Estimating Deterministic Trends with an Integrated or Stationary Noise Component*

Pierre Perron[†] **Tomoyoshi Yabu**[‡]

Boston University **University of Tsukuba**

October 30, 2004; This version: Fevruary 20, 2009

Abstract

We propose a test for the slope of a trend function when it is a priori unknown whether the series is trend-stationary or contains an autoregressive unit root. The procedure is based on a Feasible Quasi Generalized Least Squares method from an AR(1) specification with parameter α, the sum of the autoregressive coefficients. The estimate of α is the OLS estimate obtained from an autoregression applied to detrended data and is truncated to take a value 1 whenever the estimate is in a $T^{-\delta}$ neighborhood of 1. This makes the estimate "super-efficient" when $\alpha = 1$ and implies that inference on the slope parameter can be performed using the standard Normal distribution whether $\alpha = 1$ or $|\alpha| < 1$. Theoretical arguments and simulation evidence show that $\delta = 1/2$ is the appropriate choice. Simulations show that our procedure has better size and power properties than the tests proposed by Bunzel and Vogelsang (2005) and Harvey, Leybourne and Taylor (2007).

JEL Classification Number: C22.

Keywords: linear trend, unit root, median-unbiased estimates, GLS procedure, super efficient estimates.

---

*Pierre Perron acknowledges financial support from the National Science Foundation under Grant SES-0078492. We also wish to thank Eiji Kurozumi, Adam McCloskey and two referees for comments.

[†]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).

[‡]Graduate School of Systems and Information Engineering, University of Tsukuba, Tennodai, Tsukuba, Ibaraki, 305-8573, Japan (tyabu@sk.tsukuba.ac.jp).

1 Introduction

Many time series are well captured by a deterministic linear trend. With a logarithmic transformation, the slope of the trend function represents the average growth rate of the time series, a quantity of substantial interest. To be more precise, consider the following model for the time series process $\{y_t\}$:

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

where u_t are the deviations from the trend. The parameter β is then of primary interest.

Hypothesis testing on the slope of the trend function is important for many reasons. First, assessing whether a trend is present (i.e., whether $\beta = 0$ or $\beta \neq 0$) is of direct interest. One application pertains to global warming. For example, Vogelsang and Fomby (2002) found that global temperatures have increased roughly 0.5 degrees Celsius per 100 years, and Vogelsang and Franses (2005), analyzing monthly trends in temperature series for the Netherlands, found the largest positive trends in the winter. Another application relates to the Prebisch-Singer hypothesis, which states that over time the net barter terms of trade should decline between countries that primarily export commodities and countries that primarily export manufactured products. Bunzel and Vogelsang (2005) found strong evidence to support the Prebisch-Singer hypothesis. Second, Perron (1988) showed that the correct specification of the trend function is important in the context of testing for a unit root in the noise component u_t. He showed that unit root tests that omit a trend when one is present in the data are inconsistent. On the other hand, including a trend when one is not needed leads to a substantial loss of power (e.g., DeJong, Nankervis, Savin and Whiteman, 1992). Third, it is often of considerable interest to form confidence intervals on the rate of growth of series such as real GNP or other indices of real aggregate production. For example, this allows for cross-country comparisons or sub-period comparisons to determine if structural changes are present. Canjels and Watson (1997) analyzed the annual growth rates for real per-capita GDP in 128 countries. Also, Vogelsang (1998) provided estimates and confidence intervals for postwar real GNP quarterly growth rates for the G7 countries.

There is a large literature on issues pertaining to inference about the slope of a linear trend function, most of which is related to the case where the noise component is stationary, i.e., integrated of order zero, $I(0)$. A classic result due to Grenander and Rosenblatt (1957) states that the estimate of β obtained from a simple least-squares regression of the form (1) is asymptotically as efficient as that obtained from a Generalized Least Squares (GLS) regression when the process for u_t is correctly specified. However, it is now recognized that many economic time series of interest are potentially characterized as having a noise component u_t with an autoregressive root that is unity, i.e., integrated of order one, $I(1)$ (e.g., Nelson and Plosser, 1982), or a root that is close to one. In the former case, the least square estimate of β obtained from (1) is no longer asymptotically efficient. Instead, the estimate of the mean of the first-differenced series is efficient in large samples. In the case of a root close to one, the standard Grenander and Rosenblatt (1957) result still holds but the limiting Normal distribution is a poor approximation in sample sizes of interest.

It is important to realize that in all these cases the noise component can be either $I(0)$ or $I(1)$ and that in general no a priori knowledge about this is available. The limiting distribution of test statistics depends on the $I(0)$ or $I(1)$ dichotomy so that methods of inference that are robust to both possibilities are needed. Consider the standard t-statistic based on the OLS regression. The t-statistic for testing β converges in distribution to a $N(0,1)$ random variable when u_t is $I(0)$. On the other hand, as shown by Phillips and Durlauf (1988), the t-statistic normalized by $T^{-1/2}$ has a non-degenerate non-Normal limiting distribution when the errors are $I(1)$. This dichotomy is one of the major source of problems that makes inference about the slope of the trend function a difficult issue.

Several papers have tackled the issue of constructing tests and confidence intervals for the parameter β when it is not known a priori whether the noise component u_t is $I(1)$ or not. Most closely related to our work are the following papers. Sun and Pantula (1999) proposed a pre-test method which first applies a test of the unit root hypothesis and then chooses the critical value to be used for the t-statistic according to the outcome of the test. When using this method, however, the probability of using the critical values from the $I(0)$ case does not converges to zero when the errors are $I(1)$, and the simulations reported accordingly show substantial size distortions remain. Canjels and Watson (1997) considered various Feasible GLS methods. Their analysis is, however, restricted to the cases where u_t is either $I(1)$ or the autoregressive root is local to one (i.e., $u_t = \alpha_T u_{t-1} + e_t$ with $\alpha_T = 1 + c/T$ and e_t being $I(0)$). Hence, they do not allow $I(0)$ processes for the noise. Also, even with $I(1)$ or near $I(1)$ processes, their method yields confidence intervals that are substantially conservative with common sample sizes. Roy et al. (2004) considered a test based on a one-step Gauss Newton regression that uses a truncated weighted symmetric least-squares estimate of the autoregressive parameter (as suggested by Roy and Fuller, 2001) but the limit distribution of their test is not the same in the $I(1)$ and $I(0)$ cases. Vogelsang (1998), Bunzel and Vogelsang (2005) and Harvey et al. (2007) proposed tests valid with either $I(1)$ or $I(0)$ errors (see section 4). Their approach, however, uses randomly scaled versions of tests for trends so that in finite samples the good properties of such tests are lost, at least to some extent. Our approcah is different and does not involve such random scaling.

We propose a new robust test statistic that is valid with either $I(0)$ or $I(1)$ noise in the sense that the limit distribution is the same in both cases. It is based on applying a Feasible GLS procedure (e.g., the Cochrane-Orcutt procedure) with an estimate of the sum of the autoregressive coefficients truncated to take value one when the usual estimate is in a neighborhood of one. Hence, our estimate of the sum of the autoregressive coefficients is super-efficient when the true value is 1. As a result, the limiting distribution of our statistic does not depend on whether the noise is $I(0)$ or $I(1)$ and is the usual standard normal. Theoretical arguments are provided to show that the size of the neighborhood where truncation applies should be of order $T^{-1/2}$. Also, to improve the finite sample performance, we advocate the use of a truncated version of the weighted symmetric least-squares estimate as described by Roy and Fuller (2001). The resulting statistic is easy to implement and it has better size and power properties than tests previously suggested.

The outline of the paper is as follows. In section 2, we analyze the simple AR(1) case and lay out the basic framework, the test, its large sample properties and simulations for finite sample performance. In section 3, we consider extensions for a general class of processes for the noise u_t. Section 4 considers the finite sample size and power of our test including a comparison with those of Bunzel and Vogelsang (2005) and Harvey et al. (2007). Section 5 contains brief concluding comment and an appendix includes some technical derivations.

2 The AR(1) Case

We begin with the leading case of $AR(1)$ errors so that $\{y_t; t = 1, ..., T\}$, is generated by

$$y_t = \mu + \beta t + u_t, \quad u_t = \alpha u_{t-1} + e_t, \tag{2}$$

where $e_t \sim i.i.d.(0, \sigma^2)$ and $E(e_t^4) < \infty$. For simplicity, we let u_0 be some finite constant. Here $-1 < \alpha \leq 1$ so that both stationary and integrated errors are allowed.

Remark 1 *The analysis extends easily to more general trend functions of the form*

$$\sum_{i=0}^{n} \beta_i t^i + \sum_{j=1}^{m} \sum_{i=0}^{q_j} \theta_i (t - T_{B,j})^i 1(t > T_{B,j}),$$

where $1(\cdot)$ *is the indicator function. When* $m = 0$*, we have a polynomial trend function of order* n*. If* $m \neq 0$*,* m *breaks are present occuring at some breaks dates* $T_{B,j}$ $(j = 1, ..., m)$*. For* *the analysis to carry over, we need these break dates to be known. For the case of unknown break dates, the reader is refered to Perron and Yabu (2007). All of our main theoretical results remain valid in this more general case, though some equations may have a different form. For simplicity of exposition, we consider throughout the first-order linear trend, which is the leading case of interest in practice.*

The GLS estimator considered applies Ordinary Least Squares (OLS) to the regression

$$y_t - \alpha y_{t-1} = (1-\alpha)\mu + \beta[t - \alpha(t-1)] + e_t, \tag{3}$$

for $t = 2, ..., T$, together with

$$y_1 = \mu + \beta + u_1. \tag{4}$$

Consider the infeasible Generalized Least Squares (GLS) estimate of β that assumes α is known. The t-statistic for testing the null hypothesis that $\beta = \beta_0$, t_β^G, is then asymptotically distributed as $N(0,1)$ for any values of α in the permissible range. Canjels and Watson (1997) used a framework with α local to unity, i.e., $\alpha = 1 + c/T$, and an initial value u_0 having a variance that can depend on the sample size. Two methods have been advocated. First, the estimator described above is the GLS estimator under the assumption that $u_0 = 0$ and, accordingly, works best when the variance of u_0 is small. The second is the Prais-Winsten (1954) estimator which also incorporates the first observation but with

$$(1-\alpha^2)^{1/2} y_1 = (1-\alpha^2)^{1/2}\mu + (1-\alpha^2)^{1/2}\beta + (1-\alpha^2)^{1/2} u_1.$$

This estimator is superior when the variance of u_0 is not small. We shall not be concerned about the effect of the initial condition in this paper and, hence, we have assumed a fixed value for simplicity and will use the specification (4). Nevertheless, all our results remain valid using one or the other GLS estimate.

2.1 The feasible GLS estimate

We consider the Feasible GLS (FGLS) estimate of β that uses the following estimate of α:

$$\hat{\alpha} = \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} / \sum_{t=2}^{T} \hat{u}_{t-1}^2, \tag{5}$$

where $\{\hat{u}_t\}$ are the OLS residuals from a regression of y_t on $\{1, t\}$. When $|\alpha| < 1$, $T^{1/2}(\hat{\alpha} - \alpha)$ $\to^d N(0, 1-\alpha^2)$ and from Grenander-Rosenblatt (1957), the OLS and FGLS estimates of β

are asymptotically equivalent and the t-statistic from the FGLS procedure, denoted t_β^F, still has a $N(0,1)$ limit distribution. Things are different when $\alpha = 1$. From standard results,

$$T(\hat{\alpha}-1) \Rightarrow \int_0^1 W^*(r)dW(r)/\int_0^1 W^*(r)^2 dr \ \equiv \kappa, \tag{6}$$

with $W^*(r)$, $0 \le r \le 1$, the residual process from a continuous time regression of a unit Wiener process $W(r)$ on $\{1, r\}$, and where $\Rightarrow$ denotes weak convergence in distribution under the Skorohod topology. It is shown in the appendix that, provided $T(\hat{\alpha}-1) = O_p(1)$, the t-statistic from the FGLS procedure can be expressed as

$$\begin{aligned} t_\beta^F \;=\; & [[T^{-1/2}\sum_{t=2}^{T} e_t - T(\hat{\alpha}-1)T^{-3/2}\sum_{t=2}^{T} u_{t-1}] \\ & -T(\hat{\alpha}-1)[T^{-3/2}\sum_{t=2}^{T} te_t - T(\hat{\alpha}-1)T^{-5/2}\sum_{t=2}^{T} tu_{t-1}]] \\ & /\left[\sigma^2(1+T(\hat{\alpha}-1)+T^2(\hat{\alpha}-1)^2/3)\right]^{1/2} + o_p(1). \end{aligned} \tag{7}$$

Given (6), $T^{-1/2}\sum_{t=2}^{T} e_t \Rightarrow \sigma W(1)$, $T^{-3/2}\sum_{t=2}^{T} u_{t-1} \Rightarrow \sigma\int_0^1 W(r)dr$, $T^{-5/2}\sum_{t=2}^{T} tu_{t-1} \Rightarrow \sigma\int_0^1 rW(r)dr$ and $T^{-3/2}\sum_{t=2}^{T} te_t \Rightarrow \sigma\int_0^1 rdW(r)$ so the limit can be expressed as

$$t_\beta^F \Rightarrow \frac{W(1) - \kappa\int_0^1 W(r)dr - \kappa[\int_0^1 rdW(r) - \kappa\int_0^1 rW(r)dr]}{(1-\kappa+\kappa^2/3)^{1/2}}, \tag{8}$$

which is non-normal. Now suppose that κ, the limit of $T(\hat{\alpha}-1)$, is zero. Then $t_\beta^F \Rightarrow W(1) =^d N(0,1)$ and we would recover in the $I(1)$ case the same limit distribution as in the $I(0)$ case so that no discontinuity would be present. Our main idea exploits this feature.

