Testing for Flexible Nonlinear Trends with an Integrated or Stationary Noise Component*

Pierre Perron† **Mototsugu Shintani**‡ **Tomoyoshi Yabu**§

First version: January 2015
This version: October 2016

Abstract

This paper proposes a new test for the presence of a nonlinear deterministic trend approximated by a Fourier expansion in a univariate time series for which there is no prior knowledge as to whether the noise component is stationary or contains an autoregressive unit root. Our approach builds on the work of Perron and Yabu (2009a) and is based on a Feasible Generalized Least Squares procedure that uses a super-efficient estimator of the sum of the autoregressive coefficients α when $\alpha = 1$. The resulting Wald test statistic asymptotically follows a chi-square distribution in both the $I(0)$ and $I(1)$ cases. To improve the finite sample properties of the test, we use a bias corrected version of the OLS estimator of α proposed by Roy and Fuller (2001). We show that our procedure is substantially more powerful than currently available alternatives. We illustrate the usefulness of our method via an application to modeling the trend of global and hemispheric temperatures.

JEL Classification Number: C22.

Keywords: Fourier approximation, median-unbiased estimator, nonlinear trends, super-efficient estimator, unit root.

---

*We thank the associate editor, two anonymous referees and Vadim Marmer for helpful comments and Francsico Estrada for providing the data used in the empirical applications. We also thank the seminar and conference participants for helpful comments at Boston University, Osaka University, Univeristy of Alabama, 23th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, 11th International Symposium on Econometric Theory and Applications (SETA), and 32nd Meeting of the Canadian Econometric Study Group.

†Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02458; e-mail: perron@bu.edu.

‡RCAST, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan and Department of Economics, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN 37235-1819; e-mail: moto.shintani@gmail.com.

§Department of Business and Commerce, Keio University, Tokyo 108-0073, Japan; e-mail: tyabu@fbc.keio,ac.jp.

1 Introduction

It is well known that economic time series often exhibit trends and serial correlation. Since the functional form of the deterministic trend is typically unknown, there is a need to determine statistically whether a simple linear trend or a more general nonlinear one is appropriate. The main issue is that the limiting distributions of statistics to test for the presence of nonlinearities in the trend usually depend on the order of integration which is also unknown. On the other hand, testing whether the noise component is stationary, I(0), or has an autoregressive unit root, I(1), depends on the exact nature of the deterministic trend (e.g., Perron, 1989, 1990, for the cases of abrupt structural changes in slope or level). In particular, if the trend is misspecified, unit root tests will lose power and can be outright inconsistent (e.g., Perron, 1988, Campbell and Perron, 1991). This loss in power can also be present if the components of the trend function are over-specified e.g., by including an unnecessary trend; Perron, 1988. Hence, we are faced with a circular problem and what is needed is a procedure to test for nonlinearity that is robust to the possibilities of an I(1) or I(0) noise component.

We propose a Feasible Generalized Least Squares (FGLS) method to test for the presence of a smooth nonlinear deterministic trend function that is robust to the presence of I(0) or I(1) errors. A similar issue was tackled by Perron and Yabu (2009a) in the context of testing for the slope parameter in a linear deterministic trend model when the integration order of the noise component is unknown. The key idea is to make the estimate of the sum of the autoregressive (SAR) coefficients from the regression residual "super-efficient" when the error is I(1). This is achieved by replacing the least squares estimate of the SAR by unity whenever it reaches an appropriately chosen threshold. The limiting distribution of the test statistic is then standard normal regardless of the order of integration of the noise.

As a class of smooth nonlinear trend functions, we consider a Fourier expansion with an arbitrary number of frequencies, as in Gallant (1981) and Gallant and Souza (1991) among others. Its advantage is that it can capture the main characteristics of a very general class of nonlinear functions. This specification of the nonlinear trend function has been used in recent studies. For example, Becker, Enders and Hurn (2004) use a Fourier expansion to approximate the time-varying coefficients in a regression model and propose a test for parameter constancy when the frequency is unknown. Becker, Enders and Lee (2006) recommend pretesting for the presence of a Fourier-type nonlinear deterministic trend under the assumption of I(0) errors before employing their test for stationarity allowing a nonlinear trend. Similarly, Enders and Lee (2012) propose a Lagrange Multiplier (LM) type unit root test allowing for a flexible nonlinear trend using a Fourier approximation and use it along with a nonlinearity test under the assumption of I(1) errors. Rodrigues and Taylor (2012) also consider the same nonlinear trend in their local GLS detrended test for a unit root.

Our analysis is not the first to propose a nonlinear trend test using a flexible Fourier approximation while maintaining robustness to both I(0) and I(1) noise. At least two previous studies share the same motivation. Harvey, Leybourne and Xiao (2010, hereafter HLX) extend the robust linear trend test of Vogelsang (1998) to the case of a flexible Fourier-type trend function. Vogelsang's (1998) approach requires the choice of an auxiliary statistic so that the multiplicative adjustment term on the Wald statistic approaches one under I(0) errors and has a non-degenerate distribution under I(1) errors in the limit under the null hypothesis. By controlling the coefficient on the auxiliary statistic, the modified Wald test can have a critical value common to both I(0) and I(1) cases. HLX suggest employing a unit root test to be used as the required auxiliary statistic. Astill, Harvey, Leybourne and Taylor (2015, hereafter AHLT) suggest instead an adjustment to the critical values using a similar auxiliary statistic. AHLT show that their procedure is also robust to I(0) and I(1) errors, yet dominates the HLX method in terms of local asymptotic and finite sample power. We show that our FGLS approach has many advantages over these two methods.

The notable advantages of our proposed method can be summarized as follows. First, the local asymptotic power of our test uniformly dominates that of the other available tests and, in most cases, the power is also higher in finite samples. Second, unlike the other test statistics, ours asymptotically follows a standard chi-square distribution for both the I(0) and I(1) cases. Third, the degrees of freedom of the limiting distribution depends only on the number of frequencies, but not on the choice of frequencies. This characteristic is practically convenient since the same critical value can be used for any combination of frequencies as long as the total number of frequencies remains unchanged. In contrast, the tabulation of critical values for the other tests becomes complicated since the number of possible combinations increases rapidly with the total number of frequencies. Fourth, our test is also useful when used as a pretest in a unit root testing procedure designed to have power in the presence of nonlinear trends. In particular, for moderate nonlinearities, the magnitude of the power reduction is lower than when other tests are used as pretests. We also show that our procedure is robust to various forms of nonlinearity. In particular, unit root tests constructed using our estimated fitted trend maintain decent power even when the nonlinearities are not generated by a pure Fourier function showing the usefulness of a Fourier expansion as an approximation which can capture the main nonlinear features. Of interest, is the fact that contrary to the case of testing in a linear trend model as in Perron and Yabu (2009a), the FGLS methods needs to be implemented using the method of Prais and Winsten (1954). Using the Cochrane and Orcutt (1949) procedure fails to deliver a test with the same limit distribution regardless of the integration order of the noise component.

