Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

The impact of the initial condition on robust tests for a linear trend

by

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

---

Granger Centre Discussion Paper No. 09/03

The University of Nottingham

The Impact of the Initial Condition on Robust Tests for a Linear Trend

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

Granger Centre for Time Series Econometrics, University of Nottingham

July 2009

Abstract

This paper examines the behaviour of some recently proposed "robust" (to the order of integration of the data) tests for the presence of a deterministic linear trend in a univariate times series in situations where the magnitude of the initial condition of the series is non-negligible. We demonstrate that the asymptotic size and/or local power properties of these tests are extremely sensitive to the initial condition. Straightforward modifications to the trend tests are suggested, based around the use of trimmed data, which are demonstrated to greatly reduce this sensitivity.

Keywords: Trend tests; initial condition; asymptotic local power.

JEL Classification: C22.

1 Introduction

In this paper we examine the behaviour of the tests for the presence of a deterministic linear trend of Bunzel and Vogelsang (2005), Harvey *et al.* (2007) and Perron and Yabu (2009) in situations where the magnitude of the initial condition (the deviation of the first observation from the linear trend component) of the series is non-negligible. When the initial condition is negligible, these statistics are termed "robust" in the sense that the asymptotic critical values for testing hypotheses on the trend coefficient are the same regardless of whether the stochastic component of the time series contains an

*Corresponding author: School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, U.K. Email: robert.taylor@nottingham.ac.uk

autoregressive unit root (denoted $I(1)$) or is stationary (denoted $I(0)$). Here, we show that when the initial condition of the autoregressive model is non-negligible, in the sense of Elliott and Müller (2006), it can in fact exert a very substantial influence over the asymptotic size and/or local power of these tests.

As with the underlying motivation behind the analysis of, *inter alia*, Müller and Elliott (2003), Elliott and Müller (2006) and Harvey *et al.* (2009), who considered the impact of the initial condition on the size and power of unit root tests, an examination of the behaviour of trend function tests with respect to the initial condition stems from important empirical considerations. A mean-reverting process with a large (in absolute sense) initial condition could be used to characterize economic data that just happens to be observed directly after some structural episode, such as a policy shift or political regime change. Moreover, the same effect is seen as a result of the common practice in applied work of using empirical data sets which are (deliberately) chosen to begin just after a major structural episode, such as, for example, the second world war or the 1970's oil price shocks. Conversely, a more modest initial condition might be associated with data observed within a period of comparative economic stability. It is clearly important therefore to analyse the impact that the magnitude of the initial condition has on tests for the presence of a trend. Where the initial condition is non-negligible, the relative position of the initial value of the time series to the underlying trend path is also clearly likely to be important. If, for example, the initial value lay significantly below an underlying positive trend line then this would clearly be anticipated to reinforce the appearance of the trend in the data and, other things being equal, increase the power of the trend test.

We first demonstrate that, with the exception of the test of Bunzel and Vogelsang (2005) which displays under-sizing, the aforementioned robust trend tests can be (potentially very) badly over-sized when the initial condition is large, and can thereby spuriously signal the presence of a linear trend term. We then show that the power of these tests to detect a trend, other things held equal, can be either significantly increased or decreased by a large initial condition, the direction of change depending on the particular test being considered and the relative signs of the initial condition and linear trend. Of particular concern arising from this finding is that all these tests can have very low asymptotic power to detect a substantial linear trend component present in the data. Consequently, should these tests be employed in a pre-test role for deciding whether to include or exclude a trend term in a unit root test regression, they can often wrongly signal exclusion. Irrespective of any other impact the initial condition may have, the incorrect exclusion of a linear trend term in any given unit root test regression is known to have a profoundly negative impact on that unit root test's asymptotic power properties; see, e.g., Harvey *et al.* (2009). In the converse situation where an irrelevant trend term is included in the unit root regression, power is also sacrificed, although the effect is much less severe; again see Harvey *et al.* (2009).

In order to help counteract the adverse effects of the initial condition on the large sample properties of the trend tests, we suggest a simple strategy based on calculating the trend tests over a shortened (left-trimmed) series that excludes a proportion of the earliest observations. The reasoning behind this suggestion is that, outside of the exact unit root case (where the initial condition has no impact on the trend tests), the initial condition has a decreasing influence on observations the later on they occur in the series. This simple procedure is shown to perform well in helping to reduce the degree of asymptotic over-size seen in the trend tests when the initial condition is non-negligible. A fairly transparent downside of trimming the series in this way is that there is, necessarily, some reduction in a trend test's power in those situations where, other things equal, the initial condition does not effect a decrease in the power of the corresponding test based on the complete series. However, these losses are shown to be relatively minor when compared with the improvements in power seen where the initial condition does negatively impact on the power of the full sample test.

A further avenue we pursue is to investigate how discernable sample information on the magnitude of the initial condition might be used to suggest whether or not trimming the series is the appropriate course of action for the trend tests. This takes the form of employing an auxiliary statistic (itself exact invariant to the parameters characterising the trend component) which is used to detect the presence of a large initial condition and thereby indicate whether or not trimming appears warranted.

The paper is organized as follows. In the next section we present our framework of a local-to-unity autoregressive root model with local-to-zero linear trend, where, in contrast to the existing robust trend function testing literature, we allow for the possibility of a non-negligible initial condition. This is the environment in which our asymptotic analysis is conducted. Here we also provide descriptions of the robust trend tests under consideration. Section 3 examines their asymptotic size and local power as functions of the initial condition, highlighting the potential for poor size control and/or low power in the trend tests. In section 4 we show how this type of sensitivity is reduced by calculating the tests from the trimmed series and give a concrete recommendation for the degree of trimming to use in practice. Section 5 examines how the auxiliary statistic for detecting large initial conditions can be incorporated into the analysis and examines the effectiveness of this approach. Section 6 offers some conclusions.

In what follows we use the notation: '$x := y$' indicates that x is defined by y; '$\xrightarrow{d}$' denotes weak convergence and $\lfloor \cdot \rfloor$ denotes the integer part of its argument.

2 The Model and Robust Trend Tests

Consider the case where we have a sample of T observations generated according to the data generating process [DGP]:

$$
\begin{aligned}
y_t &= \mu + \beta_T t + u_t, \quad t = 1, ..., T &(1)\\
u_t &= \rho_T u_{t-1} + \varepsilon_t, \quad t = 2, ..., T. &(2)
\end{aligned}
$$

We make the following assumptions on (1)-(2):

Assumption 1 *The stochastic process* $\{\varepsilon_t\}$ *is such that*

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

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

Assumption 2 *The trend coefficient in (1) satisfies* $\beta_T := \kappa\omega_\varepsilon T^{-1/2}$, *where* $|\kappa| < \infty$.

Assumption 3 *The autoregressive parameter in (2) satisfies* $\rho_T := 1 - c/T$ *for* $0 \leq c < \infty$.

Assumption 4 *The initial condition,* u_1, *is generated according to* $u_1 = \alpha\sqrt{\omega_\varepsilon^2/(1-\rho_T^2)}$, *for* $\rho_T := 1 - c/T$, $c > 0$. *For* $c = 0$, *we may set* $u_1 = 0$, *without loss of generality, due to the exact similarity of the trend tests considered in this paper to the initial condition when* $c = 0$.

Remark 2.1. Assumption 1 renders the innovation process $\{\varepsilon_t\}$ of (2) a conventional stable and invertible linear ($I(0)$) process. Assumption 2 specifies the behaviour of the coefficient on the local-to-zero linear trend term in (1), providing an appropriate Pitman (local) drift for our subsequent asymptotic analyses. The scaling of the linear trend coefficient by ω_ε is simply a convenience measure to ensure that ω_ε does not appear in subsequent expressions for the limit distributions. Under Assumption 3, $c = 0$ corresponds to the case of an autoregressive unit root (exact $I(1)$) process, and $c > 0$ to a local-to-unity (near-integrated) autoregressive root process. In Assumption 4, α controls the magnitude of the initial condition u_1 relative to the standard deviation of a stationary $AR(1)$ process with parameter ρ_T and innovation long-run variance ω_ε^2. This form for the initial value is closely related to that given in Müller and Elliott (2003) and Elliott and Müller (2006). Notice also that, when $c > 0$, the initial value is asymptotically non-negligible when $\alpha \neq 0$, since $T^{-1/2}u_1 \to \alpha\lambda_c$, where $\lambda_c := \sqrt{\omega_\varepsilon^2/2c}$, as $T \to \infty$.