2.2 A super-efficient estimate when $\alpha = 1$

The estimate of α that we propose is a super-efficient estimate when $\alpha = 1$. It is defined by

$$\hat{\alpha}_S = \begin{cases} \hat{\alpha} & \text{if } \ |\hat{\alpha}-1| > dT^{-\delta} \\ 1 & \text{if } \ |\hat{\alpha}-1| \le dT^{-\delta} \end{cases} \tag{9}$$

for some $\delta \in (0,1)$ and $d > 0$. Hence, when $\hat{\alpha}$ is in a $T^{-\delta}$ neighborhood of 1 it is assigned a value of 1. The fact that we use a super-efficient estimate when $\alpha = 1$ warrants some comments. As is well known, such estimates have better properties only on a set of measure zero, here $\alpha = 1$, and have less precision in a neighborhood of this value. This is not a problem here since α is a nuisance parameter in the context of our testing procedure which pertains to the coefficients of the trend function, in particular β. Hence, the fact that $\hat{\alpha}_S$

is less efficient than $\hat{\alpha}$ for values of α close to one does not imply that tests related to β will be badly bahaved, as we will show. When the FGLS procedure is applied with $\hat{\alpha}_S$ as an estimate of α, we denote the resulting estimate by $\hat{\beta}^{FS}$ and the t-statistic by t_{β}^{FS}. The following theorem, proved in the appendix, shows that this indeed changes nothing in the stationary case and delivers a t-statistic with a normal limit distribution when $\alpha = 1$.

Theorem 1 *a)* $T^{1/2}(\hat{\alpha}_S - \alpha) \to^d N(0, 1 - \alpha^2)$ *when* $|\alpha| < 1$*; and b)* $T(\hat{\alpha}_S - 1) \to_p 0$ *when* $\alpha = 1$*. Hence, under* $H_0 : \beta = \beta_0$, $t_{\beta}^{FS} \to^d N(0, 1)$ *for both* $|\alpha| < 1$ *and* $\alpha = 1$*.*

Constructing the FGLS regression using the truncated estimate $\hat{\alpha}_S$ effectively bridges the gap between the $I(0)$ and $I(1)$ cases, and the normal asymptotic distribution applies.

2.3 The case with α local to unity

The result obtained in Theorem 1 is pointwise in α with $-1 < \alpha \leq 1$ and does not hold uniformly, in particular in a local neighborhood of one [1]. Hence, it is of interest to see what happens when the true value of α is close to but not equal to one. Adopting the standard local to unity approach, we have the following result proved in the appendix.

Theorem 2 *Suppose that the data are generated by (2) with* $\alpha = \alpha_T = 1 + c/T$ *for some* $c \leq 0$*, then a)* $T(\hat{\alpha}_S - 1) \to_p 0$ *and, b) under* $H_0 : \beta = \beta_0$, $t_{\beta}^{FS} \to^d N(0, (\exp(2c) - 1)/2c)$*.*

The results are fairly intuitive. Since the true value of α is in a T^{-1} neighborhood of 1, and $\hat{\alpha}_S$ truncates values of $\hat{\alpha}$ in a $T^{-\delta}$ neighborhood of 1 for some $0 < \delta < 1$ (i.e., a larger neighborhood), in large enough samples $\hat{\alpha}_S = 1$. Then, the FGLS estimate of β becomes the first-difference (FD) estimator, $\hat{\beta}^{FD} = (T-1)^{-1} \sum_{t=2}^{T} \Delta y_t$ and the t-statistic is such that

$$t_{FD} = \frac{\hat{\beta}^{FD} - \beta}{std(\hat{\beta}^{FD})} = \frac{(T-1)^{-1/2} u_T - (T-1)^{-1/2} u_0}{[(T-1)^{-1} \sum_{t=2}^{T} (\Delta u_t)^2]^{1/2}} \Rightarrow J_c(1),$$

since $T^{-1/2} u_T \Rightarrow \sigma J_c(1)$, $T^{-1/2} u_0 = o_p(1)$ and $T^{-1} \sum_{t=2}^{T} (\Delta u_t)^2 \to_p \sigma^2$. Here $J_c(1) \equiv \int_0^1 \exp(c(1-s)) dW(s) \sim N(0, (\exp(2c) - 1)/2c)$. Note that when $c = 0$, we recover the result of Theorem 1 for the $I(1)$ case. However, when $c < 0$, the variance of the limiting distribution is different from 1. In fact, it is lower than 1 so that, without modifications, a

[1] It is, however, possible to show, using the results of Mikusheva (2007), that our test has an asymptotic size no higher than the stated nominal size uniformly over $|\alpha| \leq 1$. Though, as we show below, the test is conservative for values of α local to one.

conservative test may be expected for values of α close to 1, relative to the sample size. Also, the limit of the variance as $c \to -\infty$ is 0, not 1, and we do not recover the same result that applies to the $I(0)$ case. As noted by Phillips and Lee (1996), the local to unity asymptotic framework with $c \to -\infty$ involves a doubly infinite triangular array such that the limit of the statistic depends on the relative approach to infinity of c and T. To address this issue and also provide some guidelines on the choice of the truncation parameter δ, we analyze the properties of the OLS, FD and FGLS estimate of β in a local asymptotic framework whereby c is allowed to increase with the sample size. To that effect, we specify that $c = bT^{1-h}$ for some $b < 0$ and some $0 < h < 1$, and accordingly $\alpha = \alpha_T = 1 + b/T^h$. When h approaches 1, we get back to the local to unity case, and when h approaches 0, we get closer to the fixed $|\alpha| < 1$ case. We start with the following theorem.

Theorem 3 *Suppose that $u_t = \alpha_T u_{t-1} + e_t$ with e_t as defined by (2), and $\alpha_T = 1 + b/T^h$ for some $b < 0$ and some $0 < h < 1$. a) Let $\hat{\beta}^{ols}$ be the OLS estimate of β in the regression $y_t = \mu + \beta t + u_t$, then for $0 < h < 1$, $T^{3/2-h}(\hat{\beta}^{ols} - \beta) \Rightarrow N(0, 12\sigma^2/b^2)$ and the associated t-statistic is such that $t_\beta^{ols} \Rightarrow N(0,1)$; b) for the FGLS estimate, $T^{3/2-h}(\hat{\beta}^{F} - \beta) \Rightarrow N(0, 12\sigma^2/b^2)$ if $0 < h < 1/2$, and $T^{3/2-h}(\hat{\beta}^{F} - \beta) \Rightarrow N(0, 3\sigma^2/b^2)$ if $1/2 < h < 1$, while $t_\beta^F \Rightarrow N(0,1)$ for $0 < h < 1$; c) for the FD estimate, $T^{1-h/2}(\hat{\beta}^{FD} - \beta) \Rightarrow N(0, \sigma^2/(-2b))$ for $0 < h < 1$, while $t_\beta^{FD} \Rightarrow 0$.*

This theorem is helpful for analyzing the effects of different truncations, i.e., different values of δ, on the properties of the estimate $\hat{\beta}^{FS}$ and the t-statistic t_β^{FS} in a local framework where c gets large. Note that $\hat{\beta}^{FS} = \hat{\beta}^{FD}$ when truncation applies, otherwise $\hat{\beta}^{FS} = \hat{\beta}^{F}$. Consider first the relative efficiency of the estimates for different increasing sequences for c, i.e., different values of h. When $h < 1/2$, $\hat{\beta}^{F}$ is as efficient as $\hat{\beta}^{ols}$ [2] and both are more efficient than $\hat{\beta}^{FD}$. On the other hand, $\hat{\beta}^{F}$ is more efficient than $\hat{\beta}^{ols}$ when $h > 1/2$. The FGLS estimate $\hat{\beta}^{F}$ is more efficient than $\hat{\beta}^{FD}$ for all $0 < h < 1$. Hence, if the goal were to maximize power, we would never truncate. So we must rely on size considerations since a well documented fact is that the size distortions of t_β^F increase as α approaches 1 when using the standard normal critical values. To minimize these size distortions, one needs to truncate. So the issue is when to stop truncating. We recommend to do so when the region for which no truncation applies is also the region for which we attain the asymptotic

[2] This is consistent with the result of Phillips and Lee (1996) who considered the case with $\alpha = 1 + c/T$ known. They showed that to establish the equivalence between the GLS and OLS estimate as c increases one must consider value of c such that c^2/T is large.

equivalence between $\hat{\beta}^F$ and $\hat{\beta}^{ols}$. This occurs when $h < 1/2$, which suggests that setting $\delta = 1/2$ so that truncation only applies when $h > 1/2$. The rational for not truncating more is that when $h < 1/2$, $\hat{\beta}^F$ is as efficient as $\hat{\beta}^{ols}$ and we are then in a region that can be labeled stationary and one can expect the standard normal distribution theory to be a good approximation. The obvious price to pay is that the test will be conservative when $1/2 < h < 1$ (in a neighborhood of 1). However, as c increases and we move away from the non-stationary region, we effectively ensure a test with a limiting $N(0,1)$ distribution and estimates that are as efficient as the OLS, thereby bridging the gap between the non-stationary and stationary regions. Our simulations will confirm that this is indeed the best choice to ensure small liberal size distortions when α is close to one while also restricting the region where the test is conservative to a reasonably small neighborhood near $\alpha = 1$.

2.4 Useful modifications for improved finite sample properties

In finite samples, a further problem needs to be addressed. It is well known that the OLS estimate of α is biased downward and that the bias is especially severe when the data are linearly detrended. A solution is to replace the OLS estimate with one that is less biased. We consider the truncated version of the weighted symmetric least-squares (WSLS) estimate as described by Roy and Fuller (2001) and used by Roy et al. (2004) [3]. The WSLS estimate of the autoregressive parameter α is defined by $\hat{\alpha}_W = [\sum_{t=2}^{T-1} \hat{u}_t^2 + T^{-1} \sum_{t=1}^{T} \hat{u}_t^2]^{-1} \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}$. An estimate of its variance is given by $\hat{\sigma}_W^2 = [\sum_{t=2}^{T-1} \hat{u}_t^2 + T^{-1} \sum_{t=1}^{T} \hat{u}_t^2]^{-1} (T-3)^{-1} \sum_{t=2}^{T} (\hat{u}_t - \hat{\alpha}_W \hat{u}_{t-1})^2$ and the associated t-ratio for testing that $\alpha = 1$ is $\hat{\tau}_W = (\hat{\alpha}_W - 1)/\hat{\sigma}_W$. The modification proposed by Roy and Fuller (2001) is similar in spirit to our superefficient estimate in that it also replaces estimates which are in a $T^{-1/2}$ neighborhood of one. The modification is however discontinuous and depends on the value of the t-ratio $\hat{\tau}_W$, making the procedure explicitly dependent on some pre-test. The modified value is given by $\hat{\alpha}_{TW} = \hat{\alpha}_W + C(\hat{\tau}_W)\hat{\sigma}_W$, where

$$C(\hat{\tau}_W) = \begin{cases} -\hat{\tau}_W & \text{if} \quad \hat{\tau}_W > \tau_{pct} \\ T^{-1}\hat{\tau}_W - 3[\hat{\tau}_W + K(\hat{\tau}_W + 5)]^{-1} & \text{if} \quad -5 < \hat{\tau}_W \leq \tau_{pct} \\ T^{-1}\hat{\tau}_W - 3[\hat{\tau}_W]^{-1} & \text{if} \quad -(3T)^{1/2} < \hat{\tau}_W \leq -5 \\ 0 & \text{if} \quad \hat{\tau}_W \leq -(3T)^{1/2} \end{cases}$$

[3] We also considered the median unbiased estimate as described by Andrews (1993), which provided slightly higher power in the AR(1) case. However, the extensions of Andrews and Chen (1994) to higher order autoregressions provided tests with substantial size distortions. Hence, we do not recommend their use.

with $K = [3T - \tau_{pct}^2(I_p + T)][\tau_{pct}(5 + \tau_{pct})(I_p + T)]^{-1}$, where in the case of a general noise component with an $AR(p)$ structure, I_p is the integer part of $(p+1)/2$ and τ_{pct} is a percentile of the limiting distribution of $\hat{\tau}_W$ when $\alpha = 1$. If $\tau_{pct} = -1.96$, the median of the distribution of $\hat{\tau}_W$ when $\alpha = 1$, $\hat{\alpha}_{TW}$ is approximately median unbiased, in the sense that it is nearly unbiased when $\alpha < 1$ and has a median of 1 when $\alpha = 1$. When using this value, we shall denote the corrected estimate by $\hat{\alpha}_{MU}$. As suggested by Roy et al. (2004) an alternative is to use $\tau_{pct} = -2.85$, which is approximately the 85th percentile of the distribution $\hat{\tau}_W$ when $\alpha = 1$. We shall also consider using this value, in which case the estimate is denoted by $\hat{\alpha}_{UB}$. Hence, the recommended procedure is the following: 1) detrend the data by OLS to obtain residuals $\hat{u}_t$; 2) construct the WSLS estimate $\hat{\alpha}_W$; 3) get the truncated estimate $\hat{\alpha}_{MU}$ or $\hat{\alpha}_{UB}$; 4) apply the truncation $\hat{\alpha}_{MS} = \hat{\alpha}_{MU}$ if $|\hat{\alpha}_{MU} - 1| > dT^{-\delta}$ and 1 otherwise (with $\hat{\alpha}_{MU}$ possibly replaced by $\hat{\alpha}_{UB}$); 5) apply a GLS procedure using $\hat{\alpha}_{MS}$ to obtain an estimate of the trend parameter β and construct the standard t-statistic which we shall denote by $t_\beta^{FS}(MU)$ or $t_\beta^{FS}(UB)$.

