Queen's Economics Department Working Paper No. 1185

A Powerful Test of the Autoregressive Unit Root Hypothesis Based on a Tuning Parameter Free Statistic

Morten Ørregaard Nielsen
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

10-2008

A Powerful Test of the Autoregressive Unit Root Hypothesis Based on a Tuning Parameter Free Statistic[1]

Morten Ørregaard Nielsen[2]
Queen's University and CREATES

October 24, 2008

[1]I am grateful to James Davidson, Javier Hualde, James MacKinnon, Peter Phillips, Robert Taylor, Tim Vogelsang, three anonymous referees, and seminar participants at Cornell University, Purdue University, Queen's University, the 2006 North American Summer Meeting of the Econometric Society, the 2006 Aarhus Summer Econometrics Workshop, and the 2007 Granger Centre Conference in Honour of Paul Newbold 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, Dunning Hall Room 307, 94 University Avenue, Queen's University, Kingston, Ontario K7L 3N6, Canada; phone: +1 613 533 2262; fax: +1 613 533 6668; email: mon@econ.queensu.ca

Abstract

This paper presents a family of simple nonparametric unit root tests indexed by one parameter, d, and containing Breitung's (2002) test as the special case $d = 1$. It is shown that (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, and (iii) since the asymptotic distribution depends on d and the test remains consistent for all $d > 0$, it is possible to analyze the power of the test for different values of d. The usual Phillips-Perron or Dickey-Fuller type tests are indexed by bandwidth, lag length, etc., but have none of these three properties.

It is shown that members of the family with $d < 1$ have higher asymptotic local power than the Breitung (2002) test, and 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, GLS detrending is shown to improve power when d is small, which is not the case for Breitung's (2002) test. Simulations demonstrate that when applying a sieve bootstrap procedure, the proposed variance ratio test has very good size properties, with finite sample power that is higher than that of Breitung's (2002) test 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 parameter, tuning parameter, power envelope, unit root test, variance ratio

Short Title: Variance Ratio 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 three 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) and Perron & Qu (2007). 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 power envelopes for such tests, i.e. bounds on the possible power of parametric unit root tests under conditions allowing for serial correlation, deterministic components, etc.

In the presence of serial correlation all the above tests are characterized by nuisance parameters appearing 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 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 selection of bandwidth and kernel for the estimation of the long-run variance. The performance of the tests depends 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 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 do not depend on 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, 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*.

[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.

Existing unit root tests that are free of tuning parameters include the variable addition test of Park & Choi (1988) and 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 [the bandwidth parameter] l." 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 based on tuning parameter free statistics and improve upon existing tuning parameter free tests in terms of asymptotic local power. Compared to parametric tests, the proposed tests avoid many of the issues related to nuisance parameters, at least asymptotically, while maintaining competitive power properties. The nonparametric tests are constructed as a ratio of the sample variance of the observed series and that of a fractional partial sum (fractional difference of a negative order) of the series. Recently, 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 variance ratio statistic is indexed by one parameter, d, which determines the order of the fractional partial summation, but since the asymptotic distribution depends on d this is not a tuning parameter. The test procedure itself applies a sieve bootstrap to improve finite sample performance. The lag length p in the sieve approximation does not appear in the asymptotic distribution, and the variance ratio test with the bootstrap procedure is therefore not tuning parameter free.

There are several other important differences between the parameter d indexing the variance ratio tests and the tuning parameters in ADF regressions (lag length) or Phillips-Perron type tests (bandwidth and kernel). First of all, for any member of the family with $d > 0$, the variance ratio 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 analyze the (asymptotic local) power properties of the test for different values of d, and then try to locate a member of the family which is "tailored" to maximize power against relevant alternatives. The usual ADF/ADF-GLS or Phillips-Perron type tests have none of these three properties.

The proposed procedure is nonparametric and does not rely on the specification of a particular

[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.

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 important in practical applications where 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.

When $d = 1$ the Breitung (2002) test appears as a particular member of the proposed family. However, it is shown that members of the family with parameter $d < 1$ have higher asymptotic local power than Breitung's (2002) test. Furthermore, 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 only 12% more observations would be required for the nonparametric variance ratio test with $d = 0.1$ to achieve asymptotic local power of one-half compared to the ADF-GLS test.

Finally, a simulation study is conducted where comparisons are made with the leading tuning parameter free test of Breitung (2002) and the (nearly) optimal ADF-GLS test of Elliott et al. (1996) applying the MAIC lag augmentation selection rule of Ng & Perron (2001) as modified by Perron & Qu (2007). The simulations apply a sieve bootstrap procedure to the proposed variance ratio test which is then characterized by the tuning parameter p denoting the sieve lag length which is chosen by the MAIC. The bootstrapped variance ratio test has size properties that are as good as those of the ADF-GLS test using the MAIC lag selection rule. With the sieve bootstrap procedure, the finite sample power of the variance ratio test is similar to that of the ADF-GLS test and even superior in some cases, such as the important case of a model that includes a linear time trend and has moving average errors. 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 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 Breitung's (2002) test as well as (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 model

$$y_t = \phi y_{t-1} + u_t, \quad t = 1, 2, ..., \quad y_0 = 0, \tag{1}$$

where 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,

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

where we have used the definition

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

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.

$$T^{-1/2} y_{\lfloor Ts \rfloor} \Rightarrow \sigma_y W_0(s), \quad 0 \le s \le 1, \tag{4}$$
$$T^{-1/2-d} \tilde{y}_{\lfloor Ts \rfloor} \Rightarrow \sigma_y W_d(s), \quad 0 \le s \le 1, \tag{5}$$

as $T \to \infty$ for some σ_y to be specified later. Here, $\lfloor \bullet \rfloor$ denotes the integer part of the argument, "$\Rightarrow$" means weak convergence in $D[0,1]$, and W_d is the type II fractional standard Brownian motion of order d ($> -1/2$), see e.g. Marinucci & Robinson (2000), defined as

$$W_d(r) = 0, \text{ a.s., } r = 0, \tag{6}$$
$$W_d(r) = \frac{1}{\Gamma(d+1)} \int_0^r (r-s)^d \, dW_0(s), \; r > 0. \tag{7}$$

[3] The initial condition can be replaced by other well-known conditions that yield the same functional central limit theorems (4) and (5). 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.

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

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.

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. Thus, setting $d = 1$, $\tilde{y}_t$ is the partial sum of y_t and $\rho(1)$ is then (the inverse of) the statistic proposed by Breitung (2002), which is therefore also a member of the family of tests in (10). The same idea was applied by Vogelsang (1998*a*, 1998*b*) to test for structural breaks without estimating serial correlation parameters.

In recent work, Müller (2007, 2008) 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 a deterministic linear time trend. Thus, the family of variance ratio statistics corrected for deterministic terms is defined as in (10) but with residuals $\hat{y}_t = y_t - \hat{\alpha}' \delta_t$ replacing the observed time series y_t. For now, $\hat{y}_t$ are OLS residuals, but 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 near-integrated alternatives and for an important range of d values.

The following assumption on u_t in (1) is imposed throughout.