The paper is organized as follows. In Section 2, the basic idea of our approach is explained using a simple model with a single frequency in the Fourier expansion. In Section 3, the main theoretical results are presented for the general case which allows for multiple frequencies and serial correlation of unknown form. In Section 4, Monte Carlo evidence is presented to evaluate the finite sample performance of our procedure, as well as its performance as a pretest for a unit root test allowing for a nonlinear trend. It is also shown that our test has higher power compared to existing alternative tests and that it is robust to various forms of nonlinearities. In Section 5, we illustrate the usefulness of our method via an application to modeling the trend of global and hemispheric temperatures. Some concluding remarks are made in Section 6. All technical details are available in the online appendix. A code to compute the suggested procedures is available on the authors' websites.

2 The basic model

In order to highlight the main issues involved, we start with the simple case of a Fourier series expansion with a single frequency where the noise component follows a simple autoregressive model of order one (AR(1)). The extensions to the general case are presented in Section 3. In this basic model, a scalar random variable y_t is assumed to be generated by:

$$y_t = \sum_{i=0}^{p_d} \beta_i t^i + \gamma_1 \sin(2\pi k t/T) + \gamma_2 \cos(2\pi k t/T) + u_t \tag{1}$$

$$u_t = \alpha u_{t-1} + e_t \tag{2}$$

for $t = 1, ..., T$ where e_t is a martingale difference sequence with respect to the sigma-field $\mathcal{F}_t = \sigma\text{-field}\{e_{t-s}, s \geq 0\}$, i.e., $E(e_t|\mathcal{F}_{t-1}) = 0$, with $E(e_t^2) = \sigma^2$ and $E(e_t^4) < \infty$. Also, $u_0 = O_p(1)$. For the AR(1) coefficient of the noise component u_t, we assume $-1 < \alpha \leq 1$, so that both stationary, $I(0)$ with $|\alpha| < 1$, and integrated, $I(1)$ with $\alpha = 1$, processes are allowed. The single frequency k in the Fourier series expansion is fixed and assumed to be known. We concentrate on the cases $p_d = 0$ (non-trending) and $p_d = 1$ (linear trend), though the method is applicable in the presence of an arbitrary polynomial in time.

The interest is testing the absence of nonlinear components, $H_0 : \gamma_1 = \gamma_2 = 0$, against the alternative of the presence of a nonlinear component approximated by the Fourier series expansion, $H_1 : \gamma_1 \neq 0$ and/or $\gamma_2 \neq 0$. If the AR(1) coefficient α were known, the quasi-differencing transformation $1 - \alpha L$ could be applied to (1) and the testing problem would then simply amount to using a standard Wald test based on the OLS estimates of the quasi-differenced regression. Such a GLS procedure, however, is generally infeasible since α is unknown. Below, we briefly review the integration order-robust FGLS procedure proposed by Perron and Yabu (2009a) and explain the changes needed in the current context.

2.1 The Perron-Yabu procedure for integration order-robust FGLS

There are two main steps in Perron and Yabu's (2009a) approach to have a Wald test based on a FGLS regression so that the limit distribution is standard chi-square (or normal) in both the I(0) and I(1) cases. The first involves obtaining an estimate of α that is $\sqrt{T}$ consistent in the I(0) case but is "super-efficient" in the I(1) case. The second involves the computation of the Wald test statistic based on the FGLS estimator using an estimate of α having the stated properties. For illustration purposes consider a model with a single regressor given by $y_t = \gamma \sin(2\pi kt/T) + u_t$ combined with (2). Using the residuals $\hat{u}_t$ from a first-step OLS regression of y_t on $\sin(2\pi kt/T)$, the OLS estimator of α is given by $\hat{\alpha} = \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} / \sum_{t=2}^{T} \hat{u}_{t-1}^2$. Applying a Cochrane and Orcutt (1949) transformation, the FGLS estimate can be obtained from OLS applied to a regression of the form:

$$y_t - \hat{\alpha} y_{t-1} = \gamma \left\{ \sin(2\pi kt/T) - \hat{\alpha} \sin(2\pi k(t-1)/T) \right\} + u_t - \hat{\alpha} u_{t-1} \tag{3}$$

for $t = 2, ..., T$, together with $y_1 = \gamma \sin(2\pi k/T) + u_1$. Note that this corresponds to the FGLS estimator assuming an initial condition $u_0 = 0$. When $|\alpha| < 1$, this FGLS estimator of γ is asymptotically efficient and its t-statistic, $t_{\widehat{\gamma}}$, is asymptotically standard normal under the null hypothesis of $\gamma = 0$. In contrast, the limit distribution of the FGLS estimator is different when $\alpha = 1$. From standard results, $T(\hat{\alpha} - 1) \Rightarrow \int_0^1 W^*(r) dW(r) / \int_0^1 W^*(r)^2 dr \equiv \xi$, where '$\Rightarrow$' denotes weak convergence under the Skorohod topology, $\{W^*(r), 0 \leq r \leq 1\}$ is the continuous time residual function from a projection of a Wiener process $W(r)$ on $\sin(2\pi kr)$. The limit distribution of $t_{\widehat{\gamma}}$ under $\gamma = 0$ is then given by (see the online appendix for details):

$$\begin{aligned} t_{\widehat{\gamma}} \;\Rightarrow\; & \left[(2\pi k)^2 \int_0^1 \cos^2(2\pi kr) dr + \xi^2 \int_0^1 \sin^2(2\pi kr) dr \right]^{-1/2} \times \\ & \left\{ 2\pi k \left(\int_0^1 \cos(2\pi kr) dW(r) - \xi \int_0^1 \cos(2\pi kr) W(r) dr \right) \right. \\ & \left. -\xi \left(\int_0^1 \sin(2\pi kr) dW(r) - \xi \int_0^1 \sin(2\pi kr) W(r) dr \right) \right\} \end{aligned} \tag{4}$$

In order to obtain a standard normal limit distribution with $I(1)$ errors, Perron and Yabu (2009a) suggest replacing the $\hat{\alpha}$ by a super-efficient estimator which converges to unity at a rate faster than T when $\alpha = 1$, defined by $\hat{\alpha}_S = \hat{\alpha}$ if $T^{\delta}|\hat{\alpha} - 1| > d$ and 1 otherwise, for $\delta \in (0,1)$ and $d > 0$. Thus, whenever $\hat{\alpha}$ is in a $T^{-\delta}$ neighborhood of 1, $\hat{\alpha}_S$ takes the value 1. As shown by Perron and Yabu (2009a), $T^{1/2}(\hat{\alpha}_S - \alpha) \to^d N(0, 1-\alpha^2)$ when $|\alpha| < 1$ and $T(\hat{\alpha}_S - 1) \to^p 0$ when $\alpha = 1$.[1] When constructing the FGLS estimator of γ with this super-efficient estimator $\hat{\alpha}_S$, rather than the OLS estimator $\hat{\alpha}$, ξ in (4) can be replaced by the limit of $T(\hat{\alpha}_S - 1)$ which is zero when $\alpha = 1$. Hence, when $\alpha = 1$, under the null hypothesis the FGLS t-statistic for testing $\gamma = 0$ is such that:

$$t_{\widehat{\gamma}} \Rightarrow \left[\int_0^1 \cos^2(2\pi kr)dr\right]^{-1/2} \int_0^1 \cos(2\pi kr)dW(r) =^d N(0,1) \tag{5}$$

We then recover in the unit root case the same limiting distribution as in the stationary case.