2.5 Simulation evidence for the AR(1) case

We first conducted simple Monte Carlo experiments to illustrate the size of our suggested test in the leading case of an AR(1) process for the noise component. The data generating process is $y_t = \beta t + u_t$, where $u_t = \alpha u_{t-1} + e_t$ with $e_t \sim i.i.d.\ N(0, 1)$ and $u_0 = 0$. In all cases, $50,000$ replications were used, and the nominal size is 5%. The null and alternative hypotheses are H_0: $\beta = 0$ and H_1: $\beta > 0$.

Consider first the size properties of the t-statistic t_β^{FS} without the median-unbiased adjustment. The null rejection probabilities were simulated for values of α in the range $[0, 1]$ with increments of 0.05 in the range $[0.0, 0.90]$ and in increments of 0.01 in the range $[0.90, 1.0]$. The sample sizes used are $T = 100$, 250, and 500. We considered four cases for the value of δ, namely $\delta = 0.3$, 0.4, 0.5 and 0.6. d is set to 1. The results are summarized in Figure 1. Consider first the cases with $\delta < 1/2$. The results show that the test is conservative when α is close to 1 even when the sample size is large ($T = 500$). This is in accordance with the theoretical explanations given in Section 2.3. As discussed, when $\delta < 1/2$, the result of Theorem 2 applies even for moderate values of c so that, in a local neighborhood of 1, a conservative test results. When $\delta > 1/2$, the opposite applies. The size distortions are liberal even at $\alpha = 1$. This is due to the fact that large values of δ imply that we barely truncate so that the usual size distortions that arise from using the standard OLS estimate apply. The results for the case $\delta = 1/2$ are encouraging since when $T = 500$ only very small size distortions remain. However, with a smaller sample important distortions can be seen local to $\alpha = 1$. This is due to the fact that in small to moderate sample sizes, the OLS estimate of α is biased downward so that no truncation applies when some would be desirable.

Figure 2 presents the size of the t-statistic $t_\beta^{FS}(MU)$ when the OLS estimate of α is replaced by the truncated estimate (the results for and $t_\beta^{FS}(UB)$ are similar and omitted). As expected, when $\delta < 1/2$, the test is conservative in a neighborhood of 1. When $\delta > 1/2$, the liberal size distortions are greatly reduced but remain significant. More importantly, the test constructed with $\delta = 1/2$ now shows basically no size distortion, even with $T = 100$. Note that the size simulations show that the choice of $d = 1$ is appropriate if the goal is to reduce size distortions to a minimum. For example, when $T = 100$, setting $d = 1/2$ would yield results similar to the case $\delta = 0.6$ and setting $d = 2$ would be almost the same as using $\delta = 0.3$. Hence, we shall continue to use $d = 1$. Further simulations showed two-sided 10% tests to have similar properties. With two-sided 5% tests, size distortions are somewhat higher when $T = 100$ and $\alpha = 1$, but decrease rapidly as T increases.

3 Generalization of the model

We now consider an extension of the analysis to the case where the error term u_t is allowed to have a more general structure than the simple $AR(1)$ process with $i.i.d.$ shocks assumed so far. The data generating process is now given by

$$y_t = \mu + \beta t + u_t, \quad u_t = \alpha u_{t-1} + v_t, \tag{10}$$

where $v_t = d(L)e_t$ with $d(L) = \sum_{i=0}^{\infty} d_i L^i$, $\sum_{i=0}^{\infty} i|d_i| < \infty$, $d(1) \neq 0$ and $\{e_t\}$ a martingale difference sequence with respect to a filtration $\mathcal{F}_t$ to which it is adapted. Also, $E[e_t^2|\mathcal{F}_{t-1}] = \sigma^2$ and $\sup_t E[e_t^4] < \infty$. Again, we assume for simplicity that u_0 is some constant. These conditions imply the following functional central limit theorem: $T^{-1/2} \sum_{t=1}^{[rT]} v_t \Rightarrow \sigma d(1) W(r)$. Under the stated conditions, u_t has an autoregressive representation, say $A(L)u_t = e_t$ where $A(L) = 1 - \sum_{i=1}^{\infty} a_i L^i$. In the representation (10), we wish to have the parameter α pertain to the sum of the autoregressive coefficients. Accordingly,

$$u_t = \alpha u_{t-1} + A^*(L)\Delta u_{t-1} + e_t$$

with $A^*(L) = \sum_{i=0}^{\infty} a_i^* L^i$ where $a_i^* = -\sum_{j=i+1}^{\infty} a_j$. Given that α represents the sum of the autoregressive coefficients, we cannot use the estimate (5) based on an autoregression of order one since it is inconsistent for α when the errors u_t are a general $I(0)$ process. Instead, we base our estimate on a truncated autoregression of order k. Let $\hat{u}_t$ be the residuals from a regression of y_t on $\{1, t\}$, then the estimate $\widetilde{\alpha}$ of α is obtained from the OLS regression

$$\hat{u}_t = \alpha \hat{u}_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta \hat{u}_{t-i} + e_{tk}. \tag{11}$$

The estimate $\widetilde{\alpha}$ has the following properties provided $k \to \infty$ and $k^3/T \to 0$ as $T \to \infty$, see Said and Dickey (1984) and Berk (1974). When u_t is $I(0)$, $T^{1/2}(\widetilde{\alpha} - \alpha) = O_p(1)$. On the other hand, if $\alpha = 1 + c/T$, $T(\widetilde{\alpha} - 1) \Rightarrow c + d(1) \int_0^1 J_c^*(r) dW(r) / \int_0^1 J_c^*(r)^2 dr$, where $J_c^*(r)$ is the residual function from the regression of $J_c(r) \equiv \int_0^r \exp(c(r-s)) dW(s)$ on $\{1, r\}$. The truncated estimate $\widetilde{\alpha}_S$ defined by (9), with $\widetilde{\alpha}$ replacing $\hat{\alpha}$, is therefore still superefficient under a local unit root, i.e., $T(\widetilde{\alpha}_S - 1) \to_p 0$ when $\alpha = 1 + c/T$.

As in the $AR(1)$ case, to improve the finite sample performance of the test, we consider a truncated version of the WSLS estimate of α, which for an $AR(p)$ process is described in Fuller (1996, p. 572). The truncated version of this estimate is generically denoted $\widetilde{\alpha}_{MS}$, which can be either $\widetilde{\alpha}_{MU}$ or $\widetilde{\alpha}_{UB}$ depending on the choice of τ_{pct}. The estimate of β considered is a quasi-FGLS estimate assuming $AR(1)$ errors, i.e., the OLS estimate in the transformed regression:

$$y_t - \widetilde{\alpha}_{MS} y_{t-1} = (1 - \widetilde{\alpha}_{MS})\mu + \beta[t - \widetilde{\alpha}_{MS}(t-1)] + (u_t - \widetilde{\alpha}_{MS} u_{t-1}) \tag{12}$$

for $t = 2, ..., T$, together with $y_1 = \mu + \beta + u_1$. Denote the resulting estimate of β by $\widetilde{\beta}$. Since v_t exhibits serial correlation in general, we need to correct for this. Hence, the statistic considered is (where the superscript RQF stands for Robust Quasi Feasible (GLS)) $t_\beta^{RQF} = (\widetilde{\beta} - \beta_0)/\sqrt{\hat{h}_v (X'X)^{-1}_{22}}$, where $(X'X)^{-1}_{22}$ is the (2,2) element of the matrix $(X'X)^{-1}$ with $X = [x_1, ..., x_T]'$, $x_t' = [(1 - \widetilde{\alpha}_{MS}), t - \widetilde{\alpha}_{MS}(t-1)]$ for $t = 2, ..., T$ and $x_1' = (1, 1)$. Also, $\hat{h}_v$ is a consistent estimate of 2π times the spectral density function of v_t at frequency 0 and we use an autoregressive spectral density estimate. Note that the errors from the regression (12) are $(1 - \widetilde{\alpha}_{MS} L) u_t$ and hence, in large samples, they are equivalent to $v_t = (1 - \alpha L) u_t$. Consider first the case where $|\alpha| < 1$. $A(L)$ is then invertible so that $u_t = A(L)^{-1} e_t$ and thus $v_t = (1 - \alpha L) A(L)^{-1} e_t$. Given that α is the sum of the autoregressice coefficients, $A(1) = 1 - \alpha$, and 2π times the spectral density at frequency zero of v_t is simply σ^2. To obtain a consistent estimate, we use the approximate regression

$$y_t - \widetilde{\alpha}_{MS} y_{t-1} = \mu^* + \beta^* t + \sum_{i=1}^{k} \rho_i \Delta y_{t-i} + e_{tk}$$

with $\hat{e}_{tk}$ the corresponding OLS residuals. The estimate of σ^2 is then $\hat{\sigma}^2 = \hat{h}_v = (T - k)^{-1} \sum_{t=k+1}^{T} \hat{e}_{tk}^2$, which is consistent provided $k \to \infty$ and $k^3/T \to 0$ as $T \to \infty$. Consider now the case with $\alpha = 1$. Then, in large samples, $(1 - \widetilde{\alpha}_{MS} L) u_t$ is equivalent to $v_t = \Delta u_t$ and an autoregressive spectral density estimate at frequency zero can be obtained from the regression $\hat{v}_t = \sum_{i=1}^{k} \phi_i \hat{v}_{t-i} + \eta_{tk}$, where $\hat{v}_t$ are the OLS residuals from the regression (12). Denote the estimates by $(\hat{\phi}_1, ..., \hat{\phi}_k)$, $\hat{\sigma}_{\eta k}^2 = (T-k)^{-1} \sum_{t=k+1}^{T} \hat{\eta}_{tk}^2$ and $\hat{h}_v = \hat{\sigma}_{\eta k}^2 / (1 - \sum_{i=1}^{k} \hat{\phi}_i)^2$. The decision rule to select whether to use the estimate for the case $\alpha = 1$ or $|\alpha| < 1$ is based on whether the truncated value $\widetilde{\alpha}_{MS}$ is 1 or not. Asymptotically, this results in a correct classification and a consistent estimate for both cases. Hence, the procedure we recommend is the following: 1) detrend the data by OLS to obtain residuals $\hat{u}_t$; 2) consider the autoregression (11) with k selected using an information criterion (we recommend using the MAIC of Ng and Perron, 2001, with k allowed to be in the range $[0, 12(T/100)^{1/4}]$) and construct the WSLS estimator for an $AR(p)$ process as described in Fuller (1996, p. 572). The corresponding estimate is denoted $\widetilde{\alpha}$; 3) get the truncated estimate $\widetilde{\alpha}_{MU}$ or $\widetilde{\alpha}_{UB}$; 4) apply the truncation $\widetilde{\alpha}_{MS} = \widetilde{\alpha}_{MU}$ if $|\widetilde{\alpha}_{MU} - 1| > T^{-1/2}$ and 1 otherwise (with $\widetilde{\alpha}_{MU}$ possibly replaced by $\widetilde{\alpha}_{UB}$); and 5) apply the quasi FGLS procedure using $\widetilde{\alpha}_{MS}$ to obtain the estimate of the trend parameter β and construct the t-statistic $t_{\beta}^{RQF}(MU)$ or $t_{\beta}^{RQF}(UB)$ using one of the estimates $\hat{h}_v$ suggested for constructing the spectral density function at frequency zero of v_t.

Remark 2 *Consider a sequence of local alternatives of the form $\beta = \beta_0 + dT^{-1/2}$ in the $I(1)$ and local to $I(1)$ cases, and of the form $\beta = \beta_0 + dT^{-3/2}$ in the $I(0)$ case. Given our results, it is easy to show that the local asymptotic power function of our test is the same as that of the tests of Harvey et al. (2007). We refer to their paper for more details. Hence, we shall evaluate the relative performance of the tests in finite samples via simulations.*

4 Finite sample simulations

This section considers the finite sample size and power of our test, including a comparison with those of Bunzel and Vogelsang (2005) and Harvey et al. (2007). The DGP is

$$y_t = \beta t + u_t, \quad u_t = \alpha u_{t-1} + e_t + \phi e_{t-1}, \tag{13}$$

where $e_t \sim i.i.d.\ N(0, 1)$ and $u_0 = u_{-1} = 0$. Our hypotheses are H_0: $\beta = 0$ and H_1: $\beta > 0$. Throughout, 5000 replications are used.