Assumption 1 *The zero-mean process u_t is weakly stationary and ergodic and satisfies*

(a) $0 < \sum_{k=-\infty}^{\infty} |\gamma_u(k)| < \infty$*, where* $\gamma_u(k) = E(u_t u_{t+k})$*,*
(b) $T^{-1/2} \sum_{t=1}^{\lfloor Ts \rfloor} u_t \Rightarrow \sigma_y W_0(s)$ *for* $\sigma_y > 0$ *and all* $0 \le s \le 1$*,*
(c) $T^{-1/2-d} \sum_{t=1}^{\lfloor Ts \rfloor} \Delta_+^{-d} u_t \Rightarrow \sigma_y W_d(s)$ *for* $d > 0$*,* $\sigma_y > 0$*, and all* $0 \le s \le 1$*.*

Assumption 1 is similar to Condition C in Elliott et al. (1996) and holds under a variety of regularity assumptions. Sufficient conditions for (b) are given by, e.g., Phillips (1987*a*) and Phillips & Solo (1992), and for (c) by, e.g., Akonom & Gourieroux (1987), Davidson & de Jong (2000), and Marinucci & Robinson (2000). The conditions include mixing conditions and moment conditions (existence of a moment of order greater than two), and are satisfied by, e.g., stationary and invertible ARMA models. The conditions permit conditional heteroskedasticity in $\{u_t\}$ but rule out unconditional heteroskedasticity.

Under the null hypothesis that $\phi = 1$ and under Assumption 1 on u_t, (4) and (5) clearly hold. In that case, 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 $B_0(s) = W_0(s)$ and the demeaned ($j = 1$) and detrended ($j = 2$) standard Brownian motions are defined as

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

where $D_1(s) = 1$, $D_2(s) = [1, s]'$, and also $\tilde{B}_{0,d}(s) = W_d(s)$ and

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

Note that in this theorem and below, weak convergence is for a fixed value of d. Also note that

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

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

Deterministics	γ	T	$d=0.10$	$d=0.25$	$d=0.50$	$d=0.75$	$d=1.00$
$j=0:\delta_t=0$	0.10	100	1.54	2.78	6.76	15.30	33.63
		500	1.54	2.77	6.70	15.09	33.13
	0.05	100	1.62	3.13	8.45	21.00	48.73
		500	1.62	3.14	8.44	20.70	49.42
	0.01	100	1.76	3.90	12.55	36.12	98.82
		500	1.77	3.92	12.93	38.59	106.6
$j=1:\delta_t=1$	0.10	100	1.75	3.83	12.29	32.04	70.03
		500	1.76	3.87	12.39	32.32	70.43
	0.05	100	1.81	4.18	14.49	41.55	100.4
		500	1.82	4.20	14.43	40.83	97.83
	0.01	100	1.92	4.82	19.39	64.82	186.3
		500	1.93	4.85	18.98	62.12	173.1
$j=2:\delta_t=[1,t]'$	0.10	100	1.91	4.78	19.50	69.86	227.4
		500	1.92	4.83	19.68	70.35	228.0
	0.05	100	1.96	5.09	22.07	83.94	291.3
		500	1.98	5.17	22.33	84.79	289.6
	0.01	100	2.04	5.68	27.73	119.1	455.1
		500	2.08	5.83	28.26	118.8	446.2

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.

so that the term $\int_0^s \frac{(s-r)^{d-1}}{\Gamma(d)} D_j(r)\,dr$ appearing in the definition of $\tilde{B}_{j,d}(s)$ for $j=1$ and $j=2$ corresponds to fractional powers of s. That is, $\tilde{B}_{j,d}(s)$, $j=1,2$, 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 when $j=1$ and $[s^d, s^{d+1}]$ when $j=2$.

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 values of the parameter d are presented in Table 1.

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

Theorem 2 *Under the assumptions of Theorem 1 the test that rejects H_0 in (2) when $\rho(d) >$*

$CV_{j,\gamma}(d)$, where $CV_{j,\gamma}(d)$ is found from

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

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 asymptotic distribution of the variance ratio statistic, unlike the tuning parameters, e.g. lag length and bandwidth parameters, in the Dickey-Fuller or Phillips-Perron unit root tests. Thus, the test based on Theorem 2 and the asymptotic distribution in Theorem 1 would be tuning parameter free. Another consequence is that 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 is (the inverse of) the statistic used by Breitung (2002) and Taylor (2005) to test for (seasonal) unit roots.

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 would appear 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.

3 Asymptotic Local Power Analysis

In this section, the asymptotic local power of the autoregressive unit root test described in Theorems 1 and 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 to "tailor" the test to obtain high power, i.e. to select the member of the 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 $\phi = 1$. Under (13) and Assumption 1, sample moments such as (8) have limiting distributions which are expressed in terms of the Ornstein-Uhlenbeck process

$$J_{0,c}(s) = W(s) - c\int_0^s e^{-c(s-r)} W(r)dr, \quad J_{0,c}(0) = 0. \tag{14}$$

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 asymptotic 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 (14), $J_{1,c}(s), J_{2,c}(s)$ are the demeaned ($j = 1$) and detrended ($j = 2$) Ornstein-Uhlenbeck processes,

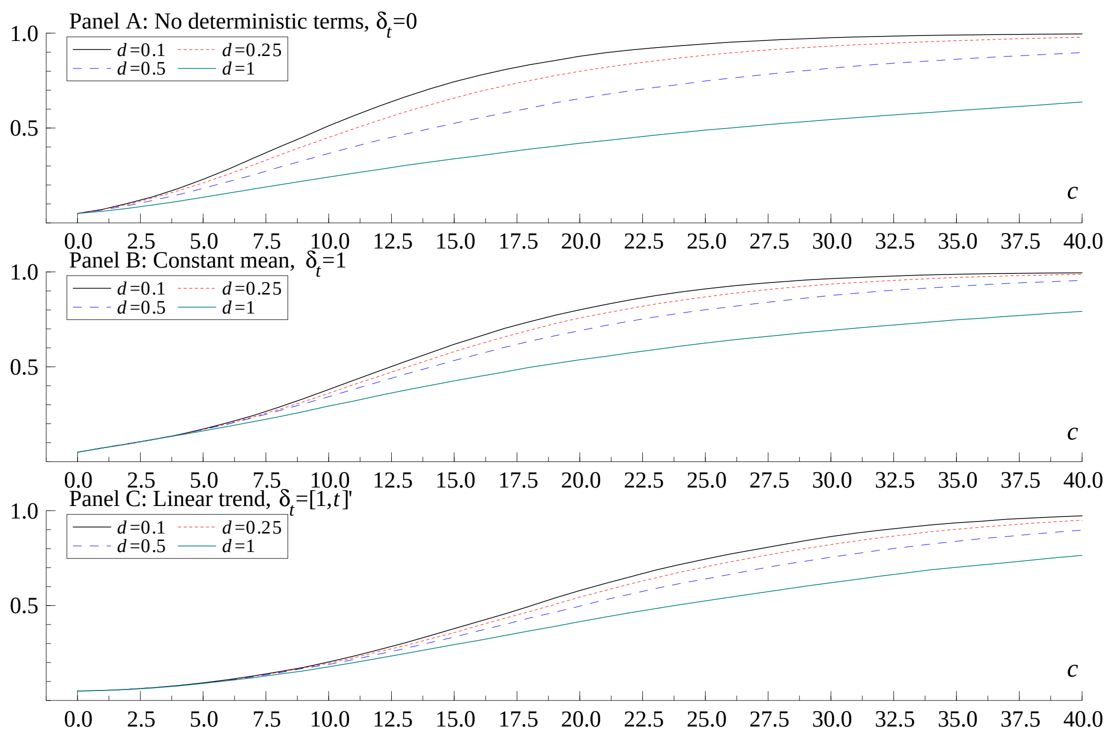
$$J_{j,c}(s) = J_{0,c}(s) - \left(\int_0^1 J_{0,c}(r) D_j(r)' dr\right)\left(\int_0^1 D_j(r) D_j(r)' dr\right)^{-1} D_j(s), \quad j = 1, 2,$$