Consider now another special case with $y_t = \gamma\cos(2\pi kt/T) + u_t$ combined with (2). While the difference between the sine and cosine functions seems minor, the same FGLS estimator combined with the super-efficient estimator $\hat{\alpha}_S$ using the Cochrane-Orcutt transformation,

$$y_t - \hat{\alpha}_S y_{t-1} = \gamma\left\{\cos\left(\frac{2\pi kt}{T}\right) - \hat{\alpha}_S \cos\left(\frac{2\pi k(t-1)}{T}\right)\right\} + u_t - \hat{\alpha}_S u_{t-1} \tag{6}$$

for $t = 2, ..., T$, together with $y_1 = \gamma\cos(2\pi k/T) + u_1$ will not yield the same limiting distribution. Instead, when $\alpha = 1$, $t_{\widehat{\gamma}} \Rightarrow \sigma^{-1}u_1 = \sigma^{-1}(u_0 + e_1)$ so that the limiting behavior of the t-statistic is dominated by the initial condition and the first value of the innovation (see the online appendix for details). This problem can be remedied using the FGLS estimator proposed by Prais and Winsten (1954), which is obtained using (6) together with

$$(1-\hat{\alpha}_S^2)^{1/2} y_1 = (1-\hat{\alpha}_S^2)^{1/2}\gamma\cos\left(\frac{2\pi k}{T}\right) + (1-\hat{\alpha}_S^2)^{1/2}u_1. \tag{7}$$

Note that it differs from the Cochrane-Orcutt FGLS estimator only in how the initial observation is transformed.[2] The null limiting distribution of the t-statistic for testing $\gamma = 0$ based on this alternative FGLS estimator is given by (see the online appendix for details):

$$t_{\widehat{\gamma}} \Rightarrow -\left[\int_0^1 \sin^2(2\pi kr)dr\right]^{-1/2} \int_0^1 \sin(2\pi kr)dW(r) =^d N(0,1) \tag{8}$$

[1] This class of the super-efficient estimator is also referred to as the Hodges estimator and has been often used as a counter-example to the efficient estimator with its asymptotic variance given by the Cramer-Rao lower bound. See, e.g., Amemiya (1985, p.124), for more discussion.

[2] See Canjels and Watson (1997) for more details on the difference between these two FGLS estimators.

when $\alpha = 1$, as required. It can easily be shown that using the Prais-Winsten FGLS estimator also delivers a null limiting distribution of the t-statistic given by (5) with the sine as well as the cosine functions. Hence, when dealing with tests related to nonlinear trends generated by Fourier expansions, one needs to modify Perron and Yabu's (2009a) procedure using the Prais-Winsten FGLS estimator instead of the FGLS estimator derived from the condition $u_0 = 0$. The limiting distribution of the test statistic is then standard normal in both the $I(0)$ and $I(1)$ cases. This is in contrast to the cases of a linear trend model considered in Perron and Yabu (2009a) and the break model considered in Perron and Yabu (2009b) since the asymptotic results for these models do not depend on the choice of the FGLS estimator.

2.2 The test statistic

We return to the basic model (1) with one frequency and express the model as:

$$y_t = x_t'\Psi + u_t \tag{9}$$

where $x_t = (z_t', f_t')'$ with $z_t = (1, t, ..., t^{p_d})'$ and $f_t = (\sin(2\pi kt/T), \cos(2\pi kt/T))'$, and the parameters are $\Psi = (\beta', \gamma')'$, $\beta = (\beta_0, ..., \beta_{p_d})'$ and $\gamma = (\gamma_1, \gamma_2)'$. Since we are interested in testing whether nonlinear trend components are present, the null hypothesis is given by $H_0 : R\Psi = 0$ where $R = [0 : I_2]$ is a $2 \times (p_d + 3)$ restriction matrix. Let $\hat{\Psi} = (\widetilde{X}'\widetilde{X})^{-}\widetilde{X}'\widetilde{y}$ be the Prais-Winsten FGLS estimator where $\widetilde{X}$ is a $T \times (p_d + 3)$ matrix of transformed data whose t^{th}-row is given by $\widetilde{x}_t' = (1 - \hat{\alpha}_S L)x_t'$ except for $\widetilde{x}_1' = (1 - \hat{\alpha}_S^2)^{1/2}x_1'$. The $T \times 1$ vector $\widetilde{y}$ is similarly defined as $\widetilde{y}_t = (1 - \hat{\alpha}_S L)y_t$ for $t = 2, ..., T$, and $\widetilde{y}_1 = (1 - \hat{\alpha}_S^2)^{1/2}y_1$. Here, $(\widetilde{X}'\widetilde{X})^{-}$ is the generalized inverse of $\widetilde{X}'\widetilde{X}$. Denote the residuals associated with this regression by $\hat{e}_t$. The Wald statistic for testing the null hypothesis is, where $s^2 = T^{-1}\sum_{t=1}^{T}\hat{e}_t^2$:

$$W_{\widehat{\gamma}} = \hat{\Psi}'R'[s^2 R(\widetilde{X}'\widetilde{X})^{-}R']^{-1}R\hat{\Psi} \tag{10}$$

Theorem 1 shows that $W_{\widehat{\gamma}}$ has a $\chi^2(2)$ distribution in both the $I(0)$ and $I(1)$ cases.

Theorem 1 *Let y_t be generated by (1) with $\gamma_1 = \gamma_2 = 0$. Then,*

$$\begin{aligned} W_{\widehat{\gamma}} \;\Rightarrow\; & [R(\int_0^1 G(r)G(r)'dr)^{-}\int_0^1 G(r)dW(r)]'[R(\int_0^1 G(r)G(r)'dr)^{-}R']^{-1} \\ & \times[R(\int_0^1 G(r)G(r)'dr)^{-}\int_0^1 G(r)dW(r)] =^d \chi^2(2) \end{aligned}$$

where $G(r) = F(r) = [1, r, ..., r^{p_d}, \sin(2\pi kr), \cos(2\pi kr)]'$ if $|\alpha| < 1$ and $G(r) = Q(r) = [0, 1, ..., p_d r^{(p_d-1)}, 2\pi k\cos(2\pi kr), -2\pi k\sin(2\pi kr)]$ if $\alpha = 1$.

Therefore, constructing the GLS regression with the super-efficient estimator, $\hat{\alpha}_S$, effectively bridges the gap between the $I(0)$ and $I(1)$ cases, and the chi-square asymptotic distribution is obtained in both cases. Note that the Wald test statistic involves a generalized inverse of $\widetilde{X}'\widetilde{X}$ due to the singularity of this matrix (the same issue occurs in Rodrigues and Taylor,2012). This is because the first column of $\widetilde{X}$ is asymptotically a zero vector when $\alpha = 1$. This poses no problem since we do not make inference about the constant term.

