesis testing in the presence of serial correlation', *Econometrica* **66**, 123–148.

White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, San Diego.

Table 1: Critical values $CV_{j,\alpha}(d)$ of the variance ratio test (10)

			d				
Deterministics	γ	T	0.10	0.25	0.50	0.75	1.00
$j=0:\delta_t=0$	0.10	100	1.5388	2.7808	6.7604	15.3013	33.6340
		500	1.5395	2.7677	6.7019	15.0938	33.1256
	0.05	100	1.6170	3.1344	8.4524	20.9946	48.7309
		500	1.6230	3.1434	8.4422	20.7033	49.4214
	0.01	100	1.7623	3.8953	12.5536	36.1244	98.8180
		500	1.7705	3.9230	12.9280	38.5938	106.6029
$j=1:\delta_t=1$	0.10	100	1.7492	3.8341	12.2919	32.0381	70.0327
		500	1.7609	3.8718	12.3902	32.3151	70.4341
	0.05	100	1.8110	4.1758	14.4893	41.5508	100.4194
		500	1.8215	4.1995	14.4260	40.8306	97.8254
	0.01	100	1.9205	4.8245	19.3901	64.8151	186.3318
		500	1.9312	4.8477	18.9774	62.1184	173.0501
$j=2:\delta_t=[1,t]'$	0.10	100	1.9064	4.7761	19.5001	69.8565	227.3997
		500	1.9245	4.8319	19.6838	70.3534	227.9921
	0.05	100	1.9563	5.0871	22.0730	83.9390	291.3214
		500	1.9785	5.1681	22.3296	84.7881	289.6125
	0.01	100	2.0441	5.6818	27.7271	119.0518	455.0747
		500	2.0780	5.8320	28.2634	118.8001	446.2204

Note: The critical values are simulated based on 20,000 Monte Carlo replications. The test rejects when the test statistic is larger than the critical values in this table.

Table 2: Values of $\bar{c}$ and $CV_{j,\alpha}(\bar{c},d)$ of the VR-GLS test (15)

			d				
Deterministics	γ	T	0.10	0.25	0.50	0.75	1.00
Panel A: Values of $\bar{c}=c$ that yield asymptotic local power of 50%							
$j=1:\delta_t=1$	0.05	500	9.4	10.6	12.8	16.3	20.8
$j=2:\delta_t=[1,t]'$	0.05	500	15.1	16.1	18.7	22.5	28.0
Panel B: Critical values $CV_{j,\gamma}(\bar{c},d)$							
$j=2:\delta_t=[1,t]'$	0.10	100	1.7988	4.0537	13.7148	41.7500	122.8662
		500	1.7737	3.8567	11.9207	31.8438	78.2120
	0.05	100	1.8512	4.3743	15.9824	52.2843	161.4733
		500	1.8340	4.1947	14.0528	40.6188	108.0809
	0.01	100	1.9533	5.0089	20.9652	76.3726	267.6498
		500	1.9523	4.8893	19.2863	65.1618	195.2196

Note: The values of $\bar{c}=c$ in Panel A of the table correspond to the (local) point alternatives against which the local asymptotic power for significance level 5% equals one-half. The critical values in Panel B apply the corresponding value of $\bar{c}$ from Panel A. The results are simulated based on 20,000 Monte Carlo replications. The test rejects when the test statistic is larger than the critical values in Panel B of this table.