and

$$\tilde{J}_{0,c,d}(s) = W_d(s) - c\int_0^s e^{-c(s-r)} W_d(r) dr,$$

$$\tilde{J}_{j,c,d}(s) = \tilde{J}_{0,c,d}(s) - \left(\int_0^1 J_{0,c}(r) D_j(r)' dr\right)\left(\int_0^1 D_j(r) D_j(r)' dr\right)^{-1} \int_0^s \frac{(s-r)^{d-1}}{\Gamma(d)} D_j(r) dr, \quad j = 1, 2.$$

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



It follows from Theorem 3 that the asymptotic local power of any member of the family of variance ratio tests 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. Buchmann & Chan (2007). 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 displays simulated asymptotic local power curves for several members of the variance ratio test family (with $\gamma = 0.05$) as functions 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 graph, the power curves are drawn for $d \in \{0.1, 0.25, 0.5, 1.0\}$, where $d = 1$ is the test of Breitung (2002).

From Figure 1 it appears that the asymptotic local power of the variance ratio test is monotonic in d, and that $d = 0.1$ is 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.[4] 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. In section 4 below, further support of the $d = 0.1$ test relative to Breitung's (2002) test ($d = 1$) 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 $d = 1$ 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

$$\hat{y}_{\bar{c},t} = y_t - \tilde{\alpha}'\delta_t, \tag{15}$$

where

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

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

[4] In fact, $\rho(0) = 1$ and $\lim_{d\to 0} d^{-1}(\rho(d)^{-1} - 1) = \lim_{d\to 0}(d\sum_{t=1}^{T}(d^{-1}(\Delta_+^{-d} - 1)y_t)^2 + 2\sum_{t=1}^{T} y_t d^{-1}(\Delta_+^{-d} - 1)y_t)/\sum_{t=1}^{T} y_t^2 = 2\sum_{j=1}^{T-1} j^{-1} r_j$, where r_j is the j'th sample autocorrelation and $\lim_{d\to 0} d^{-1}(\Delta_+^{-d} - 1) = \sum_{j=1}^{t-1} j^{-1} L^j$. The statistic $\sum_{j=1}^{T-1} j^{-1} r_j$ is well known as a test in fractionally integrated models, e.g. Robinson (1994) and Tanaka (1999), although there it is used as a test of $I(0)$, not $I(1)$. The asymptotic distribution of this statistic under the unit root null can be derived using results of Hualde (2007, Lemma 2), but simulations indicate that it is very sensitive to short-run dynamics in u_t, so this is not pursued further here.

augmentation for the ADF-GLS tests, it is assumed in the asymptotic comparisons to be chosen correctly and optimally, and has no effect on asymptotic local power.

Consider constructing the variance ratio test based on the GLS detrended series (15). That is,

$$\rho(\bar{c}, d) = T^{2d} \frac{\sum_{t=1}^{T} \hat{y}_{\bar{c},t}^2}{\sum_{t=1}^{T} \tilde{y}_{\bar{c},t}^2}, \tag{16}$$

where $\tilde{y}_{\bar{c},t} = \Delta_+^{-d} \hat{y}_{\bar{c},t}$ similarly to (3). The distribution of the GLS detrended variance ratio test (16) 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 (15) and the variance ratio statistic is given by (16). 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 $U_{0,NI}(c, d)$ is 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. 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 $P(U_{j,GLS}(\bar{c}, \bar{c}, d) > CV_{j,0.05}(\bar{c}, d))$ is equal to one-half at 5% significance level, where $CV_{j,\gamma}(\bar{c}, d)$ satisfies $P(U_{j,GLS}(\bar{c}, 0, d) > CV_{j,\gamma}(\bar{c}, d)) = \gamma$.

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.

Table 2: Values of $\bar{c}$ and $CV_{j,\gamma}(\bar{c}, d)$ of the VR-GLS test (16)

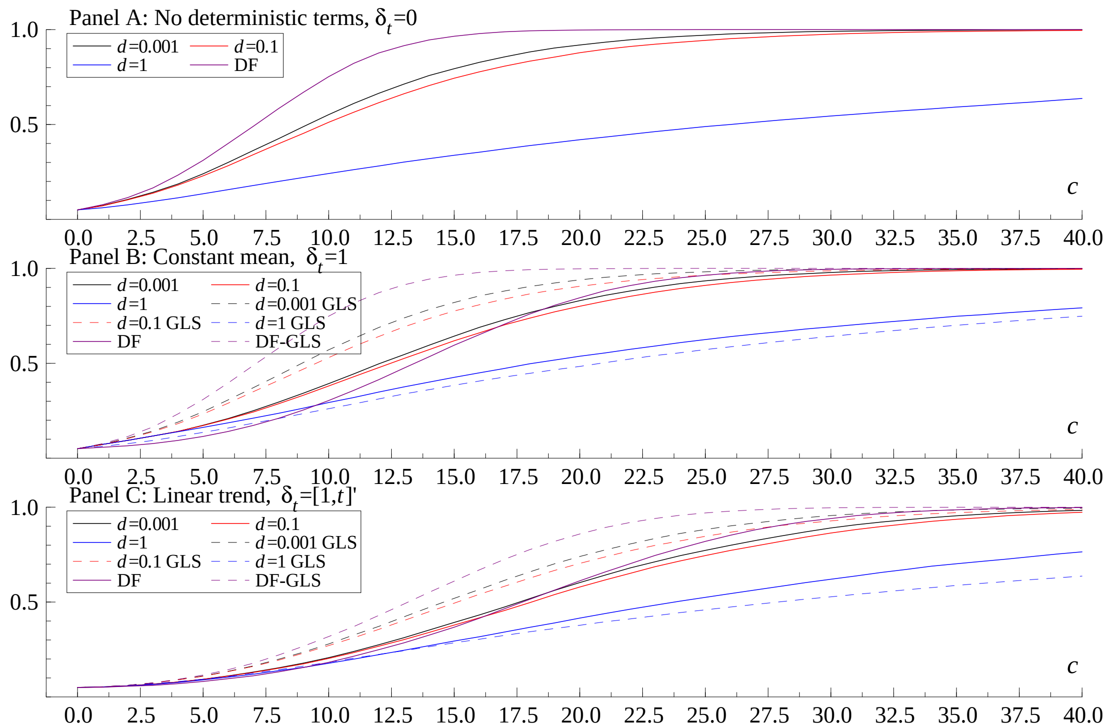
Deterministics	γ	T	$d = 0.10$	$d = 0.25$	$d = 0.50$	$d = 0.75$	$d = 1.00$
Panel A: Values of $\bar{c} = c$ that yield asymptotic local power of 50%							
$j = 1 : \delta_t = 1$	0.05	500	9.4	10.6	12.8	16.3	20.8
$j = 2 : \delta_t = [1, t]'$	0.05	500	15.1	16.1	18.7	22.5	28.0
Panel B: Critical values $CV_{j,\gamma}(\bar{c}, d)$							
$j = 2 : \delta_t = [1, t]'$	0.10	100	1.80	4.05	13.72	41.75	122.9
		500	1.77	3.86	11.92	31.84	78.21
	0.05	100	1.85	4.37	15.98	52.28	161.5
		500	1.83	4.19	14.05	40.62	108.1
	0.01	100	1.95	5.01	20.97	76.37	267.7
		500	1.95	4.89	19.29	65.16	195.2

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.