4.1 Description of alternative procedures

There are currently two procedures that offer tests with the same critical values under both the $I(0)$ and $I(1)$ cases: that of Bunzel and Vogelsang (2005), which builds on Vogelsang

(1998), and that of Harvey et al. (2007) [4]. Bunzel and Vogelsang's (2005) preferred test is labelled *Dan-J*. Their statistic in the linear trend case is the robust t-ratio weighted by a factor such that the limit distribution under the null hypothesis is the same in the $I(0)$ and $I(1)$ cases. More specifically, $Dan\text{-}J = \exp(-cUR)T^{3/2}(\hat{\beta}^{ols} - \beta)/[\hat{\sigma}^2(T^{-3}\sum_{t=1}^{T}(t-\bar{t})^2)^{-1}]^{1/2}$, where $\bar{t} = T^{-1}\sum_{t=1}^{T} t$, $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2 + T^{-1}\sum_{j=1}^{T-1}\lambda(j,m)\sum_{t=j+1}^{T}\hat{u}_t\hat{u}_{t-j}$ with $\lambda(j,m)$ the Daniel kernel function and m is chosen using a data dependent method (see Bunzel and Vogelsang, 2008, for details). The important ingredient is the scaling factor $\exp(-cUR)$, where UR is a unit root test statistic which converges to zero under the alternative of stationarity and to a well defined limit distribution under the null hypothesis. They recommend using the test proposed by Park and Choi (1988). In large samples, *Dan-J* is approximately the robust t-test in the $I(0)$ case, while in the $I(1)$ case, it is scaled by $\exp(-cUR_\infty)$, where UR_∞ is the limit distribution of the unit root test and c is chosen such that the relevant quantiles of the limit distribution of *Dan-J* are the same under both the $I(0)$ and $I(1)$ cases. An important point to note is that this scaling makes the *Dan-J* test a noisy version of a good test in the $I(1)$ case, thereby reducing its ability to discriminate between the null and the alternative hypotheses and consequently, power is low.

The approach of Harvey et al. (2007) is similar in spirit in that a scaling is also used but one which delivers asymptotically the best trend test, that based on the first-differenced data in the $I(1)$ case and that based on the data in levels in the $I(0)$ case. The test they recommend is $z_\lambda = (1-\lambda(U,S))z_0 + \lambda(U,S)z_1$, where z_0 and z_1 are the standard robust t-statistic applied to the coefficient of the slope in the model in level and to the constant in the model in differences, respectively. The scaling $\lambda(U,S)$ obeys $\lambda(U,S) = o_p(1)$ when u_t is $I(0)$ and $\lambda(U,S) = 1 + o_p(T^{-1/2})$ when u_t is $I(1)$. This is achieved by having U be a unit root test and S be stationarity test with the functional form $\lambda(U,S) = \exp(-0.00025(U/S)^2)$. For U, they recommend Elliot et al.'s (1996) DF-GLS test with the lag length selected using the MAIC of Ng and Perron (2001). For S, they recommend the so-called KPSS test (see Kwiatowski et al., 1992). There are two important differences with our procedure. First, while z_λ is asymptotically equivalent to z_0 and z_1 in the $I(0)$ and $I(1)$ cases, respectively, in any finite samples, it remains a noisy version of those best tests, while ours is not. Second, while the OLS and GLS tests are asymptotically equivalent in the $I(0)$ case, one can expect

[4] In a previous version of this paper, we also considered the method of Roy, Balk and Fuller (2004). We did so using the code they constructed. After further investigations, it turned out that the code contained a serious mistake which when corrected showed tests with very different sizes in the I(0) and I(1) cases, unlike what they claimed, and consistent with the fact that the limit distributions are not the same in both cases. Hence, the procedure is not applicable.

an increase in efficiency from using the GLS approach in finite samples. These two features can lead to relatively more size distortions and/or power losses.

4.2 Size

We consider the exact sizes at the nominal 5% level obtained for a range of values for α and ϕ and for $T = 100$, 250 and 500. The results are presented in Table 1. Consider first the leading case of an AR(1) process with $\phi = 0$. In the unit root case, the *Dan-J* test indeed achieves the best size even at $T = 100$. The z_λ test has the largest size distortions at about 12% when $T = 100$. What is more problematic is that these liberal size distortions remain even when $T = 500$, with an exact size of 8%. The $t_\beta^{RQF}(MU)$ has some liberal size distortions when $T = 100$ which, however, decrease rapidly as T increases. Those of $t_\beta^{RQF}(UB)$ are somewhat smaller and there are barely any size distortions when T reaches 250. Consider now the stationary case. When α is close to one, e.g., 0.95 or 0.90, all tests are conservative, as expected. However, the exact size approach the nominal size as T increases except for the test z_λ, which remains very conservative. When α is further away from 1, the exact sizes of all tests are close to the 5% nominal level and slightly on the conservative side.

Consider now the case of an ARMA process. We shall focus on the value $\phi = -0.8$, which is the usual problematic case. Consider first the unit root case $\alpha = 1$. Here *Dan-J* exhibits severe liberal size distortions, the size remaining as high as 24% when $T = 500$. The size of the z_λ is also liberal when $T = 100$ though it gets close to 5% when $T = 500$. The size of the test $t_\beta^{RQF}(UB)$ is similar to that of z_λ while that of $t_\beta^{RQF}(MU)$ is somewhat higher, especially when $T = 100$. In the $I(0)$ cases the following features are present. The exact sizes of the tests $t_\beta^{RQF}(UB)$ and $t_\beta^{RQF}(MU)$ are close to 5% while the size of z_λ is well below 5% and more so as α approaches 0. Finally, the exact size of *Dan-J* is substantially above 5% when $\alpha = 0.95$ or 0.90 and becomes very conservative as α gets further away from one so that when $\alpha = 0$, the exact size is 0. In summary the test with the best size properties overall is the $t_\beta^{RQF}(UB)$ while *Dan-J* has the worst. That of z_λ is best with a very negative moving-average coefficient, though it is substantially liberal in the case of a pure AR(1) process with a unit root and conservative in the $I(0)$ case.

4.3 Power

Consider now the power of the tests. We first start with the leading case of an AR(1) process to better highlight the main differences between the tests and their relative merits and drawbacks. The power curves are plotted for $\alpha = 1.0, 0.95, 0.90$ and 0.80 for a range of values of $\beta > 0$. Figures 3.1-3.3 present the results for $T = 100, 250$ and 500, respectively. Consider first the case $\alpha = 1$. The tests $t_{\beta}^{RQF}(MU)$ and $t_{\beta}^{RQF}(UB)$ have basically the same power as the infeasible GLS test (slightly higher when $T = 100$ because of small liberal size distortions). The test z_{λ} shows higher power but this is simply due to its substantial size distortions. The *Dan-J* test has very low power, consistent with the fact that in the $I(1)$ case the scaling factor implies a test quite different from the optimal one. When T increases to $T = 250$ and 500, the tests $t_{\beta}^{RQF}(MU)$ and $t_{\beta}^{RQF}(UB)$ now have exactly the same power as the infeasible GLS test. Since z_{λ} continues to show liberal size distortions, it shows power above the envelope. What is noteworthy is that the power of the test *Dan-J* remains well below that of the others.

Consider now the power at some value of α away from 1, say 0.80. When $T = 100$, $t_{\beta}^{RQF}(MU)$ has the best power while the z_{λ} test has the lowest for alternatives close to the null value and $t_{\beta}^{RQF}(UB)$ has the lowest power for distant alternatives. When T increases to 250 or 500, the $t_{\beta}^{RQF}(MU)$, $t_{\beta}^{RQF}(UB)$ and *Dan-J* have nearly the same power as the infeasible GLS test while that of z_{λ} is substantially lower. Unreported simulations show that these large differences in power remain when α gets further away from 1. This can potentially be explained by the fact that the z_{λ} test is a contaminated version of the best test, given the scaling, and also by the fact that our tests are based on a GLS procedure while that of z_{λ} uses an OLS based approach.

When α is in a neighborhood of one, e.g., 0.95 or 0.90 for $T = 100$ or 0.95 for $T = 250$ or 500, the results are more complex and no clear pattern emerges, except the fact that $t_{\beta}^{RQF}(MU)$ always has higher power than $t_{\beta}^{RQF}(UB)$. When $T = 100$, *Dan-J* has the lowest power though when $T = 250$, it is superior for alternatives close to the null value and inferior for distant alternatives. When $T = 500$, it is nearly superior overall. Comparing the power of $t_{\beta}^{RQF}(MU)$ and z_{λ}, we have that for alternatives close to the null, $t_{\beta}^{RQF}(MU)$ is more powerful and vice-versa for distant alternatives.

Remark 3 *From Figures 3.1-3.3, the power functions of our test have some unussual kinks. This is due to the fact that our procedure involves a mix of the standard FGLS (using the non-truncated estimate of α) and the first difference estimators. From unreported simulations for the case $\alpha = 0.9$ and $T = 100$, the power function of the test based on first differences is almost zero before $\beta = 0.12$ but increases rapidly thereafter. The power function of the test based on the FGLS procedure with $\hat{\alpha}$ non-truncated increases rapidly for $\beta \leq 0.06$, but afterwards the increase is small. Hence, the overall procedure has a power function with two kinks at* 0.06 *and* 0.12*. Such kinks when α is in a neighborhood of one are to be expected. This* *is the price one has to pay to have a procedure with good size and power when $\alpha = 1$ and when α is well within the stationary region. It may appear that there is room for improvements, perhaps via some smoothing, but as argued by Roy et al. (2004, p. 1088), this seems unlikely if we require a test that controls size in both the $I(1)$ and $I(0)$ cases and maintains good power (see, e.g., the modification proposed in Harvey et al., 2007).*

When $T = 100$, some of the tests have size distortions so that the rejection frequencies may not provide a good description of the relative discriminatory power of the tests. To that effect, we present in Figure 4 the size-adjusted power functions. When $\alpha = 1$, $t_{\beta}^{RQF}(UB)$ has the best power, followed by z_{λ} and $t_{\beta}^{RQF}(MU)$, with *Dan-J* a distant last. When $\alpha = 0.95$, the power of $t_{\beta}^{RQF}(UB)$ and z_{λ} are nearly equal and best while when $\alpha = 0.90$, z_{λ} is superior overall. But when $\alpha = 0.80$, z_{λ} loses power and $t_{\beta}^{RQF}(MU)$ has the best power overall.

We now turn to the power of the tests when the noise component is more general and we consider an AR(2) model as well as an ARMA(1,1). For the AR(2) model $u_t = \alpha u_{t-1} + \varphi(u_{t-1} - u_{t-2}) + e_t$, where again $e_t \sim i.i.d.\ N(0,1)$ and $u_0 = u_{-1} = 0$. The results are presented in Figure 5 for $\varphi = 0.5$, $T = 250$ and $\alpha = 1.0$, 0.95, 0.90 and 0.80. We see again that z_{λ} has substantial size distortions in the $I(1)$ case so that its power exceeds the power envelope. On the other hand, both $t_{\beta}^{RQF}(MU)$ and $t_{\beta}^{RQF}(UB)$ lack size distortions and attain power near the envelope. Again, the *Dan-J* test shows very low power. When $\alpha = 0.80$, *Dan-J* and $t_{\beta}^{RQF}(MU)$ have the best power with $t_{\beta}^{RQF}(UB)$ a close second while z_{λ} is substantialy less powerful. When $\alpha = 0.95$ or 0.90, the picture is again less clear-cut. With $\alpha = 0.95$, $t_{\beta}^{RQF}(MU)$ and $t_{\beta}^{RQF}(UB)$ have substantialy less power than *Dan-J* and z_{λ}. However, with $\alpha = 0.90$, $t_{\beta}^{RQF}(MU)$ and $t_{\beta}^{RQF}(UB)$ have the best power for alternative values of β close to the null but not for distant alternatives. As discussed above, this drop in power for values of α in a local neighborhood of 1 is the price one has to pay to have good size and power in the $I(1)$ and non-local $I(0)$ cases.

To show that our method retains its good properties for the case of a model with a long autoregressive structure, we consider now u_t generated by the ARMA(1,1) process as described in (13). We experimented with different values of ϕ and only report those with $\phi = -0.4$ (the general conclusions being the same with other values). The results are presented in Figure 6 again for $T = 250$ and $\alpha = 1.0$, 0.95, 0.90 and 0.80. When $\alpha = 1.0$ or 0.80, the pattern is similar to that for the AR(1) or AR(2) cases. Things are somewhat different when $\alpha = 0.95$ or 0.90. With $\alpha = 0.90$, now $t_{\beta}^{RQF}(MU)$ and *Dan-J* have the highest power for all values of β while z_{λ} has the lowest for small values of β and $t_{\beta}^{RQF}(UB)$ has the lowest for large values of β. With $\alpha = 0.95$, $t_{\beta}^{RQF}(MU)$ has the best power for small values of β, *Dan-J* the highest for intermediate values and z_λ the highest for large values.

Overall, we view the results of the simulations as providing evidence that the tests we propose have good size and power properties and compare favorably with those proposed by Bunzel and Vogelsang (2005) and Harvey et al. (2007). The *Dan-J* test has good size overall but very low power in the $I(1)$ case. The z_λ test of Harvey et al. (2007) has size distortions in the $I(1)$ case, which decrease very slowly as the sample size increases, and much lower power in the non-local to $I(1)$ stationary cases. Our tests exhibit some liberal size distortions in the $I(1)$ case when $T = 100$ but have the correct size for larger sample sizes. Also, they have power close to the power envelope in both the $I(1)$ and non-local to $I(1)$ stationary cases. Their power can be lower in the local to $I(1)$ case (though not uniformly) but, as argued, this seems inevitable in order to have tests with good size and power properties in the $I(1)$ and non-local to $I(1)$ stationary cases. With respect to the choice between the $t_\beta^{RQF}(MU)$ and $t_\beta^{RQF}(UB)$ versions, the former is in general preferable (higher power and good size) unless one is worried about facing a process with a strong negative moving-average component, in which case the $t_\beta^{RQF}(UB)$ has an exact size closer to the nominal level.