2.3 Local asymptotic power and the choice of δ

Under local alternative specifications as in AHLT, we can use Theorem 1 to obtain the local asymptotic power function of the test. The alternatives are given by $\gamma_1 = T^{-1/2}\gamma_0\sigma$ and $\gamma_2 = T^{-1/2}\gamma_0\sigma$ for the case of $I(0)$ errors and $\gamma_1 = T^{1/2}\gamma_0\sigma$ and $\gamma_2 = T^{1/2}\gamma_0\sigma$ for the case of $I(1)$ errors, where the scaling by σ is to factor out the variance. The details about the theoretical results on the local asymptotic power functions for our test and that of the ASW (Adaptive Scaled Wald) test of AHLT are given in the online appendix. It is easy to see that the local asymptotic power function of our test is equivalent to that of the Wald test based on the infeasible GLS procedure that assumes a known value α. Hence, it is the most powerful local test (under Gaussian errors) at least pointwise in α. To quantify the extent of the power gains over using the ASW test, Figure 1 plots the local asymptotic power functions of our test and that of the ASW test for the constant case ($p_d = 0$). Clearly, our test permits important power gains, especially in the case of I(1) errors. These power improvements will be shown to hold as well in finite samples via simulations later.

Note that the result obtained in Theorem 1 is pointwise in α for $-1 < \alpha \leq 1$ and does not hold uniformly, in particular in a local neighborhood of 1. Adopting the standard local to unity approach which is expected to provide a good approximation when α is close to but not equal to one, we have the following result proved in the online appendix.

Theorem 2 *Let y_t be generated by (1) with $\gamma_1 = \gamma_2 = 0$. Suppose that $\alpha = 1 + c/T$, then:*

$$\begin{aligned} W_{\widehat{\gamma}} \;\Rightarrow\; & [R(\int_0^1 Q(r)Q(r)'dr)^- \int_0^1 Q(r)dJ_c(r)]'[R(\int_0^1 Q(r)Q(r)'dr)^- R']^{-1} \\ & \times [R(\int_0^1 Q(r)Q(r)'dr)^- \int_0^1 Q(r)dJ_c(r)]' \end{aligned}$$

where $Q(r) = [0, 1, ..., p_d r^{(p_d-1)}, 2\pi k\cos(2\pi kr), -2\pi k\sin(2\pi kr)]$ and $J_c(r) = \int_0^r \exp(c(r-s))dW(s) \sim N(0, (\exp(2cr)-1)/2c)$.

The result is fairly intuitive. Since the true value of α is in a T^{-1} neighborhood of 1, and $\hat{\alpha}_S$ truncates the 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 $\hat{\alpha}_S = 1$. Hence, the FGLS estimator of Ψ is essentially the same as that based on first-differenced data. Note that when $c = 0$, we recover the result of Theorem 1 for the $I(1)$ case. However, when $c < 0$, the variance of $J_c(r)$ is smaller than that of $W(r)$. Hence, the upper quantiles of the limit distributions are, accordingly, smaller than those of a $\chi^2(2)$, so that, without modifications, a conservative test may be expected for values of α close to 1, relative to the sample size.

Theorem 1 is valid for the super-efficient estimator $\hat{\alpha}_S$ for any choice of $\delta \in (0, 1)$ and $d > 0$. Regarding the choice of δ, Perron and Yabu (2009a) recommend to set $\delta = 1/2$ based on local to unity arguments. We can apply the same arguments here and verified by simulations that $\delta = 1/2$ is the best choice for the tests and models considered here. Hence, we continue to use this value and will calibrate the appropriate value of d via simulations.

2.4 Bias correction for improved finite sample properties

The test statistic $W_{\hat{\gamma}}$ is constructed from the super-efficient estimator $\hat{\alpha}_S$ that is based on the OLS estimator $\hat{\alpha}$, which is known to be biased downward in finite samples especially when α is near one. Hence, in many cases, the truncation may not be used even when it would be desirable. To circumvent this problem, Perron and Yabu (2009a) recommend using Roy and Fuller's (2001) bias corrected estimator instead of the OLS estimator in the context of a linear trend model and show that such a correction improves the finite sample performance of their test without changing its asymptotic properties. The aim of this section is to suggest a similar bias correction to improve the finite sample properties of the test $W_{\hat{\gamma}}$.

Roy and Fuller (2001) proposed a class of bias corrected estimators and we consider here the one based on the OLS estimator. It is a function of a unit root test, namely the t-ratio $\hat{\tau} = (\hat{\alpha} - 1)/\hat{\sigma}_\alpha$, where $\hat{\alpha}$ is the OLS estimate and $\hat{\sigma}_\alpha$ is its standard error. The bias-corrected estimator is given by $\hat{\alpha}_M = \hat{\alpha} + C(\hat{\tau})\hat{\sigma}_\alpha$, where in the general case with errors following an autoregression of order p_T (to be considered later; for now $p_T = 0$):

$$C(\hat{\tau}) = \begin{cases} -\hat{\tau} & \text{if} \quad \hat{\tau} > \tau_{pct} \\ [(p_T + 2)/2]T^{-1}\hat{\tau} - (1 + r)[\hat{\tau} + c_2(\hat{\tau} + a)]^{-1} & \text{if} \quad -a < \hat{\tau} \leq \tau_{pct} \\ [(p_T + 2)/2]T^{-1}\hat{\tau} - (1 + r)\hat{\tau}^{-1} & \text{if} \quad -c_1^{1/2} < \hat{\tau} \leq -a \\ 0 & \text{if} \quad \hat{\tau} \leq -c_1^{1/2} \end{cases} \tag{11}$$

where τ_{pct} is some percentile of the limiting distribution of $\hat{\tau}$ when $\alpha = 1$, $c_1 = (1+r)T$, $r = p_d + 1 + 2n$ is the number of estimated parameters, $c_2 = [(1+r)T - \tau_{pct}^2([(p_T+2)/2] + T)][\tau_{pct}(a+\tau_{pct})([(p_T+2)/2]+T)]^{-1}$ and a is some constant. The parameters for which specific values are needed are τ_{pct} and a. Based on extensive simulations, we selected $a = 10$ since it leads to tests with better properties. For τ_{pct} we shall consider $\tau_{.50}$ or $\tau_{.85}$. When using $\tau_{.50}$ the version of the test is labelled as "median-unbiased" and when using $\tau_{.85}$, it is labelled as "upper-biased." The values of $\tau_{.50}$ and $\tau_{.85}$ depend on p_d and the type of frequencies included [3]. Table 1 presents values for $p_d = 0, 1$ for cases with a single frequency k taking value between 1 and 5 and for cases with multiple frequencies $k = 1, ..., n$ for n between 1 and 5. It should be noted that to obtain the super-efficient estimator $\hat{\alpha}_S$, $\hat{\alpha}$ can be replaced by $\hat{\alpha}_M$ since all that is needed is that $T(\hat{\alpha}_M - 1) = O_p(1)$ when $\alpha = 1$, and $T^{1/2}(\hat{\alpha}_M - \alpha) \to^d N(0, 1-\alpha^2)$ when $\alpha < 1$. These conditions are satisfied and thus all the large sample results, Theorems 1 and 2, continue to hold. Based on extensive simulations, we found that the value $d = 1$ combined with $\hat{\alpha}_M$ leads to the best results in finite samples. Hence, our suggested AR(1) coefficient estimator to be used in the Prais-Winsten FGLS estimator is $\hat{\alpha}_{MS}$, which takes the value $\hat{\alpha}_M$ when $|\hat{\alpha}_M - 1| > T^{-1/2}$ and 1 otherwise.