2.1 Robust Trend Tests

The statistics we consider to test the null hypothesis $\beta_T = 0$ against $\beta_T \neq 0$ in (1) are the z_λ and z_λ^{m2} statistics of Harvey *et al.* (2007), the t_β^{RQF} statistics of Perron and Yabu (2009), and the *Dan-J* statistic of Bunzel and Vogelsang (2005).

The z_λ statistic of Harvey *et al.* (2007) employs a switching-based strategy that attains the local limiting Gaussian power envelope for this testing problem (under the assumption of an asymptotically negligible initial condition) irrespective of whether u_t contains is an exact $I(1)$ process or is $I(0)$, the latter occurring where $\rho_T = \rho$ with $|\rho| < 1$. The test statistic is also asymptotically standard normal under the null in both cases. It is calculated as

$$z_\lambda := (1 - \lambda^*)z_0 + \lambda^* z_1 \tag{3}$$

where

$$z_0 := \frac{\hat{\beta}_T}{\sqrt{\hat{\omega}_u^2 / \sum_{t=1}^T (t - \bar{t})^2}} \quad \text{and} \quad z_1 := \frac{\breve{\beta}_T}{\sqrt{\breve{\omega}_v^2/(T-1)}}. \tag{4}$$

In (4), $\hat{\beta}_T$ denotes the OLS estimator of β_T from (1) and $\hat{\omega}_u^2$ is a long run variance estimator formed using $\hat{u}_t := y_t - \hat{\mu} - \hat{\beta}_T t$, $\hat{\mu}$ the corresponding OLS estimator of μ from (1), while $\breve{\beta}_T$ is the OLS estimator of β_T from (1) estimated in first differences i.e. from $\Delta y_t = \beta_T + v_t$, $t = 2, ..., T$ and $\breve{\omega}_v^2$ is a long run variance estimator based on $\breve{v}_t := \Delta y_t - \breve{\beta}_T$. The weight function λ^* is defined as

$$\lambda^* := \exp\left(-0.00025\left(\frac{ERS}{KPSS}\right)^2\right)$$

where *ERS* is the with-trend local GLS unit root test statistic of Elliott *et al.* (1996) and *KPSS* is the with-trend stationarity test statistic of Kwiatkowski *et al.* (1992).

The t_β^{RQF} statistic of Perron and Yabu (2009) takes the form of an autocorrelation-corrected t-ratio on the OLS estimate of β_T obtained from the quasi GLS regression

$$y_t - \tilde{\rho}_{MS} y_{t-1} = (1 - \tilde{\rho}_{MS})\mu + \beta_T[t - \tilde{\rho}_{MS}(t-1)] + (u_t - \tilde{\rho}_{MS} u_{t-1})$$

for $t = 2, ..., T$, along with $y_1 = \mu + \beta_T + u_1$. Here, $\tilde{\rho}_{MS}$ is defined according to the following truncation rule

$$\tilde{\rho}_{MS} := \begin{cases} 1 & \text{if } |\tilde{\rho}_{TWS} - 1| < T^{-1/2} \\ \tilde{\rho}_{TWS} & \text{otherwise} \end{cases}$$

where $\tilde{\rho}_{TWS}$ is an autocorrelation-robust weighted symmetric least squares estimate of ρ (based on the OLS residuals $\hat{u}_t$) with one of two truncations applied as described by Roy and Fuller (2001) and Roy *et al.* (2004). The t_β^{RQF} statistic is asymptotically standard normal under the null hypothesis when u_t is either exact $I(1)$ or is $I(0)$, and, as noted in Remark 2 of Perron and Yabu (2009), has the same local asymptotic power as the z_λ statistic of Harvey *et al.* (2007) in the local-to-unity autoregressive root environment that we consider in this paper.

Harvey *et al.* (2007) show that a modified variant of z_λ, denoted z_λ^{m2}, can provide a more powerful test of the trend hypothesis than z_λ when u_t is near-integrated. This replaces z_1 with $z_1^{m2} := \delta_\gamma R_2 z_1$ where

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

and $\hat{\sigma}_u^2 := (T-2)^{-1}\sum_{t=1}^{T}\hat{u}_t^2$. Here δ_γ is a constant chosen such that, at a given significance level γ, z_λ^{m2} has a standard normal critical value under both exact $I(1)$ and $I(0)$ u_t. For a two-tailed 0.05 level test, $\delta_\gamma = 0.00115$.

The *Dan-J* statistic of Bunzel and Vogelsang (2005) is essentially a modified version of the *t-PSW*^{1} test statistic of Vogelsang (1998) that employs a long run variance estimator based on the "fixed-b" asymptotics of Kiefer and Vogelsang (2005). Specifically, the statistic is

$$\textit{Dan-J} := z_0' \exp(-c_\gamma' J)$$

where z_0' is z_0 as defined in (4) but with the long run variance estimator, $\hat{\omega}_u^2$, constructed using the Daniell kernel with a data-dependent bandwidth. 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 z_0' statistic is scaled by a function of the J unit root test statistic of Park (1990) and Park and Choi (1988); see Bunzel and Vogelsang (2005) for further details. The constant c_γ' is chosen so that for a significance level γ, *Dan-J* has the same critical value under both $I(0)$ and exact $I(1)$ errors. The value of c_γ' depends on $\hat{b}_{\text{opt}}$; Bunzel and Vogelsang (2005) provide a response surface for determining c_γ' 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 what follows the limit distribution for *Dan-J* is that using $b_{\text{opt}}(c)$.

3 Asymptotic Properties of the Tests

The large sample properties of the four trend statistics are summarized in the following lemma, the proof of which is given in the Appendix. To facilitate the analysis of the trend tests based on trimmed series in section 4 below, the results are presented assuming an arbitrary degree of (left) trimming.

Lemma 1 *Let $\{y_t\}$ be generated according to (1)-(2) under Assumptions 1-4. Suppose the trend test statistics are calculated using observations $y_{\lfloor \tau T \rfloor + 1}, ..., y_T$ where $0 \leq \tau < 1$. Then,*

$$z_\lambda, t_\beta^{RQF} \xrightarrow{d} \frac{(1-\tau)\kappa + K_c(1) - K_c(\tau)}{\sqrt{1-\tau}}$$

$$z_\lambda^{m2} \xrightarrow{d} \delta_\gamma \frac{(1-\tau)\kappa + K_c(1) - K_c(\tau)}{\sqrt{1-\tau}\left\{(1-\tau)^{-2}\int_\tau^1 L_c^1(r,\tau)^2 dr\right\}^2}$$

$$\text{Dan-J} \xrightarrow{d} \frac{\kappa + \frac{12}{(1-\tau)^3}\int_\tau^1 \left\{r - \frac{1}{2}(1+\tau)\right\} K_c(r)dr}{\sqrt{\int_\tau^1 \int_\tau^1 -k''\left(\frac{r-s}{(1-\tau)b_{opt}(c)}\right)\left\{\int_\tau^r L_c^1(t,\tau)dt\right\}\left\{\int_\tau^s L_c^1(t,\tau)dt\right\} drds}} \times \frac{b_{opt}(c)(1-\tau)^3}{\sqrt{12}} \exp\{-c_\gamma' A_c(r,\tau)\}$$

$r \in [0,1]$, *where*

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

with $W_c(r) := \int_0^r e^{-(r-s)c} dW(s)$, $W(r)$ a standard Wiener process on $[0,1]$. Also, $k''(x)$ denotes the second derivative with respect to x of the Daniell kernel, and

$$A_c(r,\tau) := \frac{\int_\tau^1 L_c^1(r,\tau)^2 dr}{\int_\tau^1 L_c^9(r,\tau)^2 dr} - 1$$

with $L_c^j(r,\tau)$ used to denote the continuous time residual from the projection of $K_c(r)$, $r \geq \tau$, onto the space spanned by $\{1, r-\tau, (r-\tau)^2, ..., (r-\tau)^j\}$.

