Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

Unit root testing in practice: dealing with uncertainty over the trend and initial condition

by

David I. Harvey, Stephen J. Leybourne and A. M. Robert Taylor

Granger Centre Discussion Paper No. 07/03

The University of Nottingham

Unit Root Testing in Practice: Dealing with Uncertainty over the Trend and Initial Condition*

David I. Harvey, Stephen J. Leybourne and A. M. Robert Taylor
School of Economics and Granger Centre for Time Series Econometrics
University of Nottingham

October 2007

Abstract

In this paper we focus on two major issues that surround testing for a unit root in practice, namely: (i) uncertainty as to whether or not a linear deterministic trend is present in the data, and (ii) uncertainty as to whether the initial condition of the process is (asymptotically) negligible or not. In each case simple testing procedures are proposed with the aim of maintaining good power properties across such uncertainties. For the first issue, if the initial condition is negligible, quasi-differenced (QD) detrended (demeaned) Dickey-Fuller-type unit root tests are near asymptotically efficient when a deterministic trend is (is not) present in the data generating process. Consequently, we compare a variety of strategies that aim to select the detrended variant when a trend is present, and the demeaned variant otherwise. Based on asymptotic and finite sample evidence, we recommend a simple union of rejections-based decision rule whereby the unit root null hypothesis is rejected whenever either of the detrended or demeaned unit root tests yields a rejection. Our results show that this approach generally outperforms more sophisticated strategies based on auxiliary methods of trend detection. For the second issue, we again recommend a union of rejections decision rule, rejecting the unit root null if either of the QD and OLS detrended/demeaned Dickey-Fuller-type tests rejects. This procedure is also shown to perform well in practice, simultaneously exploiting the superior power of the QD (OLS) detrended/demeaned test for small (large) initial conditions.

Keywords: Unit root test; trend uncertainty; initial condition; asymptotic power; union of rejections decision rule.

JEL Classifications: C22.

*We are extremely grateful to Peter Phillips and five anonymous referees for their helpful and encouraging comments on the scope and content of earlier drafts of this paper. These have enabled us to make significant improvements to the paper. Correspondence to: Robert Taylor, School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, U.K. Email: robert.taylor@nottingham.ac.uk

1 Introduction

Testing for the presence of an autoregressive unit root has been an issue at the core of econometric research for the last quarter century. Since most macroeconomic time series are considered possibly to also contain some form of deterministic component, it is common practice to apply a unit root test that yields inference not dependent on whether or not a particular deterministic component is present. In the macroeconomic context, the deterministic component in question is most usually a linear trend term (taking a constant term as given). Of course, auxiliary (rather than statistical) considerations may also sometimes rule out the presence of a linear trend term at the outset. For example, few if any economists would seriously endorse the possibility that, for developed economies, macroeconomic data such as real exchange, interest or inflation rates might contain a linear trend. However, for many other macroeconomic time series such as real GDP, industrial production, money supply or consumer prices, the possibility of a linear trend certainly cannot be discounted, and unit root testing of such series needs therefore to be capable of addressing this contingency.

Adapting unit root tests to achieve linear trend-invariance, typically by including a linear trend term in the deterministic specification of the fitted regression from which the unit root test is calculated, introduces an obvious downside. Specifically, when the trend is absent, the power of the unit root test will inevitably be compromised relative to the corresponding test that would be applied if the trend term was omitted; see, for example, Marsh (2007). Moreover, as we see later, these power losses can be very significant, certainly enough to render the strategy of simply always including a linear trend in the deterministic specification a very costly one.

Similar issues arise with regard to the deviation of the first observation from its deterministic component, the so-called initial condition. As discussed in Elliott and Müller (2006,pp.286-90), while there may be situations in which one would not necessarily expect the initial condition to be unusually large or, indeed, unusually small, relative to the other data points, equally the initial condition might be relatively large in other situations. The former case occurs, for example, where the first observation in the sample is dated quite some time after the inception of a mean-reverting process, while the latter can happen if the sample data happen to be chosen to start after a break (perceived or otherwise) in the series or where the beginning of the sample coincides with the start of the process. This latter example can also allow for the case where an unusually small (even zero) initial condition occurs. In practice it is therefore hard to rule out small or large initial conditions, *a priori*. This is problematic because the magnitude of the initial condition can have a substantial impact on the power properties of unit root tests in practice and, as discussed in Elliott and Müller (2006,p.293), we observe only the initial observation rather than the initial condition.

This raises the question of whether it is possible in practice to construct unit root test strategies that: (i) retain high power irrespective of the presence or otherwise of a linear trend in the data; and (ii) maintain good power properties across both large and small initial conditions.[1] In order to make progress, in Section 2 we first introduce our reference unit root testing model and detail the unit root tests on which we focus our attention. These are the quasi difference (QD) demeaned/detrended augmented Dickey-Fuller (ADF) t-ratio tests of Elliott *et al.* (1996), and the corresponding OLS demeaned/detrended tests of Dickey and Fuller (1979) and Said and Dickey (1984), together with the first difference demeaned/detrended von-Neumann ratio tests of Bhargava (1986) and Schmidt and Phillips (1992).

Taking the presence of a constant in the data generation process (DGP) as a given and working on the assumption of an asymptotically negligible initial condition, in Section 3 we compare the performance of the aforementioned unit root tests, enumerating the asymptotic power losses that are involved in applying the detrended variant of each test when no trend is present, and show that in the reverse situation, when the demeaned variant of each test is applied to data containing either a fixed or local (in the sample size) trend, both the asymptotic power and the asymptotic type I error of the tests rapidly approach zero as the magnitude of the trend is increased. Where no trend is present, the test based on QD demeaning is near asymptotically efficient, whereas when a trend is present, it is the corresponding QD detrended ADF test that is (near) efficient.[2] Consequently, in the uncertain linear trend case, we are ideally looking for a strategy which applies the QD detrended ADF test in the presence of a trend, but revert to the QD demeaned ADF test otherwise. We suggest and compare three such strategies.

The first strategy entails *pre-testing* of the trend specification, whereby a statistic which is robust (in terms of critical values) whether the unit root is present or not (the trend tests we consider with this property are those of Harvey *et al.*, 2007, and Bunzel and Vogelsang, 2005) is used to test for the presence of a trend in the data, and subsequently applying either the QD demeaned or QD detrended ADF unit root test, depending on the significance of the initial trend test. The second strategy involves basing inference on a data-dependent *weighted average* of the QD demeaned and QD detrended ADF unit root statistics. The weight function is a continuous function based on an auxiliary trend statistic that forces a switching from the QD demeaned ADF statistic towards the QD detrended ADF statistic once a trend is present. The final strategy consists of the simple decision rule "reject the unit root null if either the QD demeaned ADF or QD detrended ADF test rejects at a given significance level". That is, we look at the *union of rejections* of the two tests. This is clearly the most straightforward strategy to apply since it does not require any explicit form of trend detection via an auxiliary statistic. In addition, this strategy has some practical

[1]We recognize that trend and initial condition uncertainty, while important, are not the only issues relevant to the empirical performance of unit root tests. For example, different distributional assumptions on the innovations can also have a significant impact; see, in particular, Jansson (2007). These, however, would require an investigation quite separate to that conducted here.

[2]Elliott *et al.* (1996) demonstrate that, although not formally efficient, in the limit these tests lie arbitrarily close to the asymptotic local power envelopes for these testing problems and, hence, with a small abuse of language we shall refer to such tests in this context as 'asymptotically efficient' in this paper.

relevance since it embodies what many applied researchers already do, albeit implicitly. We further show that this simple strategy is asymptotically identical to a sequential approach to unit root testing under trend uncertainty proposed in Ayat and Burridge (2000).

Asymptotic results reported in Section 3 demonstrate that in the case where there is either no trend or the trend is of fixed magnitude (i.e. not local in the sample size), then the weighted average test is asymptotically efficient and correctly sized. The pre-test and union of rejections strategies are all shown to lie reasonably close to the efficient test when there is no trend, exhibit little size distortion, and show themselves to be efficient and correctly sized when the trend is non-zero. However, the results for a local trend are in stark contrast to the fixed trend case, where the weighted average test now has only trivial asymptotic power. Also, the pre-test strategies can have relatively low power in this environment, while the union of rejections remains reasonably close to the efficient test throughout. Finite sample simulations are also presented in Section 3 which, for the most part, yield the same qualitative pattern as our asymptotic results. The only real exception here is that the weighted average test fails to exhibit a collapse in power to the dramatic extent that is predicted by the local trend asymptotic theory.

In Section 4 we turn attention to a consideration of the initial condition problem highlighted above. Working with a rather general formulation for the initial condition, which contains the set-up of Elliott (1999), Müller and Elliott (2003), and Elliott and Müller (2006) as special cases, we find that where the initial condition of the process is not asymptotically negligible, the QD demeaned/detrended ADF tests can perform very badly indeed with their power against a given alternative rapidly decreasing towards zero as the magnitude of the initial observation is increased. In stark contrast, the OLS demeaned/detrended ADF tests show an increase in power, other things equal, as the magnitude of the initial condition increases. Consequently, while the QD-based ADF tests are preferable when the initial condition is small, the OLS-based ADF tests are far preferable when the initial condition is large. This finding suggests that a union of rejections decision rule between the QD- and OLS-based ADF tests could again be fruitfully explored in this setting to obtain a test which maintains good power properties across both large and small initial conditions. Specifically, we propose the rule whereby the unit root null is rejected if either of the QD detrended (demeaned) ADF and OLS detrended (demeaned) ADF tests reject in the maintained (no) trend case. Asymptotic and finite sample comparisons with the $\hat{Q}$ statistics of Elliott and Müller (2006), along with statistics proposed by Harvey and Leybourne (2005, 2006), suggest that this procedure is again highly effective, despite its relative simplicity.

Firm recommendations for which tests to use in practice are made at the end of Section 3 for the trend uncertainty problem, and at the end of Section 4 for the initial condition uncertainty problem. In each case, we recommend the use of the union of rejections approach since this appears to offer the best practical solution overall. Section 5 concludes, and here we also discuss limitations of our analysis and discuss directions for further research.

Proofs of the main technical results in this paper are given in Appendix A, while other supplementary material appears in Appendices B and C. A suite of Gauss programs used to produce the figures reported in this paper is available to download from `http://www.nottingham.ac.uk/economics/grangercentre/code.htm`.

Throughout the paper we use the following notation: '$x := y$' to indicate that x is defined by y; $\lfloor \cdot \rfloor$ to denote the integer part of the argument; '$\xrightarrow{p}$' and '$\xrightarrow{d}$' denote convergence in probability and weak convergence, respectively, as the sample size diverges, and $\mathbb{I}(\cdot)$ to denote the indicator function. Finally, reference to a variable being $O_p(T^k)$ is taken to hold in its strict sense, meaning that the variable is not $o_p(T^k)$.

2 Unit Root Tests and the Linear Trend Model

As is standard in this literature, we consider the DGP given by

$$y_t = \mu + \beta t + u_t, \quad t = 1, ..., T \tag{1}$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad t = 2, ..., T. \tag{2}$$

The stochastic process $\{\varepsilon_t\}$ of (2) is taken to satisfy the following conventional (cf. Chang and Park, 2002, and Phillips and Solo, 1992, *inter alia*) stable and invertible linear process-type assumption:

Assumption 1 *Let*

$$\varepsilon_t = C(L)e_t, \; C(L) := \sum_{i=0}^{\infty} C_i L^i, \; C_0 := 1$$

with $C(z) \neq 0$ for all $|z| \leq 1$ and $\sum_{i=0}^{\infty} i|C_i| < \infty$, and where $\{e_t\}$ is a martingale difference sequence with conditional variance σ^2 and $\sup_t E(e_t^4) < \infty$. We also define $\sigma_\varepsilon^2 := E(\varepsilon_t^2)$ and $\omega_\varepsilon^2 := \lim_{T\to\infty} T^{-1}E(\sum_{t=1}^{T} \varepsilon_t)^2 = \sigma^2 C(1)^2$.

Our interest in this paper centres on discriminating between the unit root $[I(1)]$ null hypothesis $H_0 : \rho = 1$ and the alternative $H_1 : |\rho| < 1$ in (1)-(2). While many different unit roots tests are available from what is a vast and continually burgeoning literature (for a recent and comprehensive survey see Phillips and Xiao, 1998), we will restrict our attention to a selection of those that are simple to compute and widely used in practice. The unit root tests we consider are: the ADF t-ratio tests of Dickey and Fuller (1979) and Said and Dickey (1984) based on OLS demeaning or detrending ($DF\text{-}OLS^\mu$, $DF\text{-}OLS^\tau$); the ADF t-ratio tests of Elliott *et al.* (1996) based on quasi-differenced (QD) demeaning or detrending ($DF\text{-}QD^\mu$, $DF\text{-}QD^\tau$); and the von-Neumann ratio tests of Bhargava (1986) and Schmidt and Phillips (1992) based on first differencing to demean or detrend (VN^μ, VN^τ).[3]

[3]Coefficient-based versions of the ADF tests (see, for example, Xiao and Phillips, 1998), were not considered since, as demonstrated in Schmidt and Phillips (1992) and Müller and Elliott (2003), these have the property that, in contrast to t-type ADF tests, their power decreases with the magnitude of the initial condition; cf. Section 4.

The *DF-OLS*^{i} test ($i = \mu,\ \tau$) is based on the t-statistic for testing $\rho = 1$ in the fitted regression equation

$$\hat{u}_t = \rho \hat{u}_{t-1} + \sum_{j=1}^{p} \phi_j \Delta \hat{u}_{t-j} + e_t, \quad t = p+2, ..., T \tag{3}$$

where $\hat{u}_t := y_t - z_t' \hat{\theta}$ is the residual from an OLS regression of y_t on $z_t := 1$, $\theta = \mu$ (*DF-OLS*$^\mu$) or $z_t := (1,\ t)'$, $\theta = (\mu, \beta)'$ (*DF-OLS*$^\tau$). It is assumed that p is chosen according to some consistent model selection procedure, such as the MAIC procedure of Ng and Perron (2001) and Perron and Qu (2007).[4]

The *DF-QD*^{i} test ($i = \mu,\ \tau$) is based on the t-statistic for testing $\rho = 1$ in the fitted regression

$$\tilde{u}_t = \rho \tilde{u}_{t-1} + \sum_{j=1}^{p} \phi_j \Delta \tilde{u}_{t-j} + e_t, \quad t = p+2, ..., T \tag{4}$$

where, on setting $\bar{\rho}_T := 1 - \bar{c}/T$, $\tilde{u}_t := y_t - z_t' \tilde{\theta}$, where $\tilde{\theta}$ is obtained from the QD regression of $\mathbf{y}_{\bar{c}} := (y_1, y_2 - \bar{\rho}_T y_1, ..., y_T - \bar{\rho}_T y_{T-1})'$ on $\mathbf{Z}_{\bar{c}} := (z_1, z_2 - \bar{\rho}_T z_1, ..., z_T - \bar{\rho}_T z_{T-1})'$, where $z_t := 1$ for *DF-QD*$^\mu$, and $z_t := (1,\ t)'$ for *DF-QD*$^\tau$. The value of the QD parameter, $\bar{c}$, is specified according to the form of the deterministic vector z_t and the desired significance level; see Elliott *et al.* (1996) for details. For *DF-QD*$^\mu$, they suggest $\bar{c} = 7$, while for *DF-QD*$^\tau$, $\bar{c} = 13.5$, in each case for tests run at the 0.05 significance level. Different values of $\bar{c}$ should be used for tests run at other significance levels. The lag truncation in (4) again needs to be chosen via a consistent model selection procedure.

For the *VN*^{i} test ($i = \mu,\ \tau$) we construct the ratio statistic

$$VN^i := \frac{-T}{2\hat{\varpi}^2} \frac{\sum_{t=2}^{T} (\Delta S_t^i)^2}{\sum_{t=2}^{T} (S_t^i - \tilde{S}^i)^2} \tag{5}$$

where $S_t^\mu := y_t - y_1$, $\tilde{S}^\mu := 0$, and $S_t^\tau := (T-1)y_t - (t-1)y_T - (T-t)y_1$, $\tilde{S}^\tau := T^{-1} \sum_{t=2}^{T} S_t^\tau$, with in each case $\hat{\varpi}^2$ a consistent estimator of $\sigma_\varepsilon^2 / \omega_\varepsilon^2$. As discussed in Schmidt and Phillips (1992), the *VN* tests are constructed from first-difference demeaned/detrended data (that is, setting $\bar{c} = 0$), and, hence, have the practical advantage over the tests of Elliott *et al.* (1996) that their critical values do not depend on an arbitrary and significance level specific choice of $\bar{c}$.

Each of the tests *DF-OLS*$^\mu$, *DF-QD*$^\mu$ and *VN*$^\mu$ are μ-invariant, but not β-invariant. In contrast, *DF-OLS*$^\tau$, *DF-QD*$^\tau$ and *VN*$^\tau$ are β-invariant. All of the tests are left-sided and so reject the unit root null for large negative values of the associated test statistic.

[4] An alternative approach, which would not alter any of the asymptotic results presented in this paper, would be to use the semi-parametric corrected variants of the DF-type tests, as in Phillips (1987a) and Phillips and Perron (1988), *inter alia.*

3 Unit Root Tests and Uncertainty over the Trend

In this section we consider the impact on the unit root tests outlined above in the situation where we are uncertain with regard to the presence or otherwise of a linear trend; that is, whether $\beta = 0$ or $\beta \neq 0$ in (1). For the purposes of this section of the paper, the initial condition is taken to satisfy the following condition:

Assumption 2 *The initial condition* u_1 *in (2) satisfies* $T^{-1/2}u_1 \xrightarrow{p} 0$.

Assumption 2 is weaker than Condition C of Elliott *et al.* (1996). Both assumptions ensure that the initial condition has no impact on the large sample behaviour of the unit root statistics. This assumption will be subsequently relaxed in Section 4 when we discuss the impact that uncertainty about the initial condition has on unit root testing.

3.1 Asymptotic Behaviour under a Local Trend

We now consider the effect of a local linear trend on the unit root tests of Section 2 under both H_0 and the near-integrated alternative, $H_{1,c} : \rho = \rho_T = 1 - c/T$, where c is a finite non-negative constant ($0 \leq c < \infty$). Notice that $H_{1,c}$ reduces to H_0 for $c = 0$. Under $H_{1,c}$, the relevant (local) Pitman drift on the trend coefficient, β, is given by $\beta_T = \kappa\omega_\varepsilon T^{-1/2}$, with κ a finite constant.[5] The asymptotic properties of the six tests are first summarized in the following lemma.

Lemma 1 *Let* $\{y_t\}$ *be generated according to (1)-(2) and Assumptions 1 and 2, with* $\beta_T = \kappa\omega_\varepsilon T^{-1/2}$. *Then, under* $H_{1,c}$,

$$DF\text{-}OLS^\mu \xrightarrow{d} \frac{\{\frac{\kappa}{2} + W_c^\mu(1)\}^2 - \{\frac{-\kappa}{2} + W_c^\mu(0)\}^2 - 1}{2\sqrt{\int_0^1 \{\kappa(r - \frac{1}{2}) + W_c^\mu(r)\}^2 dr}}, \tag{6}$$

$$DF\text{-}OLS^\tau \xrightarrow{d} \frac{W_c^\tau(1)^2 - W_c^\tau(0)^2 - 1}{2\sqrt{\int_0^1 W_c^\tau(r)^2 dr}},$$

$$DF\text{-}QD^\mu \xrightarrow{d} \frac{\{\kappa + W_c(1)\}^2 - 1}{2\sqrt{\int_0^1 \{\kappa r + W_c(r)\}^2 dr}}, \tag{7}$$

$$DF\text{-}QD^\tau \xrightarrow{d} \frac{W_c^{\tau,\bar{c}}(1)^2 - 1}{2\sqrt{\int_0^1 W_c^{\tau,\bar{c}}(r)^2 dr}},$$

$$VN^\mu \xrightarrow{d} -\frac{1}{2\int_0^1 \{\kappa r + W_c(r)\}^2 dr}, \tag{8}$$

$$VN^\tau \xrightarrow{d} -\frac{1}{2\int_0^1 \{W_c^\mu(r) - (r - \frac{1}{2})W_c(1)\}^2 dr}$$

[5]The results which follow are invariant to replacing κ with $-\kappa$. Also, the scaling of β by ω_ε is simply a convenience measure to ensure that ω_ε does not appear in subsequent expressions for the limit distributions.

where $W_c(r) := \int_0^r e^{-(r-s)c} dW(s)$, *with* $W(r)$ *a standard Wiener process,* $W_c^{\mu}(r) := W_c(r) - \int_0^1 W_c(s)ds$, $W_c^{\tau}(r) := W_c^{\mu}(r) - 12\left(r - \frac{1}{2}\right)\int_0^1 \left(s - \frac{1}{2}\right) W_c(s)ds$, $W_c^{\tau,\bar{c}}(r) := W_c(r) - r\left\{\bar{c}^* W_c(1) + 3(1-\bar{c}^*)\int_0^1 rW_c(r)dr\right\}$, *and* $\bar{c}^* := (1+\bar{c})/(1+\bar{c}+\bar{c}^2/3)$.