Table 3: Size and Size-Adjusted Power: Constant Mean, $T = 100$

	Test	MA					AR			
ϕ	Statistic	-0.5	-0.3	0.0	0.3	0.5	-0.5	-0.3	0.3	0.5
1.00	$\rho\,(0.1)$	0.21	0.10	0.05	0.04	0.03	0.11	0.08	0.03	0.02
	$\rho\,(\bar{c}, 0.1)$	0.23	0.14	0.10	0.08	0.08	0.15	0.12	0.08	0.06
	BT	0.09	0.06	0.05	0.05	0.05	0.06	0.06	0.04	0.04
	ADF	0.07	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.05
	ADF-GLS	0.08	0.07	0.06	0.05	0.05	0.05	0.05	0.05	0.06
0.9	$\rho\,(0.1)$	0.45	0.42	0.38	0.36	0.35	0.43	0.41	0.35	0.32
	$\rho\,(\bar{c}, 0.1)$	0.67	0.63	0.57	0.54	0.54	0.63	0.60	0.53	0.49
	BT	0.33	0.31	0.29	0.28	0.28	0.31	0.30	0.28	0.27
	ADF	0.17	0.19	0.22	0.11	0.13	0.20	0.20	0.10	0.16
	ADF-GLS	0.51	0.56	0.63	0.52	0.52	0.61	0.60	0.51	0.52
0.8	$\rho\,(0.1)$	0.92	0.89	0.83	0.79	0.77	0.90	0.87	0.76	0.70
	$\rho\,(\bar{c}, 0.1)$	0.99	0.98	0.95	0.93	0.92	0.98	0.97	0.91	0.87
	BT	0.67	0.60	0.55	0.53	0.52	0.62	0.59	0.52	0.48
	ADF	0.41	0.47	0.59	0.38	0.33	0.52	0.52	0.30	0.29
	ADF-GLS	0.72	0.77	0.84	0.84	0.80	0.82	0.81	0.83	0.74
0.7	$\rho\,(0.1)$	1.00	0.99	0.98	0.96	0.95	1.00	0.99	0.95	0.90
	$\rho\,(\bar{c}, 0.1)$	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.99	0.98
	BT	0.87	0.80	0.73	0.69	0.68	0.81	0.77	0.67	0.62
	ADF	0.55	0.60	0.70	0.71	0.57	0.64	0.65	0.67	0.29
	ADF-GLS	0.79	0.81	0.86	0.89	0.86	0.84	0.84	0.90	0.83
0.6	$\rho\,(0.1)$	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.99	0.97
	$\rho\,(\bar{c}, 0.1)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	BT	0.96	0.91	0.84	0.80	0.78	0.92	0.89	0.77	0.71
	ADF	0.64	0.65	0.73	0.79	0.72	0.67	0.68	0.81	0.52
	ADF-GLS	0.83	0.83	0.87	0.89	0.88	0.86	0.85	0.90	0.89

Note: The ADF and ADF-GLS tests use the MAIC to determine the lag augmentation with $k_{\min} = 0$ and $k_{\max} = int(12(T/100)^{1/4})$ as in Ng & Perron (2001). For each statistic, entries under the rows marked $\phi = 1.00$ are the finite sample rejection frequencies under the null, i.e. the size. All other entries are size-adjusted power under the models described in each column. Based on 20,000 Monte Carlo replications.