Remark 3.1. Full sample results (i.e. those obtained when no trimming is used) are obtained by setting $\tau = 0$ in Lemma 1. For $\tau > 0$ there is no need to re-index any data; that is, the time trend regressor can always begin at 1. As noted in section 2.1, it is clear that z_λ and t_β^{RQF} share identical asymptotic properties across this local-to-unity autoregressive root environment. In what follows, therefore, we shall simply refer to z_λ, it being understood that entirely similar comments apply throughout to t_{β}^{RQF}.[1] From the representations given in Lemma 1 it is also evident that all the statistics depend in the limit on the initial condition parameter α whenever $c > 0$ (the near-integrated case), but are asymptotically invariant to α when $c = 0$ (the exact $I(1)$ case).

Figures 1 – 4 about here

We now directly simulate the limiting representations in Lemma 1, approximating the Wiener processes using $NIID(0,1)$ random variates, and approximating integrals by normalized sums of 1000 steps. The Monte Carlo simulations were programmed in Gauss 9.0 using 50,000 replications. In reporting results, we use asymptotic critical values appropriate for a nominal 0.05 level for the two-tailed tests $|z_\lambda|$, $|z_\lambda^{m2}|$ and $|Dan\text{-}J|$. It should be kept in mind, however, that these asymptotic critical values are appropriate only for the case $c = 0$; elsewhere the test sizes depend on c, even when $\alpha = 0$.

Figures 1(a)-1(d) depict the tests' asymptotic size functions under the null hypothesis $\kappa = 0$ as functions of α across $\alpha = \{-6.0, -5.8, ..., 0.0, 0.2, ... + 6.0\}$, separately for $c = \{5, 10, 20, 30\}$. Our interest here is in the full sample tests which are the solid lines on the graphs indexed $\tau = 0$. As might be expected, when $\kappa = 0$ the tests' sizes are functions of $|\alpha|$ (i.e. the functions are symmetric in the sign of α). This arises because, for a given α, equality in distribution holds between a statistic based on $K_c(r)$ and one based on $-K_c(r)$. It is immediately evident from Figures 1(a)-1(d) that both $|z_\lambda|$ and $|z_\lambda^{m2}|$ can be very badly over-sized; the former reaching a size of around 0.40 for small c and large values of $|\alpha|$; the latter reaching around 0.50 size for large c and moderate values of $|\alpha|$. In sharp contrast, while the size of $|Dan\text{-}J|$ is influenced by both c and $|\alpha|$, it is nowhere over-sized.

We next examine the tests' asymptotic local power properties under the alternative $\kappa \neq 0$. Here we repeat the above simulation experiments, now with $\kappa = \{1, 2, 3\}$ in Figures 2(a)-2(d) to Figures 4(a)-4(d), respectively.[2] By way of a general observation, the tests' powers are seen to not be symmetric in $|\alpha|$. Other things equal, the tests' powers are higher when $\alpha = -|\alpha|$ than when $\alpha = |\alpha|$. A heuristic explanation of this

[1] In this paper we do not include an asymptotic analysis of the original $t\text{-}PSW^1$ test of Vogelsang (1998). It is sufficient to note that the asymptotic properties of this test were found, throughout, to be very similar to those of *Dan-J* and so this can be considered a good proxy for the limit behaviour of $t\text{-}PSW^1$. The limit expression for the distribution of $t\text{-}PSW^1$ is, however, available upon request.

[2] In these simulations we do not need separately to consider negative values for κ. When $\alpha = 0$, our results are the same for negative and positive κ. When $\alpha \neq 0$, results for negative κ and positive (negative) α are the same as those for positive κ and negative (positive) α.

finding is offered by noting that since $\kappa > 0$ here, when α is negative (such that the initial conditional lies below the positive trend), its effect is in some sense to intensify the appearance of a positive trend in the data. Conversely, when α is positive (such that the initial conditional lies above the positive trend), the effect is to mitigate the appearance of a positive trend in the data. This asymmetry is by far and away the most evident for $|z_\lambda|$. For example, in Figure 2 ($\kappa = 1$) when $c = 5$, $|z_\lambda|$ has power around one when $\alpha = -6.0$ but this decreases rapidly in α, reaching what is effectively zero for $\alpha > 0$. In fact, it appears throughout Figures 2-4 that $|z_\lambda|$ virtually always shows power which is decreasing in α. These power asymmetries are rather less emphatic for $|z_\lambda^{m2}|$, though clearly still evident. For $|Dan\text{-}J|$ they are rather more subtle still. What is noticeable is that, in contrast to $|z_\lambda|$, for these two tests their highest powers (for a given κ) are generally associated with values of α in the region of zero, and the power drop-off away from this region can still be very rapid in both directions of α, particularly so for $|Dan\text{-}J|$. For all three tests, power, for a given c and α, increases in κ.

Ranking the tests in terms of their power is not a straightforward matter. From Figure 2 ($\kappa = 1$) it would be very difficult to make a case for $|z_\lambda|$ as across each panel its power is almost everywhere zero for $\alpha > 0$ (and zero over much of the negative α region in the bottom two panels). Either of $|z_\lambda^{m2}|$ and $|Dan\text{-}J|$ would constitute a rather better behaved test here, and between the two $|z_\lambda^{m2}|$ generally shows the higher power away from the region of large, positive α. As regards Figure 3 ($\kappa = 2$), $|z_\lambda|$ still has zero power for a substantial part of the positive α region, though for negative α it is generally the most powerful test. In addition, $|z_\lambda^{m2}|$ is starting to show considerable gains over $|Dan\text{-}J|$. In Figure 4 ($\kappa = 3$), $|z_\lambda|$ is almost everywhere the most powerful test, followed by $|z_\lambda^{m2}|$.

About the only tentative ranking available from this power analysis is that $|z_\lambda^{m2}|$ generally outperforms $|Dan\text{-}J|$. Any rankings involving $|z_\lambda|$ are largely rendered impossible due its extremes of behaviour. Once we factor in the size issues demonstrated above, the picture is even more complicated; the initial condition has arguably the most effect (in terms of over-sizing) on $|z_\lambda^{m2}|$ and certainly the least effect on $|Dan\text{-}J|$. Of the three tests, while $|z_\lambda|$ and $|z_\lambda^{m2}|$ can display more power than $|Dan\text{-}J|$, the latter is the only test that is not subject to over-sizing across α. However, what is very clear from these results is that, taking $\alpha = 0$ as a central case, any of the tests can exhibit considerable size distortions and/or very low power for values of α away from zero.

In the presence of uncertainty about the initial condition, it would seem inadvisable to use any of these trend tests as a test of $\kappa = 0$. As an example of where application of these tests could be problematic, consider their possible use as a pre-test to decide whether to include a constant or constant and linear trend in the regression specification of a unit root test. Size issues notwithstanding, when $c > 0$, for certain non-negligible initial condition configurations, each of these trend tests can exhibit near-zero power even when a substantial local trend is present. In such circumstances, therefore, non-trend unit root test regressions would be signalled, with the obvious concomitant that the resulting unit root test procedure would have only trivial power. In the next section we pursue the idea of how the effects of the initial condition on the trend tests might be countermanded.

4 Tests Based on Trimmed Series and their Asymptotic Properties

The reasoning behind using a trimmed series is as follows. First note that we can express (2) in the form:

$$\begin{aligned} u_t &= \rho_T^{t-1} u_1 + v_t \\ v_t &= \rho_T v_{t-1} + \varepsilon_t, \quad t = 2, ..., T \end{aligned}$$

with $v_1 = 0$. We therefore have that, $T^{-1/2} u_{\lfloor \tau T \rfloor} = \rho_T^{\lfloor \tau T \rfloor - 1} T^{-1/2} u_1 + T^{-1/2} v_{\tau T}$. Notice that, under the conditions of Lemma 1, $\lim_{T \to \infty} (\rho_T^{\lfloor \tau T \rfloor - 1}) = \lim_{T \to \infty} (1 - c/T)^{\lfloor \tau T \rfloor - 1} = e^{-c\tau}$, while as noted in remark 2.1, $\lim_{T \to \infty} (T^{-1/2} u_1) = \alpha \lambda_c$. Consequently, if we construct tests using all the data ($\tau = 0$) then the (scaled) initial condition $T^{-1/2} u_1$ has asymptotic magnitude $\alpha \lambda_c$, while if, on the other hand, we use data from observation $\lfloor \tau T \rfloor + 1$ ($\tau > 0$) onwards, then the (scaled) initial condition $T^{-1/2} u_{\lfloor \tau T \rfloor}$ has asymptotic magnitude $e^{-c\tau} \alpha \lambda_c + h$, where h is a $N(0, \lambda_c^2)$ variate. Since $e^{-c\tau} < 1$ and is a decreasing (towards zero) function of τ, the influence of u_1 can essentially be countermanded by the use of a (left) trimmed data set.