Remark 1. Observe that the limiting distributions of the demeaned statistics, *DF-OLS*$^{\mu}$, *DF-QD*$^{\mu}$ and *VN*$^{\mu}$, all depend on the local drift term, κ, while those of the corresponding detrended statistics do not, owing to the β-invariance. Note also that the limiting distributions of the demeaned statistics do not depend on the variance and serial correlation nuisance parameters related to the linear process of Assumption 1 e.g. σ^2, σ_{ε}^2 and ω_{ε}^2 even though a local trend term is omitted. This is because the parametric lagged difference-based and non-parametric kernel-based estimators of these quantities are still consistent under the local trend mis-specification. □

Figures 1(a)-(d) give the asymptotic powers of each of the six tests, in each case using asymptotic critical values appropriate for a nominal 0.05 significance level for a correctly specified model (i.e. $\kappa = 0.0$ for *DF-OLS*$^{\mu}$, *DF-QD*$^{\mu}$ and *VN*$^{\mu}$). We consider, for each of $\kappa = 0.0, 0.25, 0.5, 1.0$, the values $c = \{0, 1, 2, ..., 30\}$. The results were obtained by direct simulation of the limiting distributions in Lemma 1, approximating the Wiener processes using $NIID(0,1)$ random variates, and with the integrals approximated by normalized sums of 1000 steps. In the case of *DF-QD*$^{\tau}$, whose limiting distribution depends on the QD parameter $\bar{c}$, the reported results pertain to $\bar{c} = 13.5$. Here and throughout the paper, simulations were programmed in Gauss 7.0 using 50,000 Monte Carlo replications.

Figure 1(a) shows the results for the case where $\kappa = 0$. Comparing within tests, it highlights the emphatic asymptotic power gains achieved by tests which exclude linear trend terms: *DF-OLS*$^{\mu}$, *DF-QD*$^{\mu}$ and *VN*$^{\mu}$, over their detrended counterparts: *DF-OLS*$^{\tau}$, *DF-QD*$^{\tau}$ and *VN*$^{\tau}$. This is particularly marked when we compare *DF-QD*$^{\mu}$ to *DF-QD*$^{\tau}$, and *VN*$^{\mu}$ to *VN*$^{\tau}$. To put this in a very poignant perspective, as pointed out by one of the referees, applying *DF-QD*$^{\tau}$ when no trend is present yields power roughly equivalent to that which could be obtained from *DF-QD*$^{\mu}$ with a sample of only half the size.[6]

Comparing across tests, we see that *DF-QD*$^{\mu}$ and *VN*$^{\mu}$ have very similar power functions, whereas that of *DF-QD*$^{\tau}$ shows reasonable gains over *VN*$^{\tau}$. Also, it is evident that *DF-OLS*$^{\mu}$ and *DF-OLS*$^{\tau}$ are not competitive in this environment; the power curve of *DF-OLS*$^{\mu}$ is almost identical to that of *DF-QD*$^{\tau}$, and that of *DF-OLS*$^{\tau}$ lies some way below *VN*$^{\tau}$. Efficient inference is therefore provided by *DF-QD*$^{\mu}$, with *VN*$^{\mu}$ a very close competitor.

For $\kappa = 0.25$, in Figure 1(b), matters are rather less clear cut. The power curves of the demeaned and detrended variants of each test intersect. For *DF-QD*$^{\mu}$ and *DF-QD*$^{\tau}$, this occurs (approximately) when $c = 15$; when $c = 23$ for *VN*$^{\mu}$ and *VN*$^{\tau}$; and

[6]This arises from a consideration of the ratio of the Pitman distances for *DF-QD*$^{\mu}$ and *DF-QD*$^{\tau}$ at 50% asymptotic local power: viz, $7/13.5 = 0.52$.

when $c = 24$ for *DF-OLSD*$^\mu$ and *DF-OLSD*$^\tau$.[7] In each case it is the (μ-invariant) demeaned test that has the higher power before the intersection. No single test would warrant recommendation across all c; the effective power envelope being a hybrid of tests consisting of VN^μ for $c < 16$ and *DF-QD*$^\tau$ thereafter. A hybrid of *DF-QD*$^\mu$ for $c < 15$ and *DF-QD*$^\tau$ thereafter would provide a reasonably close approximation to this envelope.

Once $\kappa = 0.5$, as shown in Figure 1(c), matters are largely resolved. It is *DF-QD*$^\tau$ that essentially represents the power envelope at all but the smallest values of c, roughly $c < 4$, which is a region where the power of all tests is below 0.10 in any event. Notice that *DF-QD*$^\mu$ and VN^μ only ever have power below 0.10, and the power curve of *DF-OLS*$^\mu$ rises only very slowly. In the final graph, Figure 1(d), *DF-QD*$^\tau$ unambiguously forms the power envelope for the current analysis, and all three of the demeaned tests have power less than size, with power converging towards zero in c.

3.2 Asymptotic Behaviour Under a Fixed Trend

The behaviour for the demeaned tests in a local trend setting discussed above can also be preempted, to some extent, by considering their behaviour under a fixed non-zero trend of the form $\beta = \kappa\omega_\varepsilon \neq 0$, as is established in the next lemma. The limiting distributions of the corresponding detrended tests of course do not change from those given in Lemma 1; cf. Remark 1.

Lemma 2 *Let $\{y_t\}$ be generated according to (1)-(2) and Assumption 2, under $H_{1,c}$ with $\beta = \kappa\omega_\varepsilon \neq 0$. Moreover, let Assumption 1 hold with $c(L) = 1$ and, correspondingly, $p = 0$ in (3) and (4) and let $\hat{\varpi}^2 = 1$ in (5). Then,*

$$\begin{aligned} DF\text{-}OLS^\mu &\xrightarrow{d} \frac{\int_0^1 (r - \frac{1}{2}) dW_c(r)}{\sqrt{(\kappa^2+1)/12}}, && (9) \\ T^{-1/2} \cdot DF\text{-}QD^\mu &\xrightarrow{p} \sqrt{\frac{3\kappa^2}{\kappa^2+4}}, && (10) \\ T \cdot VN^\mu &\xrightarrow{p} -\frac{3}{2}\left(1 + \frac{1}{\kappa^2}\right). && (11) \end{aligned}$$

Remark 2. The results of Lemma 2 apply to the case where the shocks, ε_t, are serially uncorrelated and no correction for serial correlation is made in constructing the *DF-OLS*$^\mu$, *DF-QD*$^\mu$ and VN^μ statistics. Although the right members of (9), (10) and (11) change under the more general case where serially correlated shocks and/or corrections for serial correlation are made, crucially the *DF-QD*$^\mu$ statistic will still diverge to positive infinity at the rate $O_p(T^{1/2})$, VN^μ will converge to zero at rate $O_p(T^{-1})$, and *DF-OLS*$^\mu$ will remain of $O_p(1)$, and, hence, the quantitative conclusions which follow will not be altered.

[7] The power curves for *DF-OLS*$^\tau$, *DF-QD*$^\tau$ and VN^τ do not vary in κ.

Remark 3. The limiting behaviour of *DF-OLS* $^\mu$ is quite different to those of *DF-QD* $^\mu$ and *VN* $^\mu$. The former possesses a well-defined limiting distribution (which depends on c), while *DF-QD* $^\mu$ diverges to positive infinity as T diverges, and *VN* $^\mu$ converges in probability to zero. Consequently, $\lim_{T\to\infty}\Pr(DF\text{-}QD^\mu < cv_{0.05}) = \lim_{T\to\infty}\Pr(VN^\mu < cv_{0.05}) = 0$, where $cv_{0.05}$ is used in a generic sense to indicate the corresponding 0.05-level asymptotic critical value under H_0 of each test when there is no trend. As a result, the tests based on *DF-QD* $^\mu$ and *VN* $^\mu$ have asymptotic power of zero for all c. For a large enough local trend, this was also effectively seen to be the case in Figure 1.

Remark 4. As regards *DF-OLS* $^\mu$, $\int_0^1 (r-\frac{1}{2})dW_c(r)/\sqrt{1/12}$ is normally distributed with mean zero, and with a variance that is a decreasing function of c. In particular, when $c = 0$, $W_c(r) = W(r)$ and it can be shown that $\int_0^1 (r-\frac{1}{2})dW(r)/\sqrt{1/12}$ is a standard normal distribution, such that, under the conditions of Lemma 2, *DF-OLS* $^\mu \xrightarrow{d} N(0, 1/(\kappa^2+1))$. Consequently, the limiting distribution of *DF-OLS* $^\mu$ is zero-mean normal with a variance always less than unity and, as a result, $\lim_{T\to\infty}\Pr(DF\text{-}OLS^\mu < cv_{0.05}) < \Pr(Z < cv_{0.05}) = 0.002$, where Z is standard normal variate (here $cv_{0.05} = -2.86$). Thus, the test based on *DF-OLS* $^\mu$ also has only trivial asymptotic power. Again, this was evident for a large enough local trend in Figure 1.

Remark 5. As an aside, it is of some interest at this point to compare the limit behaviour of *DF-OLS* $^\mu$, which is based on prior demeaning to obtain μ-invariance, with that of the alternative procedure which constructs the Dickey-Fuller t-ratio from an OLS regression of y_t on a constant term and y_{t-1}. The properties of this latter statistic in the presence of an unattended fixed trend term are analyzed by West (1988). For the case $c = 0$, West shows that its limit distribution is standard normal, and hence invariant to $\kappa \neq 0$. As noted in Remark 4, *DF-OLS* $^\mu$ has a limiting $N(0, 1/(\kappa^2 + 1))$ distribution under H_0. Hence the two statistics, which are often used interchangeably, behave slightly differently under trend mis-specification. However the implications for test power are not dissimilar as, from Remark 4, this alternative procedure has asymptotic size of only 0.002 when using the usual 5% critical value of -2.86. □

The foregoing analysis provides a fairly unequivocal quantitative demonstration of the fact when a trend is absent the practitioner should employ *DF-QD* $^\mu$ (or perhaps *VN* $^\mu$), and would most certainly want to employ *DF-QD* $^\tau$ in place of *DF-QD* $^\mu$ in the presence of a trend. Since we are uncertain as to the presence of a trend, a risk-averse strategy, frequently used in practice, is simply to always apply *DF-QD* $^\tau$. This ensures robust inference, but is clearly at the expense of a significant loss in power when the trend is absent.

In Section 3.3 we next outline a number of other possible unit root testing strategies and analyze their behaviour. What these strategies have in common is that they aim to ensure that, as far as is possible, *DF-QD* $^\mu$ will be selected when a trend is not present and that *DF-QD* $^\tau$ will be selected otherwise. They differ, however, in how this is achieved and in the level of complexity involved. The strategies which we discuss could also be employed using the corresponding *DF-OLS* or *VN* tests, the former of which may be particularly appealing where the initial condition is large; cf. Section 4.

3.3 Test Strategies Based on $DF\text{-}QD^{\mu}$ and $DF\text{-}QD^{\tau}$

The three strategies we consider involve the following. First, *pre-testing* using a statistic to test for the presence of a trend to decide whether to apply $DF\text{-}QD^{\mu}$ or $DF\text{-}QD^{\tau}$. Second, taking a data-dependent *weighted average* of $DF\text{-}QD^{\mu}$ and $DF\text{-}QD^{\tau}$ with a weight function based on an auxiliary trend test statistic that forces a movement between $DF\text{-}QD^{\mu}$ and $DF\text{-}QD^{\tau}$ when a trend is present. Third, a very simple *union of rejections* rule of the form "reject the $I(1)$ null if either $DF\text{-}QD^{\mu}$ or $DF\text{-}QD^{\tau}$ rejects".

Pre-Testing for the Presence of a Trend

This approach involves an initial test of the trend hypothesis that $\beta = 0$ against the alternative $\beta \neq 0$ in (1), using some trend test statistic, t_β. If this two-tailed test fails to reject at a given nominal level (for expositional purposes all tests considered in the remainder of this paper will be run at the nominal 0.05 asymptotic significance level) then $DF\text{-}QD^{\mu}$ is applied; otherwise $DF\text{-}QD^{\tau}$ is applied. This strategy may be written as:

$$PT(|t_\beta|) := DF\text{-}QD^{\mu}\mathbb{I}(|t_\beta| \leq cv_{0.025}) + DF\text{-}QD^{\tau}\mathbb{I}(|t_\beta| > cv_{0.025})$$

where $cv_{0.025}$ is the 0.025 level critical value from the asymptotic distribution of $|t_\beta|$ when $\beta = 0$. Here, if $PT(|t_\beta|) = DF\text{-}QD^{\mu}$, $PT(|t_\beta|)$ is compared with the 0.05 level critical value from the asymptotic distribution of $DF\text{-}QD^{\mu}$ under H_0, i.e. -1.94, and a rejection is recorded if $PT(|t_\beta|) < -1.94$. Otherwise if $PT(|t_\beta|) = DF\text{-}QD^{\tau}$, $PT(|t_\beta|)$ is compared with the 0.05 level critical value from the asymptotic distribution of $DF\text{-}QD^{\tau}$ under H_0, i.e. -2.85, and a rejection is recorded if $PT(|t_\beta|) < -2.85$.

For t_β we will consider three candidate tests. These fall into the class of robust tests for trend in the sense that the asymptotic critical values for testing $\beta = 0$ are the same regardless of whether u_t is $I(1)$ or $I(0)$, which is clearly a highly desirable feature for t_β to possess if it is to be effective as a pre-test in this setting. Specifically, the tests which we consider for t_β are the t_λ and t_λ^{m2} trend tests of Harvey *et al.* (2007), and the *Dan-J* trend test of Bunzel and Vogelsang (2005).[8] Computational details of these three trend tests are provided in Appendix B.

Weighted Average of $DF\text{-}QD^{\mu}$ and $DF\text{-}QD^{\tau}$

We can also consider selecting between $DF\text{-}QD^{\mu}$ and $DF\text{-}QD^{\tau}$ not on the basis of a pre-test, but instead using a data-dependent weighted average of the two, where the weights are a function of the value of some auxiliary statistic mapped on to $[0, 1]$. For the present assuming that β is fixed (not local in the sample size), then as an example, consider the test statistic

$$WA(\lambda) := \lambda DF\text{-}QD^{\mu} + (1 - \lambda)\left(\frac{1.94}{2.85}\right) DF\text{-}QD^{\tau}$$

[8] Ayat and Burridge (2000) report simulation results for a corresponding strategy [S3 in their notation] based on the t-PS1 trend test of Vogelsang (1998). We do not report results for this strategy here as it was found to be dominated by those based on the trend pre-tests considered here.

where λ is a function of an auxiliary trend test statistic such that when $\beta = 0$, $\lambda \xrightarrow{p} 1$ and when $\beta \neq 0$, $\lambda \xrightarrow{p} 0$. Here, $WA(\lambda)$ is compared with the 0.05 level critical value from the asymptotic distribution of $DF\text{-}QD^{\mu}$ under H_0, i.e. -1.94. When $\beta = 0$, $\lambda \xrightarrow{p} 1$ which ensures that $WA(\lambda)$ is asymptotically correctly sized. When $\beta \neq 0$, $\lambda \xrightarrow{p} 0$, and the scaling factor $\left(\frac{1.94}{2.85}\right)$ applied to $DF\text{-}QD^{\tau}$ ensures that $WA(\lambda)$ remains correctly sized in the limit.[9]

Suitable specifications for λ are not difficult to find. Here, for the auxiliary statistic, we follow Vogelsang (1998), and consider the simple Wald statistic for testing $\beta = 0$ in the partially-summed counterpart to regression equation (1)

$$y_{t,s} = \mu t + \beta t_s + u_{t,s} \tag{12}$$

where $y_{t,s} := \sum_{i=1}^{t} y_i$, $t_s := \sum_{i=1}^{t} i$, and $u_{t,s} := \sum_{i=1}^{t} u_i$. The (unscaled) Wald statistic is given by $W := (SSR_R - SSR_U)/SSR_U$, where SSR_R and SSR_U are, respectively, the restricted and unrestricted sums of the squared OLS residuals from (12). Vogelsang (1998) shows that, when $\beta = 0$, W is $O_p(1)$ with limit distributions not depending on ω_ε^2, irrespective of whether u_t is $I(1)$ or $I(0)$ (these distributions are not the same, however). When $\beta \neq 0$, it is also easily shown that $W = O_p(T)$ under $H_{1,c}$.

Next, if we define λ as

$$\lambda(W) := \exp\left(-gT^{-1/2}W\right)$$

for some positive constant g, it is easily seen that when $\beta = 0$, $\lambda(W) \xrightarrow{p} 1$, while if $\beta \neq 0$, $\lambda(W) \xrightarrow{p} 0$. Asymptotically, then, $WA := WA(\lambda(W)) = DF\text{-}QD^{\mu}$ when $\beta = 0$, and $WA = DF\text{-}QD^{\tau}$ when $\beta \neq 0$, so that WA follows the corresponding power curves on Figure 1(a). Consequently, in the limit, WA always behaves exactly like the desired test and, importantly, also allows exact size control.[10]

Union of Rejections

Our third strategy is a very simple decision rule of the form "reject the $I(1)$ null if either $DF\text{-}QD^{\mu}$ or $DF\text{-}QD^{\tau}$ rejects", which we can write as

$$UR := DF\text{-}QD^{\mu}\mathbb{I}(DF\text{-}QD^{\mu} < -1.94) + DF\text{-}QD^{\tau}\mathbb{I}(DF\text{-}QD^{\mu} \geq -1.94).$$

Here, if $UR = DF\text{-}QD^{\mu}$, a rejection is recorded if $UR < -1.94$. Otherwise if $UR = DF\text{-}QD^{\tau}$, a rejection is recorded if $UR < -2.85$. In essence, this is exactly what many practitioners actually do, albeit informally or unconsciously.

Remark 6. Our simple UR strategy is identical, for sufficiently large sample sizes, to the considerably more involved strategy labelled $S2$ of Ayat and Burridge (2000), so

[9]Note that this scaling factor is specific to the chosen asymptotic significance level, in this case 0.05.

[10]We could equally use (suitably scaled) functions of the more complicated trend tests $|t_\lambda|$, $|t_\lambda^{m2}|$ or $|Dan\text{-}J|$ in place of W in $\lambda(\cdot)$. However, since the behaviour of λ relies on the difference in stochastic orders of magnitude of the auxiliary statistic between $\beta = 0$ and $\beta \neq 0$, and not on actual limit distributions, such embellishments are of little benefit.

far as inference on the unit root is concerned. Under the conditions of Lemma 1, t_1 of (17) (the t-statistic for testing $\beta = 0$ against $\beta \neq 0$ in (1) written in first differences), satisfies $t_1 \xrightarrow{d} \kappa + W_c(1)$. Consequently, when $\kappa = c = 0$, t_1 has a standard normal limit distribution, and so $\beta = 0$ would be rejected (at the 0.05 level) if $|t_1| > 1.96$. Now, UR can yield a different outcome on the unit root hypothesis to the $S2$ strategy of Ayat and Burridge (2000) only if $DF\text{-}QD^{\tau} \geq -2.85$ *and* $|t_1| > 1.96$ *and* $DF\text{-}QD^{\mu} < -1.94$; see the flowchart in Appendix C and the discussion therein. Regarding the latter two conditions, in the limit, they are mutually exclusive since

$$\begin{aligned}\lim_{T\to\infty} \Pr(|t_1| &> 1.96 \cap DF\text{-}QD^{\mu} < -1.94)\\ &= \Pr\left(|\kappa + W_c(1)| > 1.96 \cap \frac{\{\kappa + W_c(1)\}^2 - 1}{2\sqrt{\int_0^1 \{\kappa r + W_c(r)\}^2 dr}} < -1.94\right) = 0\end{aligned}$$

which follows since $|\kappa + W_c(1)| > 1.96$ precludes $\{\kappa + W_c(1)\}^2 - 1$ simultaneously taking a negative value. When the trend parameter β is fixed and non-zero, the same equivalence is observed since, as was noted in Remark 3, $\lim_{T\to\infty} \Pr(DF\text{-}QD^{\mu} < -1.94) = 0$ when $\beta \neq 0$. Moreover, UR and $S2$ also cannot differ in the limit under a fixed (trend-) stationary alternative, since here both will terminate at Box 1 of the flow chart in Appendix C with probability one, due to the consistency of $DF\text{-}QD^{\tau}$ in this case.[11]

3.3.1 Asymptotic Behaviour Under a Local Trend

Under the local trend specification $\beta_T = \kappa\omega_\varepsilon T^{-1/2}$, the following lemma gives the limit distribution of the three trend pre-test statistics.