In Figure 2 the asymptotic local power functions of the GLS detrended variance ratio tests with $d \in \{0.001, 0.1, 1\}$ are presented for the no deterministics case (Panel A), the constant mean case (Panel B), and the linear trend case (Panel C). The Breitung (2002) test ($d = 1$) is included for comparison with existing tuning parameter free tests, and 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. 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). All the asymptotic local power functions are simulated based on 20,000 Monte Carlo replications with $T = 500$.

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. Thus, compared to Figure 1, the asymptotic local power of the GLS detrended variance ratio test with $d < 1$ is even more superior to Breitung's (2002) test than its OLS detrended counterpart.

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



One method to measure and compare the asymptotic local power of the GLS detrended variance ratio test with that of Breitung's (2002) test and 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 such as 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 test with $d = 0.1$ and the Breitung (2002) test ($d = 1$) 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 constant mean case 122% more observations would be needed and in the linear trend case 85% more observations would be needed for the Breitung (2002) ($d = 1$) test than for the VR-GLS test with $d = 0.1$ to achieve asymptotic local power of one-half. Perhaps more surprisingly, 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 assume 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 data, using e.g. an information criterion, and 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 tests which do not require selection of tuning parameters or estimation of serial correlation parameters.

4 Finite Sample Performance

The time series y_t is simulated according to the model

$$y_t = \phi y_{t-1} + u_t, \quad t = 1, ..., T, \quad y_0 = 0. \tag{17}$$

To conserve space, only results for the linear trend case are reported[5] since most economic time series exhibit time trends. The sample sizes considered are $T = 100$ and $T = 500$, and $20,000$ Monte Carlo replications are used in the simulations. Throughout, a $\gamma = 5\%$ nominal significance level is employed. All calculations were made in Ox, see Doornik (2006).

In all the simulations, comparisons are made not only to existing tuning parameter free tests, but also 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 further modification of Perron & Qu (2007) was applied to achieve even better finite sample properties, and the lag augmentation was chosen to optimize the MAIC with $k_{\min} = 0$ and $k_{\max} = \lfloor 12(T/100)^{1/4} \rfloor$ as in Ng & Perron (2001) and Perron & Qu (2007). Note that, in the simulations, this upper bound binds rarely for the small sample size ($T = 100$) and almost 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.

Simulation results for the case where u_t is *i.i.d.* standard normal are presented in Table 3. 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

[5] More simulation results, including the constant mean case and simulations of different data generating processes, are available in the earlier working paper version, see Nielsen (2008).

Table 3: Simulated rejection frequencies when u_t is *i.i.d.*

	$T=100$					$T=500$				
ϕ	$\rho(0.1)$	$\rho(\bar{c},0.1)$	BT	ADF	ADF-GLS	$\rho(0.1)$	$\rho(\bar{c},0.1)$	BT	ADF	ADF-GLS
1.00	0.05	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.04	0.03
0.9	0.21	0.29	0.18	0.14	0.26	0.21	0.27	0.18	0.13	0.22
0.8	0.61	0.77	0.44	0.44	0.65	0.57	0.70	0.41	0.47	0.71
0.7	0.91	0.97	0.68	0.66	0.77	0.87	0.93	0.62	0.80	0.92
0.6	0.99	1.00	0.83	0.72	0.79	0.97	0.99	0.77	0.90	0.96

Note: The ADF and ADF-GLS tests use the MAIC to determine the lag augmentation with $k_{\min}=0$ and $k_{\max}=\lfloor 12(T/100)^{1/4}\rfloor$ as in Perron & Qu (2007). For each statistic, entries under the row marked $\phi=1.00$ are finite sample rejection frequencies under the null, i.e. the sizes of the tests. All other entries are finite sample power. Based on 20,000 Monte Carlo replications.

in the row marked $\phi=1.00$ are rejection frequencies under the unit root null hypothesis, i.e. the sizes of the tests, and all other entries are finite sample power.

The results in Table 3 clearly suggest that the nonparametric variance ratio test is useful and that non-trivial power gains may be obtained relative to the tuning parameter free BT test. More surprisingly, it appears that the variance ratio test even rivals 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 finite sample 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 power without the need to select any tuning parameters.

Next, consider the moving average (MA) model,

$$u_t = \varepsilon_t + \theta\varepsilon_{t-1}, \quad t = 1, ..., T, \tag{18}$$

where ε_t is *i.i.d.* standard normal. The MA model was chosen since MA errors are relevant for many economic time series as argued by Ng & Perron (2001). For instance, omitted outliers may cause MA(1) type characteristics in observed time series, c.f. Franses & Haldrup (1994). Furthermore, unreported simulations (see Nielsen (2008)) have shown that the variance ratio tests are (sometimes severely) size distorted when the errors u_t are generated according to (18). The size issue is also present in Breitung (2002) and so it is somewhat expected.

To reduce size distortion a bootstrap procedure can be applied. The sieve bootstrap algorithm applied here follows that of Chang & Park (2003).[6]

[6] This bootstrap procedure was also applied to the ADF-GLS test with the MAIC lag selection. However, only

Let $\hat{u}_t$ denote the first difference of the GLS detrended observed time series (the same procedure applies to OLS detrended series). First, fit the approximating sieve autoregression

$$\hat{u}_t = \alpha_1 \hat{u}_{t-1} + \ldots + \alpha_p \hat{u}_{t-p} + \varepsilon_t \tag{19}$$

by OLS and denote by $\{\hat{\varepsilon}_t\}_{t=1}^T$ the residuals from (19). It is important to base the bootstrap procedure on the first differences of the detrended time series since, as shown by Basawa, Mallik, McCormick, Reeves & Taylor (1991), bootstrap samples generated without the unit root restriction make bootstrap unit root testing procedures inconsistent. The lag length p in (19) is chosen to be the same as that in the ADF-GLS tests with the MAIC selection criterion.