Figure 2.a presents results about the size of the $W_{\hat{\gamma}}$ test with only a constant ($p_d = 0$) when constructed using the OLS estimator, the median unbiased estimator ($\hat{\alpha}_{MS}$ with $\tau_{pct} = \tau_{.50}$) and the upper biased estimator ($\hat{\alpha}_{MS}$ with $\tau_{pct} = \tau_{.85}$). Figure 2.b shows the corresponding results for the linear trend case ($p_d = 1$). The data are generated by the AR(1) process $y_t = \alpha y_{t-1} + e_t$ with $e_t \sim i.i.d.\ N(0,1)$ and $y_0 = 0$ (setting the constant and trend parameters to zero is without loss of generality since the tests are invariant to them). The nominal size of the tests is 5% throughout the paper and the exact size is evaluated using 10,000 replications. The sample sizes are set to T = 150, 300, and 600. The results clearly show that when using the OLS estimator $\hat{\alpha}$ the size distortions are non-negligible when α is close to 1 and remain even with T as large as 600. In contrast, the exact size of the test constructed using either the median unbiased or, especially, the upper biased estimator, is very close to the nominal size regardless of the value of α for all T. These results are encouraging and point to the usefulness of the bias correction step in our testing procedure.

[3] Roy, Falk and Fuller (2004) and Perron and Yabu (2009a) use a similar bias correction based on a weighted symmetric least-squares estimator of α instead of the OLS estimator employed here. Both lead to tests with similar properties. However, note that the test proposed by Roy, Falk and Fuller (2004) has very different sizes in the I(0) and I(1) cases; see Perron and Yabu (2012) for details.

3 The general model

Having laid out the foundation for the basic model (1), it is relatively straightforward to extend the test procedure to cover the general model which involves the possibility of more than one frequency in the Fourier expansion and a general serial correlation structure in the noise component. The general model is given by:

$$y_t = \sum_{i=0}^{p_d} \beta_i t^i + \sum_{j=1}^{n} \gamma_{1j} \sin(2\pi k_j t/T) + \sum_{j=1}^{n} \gamma_{2j} \cos(2\pi k_j t/T) + u_t \tag{12}$$

for $t = 1, ..., T$. The k_j's are nonnegative integers for $j = 1, ..., n$, and n is the total number of frequencies used in the Fourier approximation. Note that the set of k_j's can be a proper subset of all the integers between 1 and the maximum frequency k_n so that k_n need not correspond to the n^{th} frequency. For example, when $n = 2$ and $k_2 = 3$, (k_1, k_2) can be either $(1, 3)$ or $(2, 3)$. This will turn out to be useful when designing a strategy to estimate the number of frequencies to include. In vector form, (12) can also be written as (9) using $x_t = (z_t', f_t')'$ where $z_t = (1, ..., t^{p_d})'$, $f_t = (\sin(2\pi k_1 t/T), \cos(2\pi k_1 t/T), ..., \sin(2\pi k_n t/T), \cos(2\pi k_n t/T))'$and $\Psi = (\beta', \gamma')'$ where $\beta = (\beta_0, ..., \beta_{p_d})'$ and $\gamma = (\gamma_{11}, \gamma_{21}, ..., \gamma_{1n}, \gamma_{2n})'$. For the noise component, we assume that u_t is generated by one of the following two structures:

· Assumption I(0): $u_t = C(L)e_t$, $C(L) = \sum_{i=0}^{\infty} c_i L^i$, $\sum_{i=0}^{\infty} i|c_i| < \infty$, $0 < |C(1)| < \infty$;

· Assumption I(1): $\Delta u_t = D(L)e_t$, $D(L) = \sum_{i=0}^{\infty} d_i L^i$, $\sum_{i=0}^{\infty} i|d_i| < \infty$, $0 < |D(1)| < \infty$.

As before $e_t \sim (0, \sigma^2)$ is a martingale difference sequence and $u_0 = O_p(1)$. These conditions ensure that we can apply a functional central limit theorem to the partial sums of u_t in the $I(0)$ case and the partial sums of Δu_t in the $I(1)$ case. In both cases, u_t has an autoregressive representation of the form $u_t = \sum_{i=1}^{\infty} a_i u_{t-i} + e_t$, or equivalently

$$u_t = \alpha u_{t-1} + A^*(L)\Delta u_{t-1} + e_t \tag{13}$$

where α now represents the sum of the autoregressive coefficients. When u_t is I(0), $\alpha = \sum_{i=1}^{\infty} a_i$ and $A^*(L) = \sum_{i=1}^{\infty} a_i^* L^i$ where $a_i^* = -\sum_{j=i+1}^{\infty} a_j$ and $A(L) = \sum_{i=1}^{\infty} a_i L^i = C(L)^{-1}$. When u_t is I(1), $\alpha = 1$ and $A^*(L) = L^{-1}(1 - D(L)^{-1})$. The sum of the autoregressive coefficients α in (13) can be consistently estimated by the OLS applied to the regression:

$$\hat{u}_t = \alpha \hat{u}_{t-1} + \sum_{i=1}^{p_T} a_i^* \Delta \hat{u}_{t-i} + e_{pt} \tag{14}$$

where $\hat{u}_t$ are the residuals from a regression of y_t on x_t and p_T is the truncation lag order which satisfies $p_T \to \infty$ and $p_T^3/T \to 0$ as $T \to \infty$. Under this condition on the rate of p_T, the OLS estimator $\hat{\alpha}$ is consistent and $T^{1/2}(\hat{\alpha} - \alpha) = O_p(1)$ when u_t is $I(0)$

(see Berk, 1974, Ng and Perron, 1995). On the other hand, if $\alpha = 1$, $T(\hat{\alpha} - 1) \Rightarrow D(1)\int_0^1 W^*(r)dW(r)/\int_0^1 W^*(r)^2 dr$ where $W^*(r)$ is the residual function from a regression of $W(r)$ on $F(r) = [1, r, ..., r^{p_d}, \sin(2\pi k_1 r), \cos(2\pi k_1 r), ..., \sin(2\pi k_n r), \cos(2\pi k_n r)]'$. However, if we replace the OLS estimator $\hat{\alpha}$ with a super-efficient estimator similar to $\hat{\alpha}_S$ or its bias-corrected version $\hat{\alpha}_{MS}$, we have $T(\hat{\alpha}_S - 1) \to_p 0$ and $T(\hat{\alpha}_{MS} - 1) \to_p 0$ when $\alpha = 1$ so that the limiting distribution of the Prais-Winsten FGLS estimator is the same chi-square regardless of the integration order of the noise.