Lemma 3 *Let $\{y_t\}$ be generated according to (1)-(2) and Assumptions 1 and 2, under $H_{1,c}$ with $\beta_T = \kappa\omega_\varepsilon T^{-1/2}$. Then,*

$$\begin{aligned} t_\lambda &\xrightarrow{d} \kappa + W_c(1)\\ t_\lambda^{m2} &\xrightarrow{d} \gamma_\xi \left\{\int_0^1 N_c^2(r)dr\right\}^{-2} \{\kappa + W_c(1)\}\\ Dan\text{-}J &\xrightarrow{d} \sqrt{\frac{12}{\Phi_F}} \left\{\frac{\kappa}{12} + \int_0^1 rW_c(r)dr\right\} \exp\left\{-c_\xi \frac{\int_0^1 N_c^2(r)^2 dr - \int_0^1 N_c^9(r)^2 dr}{\int_0^1 N_c^9(r)^2 dr}\right\}\end{aligned}$$

where: γ_ξ and c_ξ are constants defined in Appendix B, $N_c^h(r)$ denotes the continuous time residual from the projection of $W_c(r)$ onto the space spanned by $\{1, r, ..., r^h\}$, $W_c(r)$ is as defined in Lemma 1, and Φ_F is as defined in Bunzel and Vogelsang (2005,p.385).

Proofs of the results for $|t_\lambda|$ and $|t_\lambda^{m2}|$ are given in Harvey *et al.* (2007) and that for *Dan-J* in Bunzel and Vogelsang (2005). It is clear from Lemma 3 that all three

[11] Unreported simulations also show that UR and $S2$ are all but identical even in small samples.

tests are inconsistent against local trends. However, as their limit distributions depend on κ, this does not mean that $PT(|t_\lambda|)$, $PT(|t_\lambda^{m2}|)$ and $PT(|Dan\text{-}J|)$ will necessarily have poor power under $H_{1,c}$. The situation concerning WA is rather different. This strategy relies on the consistency of the trend test, W. It is easily verified that under these local trend alternatives, W is also not consistent.[12] As a consequence,

$$\lambda(W) := \exp\left(-gT^{-1/2}W\right) \xrightarrow{p} 1$$

and so, asymptotically, $WA = DF\text{-}QD^\mu$.

Figures 2(a)-(f) give the asymptotic size and powers of the various strategies (again obtained by direct simulation methods) for $\kappa = 0.0, 0.25, 0.5, 1.0, 2.0, 4.0$ across the same sequence of local to $I(1)$ alternatives, $H_{1,c}$ as considered in Figure 1.[13] As a point of reference, we also include those of $DF\text{-}QD^\mu$ (which coincides with WA, as shown above) and $DF\text{-}QD^\tau$. Here $PT(|t_\lambda|)$, $PT(|t_\lambda^{m2}|)$ and $PT(|Dan\text{-}J|)$ are based on the two-sided trend pre-tests compared with their respective asymptotic 0.025 level critical values when no trend is present, obtained under the unit root null; that is under $H_0 = H_{1,0}$.

Examining sizes first, when $\kappa = 0.0$, as in Figure 2(a), the theoretical maximum possible asymptotic size of $PT(\cdot)$ and UR is 0.10 by the Bonferroni upper bound, yet it is clear none of the $PT(\cdot)$ tests get particularly near this upper bound, with $PT(|t_\lambda|)$, $PT(|t_\lambda^{m2}|)$ and $PT(|Dan\text{-}J|)$ having sizes of 0.053, 0.071 and 0.064, respectively. The size of UR is somewhat closer to the upper bound, at 0.089. Figures 2(b)-(f) also show that the sizes of $PT(|t_\lambda|)$, $PT(|t_\lambda^{m2}|)$, $PT(|Dan\text{-}J|)$ and UR are monotonic decreasing functions of κ (actually $|\kappa|$), all reaching 0.05 for large enough κ (as can be seen when κ reaches 4.0).

As regards asymptotic power, and abstracting from the slight size issues, for $\kappa = 0.0$ the power curves for $PT(|t_\lambda|)$, $PT(|t_\lambda^{m2}|)$ and $PT(|Dan\text{-}J|)$ are all very similar and extremely close to that of $DF\text{-}GLS^\mu$. The same is true of $PT(|t_\lambda|)$ when $\kappa = 0.25$, although $PT(|t_\lambda^{m2}|)$ and $PT(|Dan\text{-}J|)$ have slightly better power for larger c. Once $\kappa = 0.5$, while the power curve of $PT(|t_\lambda|)$ remains almost identical to that of $DF\text{-}QD^\mu$ and, as a consequence, has near-trivial power, both $PT(|t_\lambda^{m2}|)$ and $PT(|Dan\text{-}J|)$ are much more powerful, and have roughly similar power curves: both, however, fall far short of the power curve of $DF\text{-}QD^\tau$. Once $\kappa = 1.0$, while $PT(|t_\lambda|)$ still closely follows $DF\text{-}QD^\mu$ whose power is now always less than its size, both $PT(|t_\lambda^{m2}|)$ and $PT(|Dan\text{-}J|)$ have made up a reasonable amount of ground on $DF\text{-}QD^\tau$, with $PT(|t_\lambda^{m2}|)$ being the more powerful of the two. This continues through $\kappa = 2.0$, where $PT(|t_\lambda^{m2}|)$ and

[12]It seems implausible that *any* trend test could be consistent under the current local trend specification, since it is chosen precisely to possess the same order of magnitude as u_T.

[13]Harvey *et al.* (2007) estimate linear trend parameters of quarterly log real GDP for twelve industrialized countries over the period 1980:1-2005:2 ($T = 102$). The values of the local trend parameter κ implied by these (significant) estimates all lie in the interval $[3.9, 7.6]$. Since there is general consensus among economists that the presence of a trend in these in these series is a stylized fact, our choice of $\kappa = 4.0$ as an upper limit for the magnitude of local trend in the asymptotic comparison is made to allow portrayal of those situations where there is genuine uncertainty as to whether a trend is present in the data or not.

$DF\text{-}QD^{\tau}$ are now essentially identical. Here, $PT(|t_{\lambda}|)$ is also finally starting to show some power again. By $\kappa = 4.0$, $PT(|t_{\lambda}^{m2}|)$ and $PT(|t_{\lambda}|)$ display essentially the same power curves as $DF\text{-}QD^{\tau}$, with $PT(|Dan\text{-}J|)$ only very slightly behind.

For UR, when $\kappa = 0.0$, the over-sizing of UR (relative to the nominal 0.05 level) puts its power curve slightly outside that of $DF\text{-}QD^{\mu}$. When $\kappa = 0.25$, the intersection of the power curves for $DF\text{-}QD^{\mu}$ and $DF\text{-}QD^{\tau}$ results in the power curve of the union of rejections effectively taking on the higher of the two for each c. Once more, due to the over-sizing, UR then appears to dominate all other strategies. When $\kappa = 0.5$, the power curve of UR is still slightly outside that of $DF\text{-}QD^{\tau}$ for the lower values of c. For $\kappa = 1.0$, UR and $DF\text{-}QD^{\tau}$ display the same power at all but the very lowest values of c. For $\kappa = 2.0, 4.0$, UR is essentially identical to $DF\text{-}QD^{\tau}$.

If we are ambivalent about the size distortion that is evident for small values of κ, then the results from Figure 2 would suggest a clear preference for the strategy UR, followed by $PT(|t_{\lambda}^{m2}|)$, which would seem on balance to be the pre-testing strategy of choice in this local trend environment. Unlike $DF\text{-}QD^{\tau}$, these do not control size exactly, but are nonetheless never badly oversized (or undersized). Notably, however, both these strategies are clearly capable of providing very substantial power gains over the risk-averse tactic of always applying $DF\text{-}QD^{\tau}$, since they adapt to $DF\text{-}QD^{\mu}$ when $\kappa = 0$.

3.3.2 Conservative Strategies

In order to allow us to make more explicit comparisons of the relative behaviour of the unit root test strategies in the local trend environment, we also examine their size-adjusted variants. Since the sizes of the $PT(\cdot)$ and UR tests attain their maxima for $\kappa = 0$, clearly this value represents the calibration point for size adjustments to a nominal 0.05 level, such that the maximum asymptotic size for any strategy across all κ is then 0.05. Their asymptotic sizes are then conservative (below 0.05) for any $\kappa \neq 0$.[14]

Specifically, for UR with the $DF\text{-}QD^{\mu}$ and $DF\text{-}QD^{\tau}$ tests conducted at a significance level γ, their respective asymptotic critical values being denoted as cv_{γ}^{μ} and cv_{γ}^{τ} respectively, we determine a common scaling constant, τ_{γ}, such that, our decision rule

$$UR := DF\text{-}QD^{\mu}\mathbb{I}(DF\text{-}QD^{\mu} < \tau_{\gamma}cv_{\gamma}^{\mu}) + DF\text{-}QD^{\tau}\mathbb{I}(DF\text{-}QD^{\mu} \geq \tau_{\gamma}cv_{\gamma}^{\mu}),$$

that records a rejection if $UR = DF\text{-}QD^{\mu}$ and $UR < \tau_{\gamma}cv_{\gamma}^{\mu}$ or if $UR = DF\text{-}QD^{\tau}$ and $UR < \tau_{\gamma}cv_{\gamma}^{\tau}$, has an asymptotic size of γ when $\kappa = 0$. The constant τ_{γ} is unique and can be approximated via direct simulation of the limiting distributions given for $DF\text{-}QD^{\mu}$ and $DF\text{-}QD^{\tau}$ in Lemma 1, setting $c = 0$ and $\kappa = 0$, using a straightforward grid search. Table 1(a) reports τ_{γ} for the usual significance levels $\gamma = 0.10$, 0.05, 0.01, each calculated to an accuracy of within ± 0.0002 of the stated significance level. For

[14] In unreported simulations, we used a fine grid of values for κ to confirm that the largest asymptotic size of $PT(\cdot)$ and UR was always obtained for $\kappa = 0$. Details are available upon request. We do not implement any size adjustment for WA as it is correctly sized when $\kappa = 0$.

the current 0.05 significance level, $\tau_\gamma = 1.095$, reflecting the fact that the raw UR procedure is asymptotically oversized when $\kappa = 0$. A similar correction was applied to each of the three strategies based on trend pre-testing, with the correction factors again applied to the unit root tests critical values (but not to the trend pre-test critical values).

The results for the conservative testing strategies are shown in Figures 3(a)-(f). When $\kappa = 0$, Figure 3(a) shows that $PT(|t_\lambda|)$ is now the most powerful test, being virtually equivalent to $DF\text{-}QD^\mu$, followed at a little distance by, in order, $PT(|Dan\text{-}J|)$, $PT(|t_\lambda^{m2}|)$ and UR. For $\kappa = 0.25$, UR closely follows the intersection of the power curves for $DF\text{-}QD^\mu$ and $DF\text{-}QD^\tau$: while for small c there is little to choose between UR and the three pre-test procedures, for larger c, UR is clearly more powerful. The power curve for UR lies some way below that of $DF\text{-}QD^\tau$ for $\kappa = 0.5$, but is also substantially above those for $PT(|Dan\text{-}J|)$ and $PT(|t_\lambda^{m2}|)$. Here $PT(|t_\lambda|)$ is almost identical to $DF\text{-}QD^\mu$ and therefore has only trivial power. For $\kappa = 1.0$, UR and $PT(|t_\lambda^{m2}|)$ have essentially the same power curves, with that of $PT(|Dan\text{-}J|)$ slightly lower. Once $\kappa = 2.0$, $PT(|t_\lambda^{m2}|)$ and $PT(|Dan\text{-}J|)$ are slightly more powerful than UR. All three are still substantially more powerful than $PT(|t_\lambda|)$. For $\kappa = 4.0$, $PT(|t_\lambda|)$ now has power close to that of $DF\text{-}QD^\tau$, followed by $PT(|Dan\text{-}J|)$ and $PT(|t_\lambda^{m2}|)$ and UR.

As no strategy has overall dominance across κ, a useful indicator of the strategies' relative performance is to compare the asymptotic integrated conservative powers of the strategies. That is, the area under each curve in Figures 3(a)-(f) for $c = \{1, 2, ..., 30\}$. These are given in Table 2 and scaled relative to the power of the appropriate efficient test in each case. Reading row-wise, the bold entries are the *minimum* integrated relative efficiency that can obtain for a particular test strategy across the values of κ considered. On a maximin criterion, UR is clearly the preferred strategy because it has integrated relative efficiency that never drops below 0.81. For $DF\text{-}QD^\mu$ (WA) and $PT(|t_\lambda|)$ this minimum is zero, and for both $PT(|t_\lambda^{m2}|)$ and $PT(|Dan\text{-}J|)$ it is less than 0.50. Interestingly, because, $DF\text{-}QD^\tau$ has a minimum relative efficiency of 0.74, it would be reasonable to argue that this very simple risk-averse strategy provides the next best alternative after UR.

Table 3 reports a second summary measure of the asymptotic (conservative) powers. This is the *maximum* power loss relative to the efficient tests across $c = \{1, 2, ..., 30\}$. Values in bold typeface are the maximum power loss for a particular test strategy across the values of κ considered. A minimax approach to ranking the strategies leads to the same conclusions as the maximin of relative integrated powers, with UR being preferred overall.

3.3.3 Asymptotic Behaviour Under a Fixed Trend

In the situation of a fixed non-zero trend of the form $\beta = \kappa\omega_\varepsilon \neq 0$, it is straightforward to show that $|t_\lambda|$, $|t_\lambda^{m2}|$ and $|Dan\text{-}J|$ all diverge to positive infinity with increasing T. As a consequence, in the limit, $\Pr(|t_\beta| \leq c_{0.025}) = 0$ (using t_β here in a generic sense) so that $PT(|t_\lambda|)$, $PT(|t_\lambda^{m2}|)$ and $PT(|Dan\text{-}J|)$ are all asymptotically equivalent to $DF\text{-}QD^\tau$. Similarly, since $\lim_{T\to\infty}\Pr(DF\text{-}QD^\mu < -1.94) = 0$ when $\beta \neq 0$, it also holds that UR and $DF\text{-}QD^\tau$ are asymptotically equivalent. Unlike in the case where $\beta = 0$, each of these strategies will also be correctly sized. This limit behaviour is, to all intents and purposes, represented by $\kappa = 4$ in Table 2(f), with the behaviour of the corresponding conservative variants given by $\kappa = 4$ in Table 3(f). As noted above, asymptotically, $WA = DF\text{-}QD^\mu$ when $\beta = 0$ and $WA = DF\text{-}QD^\tau$ when $\beta \neq 0$, so that, being correctly sized, WA attains the power of the efficient test.

At this stage we can draw some reasonably firm conclusions about the competing testing strategies, at least as are apposite from an asymptotic standpoint. If only fixed trends are considered possible, then WA is clearly an optimal strategy as it generally delivers the power envelope from the standpoint of the tests considered here. However, it is also quite obviously a strategy that one should never adopt when considering the possibility of local trends. As we do not wish to take a specific standpoint as to which trend specification is the more plausible, it would then seem difficult to recommend WA: its extremely poor behaviour in the local trend setting far outweighs the modest power gains to be made in the fixed trend case.

We are therefore left to decide between the pre-test and union of rejections strategies. As regards the pre-test strategies, because $PT(|t_\lambda|)$ can behave similarly to $DF\text{-}QD^\mu$ in the local trend case, we must rule it out of contention. Between $PT(|Dan\text{-}J|)$ and $PT(|t_\lambda^{m2}|)$, there is relatively little to choose. The final comparison, then, is between $PT(|Dan\text{-}J|)$ ($PT(|t_\lambda^{m2}|)$) and UR. When $\beta = 0$, there is almost nothing to choose between the three. Under a local trend, however, UR can be considerably more powerful than $PT(|Dan\text{-}J|)$ and $PT(|t_\lambda^{m2}|)$, while the reverse is never true; this message being conveyed in Tables 2 and 3. We therefore conclude, on the basis of their relative asymptotic performance, that UR provides the overall best test strategy of those considered here.

Of course, asymptotic results only provide an imperfect indication of what will happen in finite samples, and this is perhaps particularly pertinent in the case of WA, since we have seen that its limit behaviour in the no trend or fixed trend case is that of the relatively efficient test $DF\text{-}QD^\mu$ or $DF\text{-}QD^\tau$, whereas under the local trend it is that of the effective pathology for this case, $DF\text{-}QD^\mu$. Hence, we now turn to an examination of the behaviour of the strategies as they might be applied in practice, and when only a relatively small sample is available.

3.3.4 Finite Sample Comparisons

Our finite sample simulations are based on the DGP (1) and (2) with $\varepsilon_t \sim NIID(0,1)$, $u_1 = \varepsilon_1$, and a sample size of $T = 100$. We set $\mu = 0$ without loss of generality and consider $\beta = 0, 0.025, 0.05, 0.1, 0.2, 0.4$, where the non-zero values of β are chosen so that for $T = 100$ they coincide with their local trend counterparts used in the asymptotic analysis above.

For our implementations of $DF\text{-}QD^{\mu}$ and $DF\text{-}QD^{\tau}$, the number of lagged difference terms, p, included in the ADF regression is determined by application of the MAIC procedure of Ng and Perron (2001) with maximum lag length set at $p_{\max} = \lfloor 12(T/100)^{1/4} \rfloor$, using the modification suggested by Perron and Qu (2007). For the long run variance estimators implicit in t_{λ} and t_{λ}^{m2}, we employ throughout the quadratic spectral kernel with Newey and West (1994) automatic bandwidth selection adopting a non-stochastic prior bandwidth of $\lfloor 4(T/100)^{2/25} \rfloor$. Finally, for the constant in $\lambda(W)$ we set $g = 3$. Asymptotic 0.05 level critical values are employed throughout.

Figures 4(a)-(f) present the results for the non-conservative variants of the five strategies, along with those for $DF\text{-}QD^{\mu}$ and $DF\text{-}QD^{\tau}$. When $\beta = 0$, $DF\text{-}QD^{\mu}$ and $DF\text{-}QD^{\tau}$ are approximately correctly sized, while the five strategies have sizes that lie very close together; all lying between 0.080 and 0.100. In terms of raw power, UR is the dominant procedure and WA the least well performing, although the latter is still far preferable to $DF\text{-}QD^{\tau}$. For $\beta \neq 0$, the sizes of all of the strategies drop towards 0.05 as β increases. None is ever more powerful than UR, and its advantage over the pre-test procedures can be quite significant; see, for example, the case of $\beta = 0.05$ in Figure 4(c). Perhaps a somewhat unexpected feature seen here is that WA is the best performing strategy aside from UR. Moreover, it is almost identical to UR for $\beta > 0.05$. As such, its behaviour here bears absolutely no resemblance to that of $DF\text{-}QD^{\mu}$, which generally has power well below size. This puts the relevance of our asymptotic results for WA in the local trend case as an indicator of finite sample behaviour into perspective.

The corresponding results using asymptotically conservative critical values for $PT(\cdot)$ and UR are given in Figures 5(a)-(f). Bearing in mind how small the sample size is, this approach works rather well for UR; its size across β is bounded by 0.057, which occurs when $\beta = 0$. For the $PT(\cdot)$ tests, this conservative approach proves less successful for this sample size, however. For example, $PT(|t_{\lambda}|)$ has size of 0.084 when $\beta = 0$ and of 0.078 when $\beta = 0.05$ (which exceed its corresponding asymptotic non-conservative sizes). As a consequence it is seen that the apparent relative improvement in performance of $PT(|t_{\lambda}|)$ compared to UR observed in Figure 1 is somewhat artificial.

Taking together our asymptotic and finite sample results, we find that, despite its lack of elaboration relative to other procedures that involve some means of trend detection, the union of rejections unit root test strategy, UR, offers very robust overall performance. It never suffers badly from size distortions and, uniquely among the strategies considered here at least, has a power profile that never lies far off that of the appropriate QD unit root test under both local and fixed trends. Further, its asymptotic size can also be made conservative in a very simple fashion, and this appears to work well in practice.

4 Unit Root Tests and the Initial Condition

A number of recent papers have highlighted the strong dependence of the power functions of certain unit root tests on the deviation of the initial observation of the series from its underlying deterministic component (see, *inter alia,* Elliott, 1999, Müller and Elliott, 2003, and Elliott and Müller, 2006, Harvey and Leybourne, 2005, 2006). Here we reconsider this issue in the current context of attempting to construct robust unit root test strategies, by replacing Assumption 2, which imposes an asymptotically negligible initial condition, with the following more general assumption.