Table 4: Size and Size-Adjusted Power: Linear Trend, $T = 100$

	Test	MA					AR			
ϕ	Statistic	−0.5	−0.3	0.0	0.3	0.5	−0.5	−0.3	0.3	0.5
1.00	$\rho\,(0.1)$	0.35	0.14	0.05	0.03	0.02	0.18	0.11	0.02	0.01
	$\rho\,(\bar{c}, 0.1)$	0.29	0.13	0.05	0.03	0.03	0.15	0.10	0.02	0.01
	BT	0.14	0.08	0.05	0.04	0.04	0.09	0.07	0.03	0.03
	ADF	0.09	0.07	0.04	0.03	0.04	0.04	0.05	0.02	0.05
	ADF-GLS	0.06	0.06	0.04	0.02	0.03	0.04	0.04	0.02	0.04
0.9	$\rho\,(0.1)$	0.24	0.22	0.21	0.20	0.19	0.23	0.22	0.19	0.18
	$\rho\,(\bar{c}, 0.1)$	0.32	0.31	0.28	0.27	0.26	0.30	0.30	0.26	0.24
	BT	0.20	0.19	0.18	0.17	0.17	0.19	0.19	0.17	0.17
	ADF	0.13	0.14	0.16	0.08	0.09	0.14	0.14	0.07	0.10
	ADF-GLS	0.23	0.24	0.28	0.21	0.21	0.27	0.26	0.19	0.20
0.8	$\rho\,(0.1)$	0.72	0.68	0.61	0.57	0.55	0.70	0.67	0.54	0.49
	$\rho\,(\bar{c}, 0.1)$	0.86	0.83	0.77	0.73	0.72	0.84	0.82	0.71	0.64
	BT	0.54	0.49	0.44	0.41	0.41	0.50	0.47	0.40	0.37
	ADF	0.33	0.36	0.48	0.26	0.18	0.42	0.40	0.19	0.15
	ADF-GLS	0.47	0.53	0.66	0.57	0.48	0.60	0.59	0.48	0.33
0.7	$\rho\,(0.1)$	0.97	0.95	0.91	0.87	0.85	0.96	0.95	0.84	0.76
	$\rho\,(\bar{c}, 0.1)$	0.99	0.99	0.97	0.95	0.94	0.99	0.99	0.94	0.89
	BT	0.82	0.76	0.68	0.63	0.62	0.78	0.73	0.64	0.55
	ADF	0.52	0.55	0.68	0.63	0.39	0.62	0.61	0.54	0.11
	ADF-GLS	0.58	0.64	0.75	0.82	0.70	0.69	0.70	0.81	0.35
0.6	$\rho\,(0.1)$	1.00	1.00	0.99	0.97	0.96	1.00	1.00	0.96	0.91
	$\rho\,(\bar{c}, 0.1)$	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.99	0.97
	BT	0.95	0.90	0.84	0.79	0.77	0.92	0.89	0.76	0.68
	ADF	0.64	0.64	0.74	0.82	0.66	0.68	0.68	0.81	0.24
	ADF-GLS	0.63	0.67	0.77	0.85	0.80	0.70	0.72	0.88	0.60

Note: The ADF and ADF-GLS tests use the MAIC to determine the lag augmentation with $k_{\min} = 0$ and $k_{\max} = int(12(T/100)^{1/4})$ as in Ng & Perron (2001). For each statistic, entries under the rows marked $\phi = 1.00$ are the finite sample rejection frequencies under the null, i.e. the size. All other entries are size-adjusted power under the models described in each column. Based on 20,000 Monte Carlo replications.