Next, bootstrap errors $\{\varepsilon_t^b\}_{t=1}^T$ are constructed by resampling from the centered residuals $\{\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}_t\}_{t=1}^T$ from (19) with replacement. Then $\{\hat{u}_t^b\}_{t=1}^T$ is generated from $\{\varepsilon_t^b\}_{t=1}^T$ using the fitted autoregression, i.e. by

$$\hat{u}_t^b = \hat{\alpha}_1 \hat{u}_{t-1}^b + \ldots + \hat{\alpha}_p \hat{u}_{t-p}^b + \varepsilon_t^b, \quad t = p+1, \ldots, T,$$

where $\hat{\alpha}_1, \ldots \hat{\alpha}_p$ are the estimated parameters from (19) and the initial values $\hat{u}_1^b, \ldots, \hat{u}_p^b$ are set to zero for simplicity. Finally, the bootstrap sample $\{y_t^b\}_{t=1}^T$ is obtained by partial summation, i.e.

$$y_t^b = y_0^b + \sum_{s=1}^{t} \hat{u}_s^b, \quad t = 1, \ldots, T,$$

with the initial condition $y_0^b = 0$. For a discussion of the initial conditions imposed on $\hat{u}_t^b$ and y_t^b, see Chang & Park (2003, p. 390).

The bootstrap sample $\{y_t^b\}_{t=1}^T$ is then GLS detrended and the variance ratio statistic, denoted $\rho_b(\bar{c}, d)$, is calculated from the GLS detrended bootstrap sample as described in the previous sections. To implement the bootstrap test, the bootstrap procedure is repeated $B = 999$ times and the simulated bootstrap p-value is then computed as the percentage of bootstrap statistics that exceed the actual statistic $\rho(\bar{c}, d)$ from the observed sample, i.e., $\hat{p} = B^{-1} \sum_{b=1}^{B} 1(\rho_b(\bar{c}, d) > \rho(\bar{c}, d))$, where $1(\cdot)$ is the indicator function. The bootstrap test rejects if $\hat{p} < \gamma$.

Note that the bootstrap procedure described here has the additional advantage relative to the test described in Theorems 1 and 2 based on the asymptotic distribution that a p-value is readily obtained as part of the procedure. Thus it might be thought of as more informative. Also note that the bootstrap procedure depends on the tuning parameter p in the sieve approximation (19), and this is only partly alleviated by the use of the MAIC to choose p. Even though the variance ratio statistic $\rho(\bar{c}, d)$ – and hence the asymptotic test of Theorems 1 and 2 – is asymptotically tuning parameter free, the bootstrap procedure implemented to approximate the finite sample distribution of $\rho(\bar{c}, d)$ depends on the tuning parameter p in (19). The critical values therefore also depend on p and hence the bootstrapped test is not tuning parameter free.

---

very small size improvements were obtained at the cost of a large loss in power, so those results are not reported here.

Table 4: Simulated rejection frequencies for model (18): $T = 100$

ϕ	Test statistic \ θ	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8
1.00	$\rho\,(0.1)$	0.17	0.07	0.05	0.05	0.04	0.03	0.03	0.02	0.02
	$\rho\,(\bar{c}, 0.1)$	0.16	0.08	0.07	0.07	0.07	0.06	0.06	0.04	0.03
	BT	0.17	0.07	0.05	0.05	0.05	0.04	0.04	0.04	0.03
	ADF	0.18	0.07	0.04	0.04	0.03	0.01	0.01	0.02	0.01
	ADF-GLS	0.19	0.09	0.07	0.06	0.04	0.02	0.02	0.02	0.02
0.9	$\rho\,(0.1)$	0.42	0.21	0.18	0.18	0.18	0.14	0.11	0.09	0.06
	$\rho\,(\bar{c}, 0.1)$	0.49	0.30	0.29	0.31	0.33	0.30	0.25	0.20	0.15
	BT	0.42	0.22	0.18	0.17	0.17	0.16	0.14	0.13	0.11
	ADF	0.44	0.21	0.17	0.15	0.12	0.04	0.04	0.05	0.03
	ADF-GLS	0.52	0.32	0.28	0.27	0.25	0.12	0.11	0.12	0.09
0.8	$\rho\,(0.1)$	0.74	0.45	0.41	0.43	0.48	0.44	0.35	0.26	0.19
	$\rho\,(\bar{c}, 0.1)$	0.79	0.57	0.57	0.62	0.68	0.69	0.60	0.49	0.38
	BT	0.75	0.46	0.41	0.39	0.40	0.38	0.34	0.29	0.25
	ADF	0.75	0.46	0.40	0.40	0.43	0.24	0.11	0.14	0.11
	ADF-GLS	0.81	0.58	0.57	0.59	0.65	0.51	0.34	0.34	0.27
0.7	$\rho\,(0.1)$	0.92	0.62	0.57	0.60	0.67	0.71	0.61	0.46	0.35
	$\rho\,(\bar{c}, 0.1)$	0.93	0.71	0.68	0.72	0.78	0.83	0.79	0.68	0.56
	BT	0.92	0.65	0.57	0.56	0.58	0.58	0.54	0.46	0.39
	ADF	0.92	0.64	0.57	0.59	0.66	0.61	0.36	0.28	0.24
	ADF-GLS	0.94	0.73	0.69	0.72	0.77	0.80	0.68	0.57	0.48
0.6	$\rho\,(0.1)$	0.98	0.75	0.65	0.66	0.72	0.79	0.76	0.63	0.49
	$\rho\,(\bar{c}, 0.1)$	0.98	0.80	0.74	0.75	0.80	0.85	0.84	0.77	0.67
	BT	0.98	0.76	0.66	0.66	0.68	0.70	0.67	0.59	0.51
	ADF	0.98	0.76	0.66	0.66	0.71	0.78	0.67	0.49	0.38
	ADF-GLS	0.98	0.81	0.75	0.76	0.79	0.84	0.83	0.74	0.62

Notes: The ADF and ADF-GLS tests use the MAIC to determine the lag augmentation with $k_{\min} = 0$ and $k_{\max} = \lfloor 12(T/100)^{1/4} \rfloor$ as in Perron & Qu (2007). The BT, $\rho(0.1)$, and $\rho(\bar{c}, 0.1)$ tests use the sieve bootstrap with same lag length as the ADF and ADF-GLS tests. For each statistic, entries under the rows marked $\phi = 1.00$ are finite sample rejection frequencies under the null, i.e. the sizes. All other entries are finite sample power. Based on 20,000 Monte Carlo replications.

Table 4 presents the simulated rejection frequencies with $T = 100$ under the MA model (18) with coefficient $\theta \in \{-0.8, -0.6, \ldots, 0.8\}$. The BT, $\rho(0.1)$, and $\rho(\bar{c}, 0.1)$ tests apply the sieve bootstrap algorithm described above. All the tests have similar size properties and show some distortion for the largest negative MA coefficient, $\theta = -0.8$, but only minor size distortion for the other values of the MA coefficient. In terms of power the variance ratio tests with $d = 0.1$, and especially the GLS detrended version, clearly dominate the BT test. Furthermore, the $\rho(\bar{c}, 0.1)$ test actually outperforms the ADF-GLS test when the MA coefficient is either small or positive. In the case of a negative coefficient the $\rho(\bar{c}, 0.1)$ and ADF-GLS tests have almost identical power. Thus, in this