Assumption 3 *The initial condition, u_1, is generated according to $u_1 = \xi$, where $\xi = \alpha\sqrt{\omega_\varepsilon^2/(1-\rho_T^2)}$ for $\rho_T = 1 - c/T$, $c > 0$, and $\alpha \sim N(\mu_\alpha \mathbb{I}(\sigma_\alpha^2 = 0), \sigma_\alpha^2)$ independently of $\{\varepsilon_t\}_{t=2}^T$.*

In Assumption 3, α controls the magnitude of the initial condition, relative to the standard deviation of a stationary $AR(1)$ process with parameter $|\rho| < 1$ and innovation long-run variance ω_ε^2. The form given for ξ allows the initial condition to be either random and of $O_p(T^{1/2})$, or fixed and of $O(T^{1/2})$. If $\sigma_\alpha^2 > 0$, then the initial condition is random; $\sigma_\alpha^2 = 1$ yields the so-called unconditional initial condition case considered in Elliott (1999). If, on the other hand, $\sigma_\alpha^2 = 0$ then ξ is non-random and of the form given in Müller and Elliott (2003), Elliott and Müller (2006). By considering both the random and fixed scenarios in this way, we try to allow for some flexibility in how the initial condition may be generated.

For the six unit root tests considered in Section 2, where we abstract from the trend issue for demeaned tests by setting $\beta = 0$, the following lemma details their asymptotic behaviour, the proof of which follows directly from results in Phillips (1987b) and Müller and Elliott (2003).

Lemma 4 *Let $\{y_t\}$ be generated according to (1)-(2) and Assumptions 1 and 3, under $H_{1,c}$ with $\beta = 0$. Then,*

$$DF\text{-}OLS^\mu \xrightarrow{d} \frac{K_c^\mu(1)^2 - K_c^\mu(0)^2 - 1}{2\sqrt{\int_0^1 K_c^\mu(r)^2 dr}},$$

$$DF\text{-}OLS^\tau \xrightarrow{d} \frac{K_c^\tau(1)^2 - K_c^\tau(0)^2 - 1}{2\sqrt{\int_0^1 K_c^\tau(r)^2 dr}},$$

$$DF\text{-}QD^\mu \xrightarrow{d} \frac{K_c(1)^2 - 1}{2\sqrt{\int_0^1 K_c(r)^2 dr}},$$

$$DF\text{-}QD^\tau \xrightarrow{d} \frac{K_c^{\tau,\bar{c}}(1)^2 - 1}{2\sqrt{\int_0^1 K_c^{\tau,\bar{c}}(r)^2 dr}},$$

$$VN^\mu \xrightarrow{d} -\frac{1}{2}\left(\int_0^1 K_c(r)^2 dr\right)^{-1},$$

$$VN^\tau \xrightarrow{d} -\frac{1}{2}\left(\int_0^1 \left\{K_c^\mu(r) - (r - \frac{1}{2})K_c(1)\right\}^2 dr\right)^{-1}$$

where

$$K_c(r) := \begin{cases} W(r) & c = 0 \\ \alpha(e^{-rc} - 1)(2c)^{-1/2} + W_c(r) & c > 0 \end{cases}, \tag{13}$$

and

$$\begin{aligned} K_c^{\mu}(r) &:= K_c(r) - \int_0^1 K_c(s)ds, \\ K_c^{\tau}(r) &:= K_c^{\mu}(r) - 12\left(r - \frac{1}{2}\right)\int_0^1 \left(s - \frac{1}{2}\right) K_c(s)ds, \\ K_c^{\tau,\bar{c}}(r) &:= K_c(r) - r\left\{\bar{c}^* K_c(1) + 3(1-\bar{c}^*)\int_0^1 rK_c(r)dr\right\}, \end{aligned}$$

with $W_c(r)$ *and* $\bar{c}^*$ *defined as in Lemma 1.*

Remark 7. Under the null hypothesis $H_0 = H_{1,0}$ the tests do not depend on u_1, so it plays no role in their asymptotic null distributions. It is under the alternative hypothesis, $c > 0$, that the initial condition has an effect. Moreover, notice that setting $\alpha = 0$ in (13) yields the corresponding large sample distributions of the statistics when the initial condition satisfies Assumption 2; cf. Lemma 1 with $\kappa = 0$. □

In Figures 6(a)-(f) we show the asymptotic powers, at the nominal 0.05 level under H_0, of each of the demeaned tests (i.e. *DF-OLS*$^{\mu}$, *DF-QD*$^{\mu}$ and *VN*$^{\mu}$) for $c = 5, 10, 15$, as a function of $\sigma_\alpha, |\mu_\alpha| = \{0.0, 0.1, 0.2, ..., 6.0\}$. We immediately see that with either random or fixed initial conditions, the power curves of *DF-QD*$^{\mu}$ and *VN*$^{\mu}$ (which are very similar throughout) exhibit monotonic decrease in σ_α or $|\mu_\alpha|$, whilst the power of *DF-OLS*$^{\mu}$ increases monotonically. In the random case, *DF-QD*$^{\mu}$ (*VN*$^{\mu}$) has higher power than *DF-OLS*$^{\mu}$ when (approximately) $\sigma_\alpha \leq 1.5, 1.3, 1.0$ for $c = 5, 10, 15$, respectively. In all three cases, the extent of power dominance of *DF-QD*$^{\mu}$ (*VN*$^{\mu}$) over *DF-OLS*$^{\mu}$ increases as σ_α shrinks towards zero. For (approximately) $\sigma_\alpha \geq 1.5, 1.3, 1.0$ and $c = 5, 10, 15$, respectively, these rankings reverse, and the power advantage of *DF-OLS*$^{\mu}$ over *DF-QD*$^{\mu}$ (*VN*$^{\mu}$) steadily grows with σ_α. In the fixed case, the results are qualitatively very similar, though the crossover points are now (approximately) $|\mu_\alpha| = 1.0, 0.9, 0.9$ for $c = 5, 10, 15$, respectively. A key feature here is the drastic speed with which the power of *DF-QD*$^{\mu}$ (*VN*$^{\mu}$) approaches zero with $|\mu_\alpha|$; it is effectively zero for $|\mu_\alpha| \geq 2$. Figures 7(a)-(f) give the results for the corresponding detrended tests (*DF-OLS*$^{\tau}$, *DF-QD*$^{\tau}$ and *VN*$^{\tau}$), for $c = 10, 15, 20$. Qualitatively, the relative performance of *DF-OLS*$^{\tau}$ and *DF-QD*$^{\tau}$ is very similar to that in the demeaned case. Minor differences are that the crossover points are slightly higher in σ_α and $|\mu_\alpha|$, and, in the fixed case, that the power of *DF-QD*$^{\tau}$ does now not reach zero until around $|\mu_\alpha| \geq 3.5$. In addition, *VN*$^{\tau}$ now has noticeably lower power than *DF-QD*$^{\tau}$ for small σ_α or $|\mu_\alpha|$, but the power of *VN*$^{\tau}$ decreases rather more slowly, not reaching zero in the fixed case until around $|\mu_\alpha| = 6.0$.

4.1 A Union of Rejections Testing Strategy

Given the clear results of Figures 6 and 7, it seems sensible to consider whether it is possible to devise a testing strategy which, for small values of σ_α or $|\mu_\alpha|$, captures the power advantages of *DF-QD*$^\mu$ (*DF-QD*$^\tau$) over *DF-OLS*$^\mu$ (*DF-OLS*$^\tau$) and, at the same time, exploits the reverse relationship that exists between the tests' power when σ_α or $|\mu_\alpha|$ is large.

An obvious, and very straightforward, candidate is to take the union of rejections formed from *DF-QD*$^\mu$ and *DF-OLS*$^\mu$ in the no-trend case, and that of *DF-QD*$^\tau$ and *DF-OLS*$^\tau$ in the trend case. Considering tests run using 0.05 level asymptotic critical values, these are respectively given by

$$UR^\mu := \textit{DF-QD}^\mu \mathbb{I}(\textit{DF-QD}^\mu < -1.94) + \textit{DF-OLS}^\mu \mathbb{I}(\textit{DF-QD}^\mu \geq -1.94)$$

where if $UR^\mu =$*DF-QD*$^\mu$, a rejection is recorded if $UR^\mu < -1.94$; otherwise if $UR^\mu =$*DF-OLS*$^\mu$, a rejection is recorded if $UR^\mu < -2.86$; and

$$UR^\tau := \textit{DF-QD}^\tau \mathbb{I}(\textit{DF-QD}^\tau < -2.85) + \textit{DF-OLS}^\tau \mathbb{I}(\textit{DF-QD}^\tau \geq -2.85)$$

where if $UR^\tau =$*DF-QD*$^\tau$, a rejection is recorded if $UR^\tau < -2.85$; otherwise if $UR^\tau =$*DF-OLS*$^\tau$, a rejection is recorded if $UR^\tau < -3.42$. Again, these strategies are not size controlled; for $c = 0$, the asymptotic size of UR^μ is 0.089 and that of UR^τ is 0.080. However, we can correct these along the same lines as Section 3.3.2. For $i = \mu, \tau$, we find the scaling constants τ^i_γ, for *DF-QD*^{i} and *DF-OLS*^{i} conducted at a significance level γ, with respective asymptotic critical values being $cv^{Q,i}_\gamma$ and $cv^{O,i}_\gamma$, such that

$$UR^i := \textit{DF-QD}^i \mathbb{I}(\textit{DF-QD}^\mu < \tau^i_\gamma cv^{Q,i}_\gamma) + \textit{DF-OLS}^i \mathbb{I}(\textit{DF-QD}^i \geq \tau^i_\gamma cv^{O,i}_\gamma)$$

which records a rejection if $UR^i =$*DF-QD*^{i} and $UR^i < \tau^i_\gamma cv^{Q,i}_\gamma$ or if $UR^i =$*DF-OLS*^{i} and $UR^i < \tau^i_\gamma cv^{O,i}_\gamma$ has an asymptotic size of γ. These constants are given in Table 1(b) for $\gamma = 0.10, 0.05, 0.01$. For the current 0.05 significance level, we find $\tau^\mu_\gamma = 1.095$ and $\tau^\tau_\gamma = 1.058$. Notice that, unlike the conservative tests of section 3.3.2, this yields testing strategies which are correctly sized in the limit, regardless of the value of σ_α or $|\mu_\alpha|$, since the (exact) null distributions of the tests involved do not depend on these parameters.

The asymptotic power curves for the union of rejections in the mean case are shown in Figure 6. Both the raw and size-corrected variants are given. Also shown are the $\hat{Q}^\mu(10, 3.8)$ test of Elliott and Müller (2006), the τ^μ_{AV} test of Harvey and Leybourne (2005), and the $\hat{Q}^\mu_{\text{AV}}$ test of Harvey and Leybourne (2006). The first of these is a preferred variant of a family of tests which maximize weighted average power over different initial conditions for a point alternative of the near integration parameter c. Asymptotically at least, this test is admissible in that no other test can have power higher than $\hat{Q}^\mu(.,.)$ across all initial conditions and c. The τ^μ_{AV} test is based on taking a data-dependent weighted average of *DF-OLS*$^\mu$ and *DF-QD*$^\mu$ where the weight function depends on an auxiliary statistic designed to capture information on the initial condition; $\hat{Q}^\mu_{\text{AV}}$ is constructed similarly, but averages *DF-OLS*$^\mu$ and $\hat{Q}^\mu(10, 3.8)$.

As we would conjecture, the power curve of UR^{μ} tends to mimic that of $DF\text{-}QD^{\mu}$ for small σ_{α} or $|\mu_{\alpha}|$, and mimics that of $DF\text{-}OLS^{\mu}$ for large σ_{α} or $|\mu_{\alpha}|$. Thus, it captures the power advantage of $DF\text{-}QD^{\mu}$ relative to $DF\text{-}OLS^{\mu}$ when σ_{α} or $|\mu_{\alpha}|$ is small, whilst avoiding the severe power losses that $DF\text{-}QD^{\mu}$ exhibits relative to $DF\text{-}OLS^{\mu}$ when σ_{α} or $|\mu_{\alpha}|$ is large. This is particularly emphasized for fixed initial conditions. The size-corrected variant of UR^{μ} obviously has somewhat lower power across all initial conditions, but the qualitative picture is the same as its uncorrected counterpart.

Comparing size-corrected UR^{μ} and the admissible test $\hat{Q}^{\mu}$, we find that neither dominates the other overall (and hence we might consider that UR^{μ} is itself "approximately" admissible). Generally speaking, UR^{μ} is the more powerful of the two when $c = 5$ and *vice versa* when $c = 15$. The ranking largely depends on the region of σ_{α} or $|\mu_{\alpha}|$ under consideration, but what does emerge reasonably clearly is that for $\sigma_{\alpha}, |\mu_{\alpha}| \leq 1$, regions of what might be considered to give rise to small (or non-extreme) initial observations, UR^{μ} is generally the more powerful of the two, particularly so for σ_{α} or $|\mu_{\alpha}|$ close to zero. In the intermediate range $\hat{Q}^{\mu}$ tends to be the more powerful, and then UR^{μ} typically dominates once more for large values. Notably, for $c = 5$, the power of $\hat{Q}^{\mu}$ appears to be a decreasing function of σ_{α} or $|\mu_{\alpha}|$ outside of small values, which suggests that, in common with $DF\text{-}QD^{\mu}$, it will in certain circumstances have power below size for extreme initial observations. This behaviour contrasts with that of UR^{μ}. Notice that τ^{μ}_{AV} displays behaviour which in general quite similar to that of the size-corrected variant of UR^{μ} , which may be expected given that both involve combinations of the same tests, $DF\text{-}OLS^{\mu}$ and $DF\text{-}QD^{\mu}$; overall $\hat{Q}^{\mu}_{\mathrm{AV}}$ behaves rather more like $\hat{Q}^{\mu}$ than UR^{μ}.

Results for the trend case are shown in Figure 7, where $\hat{Q}^{\tau}(15, 3.968)$ is the suggested test from Elliott and Müller (2006). In relation to $DF\text{-}QD^{\tau}$ and $DF\text{-}OLS^{\tau}$, we see that the behaviour of UR^{τ} is very similar to that seen in Figure 6. Comparing (size-corrected) UR^{τ} and $\hat{Q}^{\tau}$, it is again the case that UR^{τ} generally performs better for the lower and higher values of σ_{α} and $|\mu_{\alpha}|$, with the opposite being true for intermediate values, although the differences in power between the two across σ_{α} and $|\mu_{\alpha}|$ are less emphatic than in the mean case. Again, $\tau^{\tau}_{\mathrm{AV}}$ and UR^{τ} behave fairly similarly.

4.2 Finite Sample Comparisons

Finite sample simulations are again based on the DGP (1) and (2) with $\varepsilon_t \sim NIID(0, 1)$ and $T = 100$. We set $\mu = \beta = 0$, $\rho_T = 1 - c/T$ and consider the same sets of values for σ_{α}, $|\mu_{\alpha}|$and c as underlie Figures 6 and 7. The method of lagged difference determination used for $DF\text{-}QD^{i}$ and $DF\text{-}OLS^{i}$ $(i = \mu, \tau)$ is MAIC as detailed in section 3.3.4 above; the same fitted models are used to construct a (parametric) estimate of the long run variance as required by $\hat{Q}^{i}$.

In the mean case, the finite sample sizes of $DF\text{-}QD^{\mu}$, $DF\text{-}OLS^{\mu}$, UR^{μ}, size-corrected UR^{μ}, $\hat{Q}^{\mu}$, τ^{μ}_{AV} and $\hat{Q}^{\mu}_{\mathrm{AV}}$ at asymptotic 0.05 level critical values are, respectively, 0.059, 0.042, 0.090, 0.053, 0.022, 0.061 and 0.025. Notice, therefore, that $\hat{Q}^{\mu}$ and $\hat{Q}^{\mu}_{\mathrm{AV}}$ are quite badly undersized. Powers, based on these sizes, are given in Figures 8(a)-(f).

The finite sample relationships between $DF\text{-}QD^{\mu}$, $DF\text{-}OLS^{\mu}$, UR^{μ} and size-corrected UR^{μ} across σ_{α}, $|\mu_{\alpha}|$ and c very closely follow those of their asymptotic counterparts. However, the finite sample powers of $\hat{Q}^{\mu}$ and $\hat{Q}^{\mu}_{\mathrm{AV}}$ are always some way below their asymptotic levels, reflecting their undersizing here.

In the trend case, the finite sample sizes of $DF\text{-}QD^{\tau}$, $DF\text{-}OLS^{\tau}$, UR^{τ}, size-corrected UR^{τ}, $\hat{Q}^{\tau}$, τ^{μ}_{AV} and $\hat{Q}^{\mu}_{\mathrm{AV}}$ are 0.051, 0.038, 0.069, 0.045, 0.012, 0.052 and 0.019, respectively. Figure 9 demonstrates that, while the other test procedures have finite sample power curves that behave quite similarly to the asymptotic situation, its very small size for sample of size 100, dictates that the power of $\hat{Q}^{\tau}$ is actually much lower than the asymptotic theory predicts. Notice, in particular, that the size-corrected variant of UR^{τ} dominates, and often by a considerable margin, $\hat{Q}^{\tau}$ on power in every case reported.

As in the case where we are uncertain over the presence or otherwise of a linear trend, it appears that a union of rejections approach can provide a very decent practical strategy for unit root testing, this time in the context of uncertainty about the initial condition and, hence, uncertainty over whether to best employ $DF\text{-}QD^{i}$ or $DF\text{-}OLS^{i}$. In terms of asymptotic power, this simple approach seems to compete well with considerably more sophisticated testing procedures, such as those of Elliott and Müller (2006). Also, the asymptotic size-correction for the union of rejections again seems to work quite respectably in finite samples, suggesting that the strategy should be useful in practice, in particular avoiding the undersizing problems, and resultant poor finite sample power properties, seen with the tests of Elliott and Müller (2006).

5 Conclusions

In this paper we have investigated the impact that uncertainty over the presence or otherwise of a linear trend and over the magnitude of the initial condition has on commonly used unit root tests, and investigated new procedures which attempt to retain good power properties in the presence of such uncertainty. We have focused on classical methods of inference, rather than Bayesian approaches to unit root testing and the associated issues of model selection raised in this paper, for which we direct the interested reader to, *inter alia*, Phillips (1991a,b) and Phillips and Ploberger (1994).

Maintaining the assumption of an asymptotically negligible initial condition, we found that the quasi-difference (QD) detrended ADF test, $DF\text{-}QD^{\tau}$, of Elliott *et al.* (1996) outperformed the other β-invariant tests considered - the OLS detrended ADF $DF\text{-}OLS^{\tau}$ test and the first-difference detrended VN-type test - but was much less powerful than the corresponding QD demeaned test, $DF\text{-}QD^{\mu}$, in the absence of a trend (and in some cases where the trend was very small in magnitude). However, where a non-trivial trend was present all of the tests based on demeaned data had negligible power. We consequently investigated a variety of strategies that aimed to select $DF\text{-}QD^{\tau}$ when a trend was present and $DF\text{-}QD^{\mu}$ otherwise, with the aim of achieving a procedure which delivered the best possible power in both worlds. We suggested procedures involving: pre-testing for the presence of a trend, using $DF\text{-}QD^\tau$ if the pre-test rejected and $DF\text{-}QD^\mu$ otherwise; basing inference on a weighted average of $DF\text{-}QD^\tau$ and $DF\text{-}QD^\mu$, and, finally, a very simple union of rejections decision rule whereby the unit root null is rejected if either $DF\text{-}QD^\tau$ or $DF\text{-}QD^\mu$ yields a rejection. We reported asymptotic and finite sample evidence which suggested that the simple union of rejections decision rule generally outperformed the other more elaborate strategies.

Where the initial condition of the process was not asymptotically negligible, we found that the $DF\text{-}QD^\tau$ and $DF\text{-}QD^\mu$ tests performed very poorly with their power against a given alternative rapidly decreasing towards zero as the magnitude of the initial observation was increased. Similar patterns (albeit less severe in the detrended case) were also seen for the VN-type tests. In contrast, the $DF\text{-}OLS^\tau$ and OLS demeaned equivalent, $DF\text{-}OLS^\mu$, tests showed increases in power, other things equal, as the magnitude of the initial condition increased. Consequently, and in the same spirit as for our approach to the uncertain trend case, we proposed a union of rejections decision rule, whereby the unit root null was rejected if either of $DF\text{-}QD^\tau$ and $DF\text{-}OLS^\tau$ rejected in the maintained trend case, or if either of $DF\text{-}QD^\mu$ and $DF\text{-}OLS^\mu$ rejected in the no trend case. Asymptotic and finite sample evidence suggested that this procedure was again highly effective, despite its simplicity relative to other available approaches.