5 Conclusions

We proposed a new procedure to carry out inference about the slope of a trend function that is valid without prior knowledge about whether a series is $I(0)$ or $I(1)$. The test are based on a Feasible quasi GLS method with a superefficient estimate of the sum of the autoregressive coefficients α when $\alpha = 1$. Simulations have shown their usefulness and the fact that they provide improvements over existing methods. Of course, there is no test that is uniformly most powerful but we have shown that overall our tests have better size and power properties. The power of our tests are basically equivalent to that based on the infeasible GLS estimate when α is assumed known, unless α is near but not equal to one. It may appear that there is room for improvements in this case but, as argued by Roy et al. (2004, p. 1088), this seems unlikely if we require a test that controls size in both the $I(1)$ and $I(0)$ cases.

As mentioned in Section 2, our method extends with obvious modifications to the case of a general polynomial trend function that can include multiple structural changes occurring at known dates. An extension of practical importance is to consider the case where the trend function can have structural changes at unknown dates. In this case, the analysis does not extend in a straightforward fashion since the limit distribution of the t-statistic on the slope coefficient still depends on the $I(0)$-$I(1)$ dichotomy. However, some modifications can be applied to obtain useful procedures for that case as well, see Perron and Yabu (2007).

Appendix: Technical Derivations

Proof of equation (7): Under the hypothesis that the data are generated by (2) with $\alpha = 1$, using straightforward algebra we can express the t-statistic as (see also, Canjels and Watson, 1997):

$$t_\beta^F = \frac{q_{11}T^{-1/2}r_2 - T^{-1/2}q_{12}r_1}{[\hat{\sigma}^2 q_{11}(q_{11}T^{-1}q_{22} - T^{-1}q_{12}^2)]^{1/2}}, \tag{A.1}$$

where $q_{11} = 1 + (T-1)(1-\hat{\alpha})^2 = 1 + o_p(1)$,

$$T^{-1/2}q_{12} = T^{-1/2}[1 + (T-1)\hat{\alpha}(1-\hat{\alpha}) + (1-\hat{\alpha})^2 \sum_{t=2}^{T} t] = o_p(1),$$

$$\begin{aligned} T^{-1}q_{22} &= T^{-1}[1 + (T-1)\hat{\alpha}^2 + 2\hat{\alpha}(1-\hat{\alpha}) \sum_{t=2}^{T} t + (1-\hat{\alpha})^2 \sum_{t=2}^{T} t^2] \\ &= 1 + T(1-\hat{\alpha}) + T^2(1-\hat{\alpha})^2/3 + o_p(1), \end{aligned}$$

$$\begin{aligned} T^{-1/2}r_2 &= T^{-1/2}[u_1 + \hat{\alpha} \sum_{t=2}^{T} u_t^* + (1-\hat{\alpha}) \sum_{t=2}^{T} t u_t^*] \\ &= T^{-1/2} \sum_{t=2}^{T} u_t^* + T(1-\hat{\alpha})T^{-3/2} \sum_{t=2}^{T} t u_t^* + o_p(1), \end{aligned}$$

$r_1 = u_1 + (1-\hat{\alpha})\sum_{t=2}^{T} u_t^* = u_1 + o_p(1)$, $u_t^* = u_t - \hat{\alpha} u_{t-1}$ and $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T} \hat{e}_t^2$ with $\hat{e}_t$ being the residuals from a regression of $y_t^* = y_t - \hat{\alpha} y_{t-1}$ on $x_t^* = [1-\hat{\alpha},\, t - \hat{\alpha}(t-1)]'$ for $t = 2, ..., T$ and $x_1^* = [1, 1]'$, $y_1^* = y_1$. It is straightforward to show that $\hat{\sigma}^2 = \sigma^2 + o_p(1)$. Equation (7) follows by substituting these expansions into (A.1).

Proof of Theorem 1: Let $A = \{T^\delta |\hat{\alpha} - 1| > d\}$ and let $\bar{A} = \{T^\delta |\hat{\alpha} - 1| \leq d\}$. For part (a), it suffices to show that $T^{1/2}(\hat{\alpha}_S - \alpha) - T^{1/2}(\hat{\alpha} - \alpha) \to_p 0$. We have

$$\begin{aligned} \lim_{T\to\infty} \Pr(|T^{1/2}(\hat{\alpha}_S - \hat{\alpha})| &> \varepsilon) = \\ \lim_{T\to\infty} \Pr(|T^{1/2}(\hat{\alpha}_S - \hat{\alpha})| &> \varepsilon | A) \Pr(A) + \lim_{T\to\infty} \Pr(|T^{1/2}(\hat{\alpha}_S - \hat{\alpha})| > \varepsilon | \bar{A}) \Pr(\bar{A}) \end{aligned}$$

The first term is zero given that, if A is true, we have $\hat{\alpha}_S = \hat{\alpha}$ so that $\Pr(|T^{1/2}(\hat{\alpha}_S - \hat{\alpha})| > \varepsilon | A) = 0$. The second term is zero since $\hat{\alpha} \to_p \alpha \neq 1$ implies $\Pr(\bar{A}) \to 0$ as $T \to \infty$. Therefore, $\Pr(|T^{1/2}(\hat{\alpha}_S - \hat{\alpha})| > \varepsilon) \to 0$ as $T \to \infty$. For part (b), we have

$$\begin{aligned} \lim_{T\to\infty} \Pr(|T(\hat{\alpha}_S - 1)| &> \varepsilon) = \\ \lim_{T\to\infty} \Pr(|T(\hat{\alpha}_S - 1)| &> \varepsilon | A) \Pr(A) + \lim_{T\to\infty} \Pr(|T(\hat{\alpha}_S - 1)| > \varepsilon | \bar{A}) \Pr(\bar{A}). \end{aligned}$$

Now the fact that $T(\hat{\alpha}-1)=O_p(1)$ and $\delta \in (0,1)$ imply that $\Pr(A)=\Pr(T^{\delta-1}|T(\hat{\alpha}-1)|>d)\to 0$ as $T\to\infty$, so that the first term is zero. For the second term, if $\bar{A}$ is true, $\hat{\alpha}_S=1$ so that $\Pr(|T(\hat{\alpha}_S-1)|>\varepsilon|\bar{A})=0$. Thus, $\Pr(|T(\hat{\alpha}_S-1)|>\varepsilon)\to 0$ as $T\to\infty$.

Proof of Theorem 2: To prove part a), first note that with $\alpha_T=1+c/T$, $T(\hat{\alpha}-1)\Rightarrow c+\int_0^1 J_c^*(r)dW(r)/\int_0^1 J_c^*(r)^2dr$ with $J_c^*(r)$, $0\le r\le 1$ being the residual process from a continuous time regression of an Ornstein-Uhlenbeck $J_c(r)$ on $\{1,r\}$, where $J_c(r)=\int_0^r \exp(c(r-s))dW(s)$. Since the true value of α is in a T^{-1} neighborhood of 1 and $\hat{\alpha}_S$ truncates values of $\hat{\alpha}$ in a $T^{-\delta}$ neighborhood of 1 for some $0<\delta<1$ (i.e., a larger neighborhood), the results of Theorem 1 continue to apply in the local to unity case and $T(\hat{\alpha}_S-1)\to_p 0$. Now,

$$\begin{aligned} t_\beta^F \;=\; & [T^{-1/2}\sum_{t=2}^{T}(u_t-u_{t-1})-T(\hat{\alpha}_S-1)T^{-3/2}\sum_{t=2}^{T}u_{t-1} \\ & -T(\hat{\alpha}_S-1)[T^{-3/2}\sum_{t=2}^{T}t(u_t-u_{t-1})-T(\hat{\alpha}_S-1)T^{-5/2}\sum_{t=2}^{T}tu_{t-1}]] \\ & /\left[\sigma^2(1+T(\hat{\alpha}_S-1)+T^2(\hat{\alpha}_S-1)^2/3)\right]^{1/2}+o_p(1). \end{aligned}$$

We have $T^{-1/2}\sum_{t=2}^{T}(u_t-u_{t-1})=T^{-1/2}\sum_{t=2}^{T}e_t+cT^{-3/2}\sum_{t=2}^{T}u_{t-1}\Rightarrow \sigma[W(1)+c\int_0^1 J_c(r)dr]=\sigma J_c(1)\sim N(0,\sigma^2(\exp(2c)-1)/2c)$. Part (b) follows using $T(\hat{\alpha}_S-1)\to_p 0$, and that the following are $O_p(1)$: $T^{-3/2}\sum_{t=2}^{T}u_{t-1}$, $T^{-3/2}\sum_{t=2}^{T}t(u_t-u_{t-1})$, and $T^{-5/2}\sum_{t=2}^{T}tu_{t-1}$.

Proof of Theorem 3: We consider u_t as known, which allows use of results in Giraitis and Phillips (2006). The results remain valid when $\hat{\alpha}$ is constructed with the OLS residuals $\hat{u}_t$, though the derivations are more involved. We start with the following Lemma.

Lemma A.1 *Suppose that $u_t=\alpha_T u_{t-1}+e_t$ with e_t as defined by (2), and $\alpha_T=1+b/T^h$ for some $b<0$ and some $0<h<1$. Then, a) $T^{-h/2}u_{[Tr]}\Rightarrow(\sigma^2/(-2b))^{1/2}W(1)$; b) $(-b)T^{-h-1/2}\sum_{t=1}^{T}u_t\Rightarrow\sigma W(1)$; c) $(-b)T^{-h-3/2}\sum_{t=1}^{T}tu_t\Rightarrow\sigma\int_0^1 rdW(r)$; d) $(-2b)^{-1/2}T^{1/2+h/2}(\hat{\alpha}-\alpha_T)\Rightarrow N(0,1)$; e) $T^h(\hat{\alpha}-1)=b+o_p(1)$.*

Proof: For part (a), $u_{[Tr]}$ has a mean that is $o(1)$ and

$$\begin{aligned} var(T^{-\frac{h}{2}}u_{[Tr]}) &= \sigma^2T^{-h}(1+\alpha_T^2+...+\alpha_T^{2([Tr]-1)})=\frac{\sigma^2(1-\alpha_T^{2[Tr]})}{T^h(1-\alpha_T^2)} \\ &= \frac{\sigma^2(1-(1+b/T^h)^{2[Tr]})}{T^h(1-(1+b/T^h)^2)}\to\frac{\sigma^2}{-2b}. \end{aligned}$$

The result follows given the errors are i.i.d. and the weighted sums thereby converge to a normal distribution. Part (b) is from Theorem 2 of Giraitis and Phillips (2006). Their proof can be modified to prove part (c). Part (d) is from Theorem 1 of Giraitis and Phillips (2006). To prove part (e), note that

$$T^h(\hat{\alpha}-1)=T^h(\alpha_T-1)+T^h(\hat{\alpha}-\alpha_T)=b+T^{h/2-1/2}T^{1/2+h/2}(\hat{\alpha}-\alpha_T)=b+o_p(1)$$

in view of part (d) and the fact that $0 < h < 1$.■

Consider first, the limit of the OLS estimate of β. We have, with $\bar{t} = T^{-1}\sum_{t=1}^{T} t$,

$$\begin{aligned}
T^{3/2-h}(\hat{\beta}^{ols} - \beta) &= \frac{T^{-3/2-h}\sum_{t=1}^{T}(t-\bar{t})u_t}{T^{-3}\sum_{t=1}^{T}(t-\bar{t})^2} = \frac{T^{-3/2-h}\sum_{t=1}^{T} tu_t - (T^{-2}\sum_{t=1}^{T} t)(T^{-1/2-h}\sum_{t=1}^{T} u_t)}{T^{-3}\sum_{t=1}^{T} t^2 - (T^{-2}\sum_{t=1}^{T} t)^2} \\
&\Rightarrow (-12\sigma/b)W(1) - \int_0^1 rdW(r) = N(0, \sigma^2 12/b^2)
\end{aligned}$$

since $\int_0^1 rdW(r) - (1/2)W(1)$ is $N(0, 1/12)$. Consider now the Feasible GLS estimate of β, defined by $\hat{\beta}^F - \beta = (q_{11}r_2 - q_{12}r_1)/(q_{11}q_{22} - q_{12}^2)$. The following Lemma states the limit of each term when $\alpha_T = 1 + b/T^h$ for $h < 1/2$ and $h > 1/2$ (the case with $h = 1/2$ is omitted).