Recall that Lemma 1 provides the limit distributions of the trimmed statistics for $\tau > 0$. What should be clear from these representations is that the limiting distributions of the trimmed statistics still depend on the initial condition parameter α when $c > 0$. However, from the arguments made above it is anticipated that the impact of α on these distributions will be reduced for $\tau > 0$ *vis-à-vis* $\tau = 0$. This will be explored numerically below. While perhaps not immediately apparent from the limiting representations, the limit null critical values for the trend tests based on the trimmed series for a given value of τ are identical to those which obtain for $\tau = 0$, since these are evaluated under $c = 0$, where the initial condition plays no role.

Figures 1(a)-1(d) show the asymptotic sizes of the trimmed variants of the $|z_\lambda|$, $|z_\lambda^{m2}|$ and $|$ $Dan\text{-}J|$ tests under the null $\kappa = 0$ using only the last 85% of available observations, these are the dotted lines indexed $\tau = 0.15$. As with the corresponding untrimmed tests, the sizes are seen to be functions of both $|\alpha|$ and c. What can principally be see from these graphs is the result that any over-sizing seen in $|z_\lambda|$ is completely eliminated for the settings considered (it is now uniformly under-sized with size close to zero). Moreover, the trimmed $|z_\lambda^{m2}|$ test is also under-sized everywhere except for $c = 30$ where it is only very modestly over-sized. The trimmed $|Dan\text{-}J|$ test remains, like its full-sample counterpart, under-sized everywhere. For these latter two trimmed tests, their sizes stay comfortably above zero and, for a given c, are now very flat in α. It is clear therefore that trimming has, in the main, had the desired effect of reducing the impact of the initial condition on the trend tests, at least as regards test size.

Figures 2(a)-2(d) to Figures 4(a)-4(d), respectively show the power of tests based on the trimmed series. It is immediately obvious that this approach works rather well for both $|z_\lambda^{m2}|$ and $|Dan\text{-}J|$. The power profiles of both of these trimmed tests are very much flatter across α then are their full sample counterparts and while some modest levels of power are sacrificed in the central region around $|\alpha| = 0$, these would appear to be more than compensated in the regions of large negative or large positive α. Of the two, the trimmed variant of $|z_\lambda^{m2}|$ is always more powerful than $|Dan\text{-}J|$, though the differences for a given c are often much smaller than between their full sample counterparts (see, for example, Figure 2 ($\kappa = 1$) when $c = 20$). Trimming does not perform particularly well, however, where $|z_\lambda|$ is concerned. In Figure 2 ($\kappa = 1$) its power is effectively zero across all α in every panel except when $c = 5$. In Figure 3 ($\kappa = 2$) it begins to behave more in line with expectation; some loss of the very high power of its full sample counterpart for large negative α is traded off against an improvement in power for large positive α. The power profile of the trimmed variant is relatively flatter, particularly in the bottom two panels ($c = 20, 30$), but at a level well below that of the trimmed variants of $|z_\lambda^{m2}|$ and $|Dan\text{-}J|$. It does rather better in Figure 4 ($\kappa = 3$), but here, for the larger values of c, all the tests have very decent levels of power anyway. It appears therefore that trimming the series is insufficient to curb the extremes of behaviour of $|z_\lambda|$ to an acceptable degree.

Taking the size and power results together, running either of $|z_\lambda^{m2}|$ and $|Dan\text{-}J|$ on a suitably trimmed series would appear to be an expedient way of reducing the negative effects of a large initial condition on the size and power of the former and power of the latter. We cannot claim that this strategy is really suitable for $|z_\lambda|$, however. Between $|z_\lambda^{m2}|$ and $|Dan\text{-}J|$, it is probably the latter for which this approach yields the greatest benefits in terms of yielding a size-controlled and yet still powerful testing procedure.

As regards the choice of the trimming parameter, τ, we conducted corresponding analyses with values other than 0.15 (in steps of 0.05). To summarize briefly, using values of τ less than 0.15 caused the power profiles of the trimmed $|z_{\lambda}^{m2}|$ and $|Dan\text{-}J|$ tests to be less flat and more resemble the full sample case; while the over-sizing problems of $|z_{\lambda}^{m2}|$ started to resurface. Larger values of τ retained the flat power profiles, but at a reduced level; all of which is in line with what would be expected. While $\tau = 0.15$ is in no sense meant to represent any optimal degree of trimming, we would certainly consider it to represent a pragmatic choice on the basis of our experimentation. Of course, the asymptotic critical values for the tests are unaffected by the choice of τ, thereby allowing practitioners a bespoke choice for the degree of trimming, should $\tau = 0.15$ prove unappealing.

5 Detecting a Large Initial Condition

When considering $|z_{\lambda}^{m2}|$ and $|Dan\text{-}J|$, the above analysis showed that trimming the series, while leading to improved power for large values of $|\alpha|$, also leads to a power drop in the region around $|\alpha| = 0$. Obviously then, any sample information we can obtain on the magnitude of $|\alpha|$ could potentially be useful in that it could be employed to ascertain whether the use of trimming would be expected to be useful or not.

While it is not possible to estimate α consistently in our current framework, if our aim is simply to indicate the presence of a large initial condition, we can make use of the following estimator, closely related to that of Harvey and Leybourne (2005), given by:

$$\tilde{\alpha} := \frac{y_1 - \tilde{\mu} - \tilde{\beta}_T}{\tilde{\sigma}}$$

where $\tilde{\mu}$ and $\tilde{\beta}_T$ are the OLS estimators from the fitted regression

$$y_t = \tilde{\mu} + \tilde{\beta}_T t + \tilde{u}_t, \;\; t = \lfloor \tau T \rfloor + 1, ..., T$$

and $\tilde{\sigma}^2 := (\lfloor T(1-\tau) \rfloor - 2)^{-1} \sum_{t=\lfloor \tau T \rfloor + 1}^{T} \tilde{u}_t^2$.[3] Notice, crucially, that $\tilde{\alpha}$ is exact invariant to the parameters μ and β_T which characterise the trend function in (1). Consequently, the behaviour of $\tilde{\alpha}$ does not depend on whether a linear trend is present in the data or

[3] This estimator differs from that given in Harvey and Leybourne (2005) by only using the trimmed sample $t = \lfloor \tau T \rfloor + 1, ..., T$ to estimate μ and β_T as opposed to the full sample. This would seem entirely reasonable in the current context. Note also that the time trend index starts at $\lfloor \tau T \rfloor + 1$, not 1. This is important here as it ensures that y_1 is correctly de-trended.

not. Under the conditions of Lemma 1 it is straightforward to show that

$$\tilde{\alpha} \Rightarrow \frac{4\frac{1+\tau+\tau^2}{(1-\tau)^3}\int_\tau^1 K_c(r)dr - 6\frac{1+\tau}{(1-\tau)^3}\int_\tau^1 rK_c(r)dr}{\sqrt{\frac{\int_\tau^1 L_c^1(r,\tau)^2dr}{1-\tau}}}. \tag{5}$$

Remark 5.1. The proof of this result follows along very similar lines to that of Lemma 1, and is therefore omitted. Full sample results are again obtained by setting $\tau = 0$, at which point this estimator reduces to that given in Harvey and Leybourne (2005). Notice also that the limit distribution of $\hat{\alpha}$ depends on α and c, but does not depend on α when $c = 0$.