Table 5: Simulated rejection frequencies for model (18): $T = 500$

ϕ	Test statistic \ θ	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8
1.00	$\rho\,(0.1)$	0.07	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04
	$\rho\,(\bar{c}, 0.1)$	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
	BT	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	ADF	0.05	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02
	ADF-GLS	0.05	0.05	0.04	0.04	0.03	0.04	0.03	0.03	0.03
0.98	$\rho\,(0.1)$	0.22	0.20	0.20	0.19	0.19	0.19	0.18	0.17	0.14
	$\rho\,(\bar{c}, 0.1)$	0.29	0.29	0.30	0.30	0.30	0.30	0.29	0.27	0.24
	BT	0.21	0.19	0.18	0.18	0.18	0.18	0.17	0.17	0.16
	ADF	0.18	0.15	0.13	0.13	0.12	0.12	0.10	0.10	0.08
	ADF-GLS	0.28	0.25	0.24	0.23	0.22	0.21	0.19	0.18	0.16
0.96	$\rho\,(0.1)$	0.54	0.52	0.53	0.53	0.53	0.51	0.49	0.45	0.39
	$\rho\,(\bar{c}, 0.1)$	0.71	0.70	0.71	0.71	0.72	0.70	0.68	0.65	0.59
	BT	0.48	0.42	0.41	0.40	0.40	0.39	0.39	0.38	0.35
	ADF	0.49	0.44	0.45	0.46	0.46	0.43	0.38	0.35	0.29
	ADF-GLS	0.71	0.67	0.67	0.70	0.71	0.68	0.64	0.59	0.52
0.94	$\rho\,(0.1)$	0.79	0.76	0.78	0.79	0.80	0.77	0.76	0.72	0.64
	$\rho\,(\bar{c}, 0.1)$	0.91	0.88	0.90	0.90	0.91	0.89	0.89	0.86	0.82
	BT	0.70	0.64	0.61	0.60	0.60	0.59	0.58	0.56	0.53
	ADF	0.77	0.71	0.74	0.76	0.79	0.74	0.71	0.65	0.56
	ADF-GLS	0.91	0.88	0.90	0.91	0.92	0.90	0.89	0.86	0.80
0.92	$\rho\,(0.1)$	0.92	0.87	0.89	0.90	0.91	0.89	0.88	0.86	0.80
	$\rho\,(\bar{c}, 0.1)$	0.97	0.95	0.95	0.95	0.96	0.95	0.95	0.94	0.91
	BT	0.84	0.77	0.75	0.74	0.73	0.72	0.71	0.69	0.66
	ADF	0.92	0.84	0.86	0.88	0.90	0.88	0.87	0.83	0.75
	ADF-GLS	0.98	0.95	0.95	0.96	0.96	0.96	0.95	0.94	0.91

Notes: The ADF and ADF-GLS tests use the MAIC to determine the lag augmentation with $k_{\min} = 0$ and $k_{\max} = \lfloor 12(T/100)^{1/4} \rfloor$ as in Perron & Qu (2007). The BT, $\rho(0.1)$, and $\rho(\bar{c}, 0.1)$ tests use the sieve bootstrap with same lag length as the ADF and ADF-GLS tests. For each statistic, entries under the rows marked $\phi = 1.00$ are finite sample rejection frequencies under the null, i.e. the sizes. All other entries are finite sample power. Based on 20,000 Monte Carlo replications.

case with a linear time trend, it appears that the relatively close proximity of the asymptotic local power function of the $\rho\,(\bar{c}, 0.1)$ test to the parametric power envelope carries over to the simulation results.

In Table 5, laid out as the previous table, the simulated rejection frequencies for sample size $T = 500$ are reported under the same models as in Table 4. The results show that the size distortion for $\theta = -0.8$ in the smaller sample is no longer present, and all the tests now exhibit excellent size control for all the MA coefficients considered. In terms of power, the conclusions from the smaller sample size are confirmed: The $\rho\,(\bar{c}, 0.1)$ test is superior to the ADF-GLS test in the presence of small or positive MA coefficients, and the $\rho(\bar{c}, 0.1)$ and ADF-GLS tests perform very similarly in the presence of large negative MA coefficients. Thus, the $\rho(\bar{c}, 0.1)$ test remains very competitive, even with the larger sample size and therefore better lag augmentation selection by the ADF-GLS test. However, the power of the ADF-GLS test relative to the $\rho(\bar{c}, 0.1)$ test has increased somewhat for this larger sample size compared to the smaller sample size, presumably due to better lag augmentation selection, and is now generally closer to that of the $\rho(\bar{c}, 0.1)$ test. In all cases, the BT test is clearly dominated by the $\rho(0.1)$ and $\rho(\bar{c}, 0.1)$ tests.

It is clear that the application of the sieve bootstrap procedure described here delivers tests with size properties that are as good as those of the ADF-GLS tests with MAIC lag selection. Furthermore, in terms of power, the bootstrapped $\rho(\bar{c}, 0.1)$ test is clearly superior to the tuning parameter free BT test, and is generally at least as powerful as, and in some cases even superior to, the ADF-GLS test.

5 Concluding Remarks

The family of nonparametric variance ratio tests of the unit root hypothesis presented here has the property that the test statistics are free of tuning parameters. That is, there are no parameters involved in calculating the test statistic which are not reflected in the asymptotic distribution. The test statistics 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 indexed by the parameter d which determines the order of the fractional partial summation. However, unlike the choice of, 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-to-unity asymptotics, the power of each member of the family against near-integrated alternatives was derived. In particular, it was shown that members of the family with $d < 1$ have asymptotic local power that is better than that of Breitung's (2002) test; a leading tuning parameter free test. Furthermore, when d is small the asymptotic local power of the proposed test is relatively close to the parametric power envelope, especially in the case with a linear time trend.

Simulation evidence demonstrates the finite sample properties of the proposed variance ratio test. To alleviate the size distortion that is sometimes present in small samples, a sieve bootstrap procedure is applied to the test. The sieve lag length p becomes a tuning parameter in the variance ratio test in the sense that the bootstrap critical values depend on p. Applying the sieve bootstrap, the variance ratio test has size properties that are as good as those of the ADF-GLS test when the two are compared on an even footing by applying the MAIC to select both the lag augmentation of the Dickey-Fuller regressions and the sieve lag length for the bootstrap procedure. This is the case even in the presence of moving average errors with a large negative coefficient. With the sieve bootstrap procedure, the finite sample power of the variance ratio test is similar to that of the ADF-GLS test and even superior in some cases, such as the important case of a model that includes a linear time trend and has moving average errors.