Lemma A.2 *Suppose that $u_t = \alpha_T u_{t-1} + e_t$ with e_t as defined by (2) and $\alpha_T = 1 + b/T^h$ for some $b < 0$ and some $0 < h < 1$. Then, a) $T^{-1+2h}q_{11} \to b^2$ if $h < 1/2$ and $q_{11} \to 1$ if $h > 1/2$; b) $T^{-2+2h}q_{12} \to b^2/2$ for $h \lesseqgtr 1/2$; c) $T^{-3+2h}q_{22} \to b^2/3$ for $h \lesseqgtr 1/2$; d) for $h < 1/2$, $T^{-1/2+h}r_1 \Rightarrow -b\sigma W(1)$ and for $h > 1/2$, $r_1 \Rightarrow u_1$; e) for $0 < h < 1$, $T^{-3/2+h}r_2 \Rightarrow -b\sigma \int_0^1 rdW(r)$.*

The proof follows from simple calculations using the results of Lemma A.1. Using Lemma A.2, we obtain for $0 < h < 1/2$,

$$\begin{aligned}
T^{3/2-h}(\hat{\beta}^F - \beta) &= \frac{T^{-1+2h}q_{11}T^{-3/2+h}r_2 - T^{-2+2h}q_{12}T^{-1/2+h}r_1}{T^{-1+2h}q_{11}T^{-3+2h}q_{22} - T^{-4+4h}q_{12}^2} \Rightarrow \frac{(b^3/2)\sigma W(1) - b^3\sigma\int_0^1 rdW(r)}{b^4/3 - b^4/4} \\
&= (12\sigma/b)[(1/2)W(1) - \int_0^1 rdW(r)] = N(0, 12\sigma^2/b^2),
\end{aligned}$$

since $\int_0^1 rdW(r) \sim N(0, 1/3)$, which is the same limit as that of the OLS estimate. For the t-statistic, we have

$$\begin{aligned}
t_\beta^F &= \frac{T^{-1+2h}q_{11}T^{-3/2+h}r_2 - T^{-2+2h}q_{12}T^{-1/2+h}r_1}{\left[\hat{\sigma}^2 T^{-1+2h}q_{11}\left(T^{-1+2h}q_{11}T^{-3+2h}q_{22} - T^{-4+4h}q_{12}^2\right)\right]^{1/2}} \\
&\Rightarrow [b^2(b^4/3 - b^4/4)]^{-1/2}[(b^3/2)W(1) - b^3\int_0^1 rdW(r)] = N(0,1)
\end{aligned}$$

in view of the fact that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$. For $1/2 < h < 1$,

$$\begin{aligned}
T^{3/2-h}(\hat{\beta}^F - \beta) &= \frac{q_{11}T^{-3/2+h}r_2 - T^{1/2-h}T^{-2+2h}q_{12}r_1}{q_{11}T^{-3+2h}q_{22} - T^{1-2h}T^{-4+4h}q_{12}^2} = \frac{T^{-3/2+h}r_2}{T^{-3+2h}q_{22}} + o_p(1) \\
&\Rightarrow (b^2/3)^{-1}(-b\sigma\int_0^1 rdW(r)) = N(0, 3\sigma^2/b^2).
\end{aligned}$$

For the t-statistic, we have,

$$t_\beta^F = \frac{T^{-3/2+h}r_2}{(\hat{\sigma}^2 T^{-3+2h}q_{22})^{1/2}} \Rightarrow \frac{(-b)\int_0^1 rdW(r)}{(b^2/3)^{1/2}} = N(0,1).$$

Finally, using part (a) of Lemma A.1 and applying a CLT, $\hat{\beta}^{FD} = (T-1)^{-1}\sum_{t=2}^{T}\Delta y_t$ is such that $T^{1-h/2}(\hat{\beta}^{FD} - \beta) \Rightarrow N(0, \sigma^2/(-2b))$ and $t_\beta^{FD} \Rightarrow 0$.

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817-858.

Andrews, D.W.K., 1993. Exactly median-unbiased estimation of first order autoregressive/unit root models. Econometrica 61, 139-165.

Andrews, D.W.K., Chen, H.-Y., 1994. Approximately median-unbiased estimation of autoregressive models. Journal of Business and Economic Statistics 12, 187-204.

Berk, K.N., 1974. Consistent autoregressive spectral estimates. Annals of Statistics 2, 489-502.

Bunzel, H., Vogelsang, T.J., 2005. Powerful trend function tests that are robust to strong serial correlation with an application to the Prebish-Singer hypothesis. Journal of Business and Economic Statistics 23, 381-394.

Canjels, E., Watson, M.W., 1997. Estimating deterministic trends in the presence of serially correlated errors. Review of Economics and Statistics 79, 184-200.

DeJong, D.N., Nankervis, J.C., Savin, N.E., Whiteman, C.H., 1992. Integration versus trend stationary in time series. Econometrica 60, 423-433.

Elliott, G., Rothenberg, T.J., Stock, J.H., 1996. Efficient tests for an autoregressive unit root. Econometrica 64, 813-836.

Fuller, W.A., 1996. Introduction to Statistical Time Series (2nd ed.), Wiley (New York, NY).

Giraitis, L., Phillips, P.C.B., 2006. Uniform limit theory for stationary autoregression. Journal of Time Series Analysis 27, 51-60.

Grenander, U., Rosenblatt, M., 1957. Statistical Analysis of Stationary Time Series, New York: John Wiley.

Harvey, D.I., Leybourne, S.J., Taylor, A.M.R., 2007. A simple, robust and powerful test of the trend hypothesis. Journal of Econometrics 141, 1302-1330.

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

Mikusheva, A., 2007. Uniform inference in autoregressive models. Econometrica 75, 1411-1452.

Nelson, C.R., Plosser, C. 1982. Trends and random walks in macroeconomic time series: some evidence and implications. Journal of Monetary Economics 10, 139-162.

Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519-1554.

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

Perron, P., 1988. Trends and random walks in macroeconomic time series: further evidence from a new approach. Journal of Economic Dynamics and Control 12, 297-332.

Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. Econometrica 57, 1361-1401.

Perron, P., Yabu, T., 2007. Testing for a shift in trend with an integrated or stationary noise component. Forthcoming in Journal of Business and Economic Statistics.

Phillips, P.C.B., Durlauf, S., 1988. Trends versus random walks in time series analysis. Econometrica 56, 1333-1354.

Phillips, P.C.B., Lee, C.C., 1996. Efficiency gains from quasi-differencing under nonstationarity. In P.M. Robinson and M. Rosenblatt (eds), Athens Conference on Applied Probability and Time Series, Volume II: Time Series Analysis in Memory of E.J. Hannan, Lecture Notes in Statistics 115, Springer-Verlag (New York, NY), 300-313.

Prais, S.J., Winsten, C.B., 1954. Trend estimators and serial correlation. Cowles Foundation Discussion Paper 383.

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

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

Said, S.E., Dickey, D.A., 1984. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71, 599-608.

Sun, H., Pantula, S.G., 1999. Testing for trends in correlated data. Statistics and Probability Letters 41, 87-95.

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

Vogelsang, T.J., Fomby, T.B., 2002. The application of size robust trend analysis to global warming temperature series. Journal of Climate 15, 117-123.

Vogelsang, T.J., Franses, P.H., 2005. Testing for common deterministic trend slopes. Journal of Econometrics 126, 1-24.

Table 1:

Finite Sample Null Rejection Probabilities with 5% Nominal Size;

Model: $y_t = \mu + \beta t + u_t$, $u_t = \alpha u_{t-1} + e_t + \phi e_{t-1}$; H_0: $\beta = 0$

		T=100				T=250				T=500			
				QFGLS				QFGLS				QFGLS	
α	ϕ	Dan-J	Z_λ	MU	UB	Dan-J	Z_λ	MU	UB	Dan-J	Z_λ	MU	UB
1	-0.8	0.303	0.104	0.164	0.110	0.294	0.081	0.148	0.090	0.237	0.065	0.140	0.073
	-0.4	0.102	0.118	0.129	0.084	0.079	0.091	0.094	0.072	0.062	0.076	0.074	0.057
	0.0	0.050	0.118	0.093	0.076	0.050	0.090	0.073	0.060	0.053	0.080	0.056	0.062
	0.4	0.043	0.119	0.084	0.077	0.044	0.092	0.070	0.069	0.051	0.080	0.048	0.058
	0.8	0.041	0.116	0.093	0.091	0.042	0.091	0.074	0.067	0.052	0.081	0.061	0.066
0.95	-0.8	0.195	0.027	0.063	0.043	0.195	0.017	0.053	0.032	0.152	0.014	0.055	0.040
	-0.4	0.059	0.018	0.050	0.022	0.068	0.010	0.037	0.027	0.069	0.012	0.047	0.040
	0.0	0.022	0.017	0.037	0.017	0.037	0.011	0.036	0.020	0.055	0.012	0.043	0.032
	0.4	0.017	0.014	0.023	0.009	0.027	0.009	0.014	0.008	0.040	0.011	0.016	0.016
	0.8	0.017	0.012	0.007	0.006	0.026	0.007	0.005	0.001	0.033	0.009	0.005	0.005
0.9	-0.8	0.130	0.029	0.051	0.038	0.133	0.027	0.039	0.036	0.099	0.027	0.042	0.043
	-0.4	0.066	0.016	0.050	0.024	0.084	0.021	0.043	0.034	0.075	0.028	0.045	0.038
	0.0	0.028	0.017	0.044	0.019	0.051	0.021	0.045	0.033	0.061	0.032	0.044	0.042
	0.4	0.021	0.012	0.032	0.017	0.037	0.017	0.037	0.025	0.057	0.027	0.047	0.040
	0.8	0.021	0.010	0.012	0.005	0.036	0.013	0.022	0.013	0.055	0.020	0.034	0.032
0.8	-0.8	0.050	0.036	0.036	0.047	0.055	0.035	0.043	0.038	0.052	0.031	0.043	0.043
	-0.4	0.072	0.023	0.047	0.031	0.080	0.031	0.034	0.032	0.059	0.036	0.033	0.036
	0.0	0.042	0.024	0.045	0.027	0.061	0.034	0.040	0.037	0.052	0.041	0.040	0.040
	0.4	0.027	0.015	0.041	0.029	0.054	0.028	0.044	0.041	0.049	0.037	0.038	0.040
	0.8	0.025	0.014	0.025	0.013	0.050	0.022	0.042	0.032	0.048	0.032	0.043	0.038
0.7	-0.8	0.015	0.039	0.033	0.046	0.025	0.037	0.040	0.040	0.036	0.031	0.044	0.043
	-0.4	0.067	0.025	0.035	0.030	0.064	0.031	0.031	0.036	0.053	0.035	0.042	0.040
	0.0	0.049	0.029	0.041	0.030	0.054	0.038	0.035	0.032	0.047	0.040	0.040	0.043
	0.4	0.034	0.021	0.045	0.029	0.049	0.031	0.037	0.037	0.044	0.038	0.039	0.037
	0.8	0.031	0.017	0.038	0.015	0.048	0.027	0.033	0.032	0.043	0.033	0.041	0.036
0.5	-0.8	0.001	0.029	0.030	0.054	0.009	0.046	0.041	0.035	0.025	0.037	0.043	0.038
	-0.4	0.042	0.029	0.034	0.038	0.048	0.030	0.041	0.038	0.049	0.031	0.040	0.040
	0.0	0.048	0.031	0.036	0.027	0.047	0.035	0.034	0.033	0.046	0.035	0.046	0.038
	0.4	0.038	0.029	0.034	0.032	0.042	0.040	0.038	0.035	0.043	0.038	0.040	0.035
	0.8	0.035	0.020	0.035	0.023	0.041	0.026	0.040	0.029	0.042	0.031	0.036	0.038
0.3	-0.8	0.000	0.024	0.034	0.061	0.005	0.047	0.038	0.037	0.024	0.039	0.043	0.034
	-0.4	0.023	0.033	0.037	0.046	0.040	0.032	0.045	0.038	0.048	0.029	0.041	0.042
	0.0	0.038	0.031	0.035	0.028	0.044	0.032	0.035	0.032	0.046	0.032	0.045	0.039
	0.4	0.035	0.032	0.037	0.028	0.040	0.037	0.039	0.033	0.044	0.039	0.040	0.041
	0.8	0.032	0.024	0.040	0.023	0.039	0.030	0.032	0.026	0.044	0.033	0.039	0.042
0	-0.8	0.000	0.017	0.029	0.071	0.004	0.037	0.026	0.030	0.025	0.029	0.028	0.029
	-0.4	0.012	0.034	0.046	0.070	0.037	0.035	0.046	0.041	0.048	0.029	0.049	0.052
	0.0	0.030	0.029	0.035	0.043	0.044	0.030	0.043	0.033	0.048	0.030	0.044	0.046
	0.4	0.030	0.034	0.032	0.029	0.041	0.036	0.039	0.031	0.045	0.034	0.036	0.043
	0.8	0.028	0.029	0.031	0.024	0.039	0.032	0.033	0.031	0.045	0.034	0.035	0.038