Table 5: Size and Size-Adjusted Power: Constant Mean, $T = 500$

	Test	MA					AR			
ϕ	Statistic	-0.5	-0.3	0.0	0.3	0.5	-0.5	-0.3	0.3	0.5
1.00	$\rho\,(0.1)$	0.10	0.07	0.05	0.04	0.04	0.07	0.06	0.04	0.04
	$\rho\,(\bar{c}, 0.1)$	0.09	0.07	0.06	0.06	0.06	0.07	0.06	0.06	0.05
	BT	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	ADF	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	ADF-GLS	0.06	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.05
0.98	$\rho\,(0.1)$	0.42	0.40	0.38	0.37	0.37	0.40	0.39	0.37	0.36
	$\rho\,(\bar{c}, 0.1)$	0.58	0.54	0.53	0.52	0.52	0.55	0.54	0.52	0.51
	BT	0.31	0.30	0.29	0.29	0.29	0.30	0.30	0.29	0.29
	ADF	0.25	0.26	0.26	0.25	0.22	0.26	0.26	0.25	0.24
	ADF-GLS	0.69	0.71	0.72	0.71	0.69	0.72	0.72	0.71	0.70
0.96	$\rho\,(0.1)$	0.87	0.83	0.80	0.79	0.79	0.84	0.82	0.78	0.76
	$\rho\,(\bar{c}, 0.1)$	0.95	0.92	0.91	0.90	0.90	0.93	0.92	0.89	0.88
	BT	0.58	0.55	0.54	0.53	0.53	0.55	0.54	0.53	0.52
	ADF	0.67	0.71	0.75	0.70	0.66	0.74	0.74	0.72	0.70
	ADF-GLS	0.97	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.98
0.94	$\rho\,(0.1)$	0.99	0.98	0.96	0.96	0.96	0.98	0.97	0.95	0.94
	$\rho\,(\bar{c}, 0.1)$	1.00	0.99	0.99	0.98	0.98	0.99	0.99	0.98	0.98
	BT	0.75	0.71	0.69	0.69	0.69	0.71	0.70	0.68	0.67
	ADF	0.88	0.90	0.93	0.91	0.89	0.92	0.92	0.91	0.90
	ADF-GLS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
0.92	$\rho\,(0.1)$	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.99	0.99
	$\rho\,(\bar{c}, 0.1)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	BT	0.86	0.82	0.79	0.79	0.78	0.82	0.81	0.78	0.76
	ADF	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.95
	ADF-GLS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: The ADF and ADF-GLS tests use the MAIC to determine the lag augmentation with $k_{\min} = 0$ and $k_{\max} = int(12(T/100)^{1/4})$ as in Ng & Perron (2001). For each statistic, entries under the rows marked $\phi = 1.00$ are the finite sample rejection frequencies under the null, i.e. the size. All other entries are size-adjusted power under the models described in each column. Based on 20,000 Monte Carlo replications.

Table 6: Size and Size-Adjusted Power: Linear Trend, $T = 500$

ϕ	Test Statistic	MA -0.5	MA -0.3	MA 0.0	MA 0.3	MA 0.5	AR -0.5	AR -0.3	AR 0.3	AR 0.5
1.00	$\rho\,(0.1)$	0.16	0.08	0.05	0.04	0.04	0.09	0.07	0.04	0.03
	$\rho\,(\bar{c}, 0.1)$	0.12	0.07	0.05	0.04	0.04	0.07	0.06	0.04	0.03
	BT	0.07	0.06	0.05	0.05	0.05	0.06	0.05	0.05	0.04
	ADF	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	ADF-GLS	0.04	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.04
0.98	$\rho\,(0.1)$	0.22	0.21	0.20	0.20	0.20	0.22	0.21	0.20	0.19
	$\rho\,(\bar{c}, 0.1)$	0.29	0.28	0.27	0.27	0.27	0.28	0.27	0.27	0.26
	BT	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.17
	ADF	0.15	0.16	0.17	0.15	0.14	0.16	0.16	0.16	0.15
	ADF-GLS	0.28	0.29	0.30	0.29	0.28	0.30	0.30	0.29	0.28
0.96	$\rho\,(0.1)$	0.66	0.62	0.58	0.56	0.56	0.62	0.60	0.55	0.53
	$\rho\,(\bar{c}, 0.1)$	0.77	0.73	0.70	0.69	0.69	0.74	0.72	0.69	0.67
	BT	0.45	0.43	0.42	0.41	0.41	0.43	0.42	0.41	0.40
	ADF	0.44	0.47	0.53	0.48	0.44	0.51	0.51	0.49	0.47
	ADF-GLS	0.71	0.75	0.79	0.75	0.73	0.78	0.78	0.76	0.74
0.94	$\rho\,(0.1)$	0.93	0.90	0.86	0.85	0.84	0.90	0.89	0.84	0.81
	$\rho\,(\bar{c}, 0.1)$	0.97	0.95	0.93	0.92	0.92	0.95	0.94	0.91	0.90
	BT	0.68	0.65	0.62	0.61	0.61	0.65	0.64	0.61	0.59
	ADF	0.71	0.76	0.83	0.78	0.74	0.81	0.81	0.79	0.76
	ADF-GLS	0.91	0.92	0.94	0.93	0.93	0.94	0.94	0.93	0.93
0.92	$\rho\,(0.1)$	1.00	0.99	0.97	0.97	0.96	0.99	0.98	0.96	0.95
	$\rho\,(\bar{c}, 0.1)$	1.00	0.99	0.99	0.99	0.98	0.99	0.99	0.98	0.98
	BT	0.84	0.79	0.76	0.75	0.75	0.80	0.78	0.75	0.73
	ADF	0.84	0.87	0.91	0.89	0.89	0.90	0.90	0.90	0.88
	ADF-GLS	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.96