Appendix: Proofs

Proof of Theorem 1. In the case with no deterministic terms the result follows immediately by the continuous mapping theorem since (4) and (5) hold under Assumption 1. 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

$$T^{-1/2} z_{\lfloor Ts \rfloor} \Rightarrow \sigma_y W_0\left(s\right)$$

holds. Now define $N_1(T) = 1$ and $N_2(T) = diag(1, T^{-1})$ and write

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

where

$$\begin{aligned} T^{-1/2}\left(\hat{\alpha}-\alpha\right)' N_j(T)^{-1} &= \left(T^{-1}\sum_{s=1}^{T} T^{-1/2} z_s \delta_s' N_j(T)\right)\left(T^{-1}\sum_{s=1}^{T} N_j(T)\delta_s\delta_s' N_j(T)\right)^{-1} \qquad (20) \\ &= \left(T^{-1}\sum_{s=1}^{T} T^{-1/2} z_s D_j(s/T)\right)\left(T^{-1}\sum_{s=1}^{T} D_j(s/T) D_j(s/T)'\right)^{-1} \\ &\Rightarrow \sigma_y \left(\int_0^1 W_0\left(s\right) D_j\left(s\right)' ds\right)\left(\int_0^1 D_j\left(s\right) D_j\left(s\right)' ds\right)^{-1} \end{aligned}$$

by application of (4) and the continuous mapping theorem, and

$$N_j(T)\delta_{\lfloor Ts \rfloor} = D_j(\lfloor Ts \rfloor / T) \rightarrow D_j(s) \text{ as } T \rightarrow \infty. \qquad (21)$$

It thus follows that

$$\hat{y}_T\left(s\right) = T^{-1/2}\hat{y}_{\lfloor Ts \rfloor} \Rightarrow \sigma_y B_j\left(s\right), \quad j = 0, 1, 2. \qquad (22)$$

Next, for $\tilde{y}_T(s) = T^{-d}\Delta_+^{-d}\hat{y}_T(s) = T^{-1/2-d}\sum_{k=0}^{\lfloor Ts \rfloor -1}\pi_k\left(d\right)\hat{y}_{\lfloor Ts \rfloor - k} = T^{-1/2-d}\sum_{k=1}^{\lfloor Ts \rfloor}\pi_{\lfloor Ts \rfloor - k}\left(d\right)\hat{y}_k$, where $\hat{y}_t = z_t - (\hat{\alpha} - \alpha)' \delta_t$ and $\pi_k(d) = \Gamma(k+d)/(\Gamma(d)\Gamma(k+1))$, the convergence

$$T^{-1/2-d}\sum_{k=1}^{\lfloor Ts \rfloor}\pi_{\lfloor Ts \rfloor - k}\left(d\right) z_k \Rightarrow \sigma_y W_d(s)$$

holds from (5). For the remaining term, i.e.,

$$T^{-1/2-d}\sum_{k=1}^{\lfloor Ts\rfloor}\pi_{\lfloor Ts\rfloor-k}(d)(\hat{\alpha}-\alpha)'\delta_k=\left(T^{-1/2}(\hat{\alpha}-\alpha)'N_j(T)^{-1}\right)\left(T^{-d}\sum_{k=1}^{\lfloor Ts\rfloor}\pi_{\lfloor Ts\rfloor-k}(d)N_j(T)\delta_k\right),$$

the first factor converges by (20) and the last factor is deterministic and satisfies the convergence

$$\begin{aligned}T^{-d}\sum_{k=1}^{\lfloor Ts\rfloor}\pi_{\lfloor Ts\rfloor-k}(d)N_j(T)\delta_k&=T^{-d}\sum_{k=1}^{\lfloor Ts\rfloor}\pi_{\lfloor Ts\rfloor-k}(d)D_j(k/T)\\&=T^{-d}\sum_{k=1}^{\lfloor Ts\rfloor}\frac{(\lfloor Ts\rfloor-k)^{d-1}}{\Gamma(d)}D_j(k/T)+o(1)\\&=T^{-1}\sum_{k=1}^{\lfloor Ts\rfloor}\frac{\left(\frac{\lfloor Ts\rfloor}{T}-\frac{k}{T}\right)^{d-1}}{\Gamma(d)}D_j(k/T)+o(1)\\&\to\int_0^s\frac{(s-r)^{d-1}}{\Gamma(d)}D_j(r)\,dr\text{ as }T\to\infty.\end{aligned}\tag{23}$$

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 1, y_t is stationary and ergodic such that $\omega_y^2=Ey_t^2<\infty$, and

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

by the law of large numbers for stationary ergodic time series, e.g. White (1984, p. 42).

Under H_1 and Assumption 1 it also holds that $0<\sum_{k=-\infty}^{\infty}|\gamma_y(k)|<\infty$, where $\gamma_y(k)=E(y_ty_{t+k})$. If $d<1/2$ note that

$$Var(\tilde{y}_t)=\sum_{m=0}^{t-1}\sum_{k=0}^{t-1}\pi_m(d)\pi_k(d)\gamma_y(m-k)\le C\sum_{m=0}^{t-1}|\gamma_y(m)|\sum_{k=0}^{t-1-m}(k+m)^{d-1}k^{d-1}\le C\sum_{m=0}^{\infty}|\gamma_y(m)|<\infty.$$

Here, and throughout, $C>0$ denotes a generic constant which may take different values in different places. If $d>1/2$,

$$T^{1-2d}Var(\tilde{y}_t)\le CT^{1-2d}\sum_{m=0}^{t-1}|\gamma_y(m)|\sum_{k=0}^{t-1-m}(k+m)^{d-1}k^{d-1}.$$

The evaluations $(k+m)^{d-1}k^{d-1} \leq k^{2d-2}$ if $d<1$ and $(k+m)^{d-1}k^{d-1} \leq T^{2d-2}$ if $d \geq 1$ then give

$$T^{1-2d}Var(\tilde{y}_t) \leq C \sum_{m=0}^{\infty} |\gamma_y(m)| \begin{cases} T^{-1}\sum_{k=0}^{T}(k/T)^{2d-2} \to \int_0^1 x^{2d-2}dx < \infty, & d<1, \\ T^{1-2d}\sum_{k=0}^{T} T^{2d-2} < \infty, & d \geq 1. \end{cases}$$

Hence, under H_1 and Assumption 1, $\rho(d)^{-1} = O_P\left(T^{-\min(1,2d)}\right)$ and thus $\rho(d)$ diverges in probability to $+\infty$ when $d>0$, noting that $\rho(d)>0$ 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

$$T^{-1/2} z_{\lfloor Ts \rfloor} \Rightarrow \sigma_y J_{0,c}(s), \tag{24}$$

where $J_{0,c}(s) = W(s) - c\int_0^s e^{-c(s-r)}W(r)dr$, $J_{0,c}(0)=0$, 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 (20) it follows that

$$T^{-1/2}(\hat{\alpha}-\alpha)' N_j(T)^{-1} \Rightarrow \sigma_y \left(\int_0^1 J_{0,c}(r) D_j(r)' dr\right)\left(\int_0^1 D_j(r) D_j(r)' dr\right)^{-1}, \tag{25}$$

which combined with (21) and (24) implies that

$$\hat{y}_T(s) = T^{-1/2}\hat{y}_{\lfloor Ts \rfloor} \Rightarrow \sigma_y J_{j,c}(s), \tag{26}$$

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

As in the proof of Theorem 1, define $\tilde{y}_T(s) = T^{-d}\Delta_+^{-d}\hat{y}_T(s) = T^{-1/2-d}\sum_{k=1}^{\lfloor Ts \rfloor}\pi_{\lfloor Ts \rfloor - k}(d)\,\hat{y}_k$. 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 $y_t = \sum_{k=1}^{t} e^{-c(t-k)/T}u_k$ (where a negligible remainder term has been left out), and using summation by parts the representation

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

is obtained. By the mean value theorem, for $0 \leq x \leq 1$,

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

which implies that

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

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, $\tilde{y}_T(s) = T^{-1/2-d}\sum_{k=1}^{\lfloor Ts\rfloor}\pi_{\lfloor Ts\rfloor-k}(d)\,y_k$ is