Figure 1: Finite Sample Size of t_{β}^{FS}

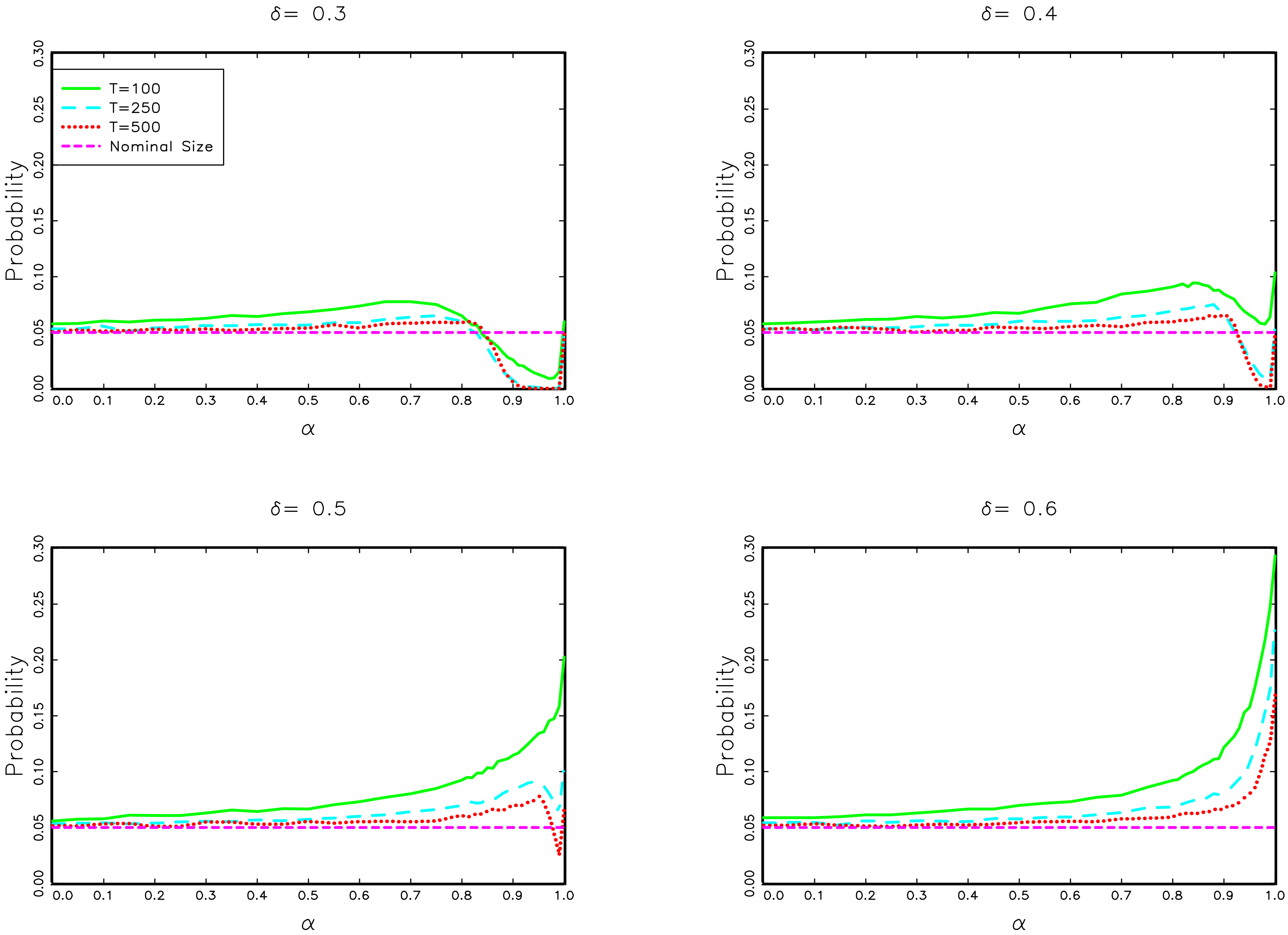

Figure 2: Finite Sample Size of $t_{\beta}^{FS}(MU)$

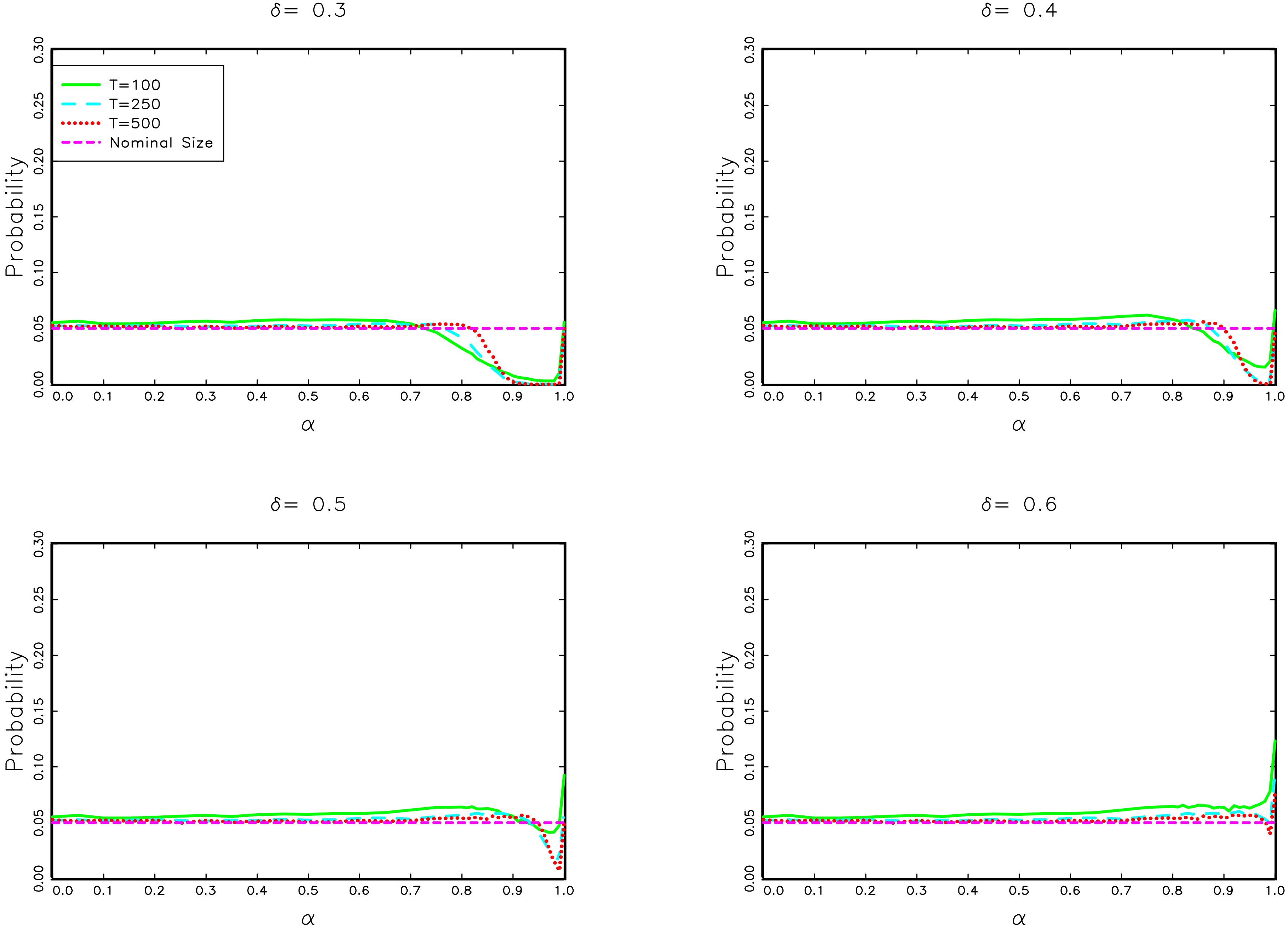

Figure 3.1: Finite Sample Power Comparisons, T=100

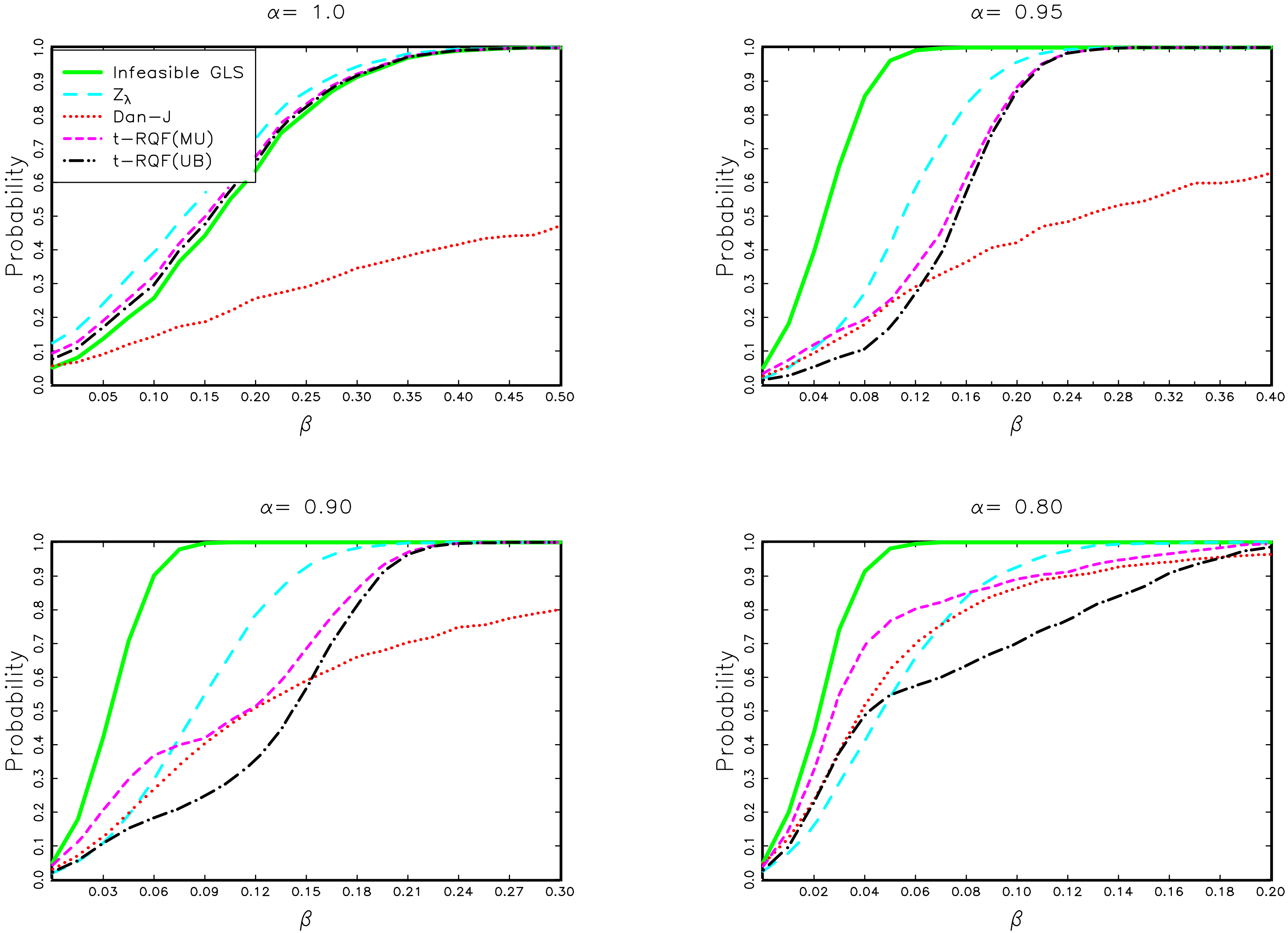

Figure 3.2: Finite Sample Power Comparisons, T=250

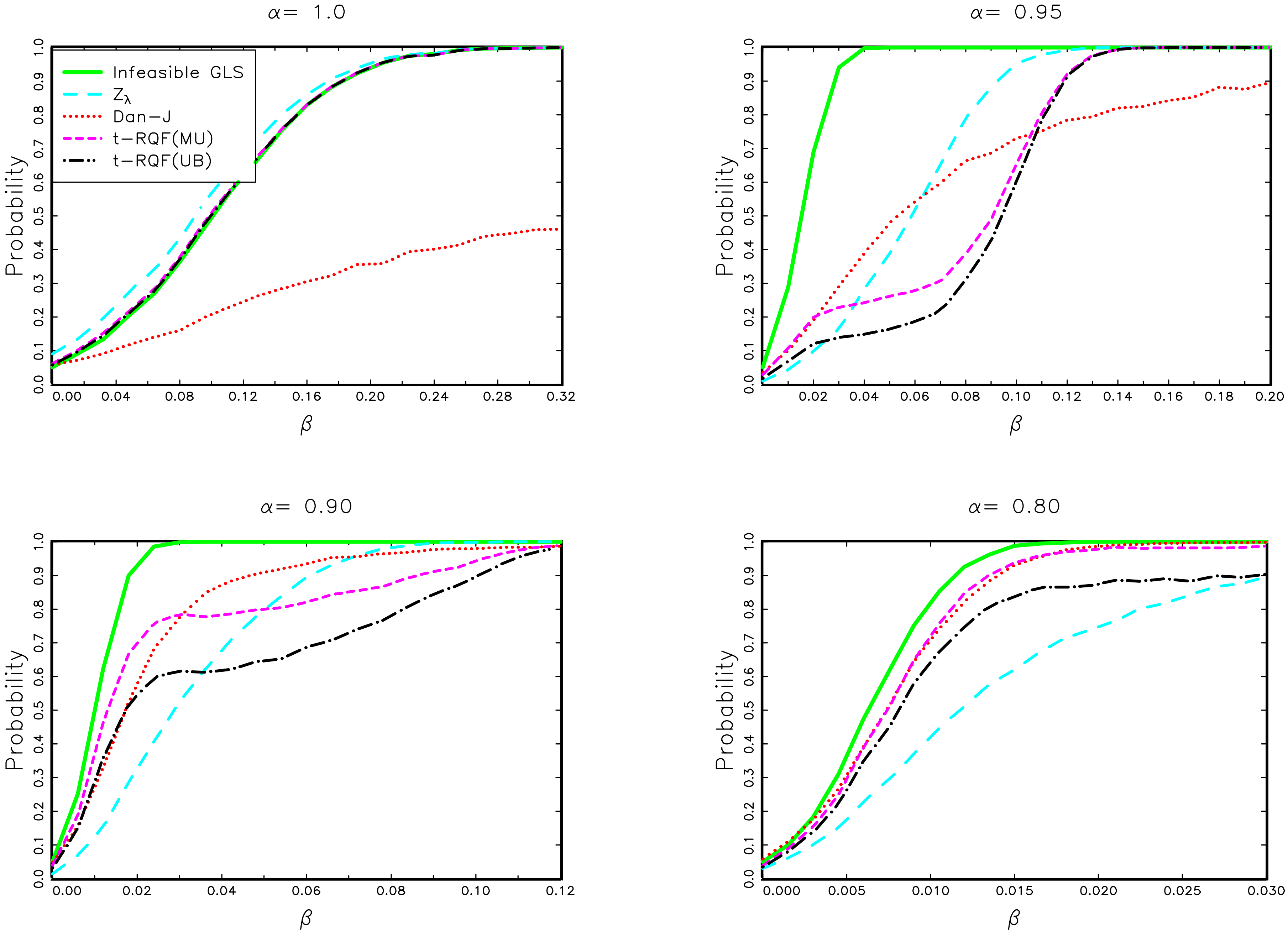