Note: The ADF and ADF-GLS tests use the MAIC to determine the lag augmentation with $k_{\min} = 0$ and $k_{\max} = int(12(T/100)^{1/4})$ as in Ng & Perron (2001). For each statistic, entries under the rows marked $\phi = 1.00$ are the finite sample rejection frequencies under the null, i.e. the size. All other entries are size-adjusted power under the models described in each column. Based on 20,000 Monte Carlo replications.

Table 7: Size and Size-Adjusted Power: Linear Trend, $T = 100$

	Test	GARCH		Bilin		VCM		TAR		STUR	
ϕ	Statistic	0.65	0.85	-0.8	0.8	-0.8	0.8	0.5	0.8	0.1	0.5
1.00	$\rho\,(0.1)$	0.06	0.05	0.06	0.05	0.01	0.06	0.03	0.05	0.05	0.04
	$\rho\,(\bar{c}, 0.1)$	0.07	0.07	0.05	0.05	0.03	0.02	0.03	0.05	0.06	0.05
	BT	0.05	0.05	0.05	0.05	0.03	0.05	0.04	0.05	0.04	0.05
	ADF	0.03	0.02	0.05	0.04	0.01	0.14	0.02	0.02	0.03	0.03
	ADF-GLS	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.06	0.03
0.9	$\rho\,(0.1)$	0.21	0.22	0.17	0.18	0.15	0.15	0.21	0.22	1.00	0.51
	$\rho\,(\bar{c}, 0.1)$	0.25	0.26	0.12	0.13	0.20	0.26	0.28	0.28	0.95	0.57
	BT	0.18	0.18	0.16	0.17	0.13	0.15	0.18	0.19	0.94	0.37
	ADF	0.17	0.19	0.15	0.15	0.08	0.01	0.12	0.11	0.83	0.48
	ADF-GLS	0.23	0.25	0.14	0.14	0.13	0.19	0.27	0.23	0.58	0.57
0.8	$\rho\,(0.1)$	0.58	0.61	0.54	0.55	0.38	0.38	0.59	0.61	1.00	0.91
	$\rho\,(\bar{c}, 0.1)$	0.68	0.71	0.46	0.47	0.48	0.64	0.75	0.76	0.97	0.88
	BT	0.44	0.44	0.43	0.44	0.29	0.28	0.43	0.45	0.99	0.67
	ADF	0.46	0.51	0.40	0.40	0.19	0.01	0.43	0.34	0.85	0.75
	ADF-GLS	0.55	0.59	0.48	0.48	0.28	0.39	0.71	0.63	0.49	0.72
0.7	$\rho\,(0.1)$	0.87	0.90	0.85	0.86	0.63	0.64	0.88	0.90	1.00	0.99
	$\rho\,(\bar{c}, 0.1)$	0.93	0.94	0.83	0.84	0.72	0.87	0.95	0.96	0.97	0.94
	BT	0.68	0.68	0.67	0.68	0.44	0.41	0.65	0.68	0.99	0.82
	ADF	0.64	0.68	0.64	0.64	0.36	0.04	0.73	0.64	0.86	0.80
	ADF-GLS	0.68	0.70	0.67	0.67	0.46	0.58	0.85	0.82	0.45	0.70
0.6	$\rho\,(0.1)$	0.97	0.98	0.96	0.96	0.80	0.82	0.98	0.98	1.00	1.00
	$\rho\,(\bar{c}, 0.1)$	0.98	0.99	0.96	0.97	0.86	0.96	0.99	1.00	0.97	0.96
	BT	0.83	0.84	0.83	0.84	0.56	0.53	0.80	0.83	1.00	0.89
	ADF	0.71	0.68	0.76	0.76	0.51	0.10	0.82	0.78	0.87	0.82
	ADF-GLS	0.72	0.71	0.72	0.73	0.59	0.69	0.85	0.84	0.43	0.67