To conclude we briefly discuss some limitations of our approach with associated directions for further research. First, in our analysis we have treated the uncertainty over the presence or otherwise of a linear trend and over the magnitude of the initial value in isolation from each other which may be undesirable in practice. It would be interesting therefore to develop testing procedures which could simultaneously maintain good power properties under both forms of uncertainty. It seems possible, given the results reported in this paper, that an approach based, for example, on a simple four-way union of rejections of tests could usefully be explored here. This would entail the decision rule "reject the $I(1)$ null if either $DF\text{-}QD^\mu$ or $DF\text{-}QD^\tau$ or $DF\text{-}OLS^\mu$ or $DF\text{-}OLS^\tau$ rejects". Second, we have limited our deterministic trend function to be either a linear trend or a constant mean. In reality trend behaviour will inevitably be more complex than this and it would be interesting to extend the procedures outlined in this paper to allow for the possible presence of more general deterministic trend cases. Harris *et al.* (2007) consider the problem of unit root testing in the presence of a possible broken trend and show that approaches based on pre-testing (in their case for the presence of a break in trend) carry over well into that context. Further extensions to a more general class of potential trend functions including, for example, the possibility of higher-order polynomial trends would constitute a far from trivial extension of the results in this paper, and the level of generality attainable would clearly be limited by the results in Phillips (1998). This topic is currently being investigated by the authors.

Appendix A

Due to invariance, in (1) we may set $\mu = 0$ in what follows. Also, nothing of asymptotic consequence is lost if, in the following algebra, we make the simplifying assumption that $\varepsilon_t = e_t$, such that $\omega_\varepsilon^2 = \sigma_\varepsilon^2 = \sigma^2$, allowing us to impose $p = 0$ in in (3) and (4) and $\hat{\varpi}^2 = 1$ in (5).

Proof of Lemma 1: The proofs of the limit distributions for *DF-OLS*$^\tau$, *DF-QD*$^\tau$ and *VN*$^\tau$ are standard; see, for example, Hamilton (1994), Elliott *et al.* (1996) and Schmidt and Phillips (1992). We will derive the result for *DF-QD*$^\mu$ given in (7). Those for *DF-OLS*$^\mu$ and *VN*$^\mu$ in (6) and (8) then follow using very similar algebra. First write *DF-QD*$^\mu$ as

$$DF\text{-}QD^\mu := T(\tilde{\rho} - 1)(T^{-2}\sum_{t=2}^{T} \tilde{u}_{t-1}^2)^{1/2}/\tilde{\sigma}.$$

where

$$T(\tilde{\rho} - 1) := \frac{T^{-2}\sum_{t=2}^{T} \tilde{u}_{t-1}\Delta\tilde{u}_t}{T^{-3}\sum_{t=2}^{T} \tilde{u}_{t-1}^2}$$

and $\tilde{\sigma}^2 := T^{-1}\sum_{t=2}^{T}(\tilde{u}_t - \tilde{\rho}\tilde{u}_{t-1})^2$. We therefore need to establish first the large sample behaviour of $T(\tilde{\rho} - 1)$. Now,

$$T(\tilde{\rho} - 1) = \frac{T^{-1}\sum_{t=2}^{T}\{T^{-1/2}\kappa\sigma(t-1) + u_{t-1}\}(T^{-1/2}\kappa\sigma + \Delta u_t)}{T^{-2}\sum_{t=2}^{T}\{T^{-1/2}\kappa\sigma(t-1) + u_{t-1}\}^2} + o_p(1)$$

$$= \frac{\kappa^2\sigma^2 T^{-2}\sum_{t=2}^{T} t + \kappa\sigma T^{-3/2}\sum_{t=2}^{T} u_{t-1} + T^{-1}\sum_{t=2}^{T} u_{t-1}\Delta u_t + \kappa\sigma T^{-3/2}\sum_{t=2}^{T} t\Delta u_t}{\kappa^2\sigma^2 T^{-3}\sum_{t=2}^{T} t^2 + 2\kappa\sigma T^{-5/2}\sum_{t=2}^{T} t u_{t-1} + T^{-2}\sum_{t=2}^{T} u_{t-1}^2} + o_p(1).$$

Using the fact that under the conditions of the lemma $T^{-1/2}u_{\lfloor Tr\rfloor} \xrightarrow{d} \sigma W_c(r)$, applications of the continuous mapping theorem [CMT] shows that

$$T(\tilde{\rho} - 1) \xrightarrow{d} \frac{\kappa^2/2 + \kappa\int_0^1 W_c(r)dr + \int_0^1 W_c(r)dW_c(r) + \kappa\int_0^1 r dW_c(r)}{\kappa^2/3 + 2\kappa\int_0^1 rW_c(r)dr + \int_0^1 W_c(r)^2 dr}. \tag{14}$$

The right member of (14) can be simplified, using an application of the Itô integral, to yield the result that

$$T(\tilde{\rho} - 1) \xrightarrow{d} \frac{\frac{1}{2}[\{\kappa + W_c(1)\}^2 - 1]}{\int_0^1\{\kappa r + W_c(r)\}^2 dr}. \tag{15}$$

Next we obtain from the CMT that $T^{-2}\sum_{t=2}^{T}\tilde{u}_{t-1}^2 \xrightarrow{d} \int_0^1\{\kappa r + W_c(r)\}^2 dr$, which is the denominator of the right member of (15). The result in (7) then follows directly from an application of the CMT and the fact that $\tilde{\sigma}^2 \xrightarrow{p} \sigma^2$, which is straightforward to show.

Proof of Lemma 2: We derive the results for *DF-OLS*$^\mu$ in (9) and *DF-QD*$^\mu$ in (10). That for *VN*$^\mu$ in (11) follows in a very similar fashion to (10).
First write *DF-OLS*$^\mu$ as

$$\textit{DF-OLS}^\mu := T^{3/2}(\hat{\rho}-1)(T^{-3}\sum_{t=2}^{T}\hat{u}_{t-1}^2)^{1/2}/\hat{\sigma}$$

where $\hat{u}_t = y_t - \bar{y} = \kappa\sigma(t-\bar{t}) + u_t - \bar{u}$,

$$T^{3/2}(\hat{\rho}-1) := \frac{T^{-3/2}\sum_{t=2}^{T}\hat{u}_{t-1}\Delta\hat{u}_t}{T^{-3}\sum_{t=2}^{T}\hat{u}_{t-1}^2}$$

and $\hat{\sigma}^2 := T^{-1}\sum_{t=2}^{T}(\hat{u}_t - \hat{\rho}\hat{u}_{t-1})^2$. We find that

$$\begin{aligned} T^{3/2}(\hat{\rho}-1) &= \frac{T^{-3/2}\sum_{t=2}^{T}\{\kappa\sigma(t-\bar{t}-1)+u_{t-1}-\bar{u}\}(\kappa\sigma+\Delta u_t)}{T^{-3}\sum_{t=2}^{T}\{\kappa\sigma(t-\bar{t}-1)+u_{t-1}-\bar{u}\}^2} \\ &= \frac{\kappa\sigma T^{-3/2}\sum_{t=2}^{T}(t-\bar{t})\Delta u_t)}{\kappa^2\sigma^2 T^{-3}\sum_{t=2}^{T}(t-\bar{t})^2} + o_p(1) \\ &\xrightarrow{d} 12\kappa^{-1}\int_0^1 (r-\frac{1}{2})dW_c(r). \end{aligned}$$

In similar fashion we can show that

$$\begin{aligned} \hat{\sigma}^2 &= T^{-1}\sum_{t=2}^{T}\{\Delta\hat{u}_t - (\hat{\rho}-1)\hat{u}_{t-1}\}^2 \\ &= T^{-1}\sum_{t=2}^{T}\{\kappa\sigma + \Delta u_t - T^{3/2}(\hat{\rho}-1)T^{-3/2}\hat{u}_{t-1}\}^2 \\ &= T^{-1}\sum_{t=2}^{T}(\kappa\sigma+\Delta u_t)^2 + o_p(1) \xrightarrow{p} \sigma^2(\kappa^2+1). \end{aligned}$$

Taken together, these results yield that

$$\begin{aligned} \textit{DF-OLS}^\mu &\xrightarrow{d} \frac{12\kappa^{-1}\int_0^1(r-\frac{1}{2})dW_c(r)(\kappa^2\sigma^2/12)^{1/2}}{\{\sigma^2(\kappa^2+1)\}^{1/2}} \\ &= \frac{1}{(\kappa^2+1)^{1/2}}\frac{\int_0^1(r-\frac{1}{2})dW_c(r)}{(1/12)^{1/2}}, \end{aligned}$$

establishing the result in (9).
For *DF-QD*$^\mu$, first write

$$T^{-1/2}\textit{DF-QD}^\mu := T(\tilde{\rho}-1)(T^{-3}\sum_{t=2}^{T}\tilde{u}_{t-1}^2)^{1/2}/\tilde{\sigma}.$$

Now, since,

$$\begin{aligned}\tilde{\mu} &= \frac{y_1 + (1-\bar{\rho}_T)\sum_{t=2}^{T}(y_t - \bar{\rho}_T y_{t-1})}{1+(T-1)(1-\bar{\rho}_T)^2} \\ &= \frac{y_1 + \bar{c}T^{-1}\sum_{t=2}^{T}(\Delta y_t + \bar{c}T^{-1}y_{t-1})}{1+(T-1)(\bar{c}T^{-1})^2} = y_1 + o_p(1)\end{aligned}$$

we have that $\tilde{u}_t = y_t - y_1 + o_p(1)$. Consequently,

$$\begin{aligned}T(\tilde{\rho}-1) &= \frac{T^{-2}\sum_{t=2}^{T} y_{t-1}\Delta y_t}{T^{-3}\sum_{t=2}^{T} y_{t-1}^2} + o_p(1) \\ &= \frac{T^{-2}\sum_{t=2}^{T}\{\kappa\sigma(t-1)+u_{t-1}\}(\kappa\sigma+\Delta u_t)}{T^{-3}\sum_{t=2}^{T}\{\kappa\sigma(t-1)+u_{t-1}\}^2} + o_p(1) \\ &= \frac{\kappa^2\sigma^2 T^{-2}\sum_{t=2}^{T}(t-1)}{\kappa^2\sigma^2 T^{-3}\sum_{t=2}^{T}(t-1)^2} + o_p(1) \xrightarrow{p} \frac{3}{2}.\end{aligned}$$

Moreover,

$$\begin{aligned}\tilde{\sigma}^2 &= T^{-1}\sum_{t=2}^{T}\{\Delta y_t - (\tilde{\rho}-1)y_{t-1}\}^2 + o_p(1) \\ &= T^{-1}\sum_{t=2}^{T}\{\kappa\sigma + \Delta u_t - T(\tilde{\rho}-1)T^{-1}y_{t-1}\}^2 + o_p(1) \\ &= T^{-1}\sum_{t=2}^{T}\{\kappa\sigma + \Delta u_t - 3/2.\kappa\sigma T^{-1}t\}^2 + o_p(1) \xrightarrow{p} \sigma^2(\kappa^2/4+1).\end{aligned}$$

So, we have that $T^{-1/2}DF\text{-}QD^{\mu} \xrightarrow{p} \frac{3}{2}(\kappa^2\sigma^2/3)^{1/2}/\{\sigma^2(\kappa^2/4+1)\}^{1/2}$ which simplifies to the expression in (10) of the main text.

Appendix B

t_λ and t_λ^{m2}

The t_λ statistic of Harvey *et al.* (2007) is a switching-based strategy that attains the local limiting Gaussian power envelope for testing $\beta = 0$ against $\beta \neq 0$ in (1) irrespective of whether u_t is $I(1)$ or $I(0)$. The test statistic is also asymptotically standard normal under both $I(1)$ and $I(0)$ errors for u_t. It is calculated as

$$t_\lambda := (1 - \lambda^*)t_0 + \lambda^* t_1$$

where, if $\hat{\mu}$ and $\hat{\beta}$ denote the OLS estimators from (1), t_0 is the t-ratio

$$t_0 := \frac{\hat{\beta}}{\sqrt{\hat{\omega}_u^2 / \sum_{t=1}^T (t - \bar{t})^2}} \tag{16}$$

with $\hat{\omega}_u^2$ a long run variance estimator formed using $\hat{u}_t := y_t - \hat{\mu} - \hat{\beta}t$. Moreover, t_1 is the t-ratio

$$t_1 := \frac{\breve{\beta}}{\sqrt{\breve{\omega}_v^2 / T}} \tag{17}$$

where $\breve{\beta}$ is the OLS estimator of β in (1) estimated in first differences i.e. from

$$\Delta y_t = \beta + v_t, \quad t = 2, ..., T$$

and $\breve{\omega}_v^2$ is a long run variance estimator based on $\breve{v}_t := \Delta y_t - \breve{\beta}$. The switch function λ^* is defined as

$$\lambda^* := \exp\left(-0.00025\left(\frac{DF\text{-}QD^\tau}{\hat{\eta}_\tau}\right)^2\right),$$

where $\hat{\eta}_\tau$ is the Kwiatkowski *et al.* (1992) [KPSS] stationarity test statistic

$$\hat{\eta}_\tau := \frac{\sum_{t=1}^T \left(\sum_{i=1}^t \hat{u}_i\right)^2}{T^2 \hat{\omega}_u^2}$$

Harvey *et al.* (2007) show that a modified variant of t_λ, denoted t_λ^{m2}, can provide a more powerful test of the trend hypothesis than t_λ when u_t is local to $I(1)$. This is given by

$$t_\lambda^{m2} := (1 - \lambda^*)t_0 + \lambda^* t_1^{m2}$$

where

$$t_1^{m\,2} := \gamma_\xi R_2 t_1$$

and

$$R_2 := \left(\frac{\breve{\omega}_v^2}{T^{-1}\hat{\sigma}_u^2} \right)^2$$

with $\hat{\sigma}_u^2 := (T-2)^{-1} \sum_{t=1}^T \hat{u}_t^2$. Here γ_ξ is is a constant chosen so that, at a given significance level ξ, t_λ^{m2} has the same standard normal critical value under both $I(0)$ and $I(1)$ errors. For a two tailed 0.05 level test, $\gamma_\xi = 0.00115$.

Dan-J

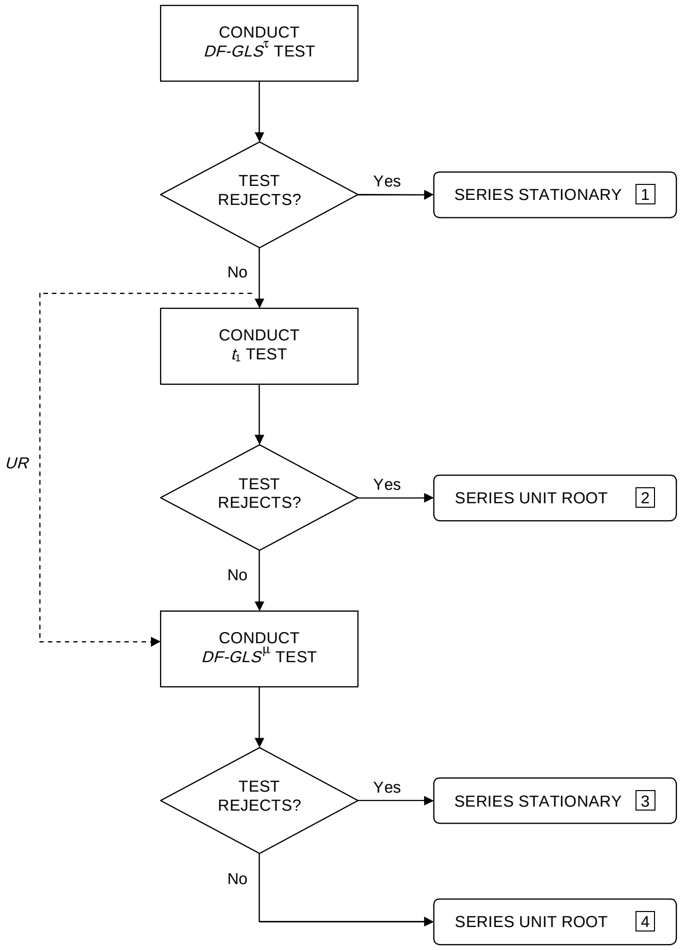
This is calculated as

$$Dan\text{-}J := t_0' \exp(-c_\xi J)$$

where t_0' is t_0 as defined in (16) but with the long run variance estimator, $\hat{\omega}_u^2$, constructed using the Daniell kernel with a data-dependent bandwidth. Specifically, the bandwidth is given by $\max(\hat{b}_{\text{opt}}T, 2)$, where $\hat{b}_{\text{opt}} = b_{\text{opt}}(\hat{c})$. Here, $\hat{c} := T(1-\hat{\rho})$ with $\hat{\rho}$ obtained by OLS estimation of (1) and (2); and $b_{\text{opt}}(.)$ is a step function given in Bunzel and Vogelsang (2005). In the expressions for *Dan-J*, the t_0' statistic is scaled by a function of the J unit root test statistic of Park (1990) and Park and Choi (1988). Again c_ξ is is a constant chosen so that for a significance level ξ, *Dan-J* has the same critical value under both $I(0)$ and $I(1)$ errors. The value of c_ξ depends on $\hat{b}_{\text{opt}}$; Bunzel and Vogelsang (2005) provide a response surface for determining c_ξ for a given significance level, and $\hat{b}_{\text{opt}}$. The critical values for the test also depend on $\hat{b}_{\text{opt}}$, and again a response surface is provided by the authors for a variety of significance levels. Because c is not consistently estimated using $\hat{c}$, Bunzel and Vogelsang (2005) only provide a limiting distribution for *Dan-J* when it is assumed that c is known in the calculation of $\hat{b}_{\text{opt}}$. That is, when $\hat{b}_{\text{opt}} = b_{\text{opt}}(\hat{c})$ is replaced by $b_{\text{opt}}(c)$. Although this strictly means that their asymptotic results are based on the limiting behaviour of an infeasible test, for the purposes of making comparisons tractable, in Lemma 3 of the main text the limit distribution for *Dan-J* is that using $b_{\text{opt}}(c)$.

Appendix C

The Ayat and Burridge (2000) $S2$ strategy for detecting a unit root can be represented by the following flowchart:



The simplified strategy UR follows the same steps as $S2$, but bypassing the t_1 test stage, as shown by the dotted line above. Now it is clear that UR and $S2$ can only differ in terms of unit root inference if $S2$ terminates at Box 2 and UR terminates at Box 3. For this to occur, it must be true that both *DF-QD*$^{\mu}$ and $|t_1|$ reject their respective null hypotheses while *DF-QD*$^{\tau}$ fails to reject.

References

Ayat, L. and Burridge, P. (2000). Unit root tests in the presence of uncertainty about the non-stochastic trend. Journal of Econometrics 95, 71-96.

Bunzel, H. and Vogelsang, T.J. (2005). Powerful trend function tests that are robust to strong serial correlation with an application to the Prebisch-Singer hypothesis. Journal of Business and Economic Statistics 23, 381-94.

Chang, Y. and Park, Y.J. (2002). On the asymptotics of ADF tests for unit roots. Econometric Reviews 21, 431–47.

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–31.

Elliott, G. (1999). Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. International Economic Review 40, 767–83.

Elliott, G. and Müller, U.K. (2006). Minimizing the impact of the initial condition on testing for unit roots. Journal of Econometrics 135, 285-310.

Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient tests for an autoregressive unit root. Econometrica 64, 813-36.

Harris, D., Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2007). Testing for a unit root in the presence of a possible break in trend. Granger Centre Discussion Paper Number 07/04. Downloadable from http://www.nottingham.ac.uk/economics/grangercentre/publications.htm

Harvey, D.I., and Leybourne, S.J. (2005). On testing for unit roots and the initial observation. The Econometrics Journal 8, 97-111.

Harvey, D.I., and Leybourne, S.J. (2006). Unit root test power and the initial condition. Journal of Time Series Analysis 27, 739-52.

Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2007). A simple, robust and powerful test of the trend hypothesis, Journal of Econometrics, 141, 1302-30.

Jansson, M. (2007). Semiparametric power envelopes for tests of the unit root hypothesis. Downloadable from http://www.econ.berkeley.edu/~mjansson/Papers/SemiparametricUnitRoot.pdf

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and 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, 159-78.

Marsh, P. (2007). The available information for invariant tests of a unit root. Econometric Theory 23, 686-710.

Müller, U.K. and Elliott, G. (2003). Tests for unit roots and the initial condition. Econometrica 71, 1269–86.

Newey, W.K. and West., K.D. (1994). Automatic lag selection in covariance matrix estimation. Review of Economic Studies 61, 631-53.

Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519-54.

Park, J.Y. (1990). Testing for unit roots and cointegration by variable addition, in Fomby, T. and Rhodes, F. (eds.) Advances in Econometrics: Cointegration, Spurious Regression and Unit Roots. Jai Press: Greenwich.

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

Perron, P. and 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. (1987a). Time series regression with a unit root. Econometrica 55, 277-301.

Phillips, P.C.B. (1987b). Towards a unified asymptotic theory for autoregression. Biometrika 74, 535–47.