Figure 3.3: Finite Sample Power Comparisons, T=500

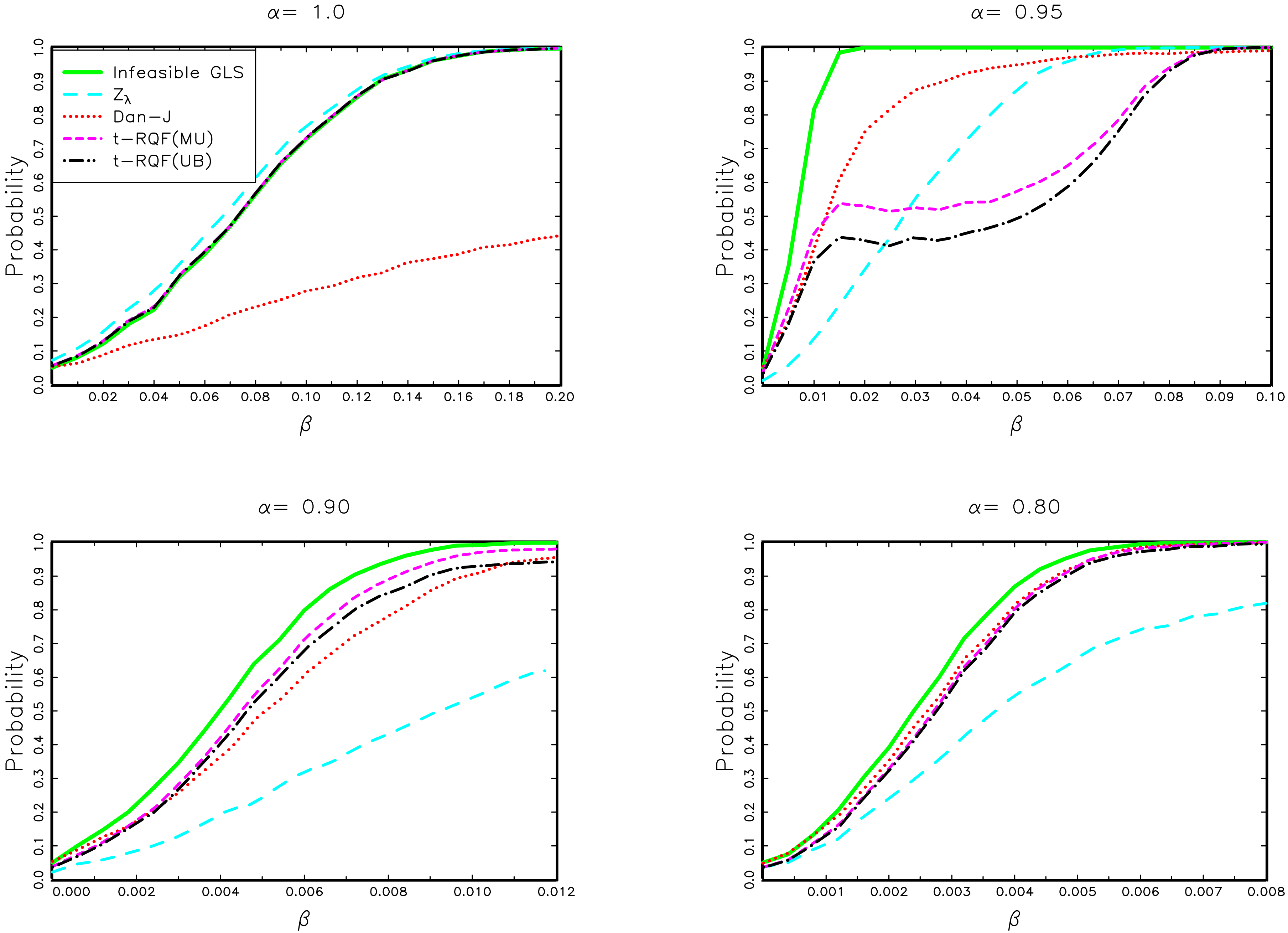

Figure 4: Finite Sample Power Comparisons (Size-Adjusted), T=100

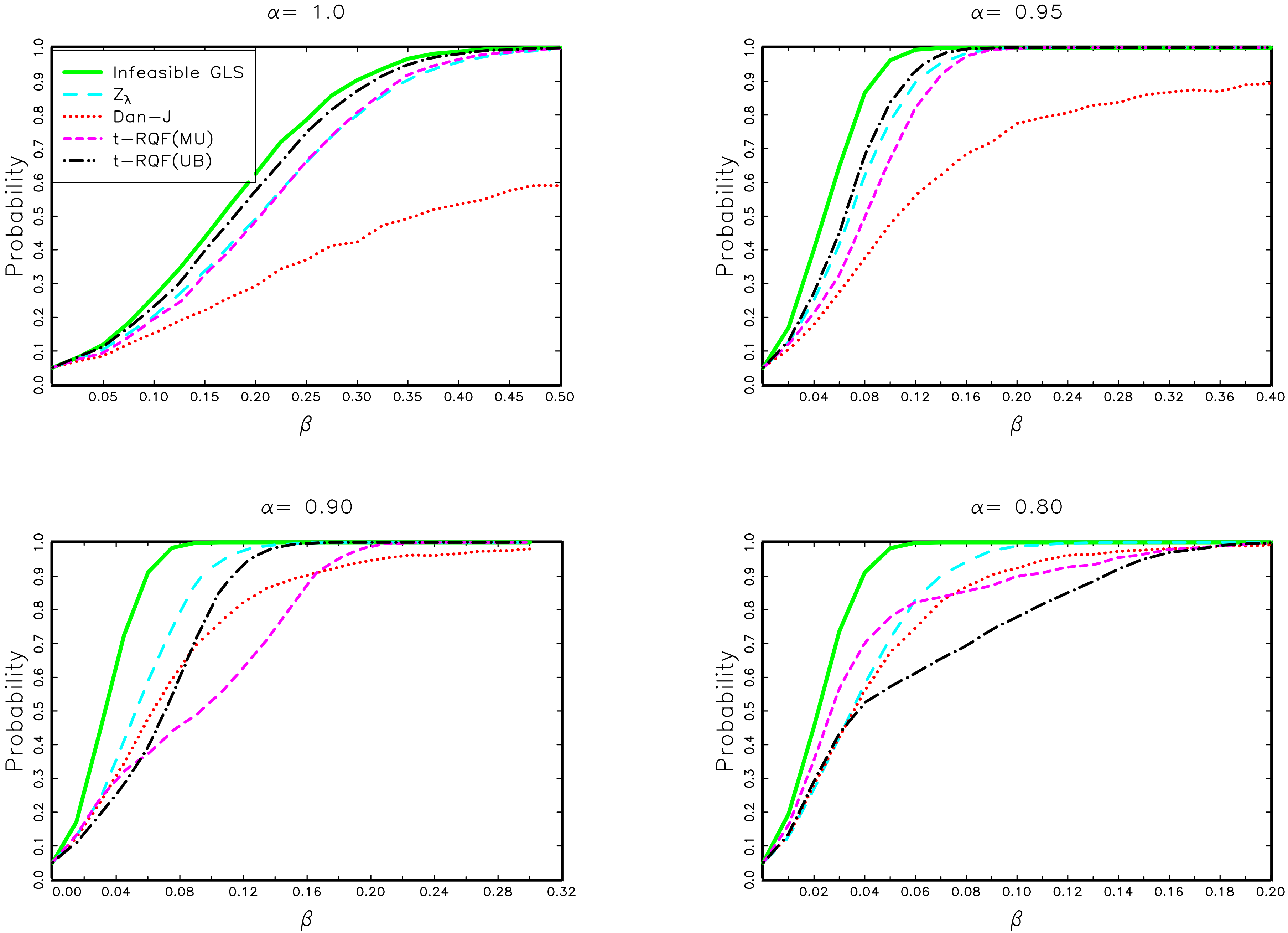

Figure 5: Finite Sample Power, AR(2) Case: T=250, φ=0.5

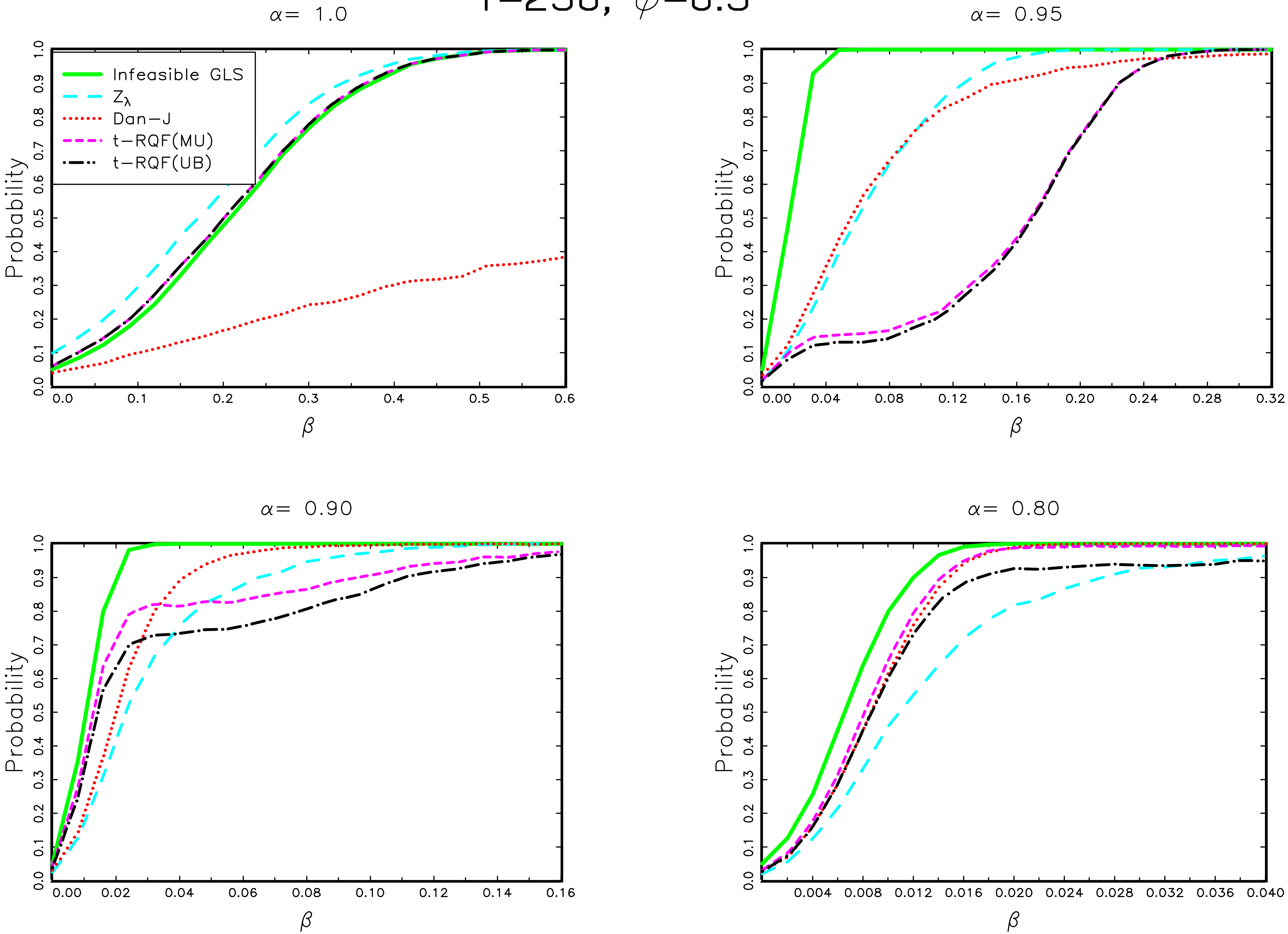

Figure 6: Finite Sample Power, ARMA Case: T=250, ϕ=-0.4