Despite the fact that, as (5) demonstrates, $\tilde{\alpha}$ is not a consistent estimator of α under our local-to-unity autoregressive root framework, it is still the case that $\tilde{\alpha}$ contains potentially useful information about α, at least to the extent that large values of $|\tilde{\alpha}|$ might be associated with large values of $|\alpha|$. We therefore consider an heuristic rule of the form:

If $|\tilde{\alpha}| \leq \alpha^*$ *then construct the trend tests using the full sample, while if* $|\tilde{\alpha}| > \alpha^*$ *construct the trend tests using a trimmed sample.*

In employing this auxiliary procedure we will need, in addition to the choice of trimming parameter, τ, to select a value for α^* in the rule above. We report results for $\alpha^* = 2$ (again with $\tau = 0.15$), this choice being motivated by reference to the power profiles for $|z_\lambda^{m2}|$ and $|Dan\text{-}J|$ in Figures 2 to 4, where the negative effects of the initial condition appear to become most pronounced for $|\alpha| > 2$. Figures 1(a)-1(d) show asymptotic sizes of the resulting trend tests under the null $\kappa = 0$, these are the dashed lines indexed $\tau = 0.15$, $\alpha^* = 2$. Although the auxiliary procedure of this section is intended only for $|z_\lambda^{m2}|$ and $|Dan\text{-}J|$, for completeness we include results for $|z_\lambda|$ also. It is clear that the auxiliary procedure has almost no impact on the size of $|Dan\text{-}J|$ (nor $|z_\lambda|$) but does result in some over-sizing of $|z_\lambda^{m2}|$ for intermediate values of $|\alpha|$.

As regards test power, Figures 2-4 show that this auxiliary procedure works well in restoring the power of $|z_\lambda^{m2}|$ and $|Dan\text{-}J|$ towards the levels displayed by the tests based on the untrimmed series for the case of small and intermediate values of $|\alpha|$, while retaining the flat and high power levels of the tests based on the trimmed series for the larger values of $|\alpha|$ (again this procedure does not seem particularly satisfactory for $|z_\lambda|$). We also experimented with different values of the cut-off value α^*. Again, while the choice $\alpha^* = 2$ is by no means meant to represent an optimal one, its properties appeared as satisfactory as those for any other value in the region of $\alpha^* = 2$.

The auxiliary procedure would therefore seem particularly relevant for $|Dan\text{-}J|$. Of course, while all of the full sample and trimmed sample trend tests are exactly sized controlled under the null $\kappa = 0$ in the case where $c = 0$, this is not the case for those which also employ the auxiliary procedure. In Figure 5 we show the asymptotic size and power of all the tests when $c = 0$ as a function of κ across $\kappa = \{0.0, 0.1, ..., 3.0\}$. When $\kappa = 0$ there is only very modest over-sizing evident for any of the trimmed series trend tests employing the auxiliary procedure; specifically, the asymptotic sizes of $|z_\lambda|$, $|z_\lambda^{m2}|$ and $|Dan\text{-}J|$ are 0.053, 0.066 and 0.069, respectively. This might, *a priori*, be expected to be the case because the auxiliary procedure is always selecting between two tests which are both (marginally) correctly sized. It is also comforting to note that the tests based on the trimmed series tests lose little in terms of power relative to their full series counterparts across different $\kappa > 0$ when $c = 0$, the situation where the initial condition plays no role.

6 Conclusions

In this paper we have analysed the impact of the initial condition on the large sample behaviour of the robust (to the order of integration of the data) trend tests of Bunzel and Vogelsang (2005), Harvey *et al.* (2007) and Perron and Yabu (2009). The initial condition, formulated as in Müller and Elliott (2003), was shown to have a substantial impact on the asymptotic size and/or local power of these tests. Specifically, the (asymptotically equivalent) z_λ and t_β^{RQF} tests of Harvey *et al.* (2007) and Perron and Yabu (2009), respectively, were shown to be highly unreliable showing either very significant over or under size (depending on the degree of persistence present in the data, as measured by the local-to-unity autoregressive parameter) with these effects becoming more pronounced the larger the magnitude of the initial condition. The $|z_\lambda^{m2}|$ test of Harvey *et al.* (2007) was also shown to be unreliable displaying considerable over-size in many cases. Of all the tests considered the $|Dan\text{-}J|$ test of Bunzel and Vogelsang (2005) appeared to be the most reliable in terms of its size properties being nowhere over-sized for the experimental designs considered. Large initial values were also shown to compromise badly the asymptotic local power of all of the tests. The local power functions of the z_λ and t_β^{RQF} tests were shown to be especially erratic in the presence of large initial conditions.

The influence of the initial condition on the trend tests cannot be removed simply by dummying out (conditioning on) the first observation. Indeed, asymptotically, such an approach would be no different from using the full sample. We have shown, however, that the impact of the initial condition on the behaviour of the tests can be ameliorated by trimming the sample of the first $\lfloor \tau T \rfloor$ observations, $\tau > 0$. This method was shown to work well for the $|z_\lambda^{m2}|$ and $|Dan\text{-}J|$ tests greatly improving the size properties of both whilst also reducing the large decreases in power experienced by both of these tests in the presence of large initial conditions. A drawback of trimming is that power is unnecessarily lost in doing so when the initial condition is small. In order to provide procedures that account for the uncertainly that will exist in practice over whether the initial condition is sufficiently large to warrant trimming, we have also proposed an approach based on an auxiliary statistic which proxies the magnitude of the initial condition. Where the proxy is considered large (small) trimming is (is not) employed. Numerical evidence suggests that this method works well and appeared particularly suited to use with the $|Dan\text{-}J|$ test.

In summary the results in this paper suggest that the outcomes from robust trend tests should be treated with caution in cases where uncertainty exists over the magnitude of the initial condition. In particular, we strongly advise against the practice of using full sample robust trend tests as pre-tests for choosing whether to use a de-meaned or de-trended variant of a given unit root test. Although the influence of the initial condition on the behaviour of the robust trend tests can never be completely removed, for practical purposes we strongly recommend the use of the modified versions of these tests based on trimmed data.

Appendix

Under the conditions of Lemma 1, z_λ and t_β^{RQF} are both asymptotically equivalent to z_1 and therefore it is sufficient to examine the limiting behaviour of z_1. Write

$$\begin{aligned} z_1 &= \frac{\sum_{\lfloor \tau T \rfloor+1}^{T} \Delta y_t / \{T(1-\tau)\}}{\sqrt{\breve{\omega}_v^2 / \{T(1-\tau)\}}} \\ &= \frac{(1-\tau)^{-1/2} T^{-1/2} (y_T - y_{\lfloor \tau T \rfloor})}{\sqrt{\breve{\omega}_v^2}} \\ &= \frac{(1-\tau)^{-1/2} T^{-1/2} \{(\kappa \omega_\varepsilon T^{-1/2} T(1-\tau) + (u_T - u_{\lfloor \tau T \rfloor})\}}{\sqrt{\breve{\omega}_v^2}} \\ &= \frac{(1-\tau)\kappa\omega_\varepsilon + T^{-1/2}(u_T - u_{\lfloor \tau T \rfloor})}{(1-\tau)^{1/2}\sqrt{\breve{\omega}_v^2}}. \end{aligned}$$

Next, observe that under the conditions of Lemma 1

$$T^{-1/2}(u_T - u_{\lfloor \tau T \rfloor}) = T^{-1/2}(u_T - u_1) - T^{-1/2}(u_{\lfloor \tau T \rfloor} - u_1) \xrightarrow{d} K_c(1) - K_c(\tau)$$

since $T^{-1/2}(u_{\lfloor rT \rfloor} - u_1) \xrightarrow{d} K_c(r)$. The first result stated in Lemma 1 then follows from applications of the continuous mapping theorem and using the fact that $\breve{\omega}_v^2 \xrightarrow{p} \omega_\varepsilon^2$. The remaining limiting results for z_λ^{m2} and *Dan-J* can be established along very similar lines and are omitted in the interests of brevity.

References

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.

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.

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

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.

Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2009). Unit root testing in practice: dealing with uncertainty over the trend and initial condition (with commentaries and rejoinder). Econometric Theory 25, 583-586.

Kiefer, N. M. and Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory 21, 11301164.

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.

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

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.