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

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

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

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

$$\begin{aligned}\tilde{y}_{2T}(s) &= \frac{c}{T}\sum_{k=1}^{\lfloor Ts\rfloor-1}e^{-c(\lfloor Ts\rfloor-k)/T}\tilde{y}_{1T}(k/T) \\ &= c\sum_{k=1}^{\lfloor Ts\rfloor-1}e^{-c(\lfloor Ts\rfloor-k)/T}\int_{k/T}^{(k+1)/T}\tilde{y}_{1T}(r)dr \\ &= c\sum_{k=1}^{\lfloor Ts\rfloor-1}\int_{k/T}^{(k+1)/T}e^{-c(\lfloor Ts\rfloor-\lfloor Tr\rfloor)/T}\tilde{y}_{1T}(r)dr \\ &= c\int_{1/T}^{\lfloor Ts\rfloor/T}e^{-c(\lfloor Ts\rfloor-\lfloor Tr\rfloor)/T}\tilde{y}_{1T}(r)dr \\ &= c\int_0^s e^{-c(s-r)}\tilde{y}_{1T}(r)dr + R_T(s).\end{aligned}$$

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

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

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

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

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

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

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

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

which is $O_P(T^{-1})$ by (5) and the continuous mapping theorem. Hence $\sup_{0 \le s \le 1} R_T(s) \xrightarrow{P} 0$.

Finally, the result for the fractional partial sum of the detrended process, i.e. the result for $\tilde{y}_T(s) = T^{-1/2-d} \sum_{k=1}^{\lfloor Ts \rfloor} \pi_{\lfloor Ts \rfloor - k}(d) \hat{y}_k$, can easily be proven. Writing $\hat{y}_t = z_t - (\hat{\alpha} - \alpha)' \delta_t$, where z_t is generated by (13), it has already been shown that

$$
T^{-1/2-d} \sum_{k=1}^{\lfloor Ts \rfloor} \pi_{\lfloor Ts \rfloor - k}(d) z_k \Rightarrow \sigma_y \tilde{J}_{0,c,d}(s), \tag{27}
$$

and it only remains to combine this result with (23) and (25) to get

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

■

Proof of Theorem 4. From Elliott et al. (1996, pp. 834-835) it follows that $T^{-1/2}\hat{y}_{\bar{c},\lfloor Ts\rfloor} \Rightarrow \sigma_y J_{0,c}(s)$ if $j=1$ and $T^{-1/2}\hat{y}_{\bar{c},\lfloor Ts\rfloor} \Rightarrow \sigma_y V_{\bar{c},c}(s)$ if $j=2$. Now $\tilde{y}_{\bar{c},t}$ is based on

$$\hat{y}_{\bar{c},t} = z_t - (\tilde{\alpha}_0 - \alpha_0) - (\tilde{\alpha}_1 - \alpha_1)t, \tag{28}$$

which is GLS detrended as in (15). Note that (28) applies to the case $j=2$; when $j=1$ the last term is not present. From Elliott et al. (1996, p. 835) it is known that $\tilde{\alpha}_0 = O_P(1)$ and

$$\sqrt{T}(\tilde{\alpha}_1 - \alpha_1) \Rightarrow \sigma_y \frac{(1+\bar{c})}{1+\bar{c}+\bar{c}^2/3} J_{0,c}(1) + \sigma_y \frac{\bar{c}^2}{1+\bar{c}+\bar{c}^2/3}\int_0^1 r J_{0,c}(r)\,dr = \sigma_y b_1. \tag{29}$$

It follows immediately that $T^{-1/2-d}\tilde{y}_{\bar{c},\lfloor Ts\rfloor} \Rightarrow \sigma_y \tilde{J}_{0,c,d}(s)$ for $j=1$ when $\tilde{y}_{\bar{c},t}$ is based on $\hat{y}_{\bar{c},t}$.

It only remains to be shown that $T^{-1/2-d}\tilde{y}_{\bar{c},\lfloor Ts\rfloor} \Rightarrow \sigma_y \tilde{V}_{\bar{c},c,d}(s)$ when $j=2$. Following the steps in the proofs of Theorems 1 and 3, write

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

and use (28) to obtain the representation

$$T^{-1/2-d}\tilde{y}_{\bar{c},\lfloor Ts\rfloor} = T^{-1/2-d}\sum_{k=1}^{\lfloor Ts\rfloor}\pi_{\lfloor Ts\rfloor-k}(d)\,z_k \tag{30}$$

$$-(\tilde{\alpha}_0-\alpha_0)T^{-1/2-d}\sum_{k=1}^{\lfloor Ts\rfloor}\pi_{\lfloor Ts\rfloor-k}(d) \tag{31}$$

$$-\sqrt{T}(\tilde{\alpha}_1-\alpha_1)T^{-d}\sum_{k=1}^{\lfloor Ts\rfloor}\pi_{\lfloor Ts\rfloor-k}(d)\frac{\lfloor Ts\rfloor}{T}. \tag{32}$$

It has already been shown in (27) that (30) converges weakly to $\sigma_y\tilde{J}_{0,c,d}(s)$. Next,

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

and since $\tilde{\alpha}_0 = O_P(1)$ this implies that (31) is $O_P(T^{-1/2})$. For (32) use (29) and

$$T^{-d}\sum_{k=1}^{\lfloor Ts\rfloor}\pi_{\lfloor Ts\rfloor-k}(d)\frac{\lfloor Ts\rfloor}{T} \to \frac{s^{d+1}}{\Gamma(d+2)},$$

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

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.

Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H. & Taylor, R. L. (1991), 'Bootstrapping unstable first-order autoregressive processes', *Annals of Statistics* **19**, 1098–1101.

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.

Buchmann, B. & Chan, N. H. (2007), 'Asymptotic theory of least squares estimators for nearly unstable processes under strong dependence', *Annals of Statistics* **35**, 2001–2017.

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

Chang, Y. & Park, J. Y. (2003), 'A sieve bootstrap for the test of a unit root', *Journal of Time Series Analysis* **24**, 379–400.

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. (2006), *Ox: An Object-Oriented Matrix Programming Language*, Timberlake Consultants Press, London.

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

Franses, P. H. & Haldrup, N. (1994), 'The effects of additive outliers on tests for unit roots and cointegration', *Journal of Business and Economic Statistics* **12**, 471–478.

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.

Hualde, J. (2007), 'Estimation of long-run parameters in unbalanced cointegration', *Working Paper, Universidad de Navarra* .

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.

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

Müller, U. K. (2007), 'A theory of robust long-run variance estimation', *Journal of Econometrics* **141**, 1331–1352.

Müller, U. K. (2008), 'The impossibility of consistent discrimination between I(0) and I(1) processes', *Econometric Theory* **24**, 616–630.

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.

Nielsen, M. Ø. (2008), 'A powerful tuning parameter free test of the autoregressive unit root hypothesis', *QED Working Paper 1175, Queen's University* .

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* .

Perron, P. & Qu, Z. (2007), 'A simple modification to improve the finite sample properties of Ng and Perron's unit root tests', *Economics Letters* **94**, 12–19.

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.