Note: The ADF and ADF-GLS tests use the MAIC to determine the lag augmentation with $k_{\min} = 0$ and $k_{\max} = int(12(T/100)^{1/4})$ as in Ng & Perron (2001). For each statistic, entries under the rows marked $\phi = 1.00$ are the finite sample rejection frequencies under the null, i.e. the size. All other entries are size-adjusted power under the models described in each column. Based on 20,000 Monte Carlo replications.

Table 8: Size and Size-Adjusted Power: Linear Trend, $T = 500$

	Test	GARCH		Bilin		VCM		TAR		STUR	
ϕ	Statistic	0.65	0.85	-0.8	0.8	-0.8	0.8	0.5	0.8	0.1	0.5
1.00	$\rho\,(0.1)$	0.06	0.05	0.05	0.05	0.03	0.16	0.04	0.05	0.22	0.06
	$\rho\,(\bar{c}, 0.1)$	0.06	0.06	0.05	0.05	0.07	0.03	0.05	0.05	0.33	0.06
	BT	0.05	0.05	0.04	0.05	0.05	0.10	0.05	0.05	0.16	0.06
	ADF	0.03	0.03	0.04	0.04	0.04	0.24	0.03	0.04	0.13	0.03
	ADF-GLS	0.04	0.04	0.03	0.03	0.07	0.04	0.03	0.04	0.35	0.04
0.98	$\rho\,(0.1)$	0.21	0.21	0.11	0.11	0.14	0.11	0.20	0.21	0.84	0.47
	$\rho\,(\bar{c}, 0.1)$	0.25	0.25	0.00	0.00	0.20	0.24	0.26	0.26	0.87	0.55
	BT	0.18	0.18	0.08	0.08	0.13	0.12	0.18	0.18	0.75	0.34
	ADF	0.17	0.19	0.23	0.27	0.13	0.01	0.15	0.15	0.85	0.51
	ADF-GLS	0.26	0.28	0.00	0.01	0.17	0.22	0.29	0.29	0.75	0.62
0.96	$\rho\,(0.1)$	0.56	0.58	0.51	0.51	0.36	0.27	0.56	0.57	1.00	0.93
	$\rho\,(\bar{c}, 0.1)$	0.67	0.67	0.02	0.02	0.48	0.53	0.68	0.69	0.95	0.91
	BT	0.42	0.42	0.34	0.34	0.30	0.23	0.41	0.41	0.98	0.69
	ADF	0.46	0.55	0.57	0.64	0.33	0.04	0.47	0.47	0.97	0.92
	ADF-GLS	0.66	0.73	0.08	0.12	0.42	0.48	0.77	0.75	0.75	0.94
0.94	$\rho\,(0.1)$	0.84	0.86	0.85	0.85	0.59	0.47	0.86	0.86	1.00	1.00
	$\rho\,(\bar{c}, 0.1)$	0.90	0.91	0.08	0.09	0.71	0.74	0.92	0.92	0.95	0.97
	BT	0.63	0.64	0.58	0.59	0.45	0.34	0.61	0.62	1.00	0.86
	ADF	0.72	0.81	0.83	0.87	0.53	0.09	0.79	0.76	0.99	0.96
	ADF-GLS	0.87	0.91	0.37	0.49	0.62	0.67	0.95	0.93	0.70	0.96
0.92	$\rho\,(0.1)$	0.96	0.97	0.97	0.97	0.75	0.66	0.96	0.97	1.00	1.00
	$\rho\,(\bar{c}, 0.1)$	0.98	0.98	0.22	0.23	0.84	0.86	0.98	0.98	0.95	0.98
	BT	0.77	0.77	0.74	0.76	0.56	0.44	0.75	0.76	1.00	0.93
	ADF	0.85	0.90	0.92	0.94	0.67	0.17	0.93	0.89	1.00	0.98
	ADF-GLS	0.94	0.96	0.64	0.74	0.75	0.78	0.98	0.97	0.65	0.97