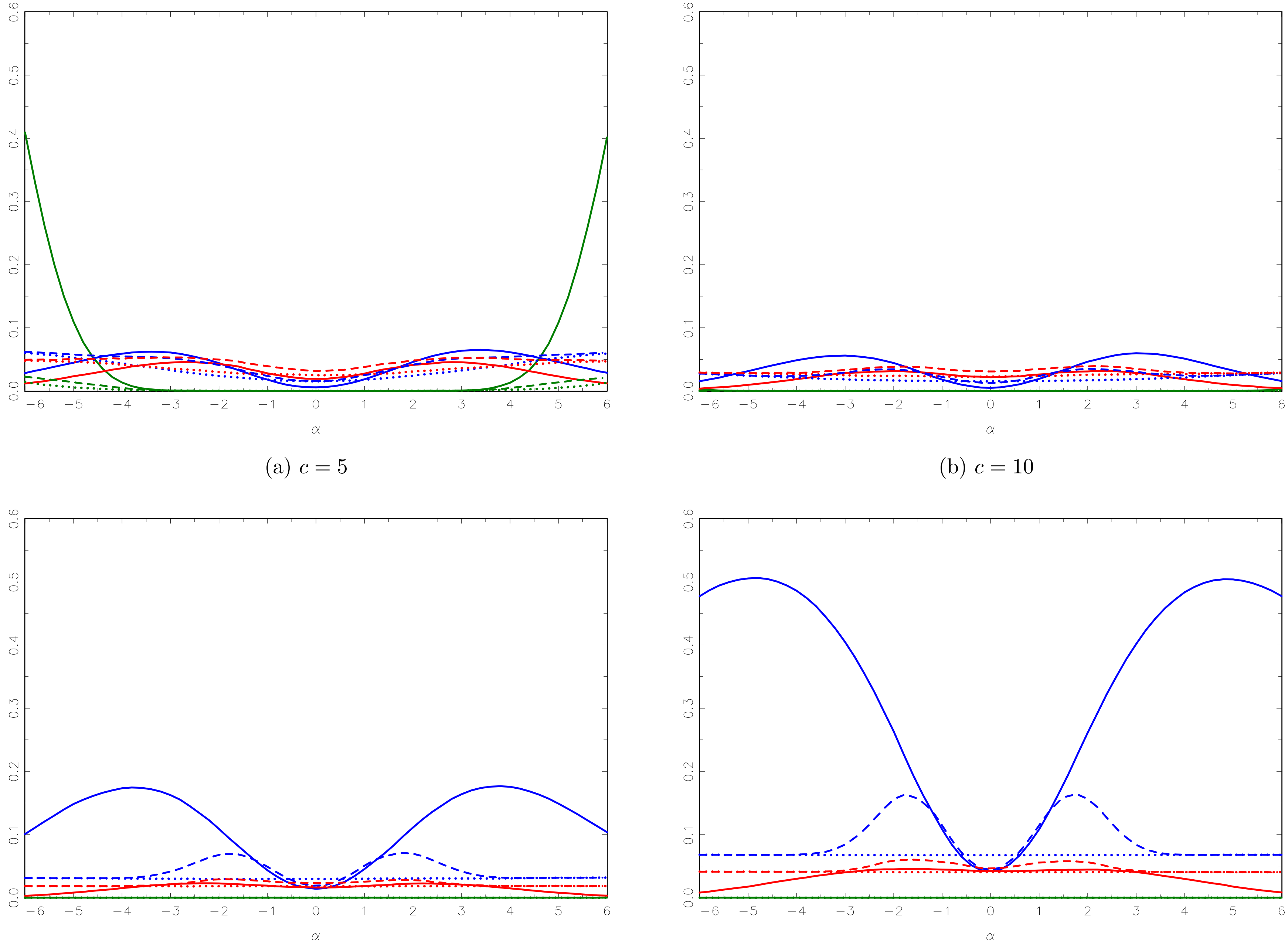
Park, J.Y. and Choi, B. (1988). A new approach to testing for a unit root. CAE Working Paper 88-23, Cornell University.

Perron, P. and Yabu, T. (2009). Estimating deterministic trends with an integrated or stationary noise component. Journal of Econometrics 151, 56-69.

Roy, A., Falk, B. and Fuller, W.A. (2004). Testing for trend in the presence of autoregressive errors. Journal of the American Statistical Association 99, 1082-1091.

Roy, A. and Fuller, W.A. (2001). Estimation for autoregressive processes with a root near one. Journal of Business and Economic Statistics 19, 482-493.

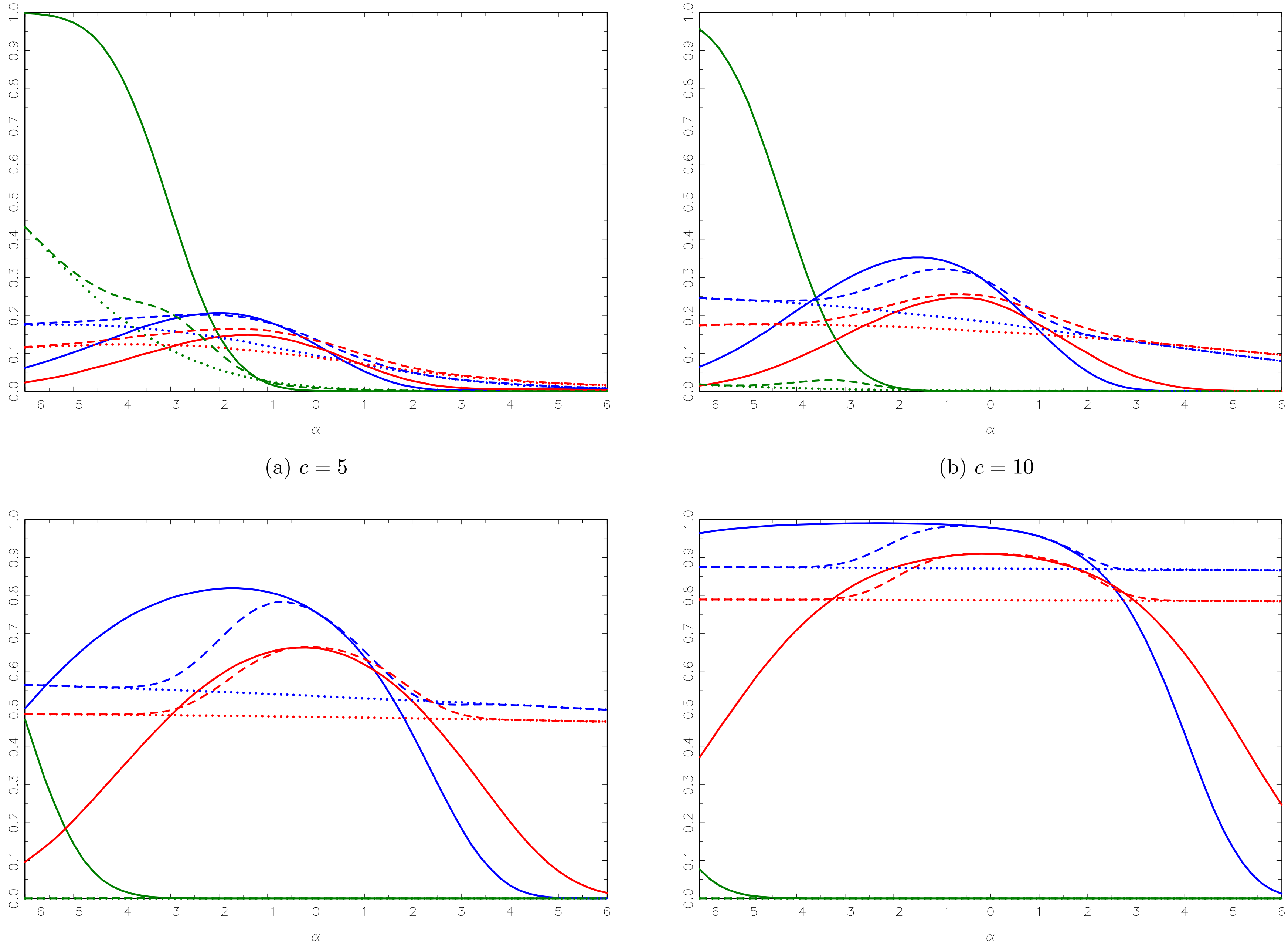
Vogelsang, T.J. (1998). Trend function hypothesis testing in the presence of serial correlation. Econometrica 66, 123-148.