Phillips, P.C.B. (1991a). To criticize the critics: an objective Bayesian analysis of stochastic trends. Journal of Applied Econometrics 6, 333-64,

Phillips, P.C.B. (1991b). Bayesian routes and unit roots: de rebus prioribus semper est disputandum. Journal of Applied Econometrics 6, 435-73.

Phillips, P.C.B. (1998). New tools for understanding spurious regressions. Econometrica 66, 1299-1326.

Phillips, P.C.B. and Perron, P. (1988). Testing for a unit root in time series regression. Biometrika 75, 335-46.

Phillips, P.C.B. and Ploberger, W. (1994). Posterior odds testing for a unit root with data-based model selection. Econometric Theory 10, 774-808.

Phillips P.C.B. and Solo, V. (1992). Asymptotics for linear processes. Annals of Statistics 20, 971-1001.

Phillips, P.C.B. and Xiao, Z. (1998). A primer on unit root testing. Journal of Economic Surveys 12, 423-70.

Schmidt, P. and Phillips, P.C.B. (1992). LM tests for a unit root in the presence of deterministic trends. Oxford Bulletin of Economics and Statistics 54, 257-87.

Vogelsang, T.J. (1998). Trend function hypothesis testing in the presence of serial correlation. Econometrica 66, 123-48.

West, K.D. (1988). Asymptotic normality, when regressors have a unit root. Econometrica 56, 1397-1417.

Xiao, Z. and Phillips, P.C.B. (1998). An ADF coefficient test for a unit root in ARMA models of unknown order with empirical applications to the US economy. Econometrics Journal 1, 27-43.

Table 1. Asymptotic scaling constants τ_γ for γ-level UR tests.

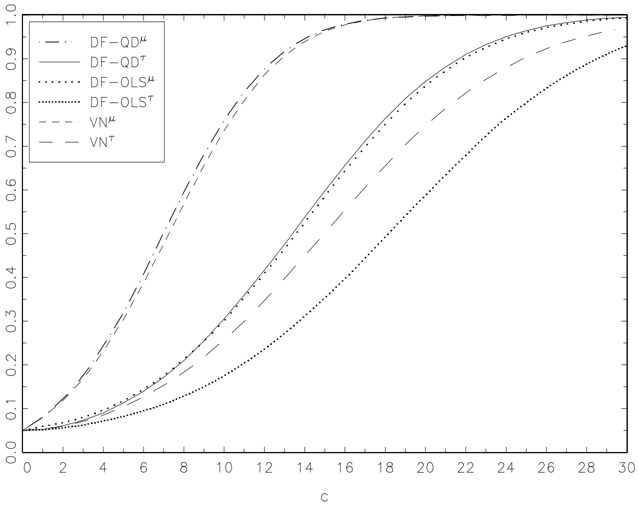
Panel A. Union of $DF\text{-}QD^\mu$ and $DF\text{-}QD^\tau$ rejections			
	$\gamma = 0.10$	$\gamma = 0.05$	$\gamma = 0.01$
UR	1.123	1.095	1.064
Panel B. Union of $DF\text{-}QD$ and $DF\text{-}OLS$ rejections			
	$\gamma = 0.10$	$\gamma = 0.05$	$\gamma = 0.01$
UR^μ	1.126	1.095	1.065
UR^τ	1.070	1.058	1.043

Table 2. Relative asymptotic integrated powers over $c = \{1, 2, ..., 30\}$, $\beta_T = \kappa\omega_\varepsilon T^{-1/2}$.

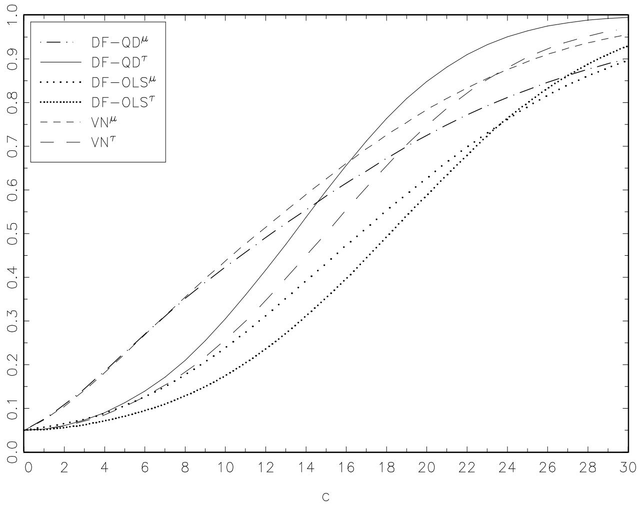
	κ					
	0.00	0.25	0.50	1.00	2.00	4.00
$DF\text{-}QD^\mu$, WA	1.00	0.97	0.06	**0.00**	**0.00**	**0.00**
$DF\text{-}QD^\tau$	**0.74**	1.00	1.00	1.00	1.00	1.00
UR	0.95	1.02	0.83	**0.81**	**0.81**	**0.81**
$PT(\lvert t_\lambda \rvert)$	0.99	0.95	0.05	**0.00**	0.59	0.98
$PT(\lvert t_\lambda^{m2} \rvert)$	0.96	0.88	**0.48**	0.81	0.86	0.86
$PT(\lvert Dan\text{-}J \rvert)$	0.97	0.93	**0.49**	0.70	0.84	0.89

Table 3. Maximum relative asymptotic power losses over $c = \{1, 2, ..., 30\}$, $\beta_T = \kappa\omega_\varepsilon T^{-1/2}$.

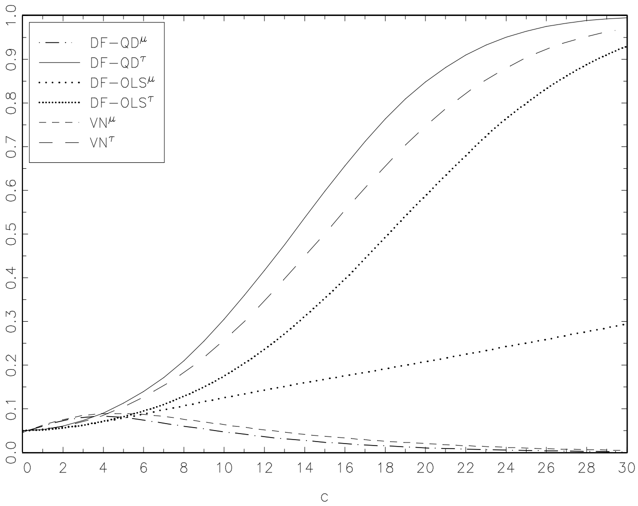
	κ					
	0.00	0.25	0.50	1.00	2.00	4.00
$DF\text{-}QD^\mu$, WA	0.00	0.14	0.99	**1.00**	**1.00**	**1.00**
$DF\text{-}QD^\tau$	**0.47**	0.00	0.00	0.00	0.00	0.00
UR	0.12	0.05	**0.20**	**0.20**	**0.20**	**0.20**
$PT(\lvert t_\lambda \rvert)$	0.01	0.15	0.99	**1.00**	0.38	0.02
$PT(\lvert t_\lambda^{m2} \rvert)$	0.10	0.19	**0.52**	0.20	0.15	0.15
$PT(\lvert Dan\text{-}J \rvert)$	0.07	0.16	**0.52**	0.29	0.15	0.11



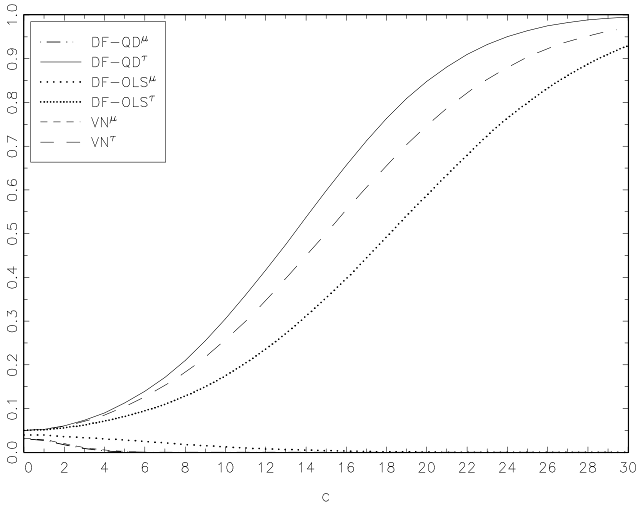
(a) $\kappa = 0.00$



(b) $\kappa = 0.25$

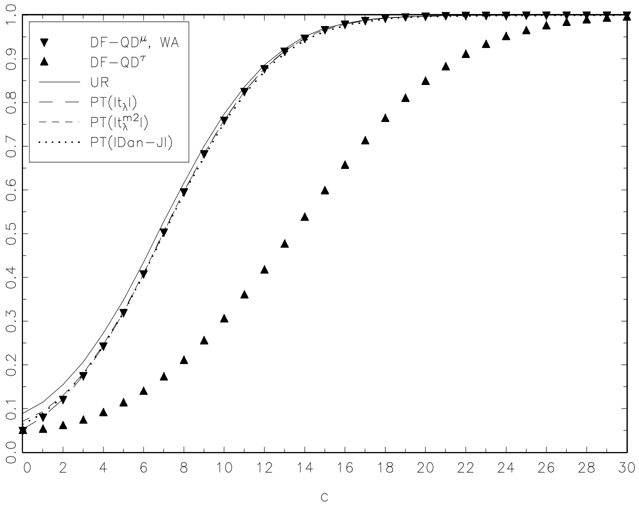


(c) $\kappa = 0.50$

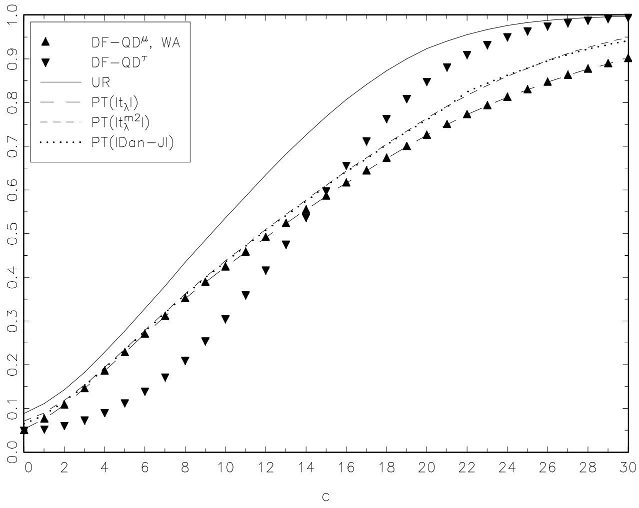


(d) $\kappa = 1.00$

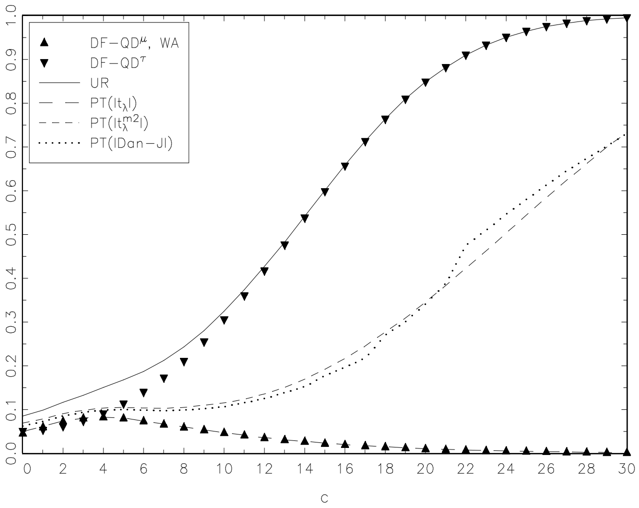
Figure 1. Asymptotic size and local power: $\beta_T = \kappa \omega_\varepsilon T^{-1/2}$



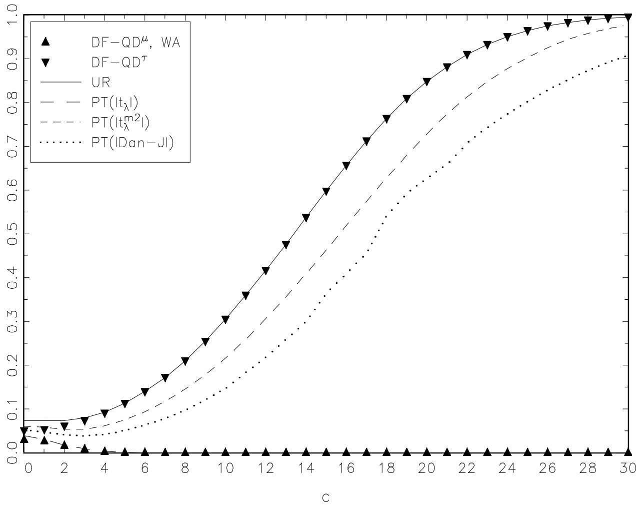
(a) $\kappa = 0.00$



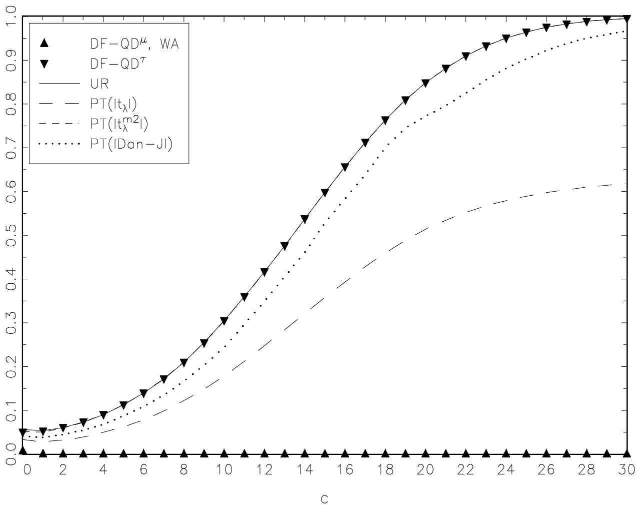
(b) $\kappa = 0.25$



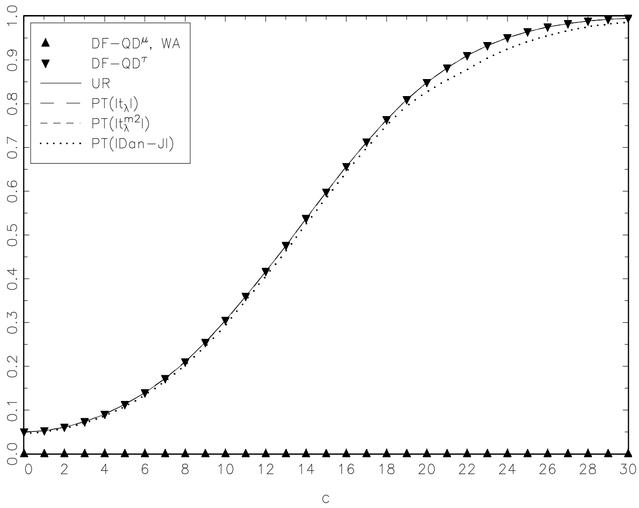
(c) $\kappa = 0.50$



(d) $\kappa = 1.00$

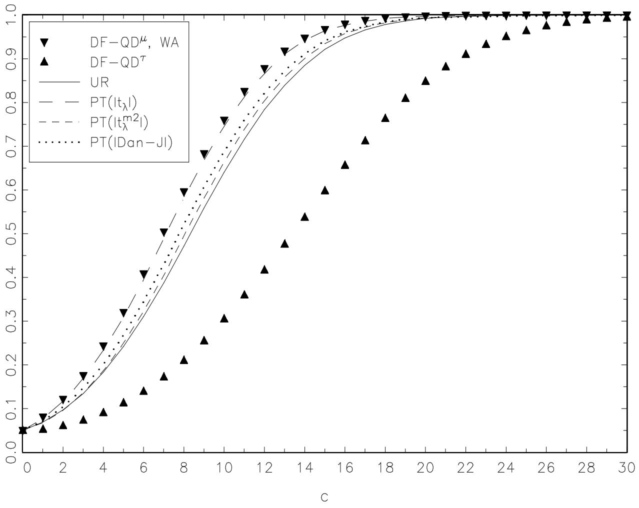


(e) $\kappa = 2.00$

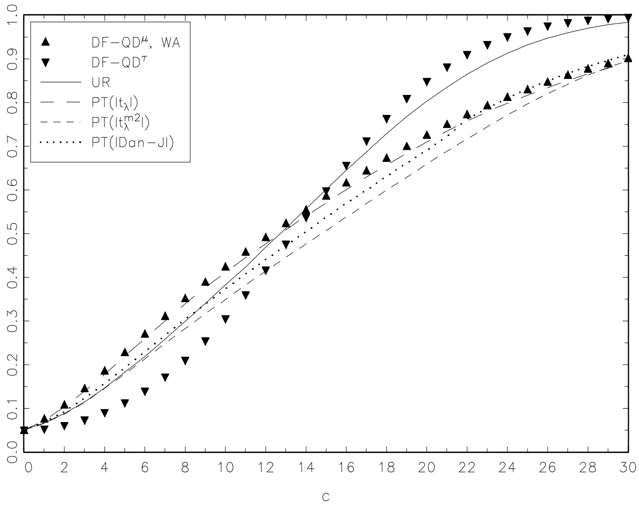


(f) $\kappa = 4.00$

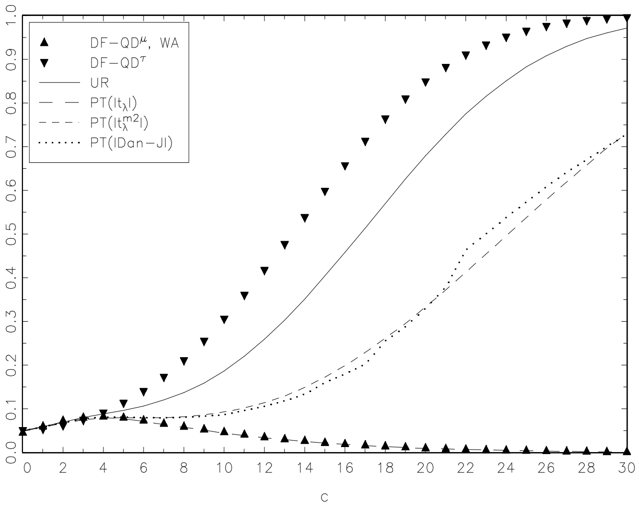
Figure 2. Asymptotic size and local power: $\beta_T = \kappa\omega_\varepsilon T^{-1/2}$