Note: The ADF and ADF-GLS tests use the MAIC to determine the lag augmentation with $k_{\min} = 0$ and $k_{\max} = int(12(T/100)^{1/4})$ as in Ng & Perron (2001). For each statistic, entries under the rows marked $\phi = 1.00$ are the finite sample rejection frequencies under the null, i.e. the size. All other entries are size-adjusted power under the models described in each column. Based on 20,000 Monte Carlo replications.

Figure 1: Asymptotic local power functions of $\rho(d)$ against near-integrated alternatives

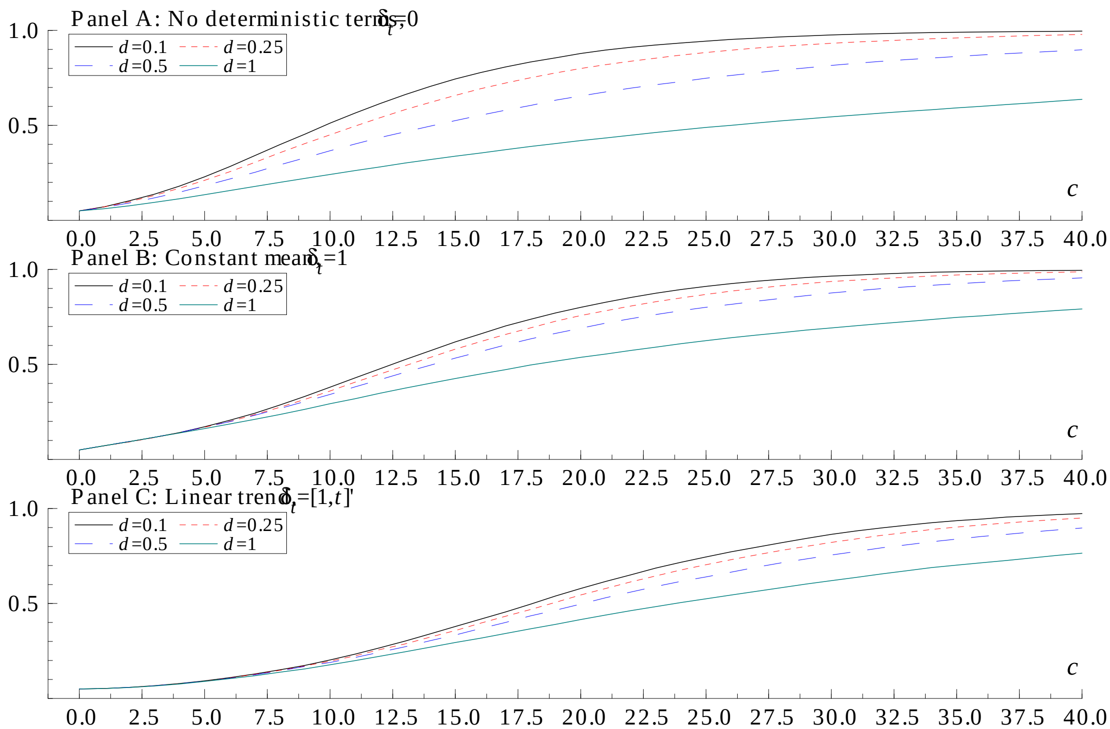

Figure 2: Asymptotic local power functions of GLS detrended tests against near-integrated alternatives