(a) $c = 5$ (b) $c = 10$

(c) $c = 20$ (d) $c = 30$

Figure 1. Asymptotic size of trend tests. $|z_\lambda|\,(\tau = 0)$: —, $|z_\lambda^{m2}|\,(\tau = 0)$: —, $|Dan\text{-}J|\,(\tau = 0)$: —, $|z_\lambda|\,(\tau = 0.15)$: · · ·, $|z_\lambda^{m2}|\,(\tau = 0.15)$: · · ·, $|Dan\text{-}J|\,(\tau = 0.15)$: · · ·, $|z_\lambda|\,(\tau = 0.15, \alpha^* = 2)$: – –, $|z_\lambda^{m2}|\,(\tau = 0.15, \alpha^* = 2)$: – –, $|Dan\text{-}J|\,(\tau = 0.15, \alpha^* = 2)$: – –.



(a) $c = 5$ (b) $c = 10$

(c) $c = 20$ (d) $c = 30$

Figure 2. Asymptotic local power of trend tests: $\kappa = 1$. $|z_\lambda|\,(\tau = 0)$: green solid line, $|z_\lambda^{m2}|\,(\tau = 0)$: blue solid line, $|Dan\text{-}J|\,(\tau = 0)$: red solid line, $|z_\lambda|\,(\tau = 0.15)$: green dotted line, $|z_\lambda^{m2}|\,(\tau = 0.15)$: blue dotted line, $|Dan\text{-}J|\,(\tau = 0.15)$: red dotted line, $|z_\lambda|\,(\tau = 0.15, \alpha^* = 2)$: green dashed line, $|z_\lambda^{m2}|\,(\tau = 0.15, \alpha^* = 2)$: blue dashed line, $|Dan\text{-}J|\,(\tau = 0.15, \alpha^* = 2)$: red dashed line.

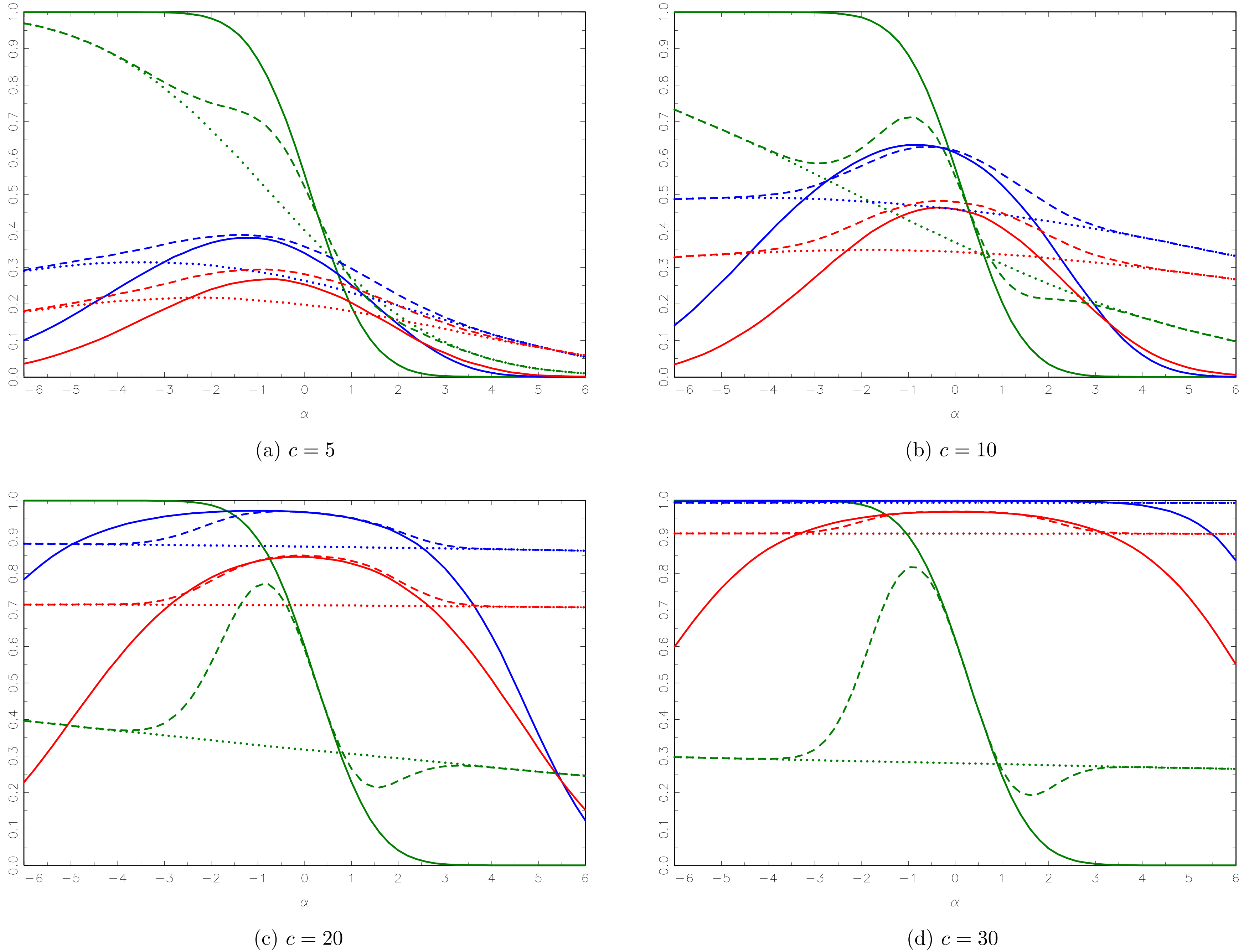
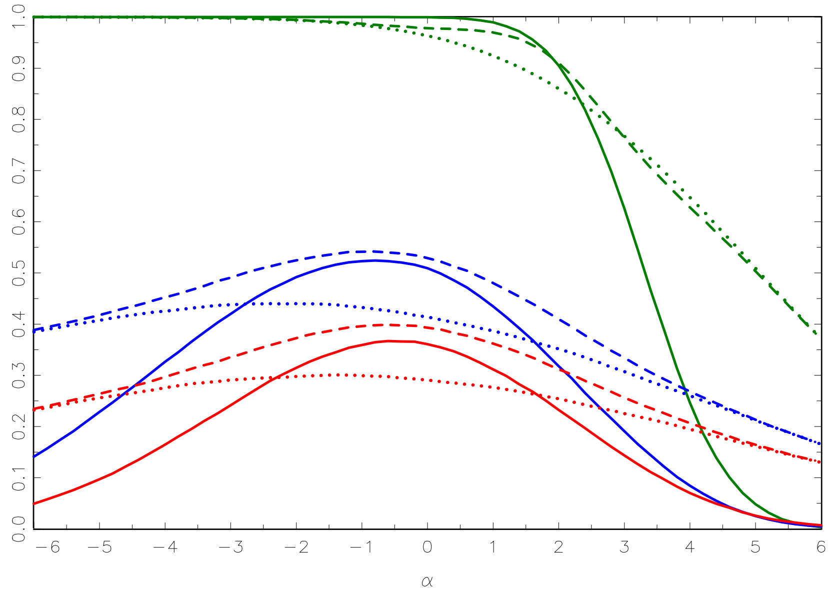
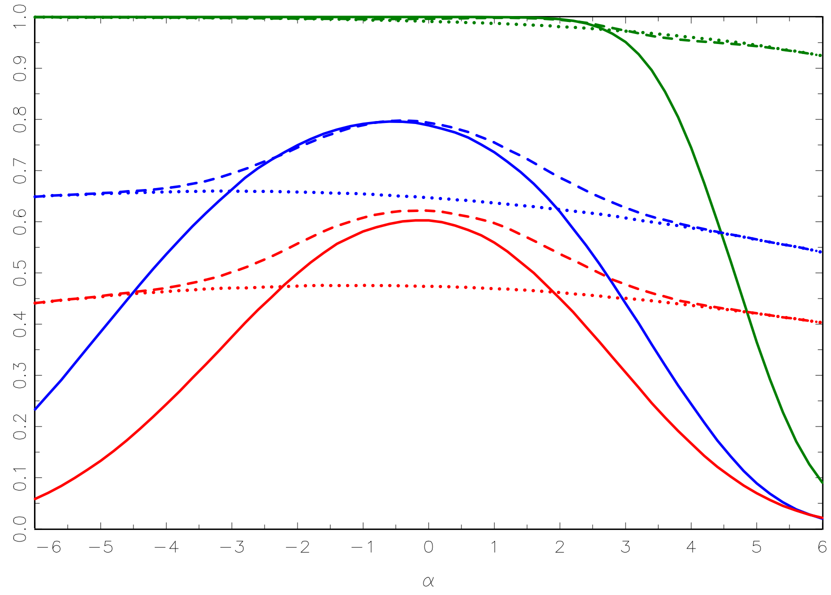