3.1 The test statistic

The null hypothesis for the absence of nonlinear components for the general case is now given by $R\Psi = 0$ where $R = [0 : I_{2n}]$ is a $2n \times (p_d + 1 + 2n)$ restriction matrix. We again use the Prais-Winsten FGLS estimator $\hat{\Psi}$ by running the transformed regression:

$$(1 - \hat{\alpha}_{MS}L)y_t = (1 - \hat{\alpha}_{MS}L)x_t'\Psi + (1 - \hat{\alpha}_{MS}L)u_t \tag{15}$$

for $t = 2, ..., T$, together with $(1 - \hat{\alpha}_{MS}^2)^{1/2}y_1 = (1 - \hat{\alpha}_{MS}^2)^{1/2}x_1'\Psi + (1 - \hat{\alpha}_{MS}^2)^{1/2}u_1$. Since the residuals from this regression now approximates to $v_t \equiv (1 - \alpha L)u_t$ instead of e_t, we denote them by $\hat{v}_t$ instead of $\hat{e}_t$. The Wald statistic robust to serial correlation in v_t is:

$$W_{\widehat{\gamma}} = \hat{\Psi}'R'[\hat{\omega}^2 R(\widetilde{X}'\widetilde{X})^{-}R']^{-1}R\hat{\Psi} \tag{16}$$

where $\widetilde{X}$ is a $T \times (p_d + 1 + 2n)$ matrix of transformed data whose t^{th}-row is given by $\widetilde{x}_t' = (1-\hat{\alpha}_{MS}L)x_t'$ except for $\widetilde{x}_1' = (1-\hat{\alpha}_{MS}^2)^{1/2}x_1'$. Here, $\hat{\omega}^2$ is a long-run variance estimator of $v_t = (1-\alpha L)u_t$ which replaces s^2 in (10). More specifically, $\hat{\omega}^2$ is a consistent estimator of (2π times) the spectral density function at frequency zero of v_t, given by $\omega^2 = (1-\alpha)^2 A(1)^{-2}\sigma^2 = \sigma^2$ when u_t follows an I(0) process, and $\omega^2 = D(1)^2\sigma^2$ when u_t follows an I(1) process. Accordingly, we use the following long-run variance estimator:

$$\hat{\omega}^2 = \begin{cases} (T - p_T)^{-1}\sum_{t=p_T+1}^{T} \hat{e}_{pt}^2 & \text{if} \quad T^{1/2}|\hat{\alpha}_M - 1| > 1 \\ T^{-1}\sum_{t=1}^{T} \hat{v}_t^2 + T^{-1}\sum_{j=1}^{T-1} w(j, m_T)\sum_{t=j+1}^{T} \hat{v}_t\hat{v}_{t-j} & \text{if} \quad T^{1/2}|\hat{\alpha}_M - 1| \leq 1 \end{cases} \tag{17}$$

where $\hat{e}_{pt}$ are the residuals from (14) and $w(j, m_T)$ is a weight function with bandwidth m_T. We use the Andrews' (1991) automatic selection procedure for m_T along with the quadratic spectral window. Note that this long-run variance estimator can be viewed as a combination of parametric and nonparametric estimators depending on the threshold used to construct the super-efficient estimator $\hat{\alpha}_S$. The following theorem, whose proof is similar to that of Theorem 1, and hence omitted, shows that the test based on the FGLS procedure using $\hat{\alpha}_{MS}$ has a $\chi^2(2n)$ distribution in both the $I(0)$ and $I(1)$ cases.

Theorem 3 *Let y_t be generated by (12). Then,*

$$W_{\hat{\gamma}} \Rightarrow [R(\int_0^1 G(r)G(r)'dr)^- \int_0^1 G(r)dW(r)]'[R(\int_0^1 G(r)G(r)'dr)^- R']^{-1} \times [R(\int_0^1 G(r)G(r)'dr)^- \int_0^1 G(r)dW(r)] =^d \chi^2(2n)$$

where $G(r) = F(r) = [1, r, ..., r^{p_d}, \sin(2\pi k_1 r), \cos(2\pi k_1 r), ..., \sin(2\pi k_n r), \cos(2\pi k_n r)]'$ *if* $|\alpha| < 1$ *and if* $\alpha = 1$, $G(r) = Q(r) = [0, 1, ..., p_d r^{(p_d - 1)}, 2\pi k_1 \cos(2\pi k_1 r), -2\pi k_1 \sin(2\pi k_1 r), ..., 2\pi k_n \cos(2\pi k_n r), -2\pi k_n \sin(2\pi k_n r)]$.

Remarks: 1) It remains in the general case that constructing the GLS regression with the super-efficient estimator, $\hat{\alpha}_{MS}$, effectively bridges the gap between the $I(0)$ and $I(1)$ cases, and the chi-square asymptotic distribution is common to both. 2) The degrees of freedom of the limiting chi-square distribution is $2n$ so that it depends only on the number of frequencies, but not on the choice of the frequencies itself. This is convenient since the same critical values can be used for any combination of frequencies as long as the total number of frequencies remains unchanged. In contrast, the limiting distribution of the MW (Modified Wald) test proposed by HLX, and that of the ASW test proposed by AHLT is non-standard and depends on the choice of the frequencies, which makes inference difficult, especially as the number of frequencies increases. 3) While we are mainly interested in testing the restriction that all the coefficients of the nonlinear trend components are zero, the test statistic can easily be modified to test zero restrictions on a subset of the coefficients. If $m(< n)$ denotes the number of frequencies of interest, we can use a $2m \times (p_d + 1 + 2n)$ restriction matrix $R = [0 : S]$ where S is a $2m \times 2n$ selection matrix constructed by excluding unrelated row vectors from I_{2n}. Under the null hypothesis, the Wald test statistic now asymptotically follows a chi-square distribution with $2m$ degrees of freedom. This version is convenient for model selection purposes when the form of the Fourier expansion is unknown.

4 Monte Carlo experiments

In this section, we conduct Monte Carlo experiments with two objectives in mind. The first is to evaluate the power of our test, both the median-unbiased and upper-biased versions, and compare it with that of previously proposed procedures to test for the presence of nonlinear trends robust to having either I(0) and I(1) errors. Such tests include the MW test statistic proposed by HLX, and the ASW test statistic proposed by AHLT. Since AHLT have already shown that the power performance of the ASW test statistic dominates that of the MW test, we only report comparisons with the former (the test is described in the online appendix). The second objective is to evaluate the performance of our test when it is used as a pretest for a unit root test. We combine our procedure and the LM unit root test of Enders and Lee (2012) that allows for a flexible nonlinear trend using a Fourier series approximation.

Before describing the simulation design, we review each step of our recommended testing procedure for the general case. 1) Run the OLS regression (12) and obtain residuals $\hat{u}_t$. 2) Run the regression (14) and obtain $\hat{\alpha}$ with p_T selected using the MAIC proposed by Ng and Perron (2001), with $p_T \in [0, 12(T/100)^{1/4}]$. 3) Construct the bias corrected estimator given by $\hat{\alpha}_M = \hat{\alpha} + C(\hat{\tau})\hat{\sigma}_\alpha$, where $C(\hat{\tau})$ is defined by (11). For the median-unbiased version use $\tau_{0.5}$ and for the upper-biased version use $\tau_{0.85}$, whose values are given in Table 1. 4) Construct the super-efficient estimator given by $\hat{\alpha}_{MS} = \hat{\alpha}_M$ if $|\hat{\alpha}_M - 1| > T^{-1/2}$ and 1 otherwise. 5) Construct the Prais-Winsten FGLS estimate $\hat{\Psi}$ and residuals $\hat{v}_t$ from the regression (15) using $\hat{\alpha}_{MS}$ and construct the Wald test statistic (16) using $\hat{\omega}^2 = (T - p_T)^{-1} \sum_{t=p_T+1}^{T} \hat{e}_{pt}^2$ if $|\hat{\alpha}_M - 1| > T^{-1/2}$ and $\hat{\omega}^2 = T^{-1} \sum_{t=1}^{T} \hat{v}_t^2 + T^{-1} \sum_{j=1}^{T-1} w(j, m_T) \sum_{t=j+1}^{T} \hat{v}_t \hat{v}_{t-j}$, otherwise.