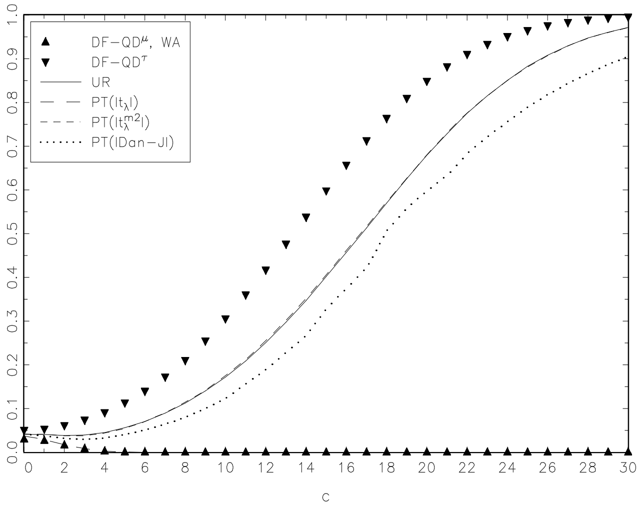
(a) $\kappa = 0.00$



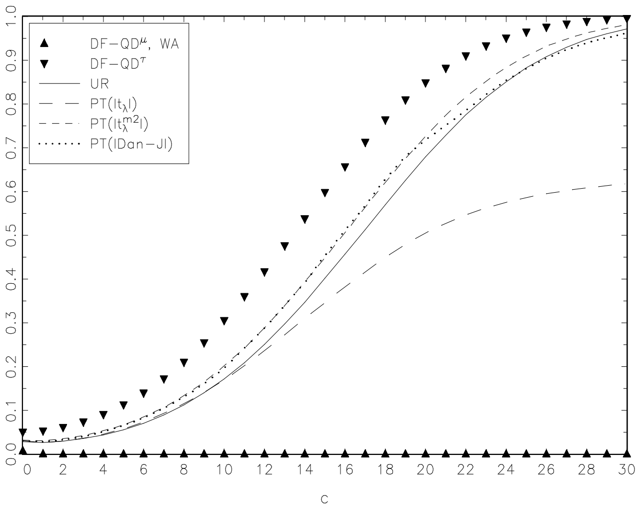
(b) $\kappa = 0.25$



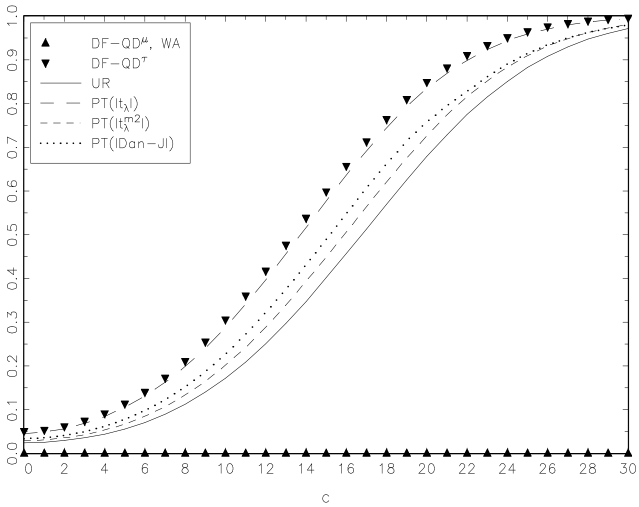
(c) $\kappa = 0.50$



(d) $\kappa = 1.00$

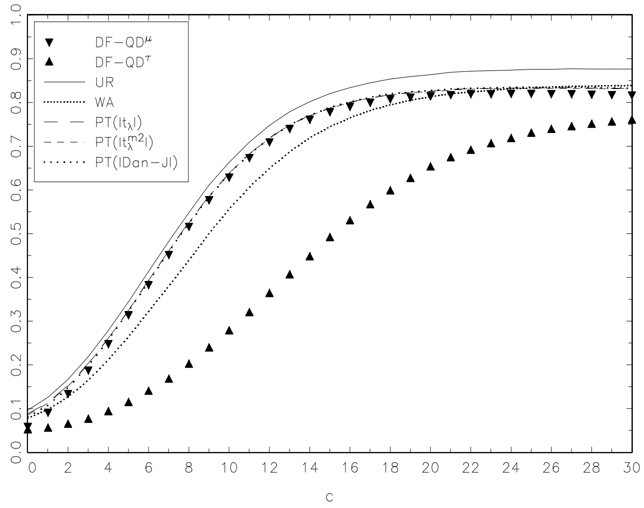


(e) $\kappa = 2.00$

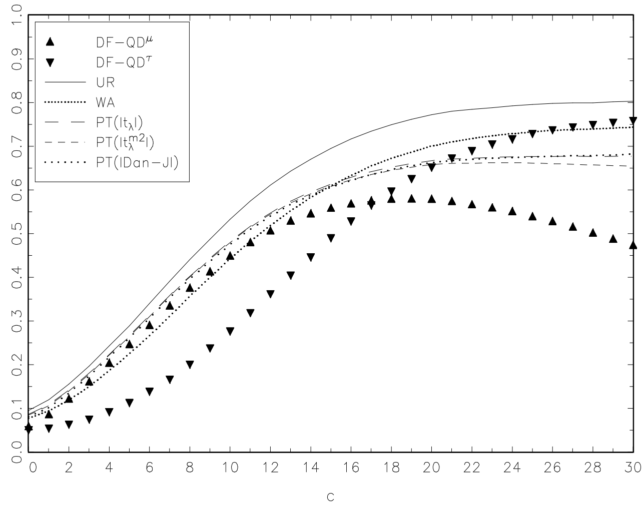


(f) $\kappa = 4.00$

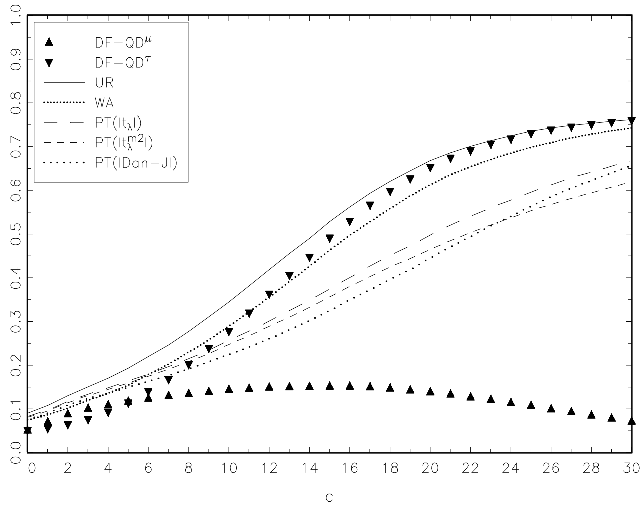
Figure 3. Asymptotic size and local power: conservative strategies, $\beta_T = \kappa \omega_\varepsilon T^{-1/2}$



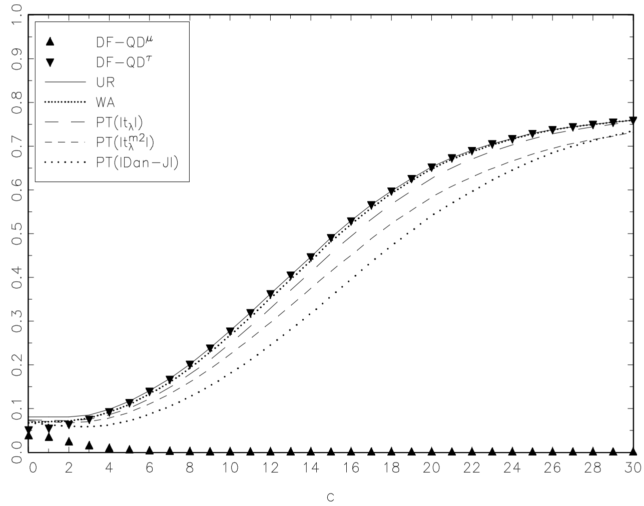
(a) $\beta = 0.000$



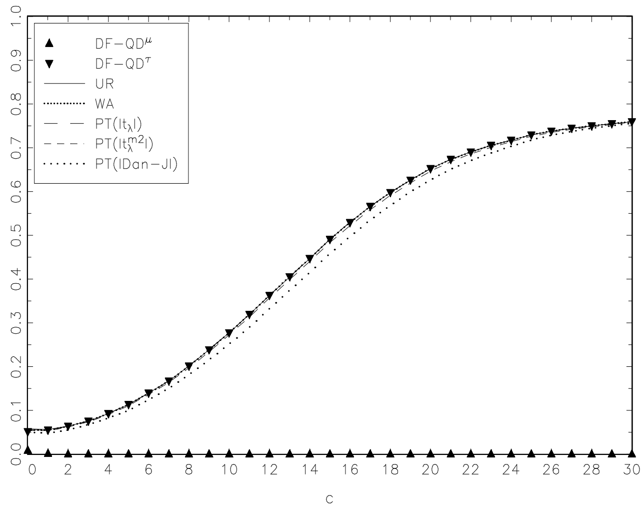
(b) $\beta = 0.025$



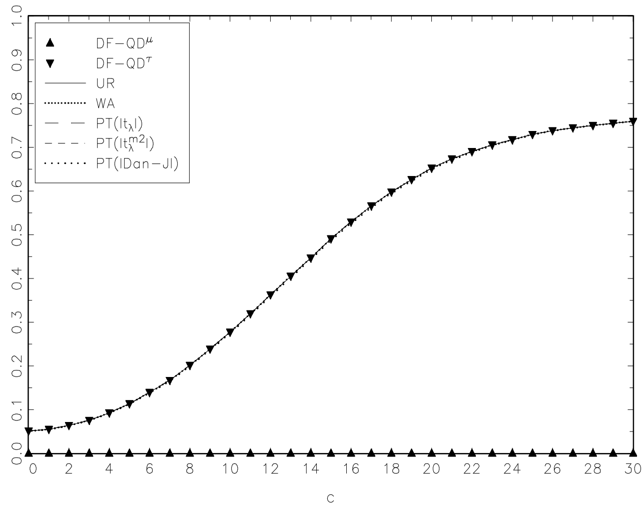
(c) $\beta = 0.050$



(d) $\beta = 0.100$

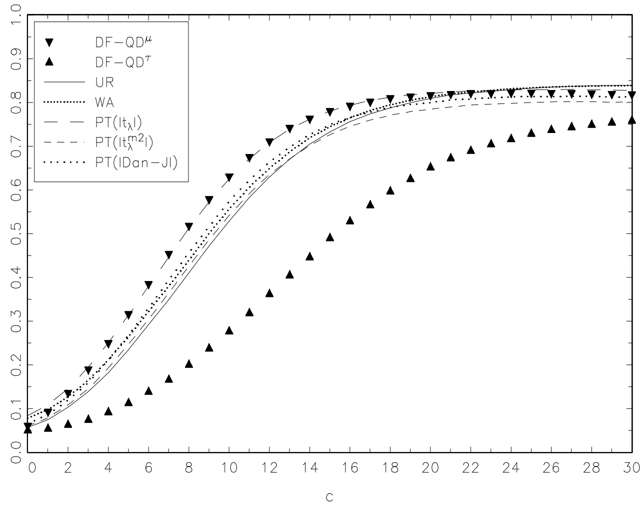


(e) $\beta = 0.200$

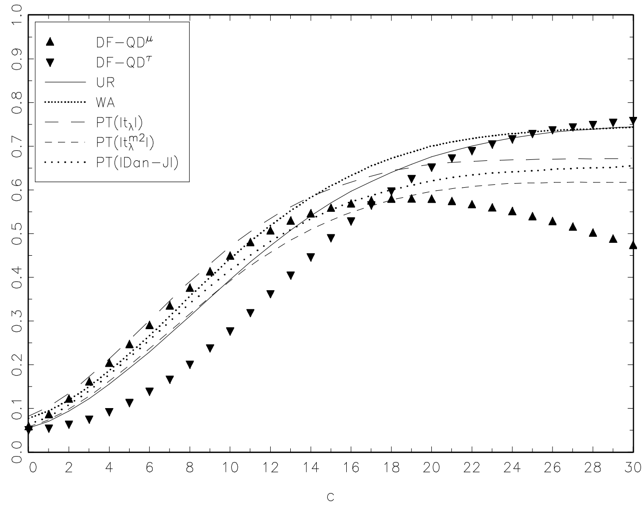


(f) $\beta = 0.400$

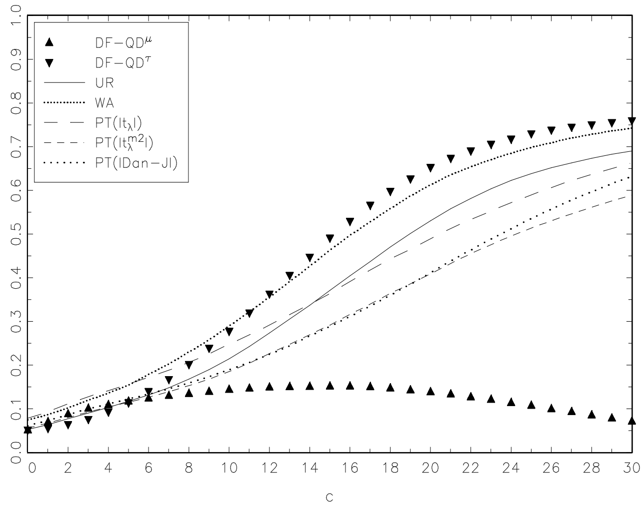
Figure 4. Finite sample size and power: $T = 100$



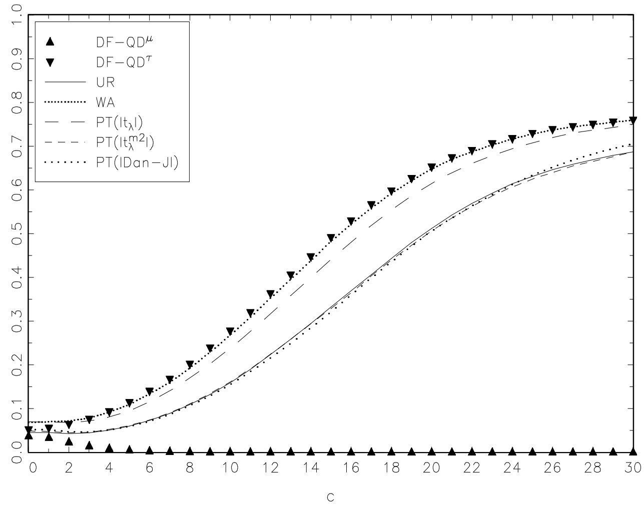
(a) $\beta = 0.000$



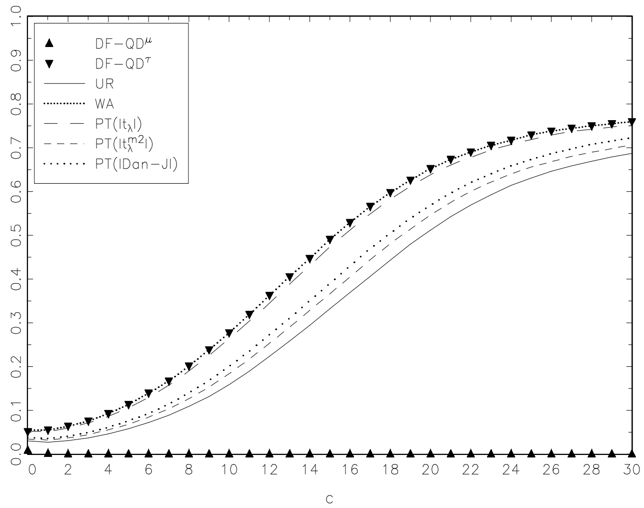
(b) $\beta = 0.025$



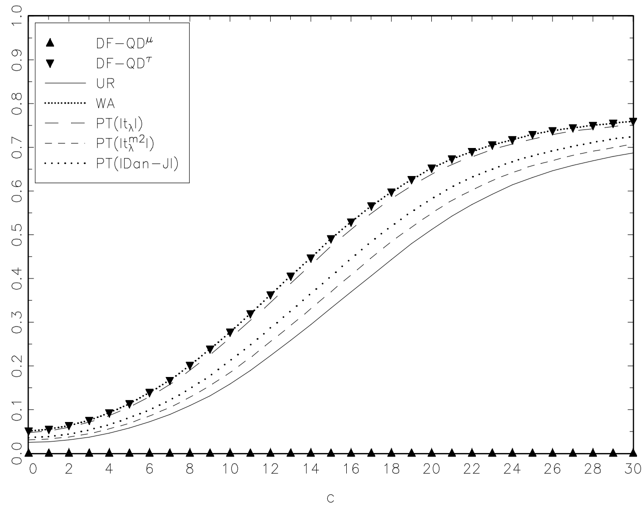
(c) $\beta = 0.050$



(d) $\beta = 0.100$

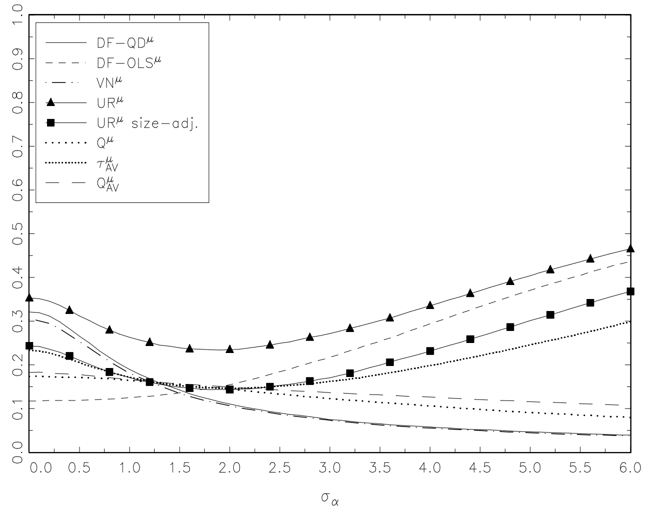


(e) $\beta = 0.200$

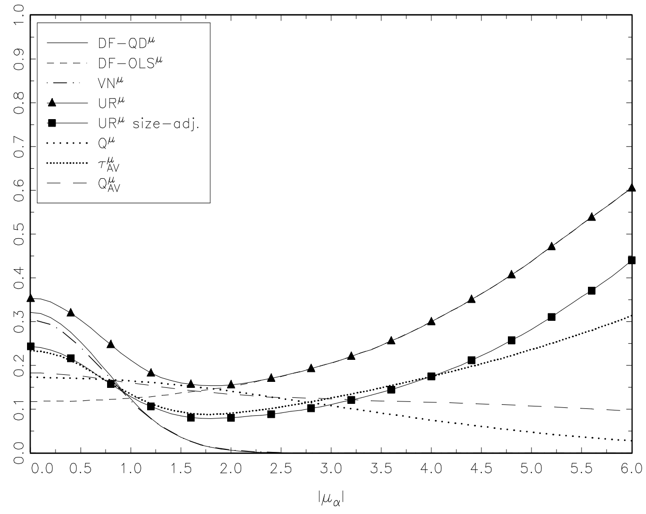


(f) $\beta = 0.400$

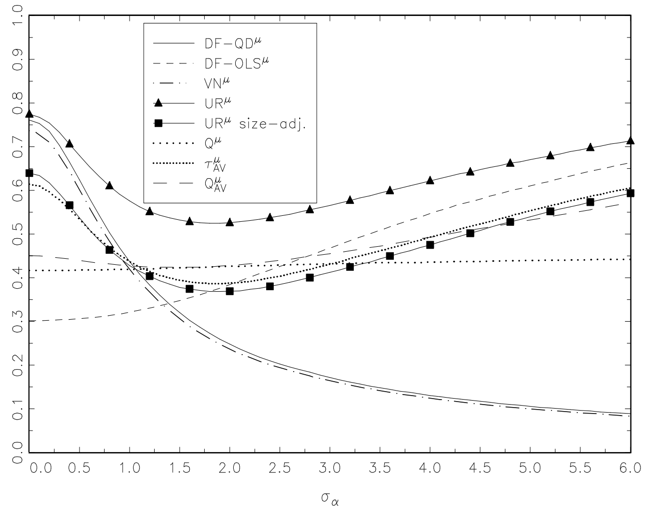
Figure 5. Finite sample size and power: conservative strategies, $T = 100$