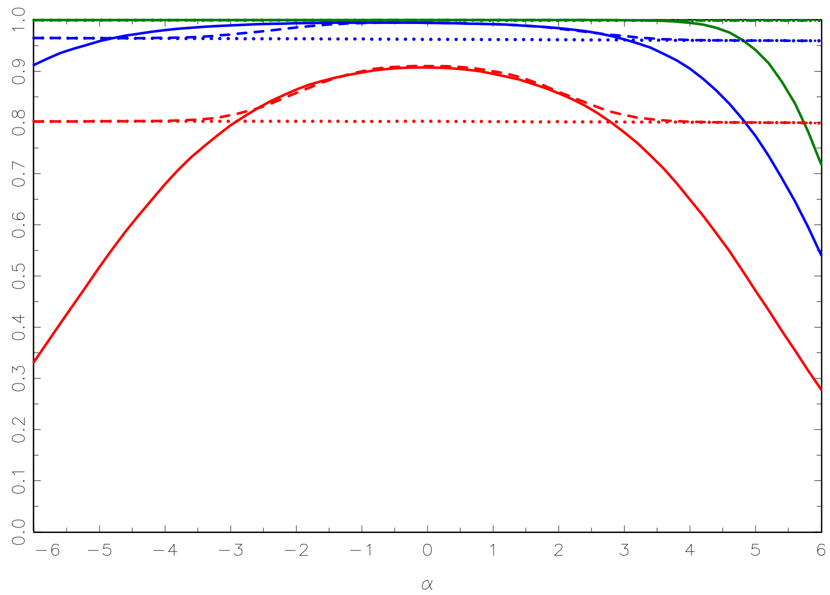
Figure 3. Asymptotic local power of trend tests: $\kappa = 2$. $|z_\lambda|\,(\tau = 0)$: , $|z_\lambda^{m2}|\,(\tau = 0)$: , $|Dan\text{-}J|\,(\tau = 0)$: , $|z_\lambda|\,(\tau = 0.15)$: , $|z_\lambda^{m2}|\,(\tau = 0.15)$: , $|Dan\text{-}J|\,(\tau = 0.15)$: , $|z_\lambda|\,(\tau = 0.15, \alpha^* = 2)$: , $|z_\lambda^{m2}|\,(\tau = 0.15, \alpha^* = 2)$: , $|Dan\text{-}J|\,(\tau = 0.15, \alpha^* = 2)$: .



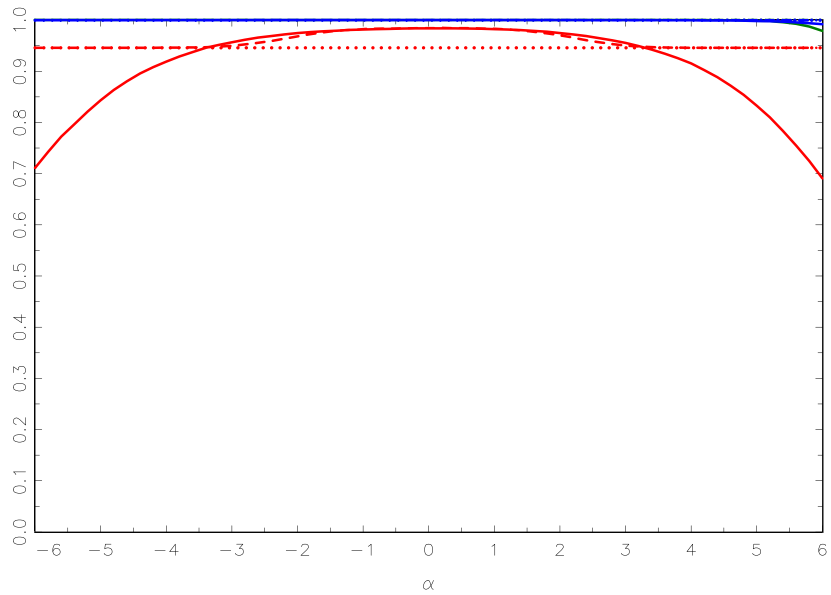
(a) $c = 5$



(b) $c = 10$



(c) $c = 20$



(d) $c = 30$

Figure 4. Asymptotic local power of trend tests: $\kappa = 3$. $|z_\lambda|\,(\tau = 0)$: ——, $|z_\lambda^{m2}|\,(\tau = 0)$: ——, $|Dan\text{-}J|\,(\tau = 0)$: ——, $|z_\lambda|\,(\tau = 0.15)$: $\cdots$, $|z_\lambda^{m2}|\,(\tau = 0.15)$: $\cdots$, $|Dan\text{-}J|\,(\tau = 0.15)$: $\cdots$, $|z_\lambda|\,(\tau = 0.15, \alpha^* = 2)$: – –, $|z_\lambda^{m2}|\,(\tau = 0.15, \alpha^* = 2)$: – –, $|Dan\text{-}J|\,(\tau = 0.15, \alpha^* = 2)$: – –.

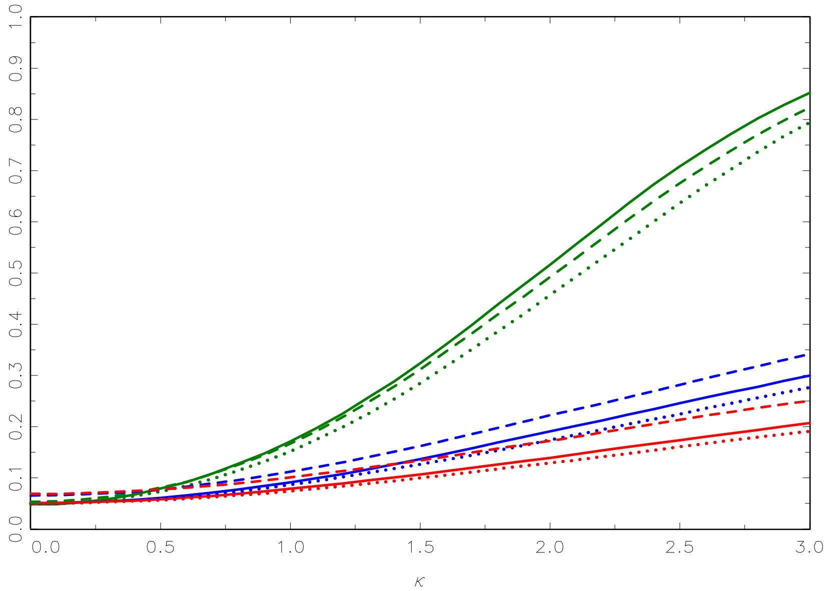


Figure 5. Asymptotic size and local power of trend tests: $c = 0$. $|z_\lambda|$ $(\tau = 0)$: ——, $|z_\lambda^{m2}|$ $(\tau = 0)$: ——, $|Dan\text{-}J|$ $(\tau = 0)$: ——, $|z_\lambda|$ $(\tau = 0.15)$: $\cdots$, $|z_\lambda^{m2}|$ $(\tau = 0.15)$: $\cdots$, $|Dan\text{-}J|$ $(\tau = 0.15)$: $\cdots$, $|z_\lambda|$ $(\tau = 0.15, \alpha^* = 2)$: – –, $|z_\lambda^{m2}|$ $(\tau = 0.15, \alpha^* = 2)$: – –, $|Dan\text{-}J|$ $(\tau = 0.15, \alpha^* = 2)$: – –.