4.1 The size and power of the tests

We first report the empirical size of the ASW test and ours with data generated by

$$y_t = u_t, \qquad (1 - \phi L)u_t = (1 + \theta L)e_t \tag{18}$$

where $e_t \sim i.i.d.\ N(0, 1)$ and $u_0 = 0$. We set $\phi = 1, 0.95, 0.9, 0.8$ and $\theta = -0.8, -0.4, 0.0, 0.4, 0.8$. The exact size is computed as the frequency of rejecting the null from $10,000$ replications when using a 5% nominal size. The sample sizes considered are $T = 150$, 300 and 600. Note that, when $\phi = 1$, the error term follows an I(1) process with the sum of the AR coefficients $\alpha = 1$. For the other choice of ϕ, the error term follows an I(0) process with the sum of AR coefficients given by $\alpha = 1 - (1 - \phi)(1 + \theta)^{-1}$. We only consider positive AR coefficients since this is the most relevant case in practice [4].

The size of our test using a single frequency $k = 1$ is reported in Table 2.a (with a constant only; $p_d = 0$) and Table 2.b (with a linear trend; $p_d = 1$). The results show that our test has reasonable size properties for both the I(0) and I(1) cases. This is especially the case for the upper-biased version of the test. The size of the ASW test is also adequate though some liberal size distortions are present in the case of a large negative moving-average coefficient, unlike our test which maintains nearly the correct size when using the upper-biased version.

[4] The combinations of ϕ and θ require some attention; e.g., when $\phi = 0.8$ and $\theta = -0.8$, the process is not ARMA(1,1) but rather a simple i.i.d. process with the true sum of the AR coefficients being 0.

To evaluate the power of the tests, the data are generated from the nonlinear process:

$$y_t = \gamma(\sin(2\pi t/T) + \cos(2\pi t/T)) + u_t \tag{19}$$

where $\gamma > 0$. The error term is generated from $u_t = \alpha u_{t-1} + e_t$ with $e_t \sim i.i.d.\ N(0,1)$ and $u_0 = 0$ for $\alpha = 1.0,\ 0.95, 0.9$ and 0.8. Here, we consider the case with the frequency $k = 1$ known. However, we continue to use the test which allows for general serial correlation and does not rely on the knowledge of the AR(1) error structure. We will later consider the case with an unknown frequency structure. The results are presented in Figures 3.a to 3.c for the case with a constant only ($p_d = 0$) and Figures 4.a to 4.c for the case with a linear trend ($p_d = 1$). The first thing to note is that the power of both versions of our test is close to that achievable by the infeasible GLS estimate with a known value of α (the upper bound with Gaussian errors) when $\alpha = 1$. In that case, the power of the ASW test is substantially lower. The same features hold approximately when α is far from one (relative to the sample size, i.e., not local to one) as shown in the case with $T = 600$ and $\alpha = 0.8$.

Things are different when α is local to 1. The power of the median-unbiased version is then higher than that of the upper-biased version. Some of the differences can be explained by the fact that the median-unbiased version tends to have higher size than the upper-biased version, which tends to be conservative. In general, the power of the ASW test is lower than either version of our test, especially the median-unbiased version. There are cases for which the ASW test is more powerful though never uniformly in the value of the alternative. This is mainly due to the fact that both versions of our test can exhibit a "kinked" power curve when α is local to 1. When comparing to the median-unbiased version, the power of the ASW test is higher in the following cases when considering a constant only ($p_d = 0$): $T = 150$, $\alpha = 0.8$ and $T = 300$, $\alpha = 0.9$ for large alternatives (though the differences are minor), $T = 300$, $\alpha = 0.95$ for medium alternatives, $T = 600$, $\alpha = 0.95$ for large alternatives. When considering a fitted linear trend ($p_d = 1$), the ASW test has lower power in all cases, with very minor exceptions. In summary, in terms of power the median-unbiased version of our test is clearly preferable. This may be counter-balanced by the fact that it is also the test most prone to having liberal size distortions (though relatively minor), occur mostly when α is close to or equal to 1 with a large moving-average coefficient, and reduce noticeably as the sample size increases. Since the upper-biased version has smaller size distortions than the median-unbiased version and the power is comparable unless the autoregressive parameter is local to one though with small differences, we tend to recommend the use of the upper-biased version. Henceforth, we only consider this version of the test.

4.2 The relative performance in choosing the number of frequencies

We now turn to the issue of choosing the number of frequencies. To simplify, we let $k_n = n$ and the data are generated from, with the same AR(1) error term as before,

$$y_t = \gamma \sum_{k=1}^{2} (\sin(2\pi kt/T) + \cos(2\pi kt/T)) + u_t \tag{20}$$

whose structure is, for simplicity, assumed to be known. Therefore, the true number of frequencies is given by $n = 2$, whenever $\gamma \neq 0$, and $n = 0$ when $\gamma = 0$. We consider experiments with $\gamma = 0, 1, 2, 3, 4$ and 5. We use a general-to-specific procedure based on the sequential application of the variant of our test for subsets of coefficients. We first set the total number of frequencies at $n = 3$ and test the null hypothesis that the coefficients related to the maximum frequency $k = 3$ are zero. If the null hypothesis is rejected, we select $n = 3$. If not, we set $n = 2$, and test whether the coefficients related to $k = 2$ are zero. We continue the procedure until we reject the null or reach $n = 0$. Note that the number of restrictions in each step is 2 ($= 2m$) so that all the tests share the same critical value from the chi-square distribution with 2 degrees of freedom. We compare the selection frequencies of this procedure with the one based on the ASW test combined with the frequency selection algorithm proposed in HLX (p. 388), as advocated by AHLT. For the ASW test, results using tests at the 5% significance level are reported. For our test, results with both 1% and 5% significance levels are reported. Table 3 reports the relative frequency of choosing each of $n = 0, 1, 2$ and 3 when a trend term is included ($p_d = 1$). Having no time trend in the DGP is irrelevant since the tests are invariant to its specification. Compared to the procedure based on the ASW test, our procedure is substantially better at selecting the true number of frequencies $n = 2$ when $\gamma \neq 0$. Note that the procedure based on the ASW test has very little power so that $n = 0$ is the value most often selected even when γ is large. With respect to the size of the test for our procedure, using a 1% significance level leads to better selection when $\gamma = 0$ or when γ is very large, otherwise using a 5% significance level is preferred.