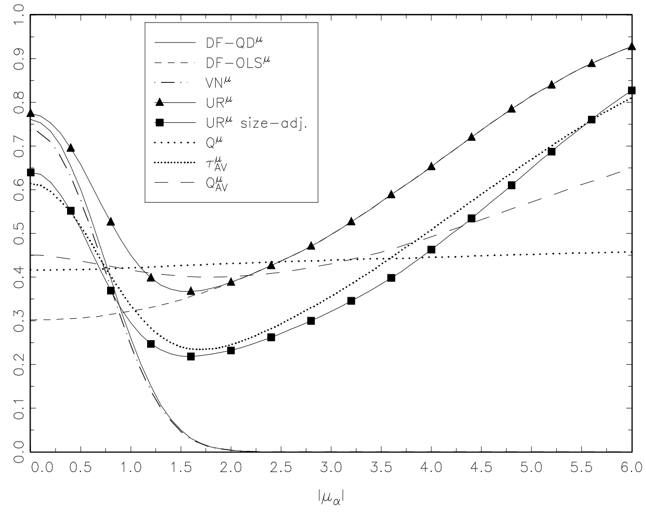
(a) $c = 5$, α random



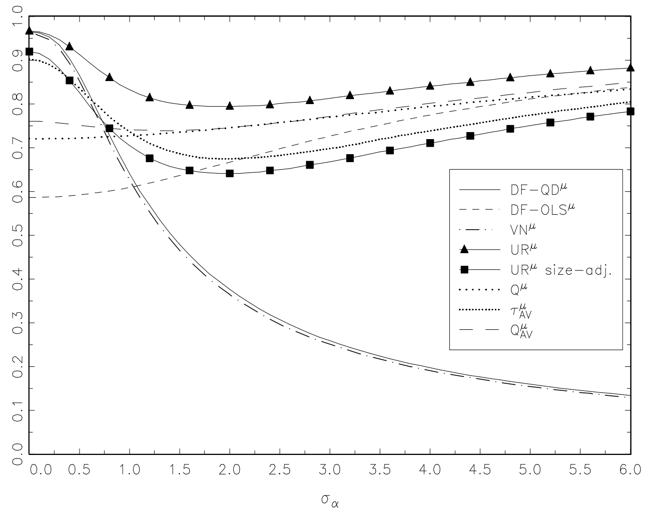
(b) $c = 5$, α fixed



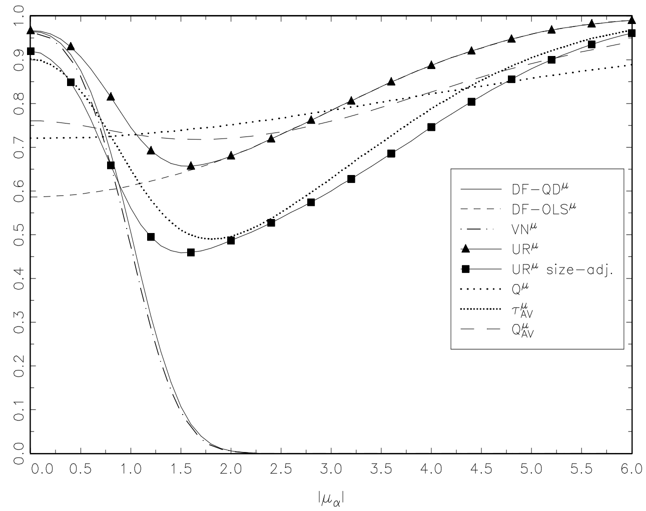
(c) $c = 10$, α random



(d) $c = 10$, α fixed

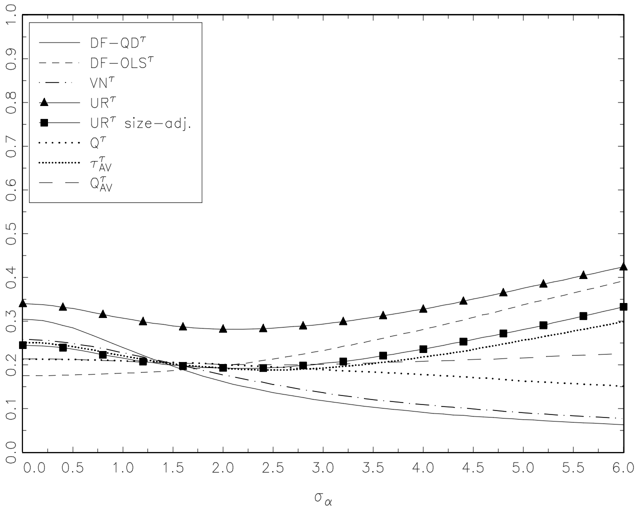


(e) $c = 15$, α random

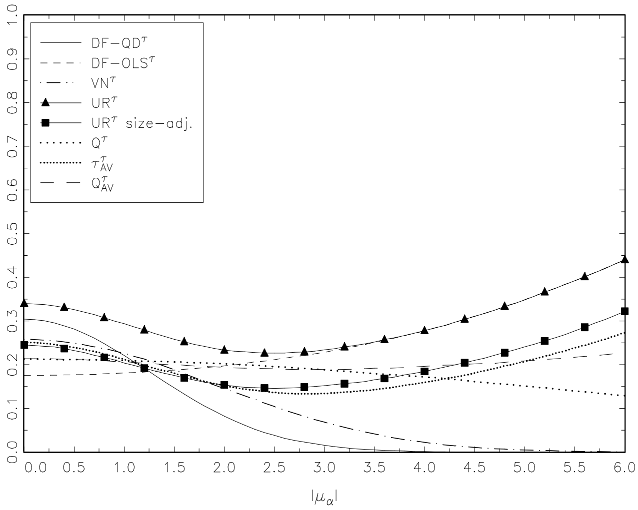


(f) $c = 15$, α fixed

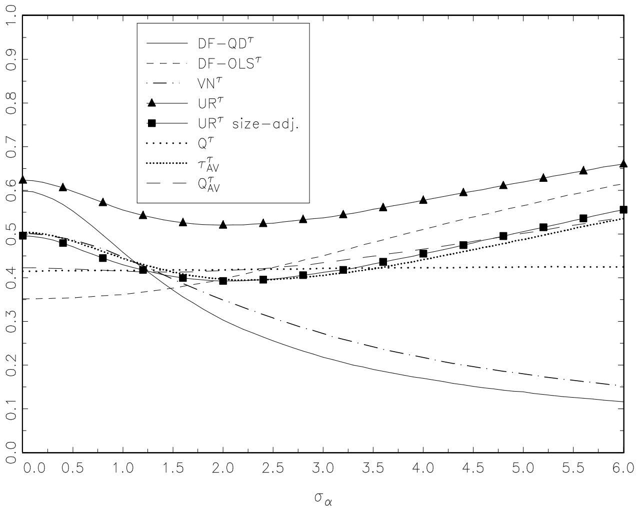
Figure 6. Asymptotic local power with varying initial values: mean case



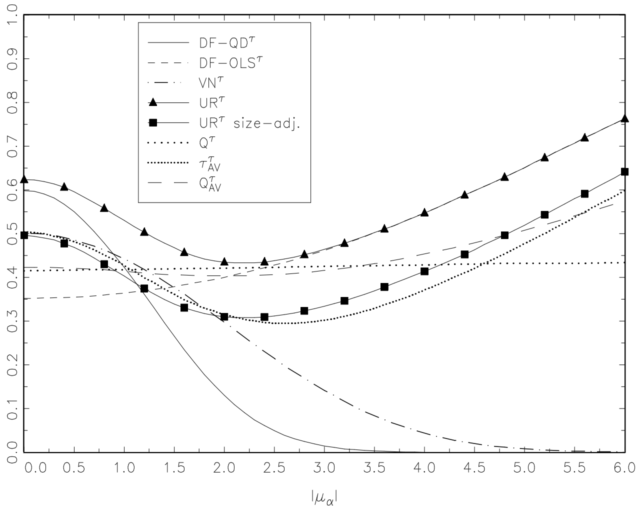
(a) $c = 10$, α random



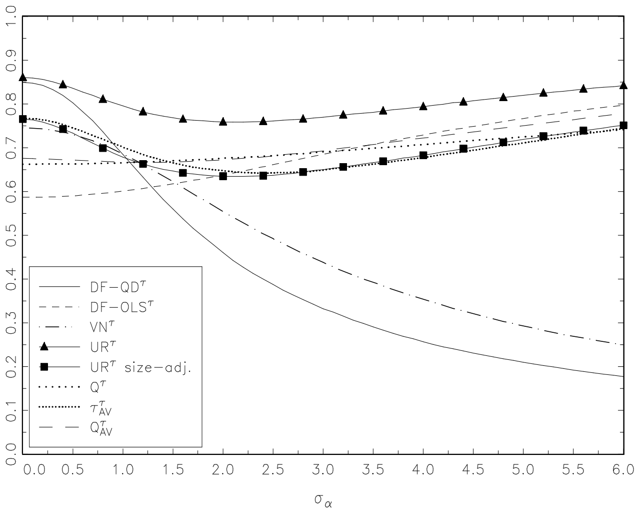
(b) $c = 10$, α fixed



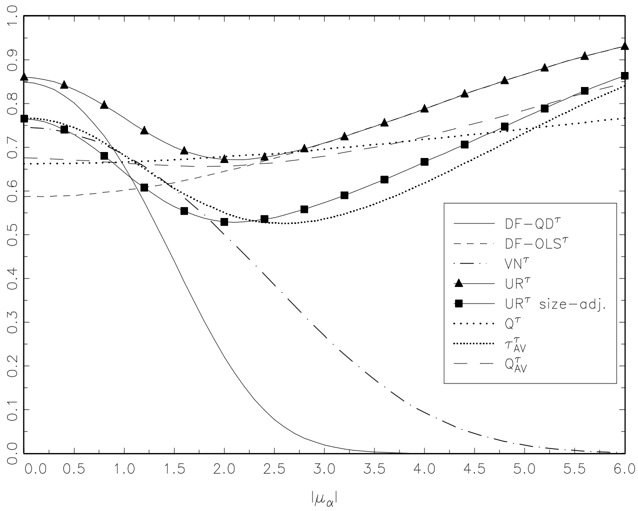
(c) $c = 15$, α random



(d) $c = 15$, α fixed

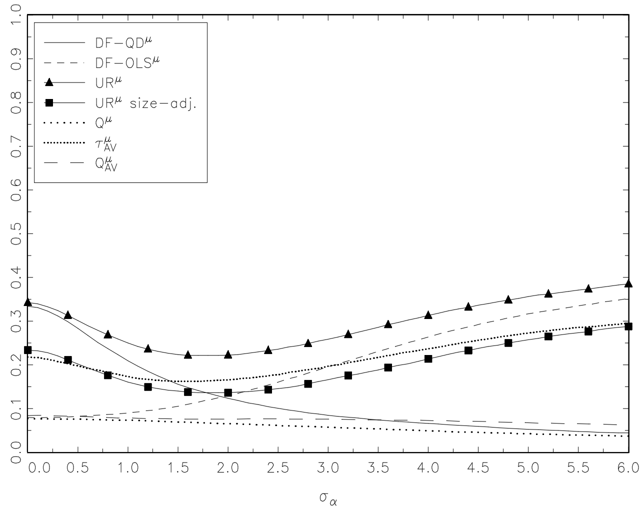


(e) $c = 20$, α random

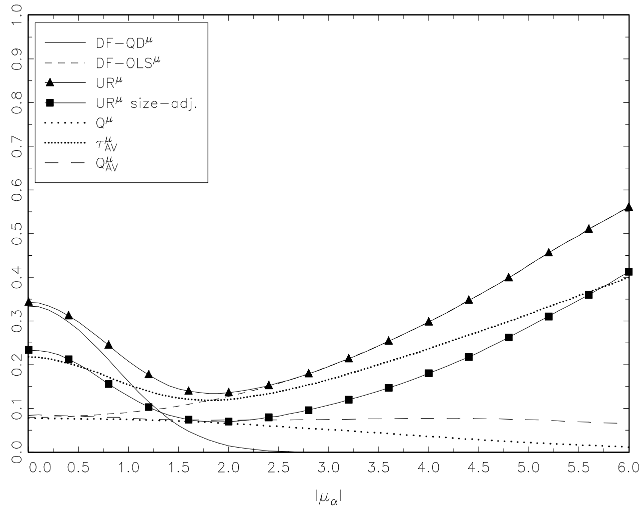


(f) $c = 20$, α fixed

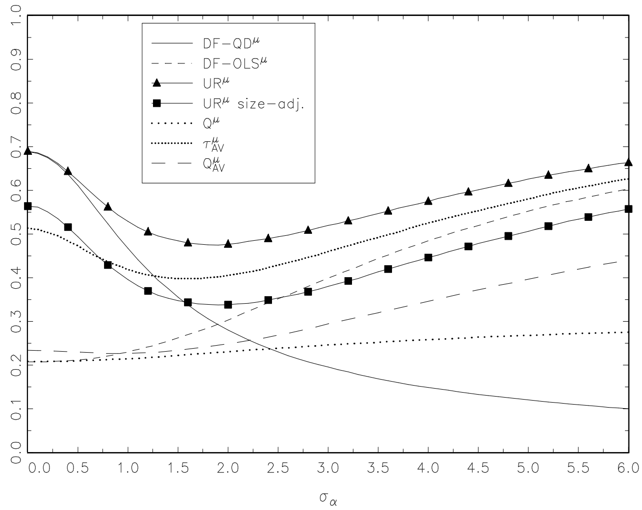
Figure 7. Asymptotic local power with varying initial values: trend case



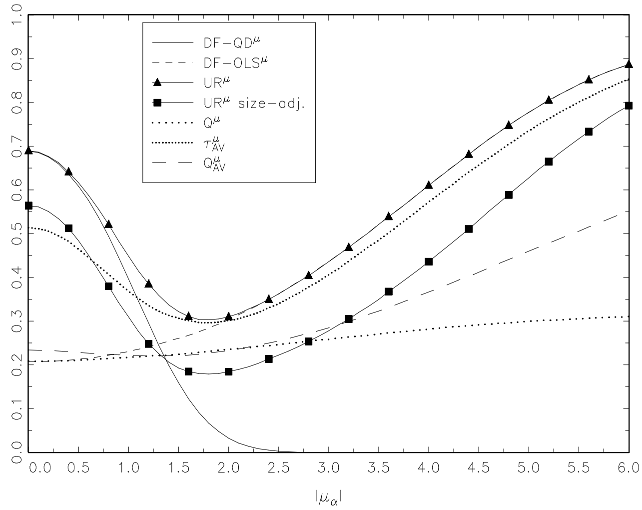
(a) $c = 5$, α random



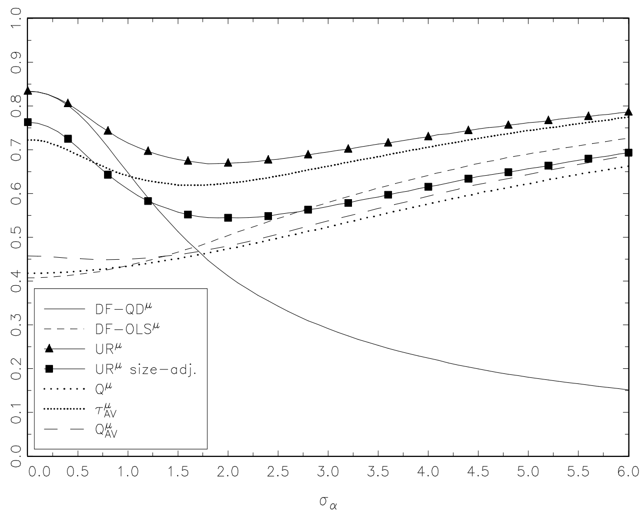
(b) $c = 5$, α fixed



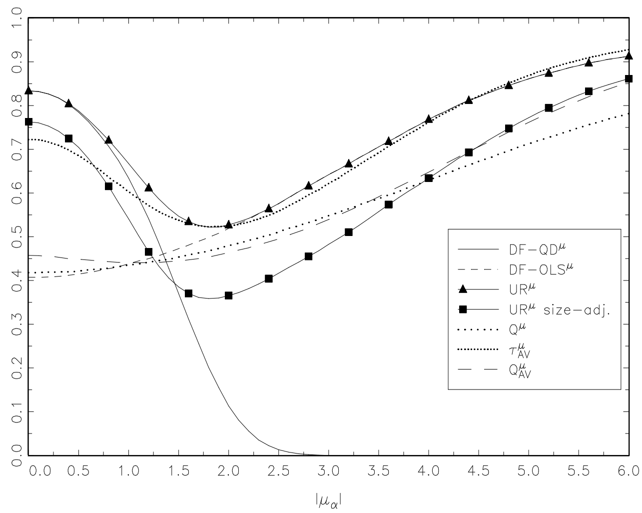
(c) $c = 10$, α random



(d) $c = 10$, α fixed

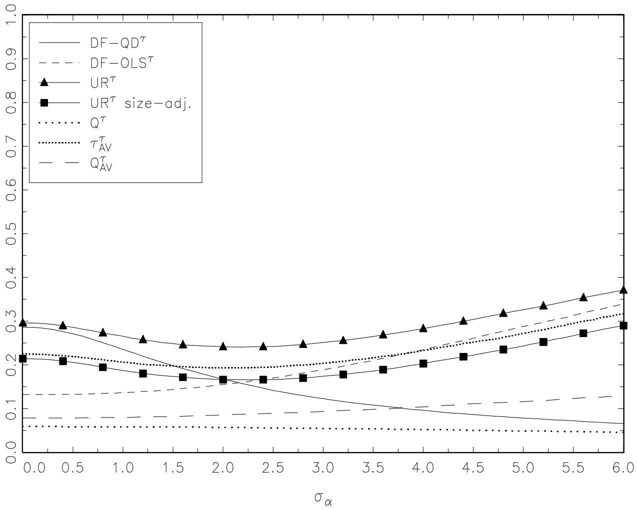


(e) $c = 15$, α random

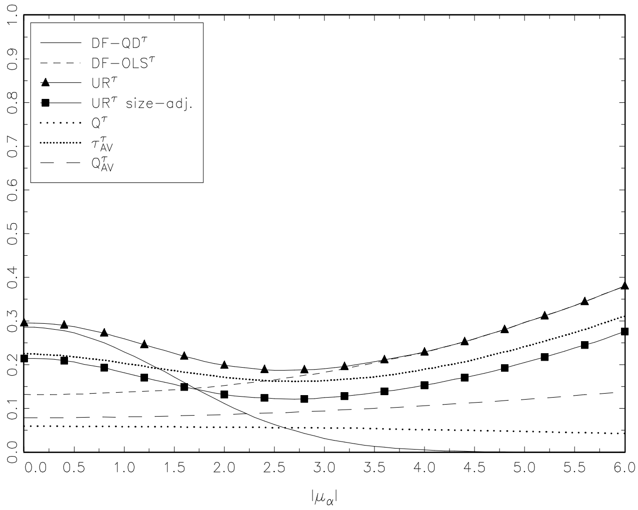


(f) $c = 15$, α fixed

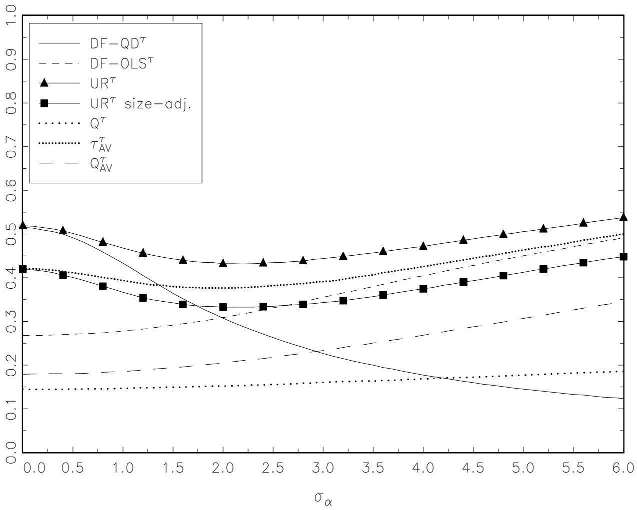
Figure 8. Finite sample power with varying initial values: mean case



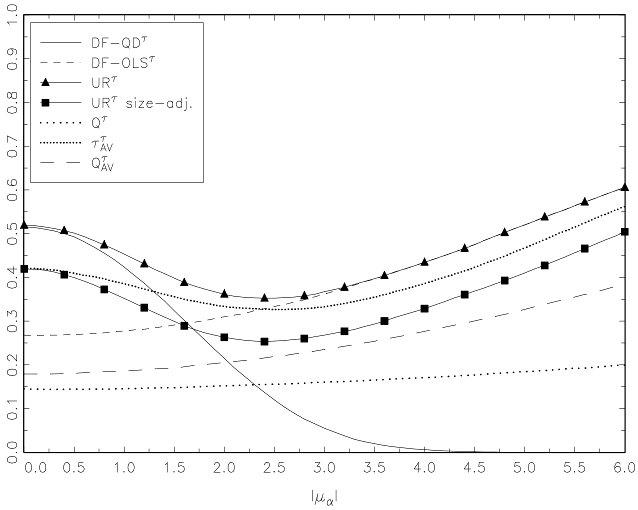
(a) $c = 10$, α random



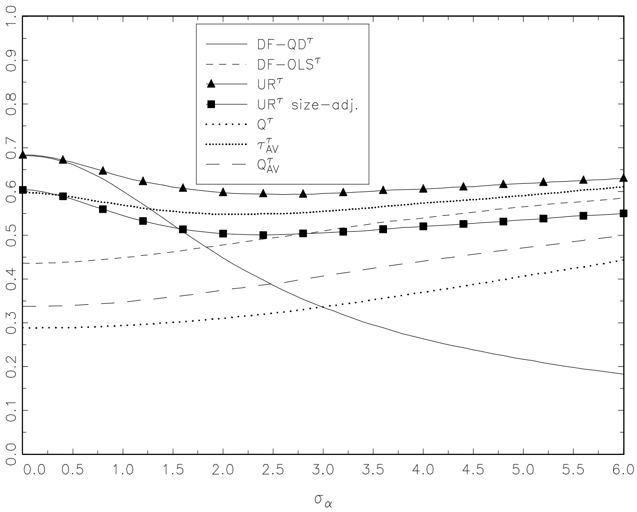
(b) $c = 10$, α fixed



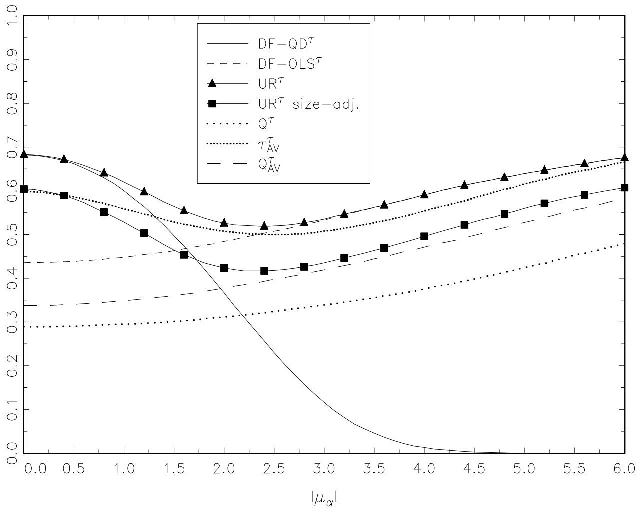
(c) $c = 15$, α random



(d) $c = 15$, α fixed



(e) $c = 20$, α random



(f) $c = 20$, α fixed

Figure 9. Finite sample power with varying initial values: trend case