4.3 The performance as pre-tests for a unit root test

Let us now investigate the performance of our test when it is used as a pretest before applying the unit root test of Enders and Lee (2012). In all cases, we consider the case with a fitted time trend ($p_d = 1$). The simulation design follows that of Enders and Lee (2012). The exact size and power of their unit root test are evaluated when the number of frequencies in the nonlinear trend function is unknown. To evaluate the size of the test, the data are generated from (20) with I(1) errors generated by a random walk with $i.i.d.$ $N(0, 1)$ errors. We set

$T = 150$, 300 and 600 and $\gamma = 0, 1, 2, 3, 4, 5$ and the nominal size of the unit root test is 5%. Table 4 shows the empirical size of the unit root tests when (i) the number of frequencies is incorrectly specified at $n = 0$ (unless $\gamma = 0$), (ii) when the number of frequencies is correctly specified at $n = 2$ (unless $\gamma = 0$), (iii) when the number of frequencies is selected based on the sequential application of the ASW test, and (iv) when the number of frequencies is selected based on the sequential application of our test. As before, results using a 5% significance level are reported for the ASW-based procedure and using both 1% and 5% levels for ours. When the number of frequencies is incorrectly specified at $n = 0$, the unit root test is clearly undersized. The exact sizes of the unit root test with n selected by the ASW-based procedure and our test are comparable to that of the correctly specified case.

The advantage of employing our procedure becomes evident when considering the power of the unit root test. In what follows, we report results for $T = 150$. Figures 5.a and 5.b present the power results when the data are generated from (20) with the I(0) error generated as AR(1) processes with coefficients $\alpha = 0.9$ and 0.8, and innovations that are i.i.d. $N(0, 1)$. For all cases, a U-shaped non-monotonic power function is observed when plotted as a function of γ. However, using our test, the reduction in power is less pronounced, especially with $\alpha = 0.8$. This feature can be understood by comparing these results with those in Figures 6.a and 6.b, which plot the power of the unit root test for the cases of fixed total number of frequencies at $n = 0$ and $n = 2$. When the unit root test is applied with an incorrect total number of frequencies of $n = 0$ its power monotonically decreases with γ. In contrast, if n is correctly specified, its power becomes invariant to γ. The results in Table 3 show that the ASW-based procedure tends to select $n = 0$ much more frequently than our test when γ is not very large. Hence, this lack of power in rejecting the null of the absence of nonlinear components directly translates into a lack of power for the unit root test. Our test being more powerful also ensures a unit root test with higher power.

4.4 Robustness of Fourier tests against various trend functions

The main reason to adopt a Fourier expansion to construct the test is that, as stated in Gallant (1981), it can approximate a wide class of nonlinear models. In this section, we assess whether our test maintains good power when the trend is generated by various types of nonlinear models and whether the unit root test constructed using the fitted Fourier expansion maintains good size and power. We consider the class of models $y_t = \eta d_t + u_t$, where d_t is a nonlinear deterministic trend component defined below and the error term is given by $u_t = \alpha u_{t-1} + e_t$ with $e_t \sim i.i.d.N(0, 1)$ for $\alpha = 1, 0.95, 0.9$ and 0.8. To make the comparison among cases with different values of α easier, we let the coefficient η depend on α by setting $\eta = \eta_0/\sqrt{1-\alpha^2}$ with $\eta_0 = 0, 1, 2, 3, 4$ and 5 (when $\alpha = 1$, we set $\eta = \eta_0/\sqrt{1-0.95^2}$).

The following five nonlinear trend components d_t are considered in the experiment.[5]

- Model 1 (Single LSTAR break): $d_t = 3/[1 + \exp(0.05(t - 0.5T))]$
- Model 2 (Single ESTAR break): $d_t = 3[1 - \exp(-0.0002(t - 0.75T)^2)]$
- Model 3 (Offsetting two LSTAR breaks): $d_t = 2+3/[1+\exp(0.05(t-0.2T))]-1.5/[1+\exp(0.05(t-0.75T))]$
- Model 4 (Reinforcing two LSTAR breaks): $d_t = 1.5/[1+\exp(0.15(t-0.2T))]+1.5/[1+\exp(0.15(t-0.75T))]$
- Model 5 (Two ESTAR breaks): $d(t) = 2 + 1.8[1 - \exp(-0.0003(t - 0.2T)^2)] - 1.5[1 - \exp(-0.0003(t - 0.75T)^2)]$

To assess the power of the tests, we compute the probability of selecting a nonlinear model ($n > 0$) using the sequential procedure described in Section 4.2 to determine the number of frequencies using a 5% significance level at each step. We only report results for the case when a trend term is included ($p_d = 1$). This is done for both our test and the ASW test of AHLT. The results with $\eta_0 = 1, 3$ and 5 are presented in Table 5, which shows the relative frequencies of choosing $n > 0$, namely, the probability of correctly detecting the existence of nonlinearity in the trend function. Overall, the results show that the FGLS test can detect nonlinearity with relatively high probability and that it is more powerful than the ASW test for all cases, except for Model 3 when $\alpha = 0.8, 0.9$ with a large value of η_0.

To assess whether the fitted nonlinear trend captures the main nonlinear features of the trend function and provides a good approximation, we evaluate the size and power of the unit root test of Enders and Lee (2012). Here the number of frequencies included in the Fourier expansion is based on the outcome of the sequential testing procedure. The results with $\eta_0 = 0, 1, 2, 3, 4$ and 5 are presented in Table 6. In all cases, the exact size is close to the 5% nominal size. Also, the power of the unit root test is higher when the trend is constructed using our sequential testing procedure compared to when it is constructed

[5] The models are identical to those considered in Jones and Enders (2014) except for Model 4 where we replace one of the parameter values (0.15) in the logistic function since their original choice of parameter value makes the model nearly linear so that both the ASW and FGLS tests fail to detect nonlinearity.

using the ASW test of AHLT. Though the power can decrease as the magnitude of the nonlinearities increase (i.e., as η_0 increases), it remains reasonably high. This indicates that our fitted Fourier expansion provides a good approximation to various forms of nonlinear processes. This is so because otherwise when the trend is substantially misspecified the unit root test would have very little power (e.g., Campbell and Perron, 1991).

5 Empirical applications

To illustrate the usefulness of our test procedure and method, we consider the trend function of global and hemispheric temperature series. The data series used are from the HadCRUT3 database (http://www.metoffice.gov.uk/hadobs/hadcrut3/) and cover the period 1850-2010 with annual observations. Three series are considered: global, Northern Hemisphere (NH) and Southern Hemisphere (SH). These are the same data used by Estrada, Perron and Martínez-López (2013) (henceforth EPM), which is the motivation for the analysis to be presented (see also, Estrada, Perron, Gay-García and Martínez-López, 2013). Based on various statistical methods, they documented that anthropogenic factors were responsible for the following features in temperature series: a marked increase in the growth rates of both temperatures and radiative forcing occurring near 1960, marking the start of sustained global warming; the impact of the Montreal Protocol (in reducing the emission of chlorofluorocarbons, CFC) and a reduction in methane emissions contributed to the recent so-called hiatus in the growth of temperatures since the mid-90s; the two World Wars and the Great Crash contributed to the mid-20th century cooling via important reductions in CO2 emissions. While the presence of the break in the slope of the trend in temperatures is well established using the test of Perron and Yabu (2009b), the statistical evidence about the two slowdowns or hiatus periods has not been statistically documented though they are well recognized in the climate change literature; see Maher, Gupta and England (2014). Our aim is to see if our method can detect the change in growth following 1960 and the two nonlinearities taking the form of a slowdown in growth during the 40s-mid-50s and the post mid-90s.

It is well known in the climate change literature that the Atlantic Multidecadal Oscillation (AMO) represents ocean-atmosphere processes naturally occurring in the North Atlantic with a large influence over NH and global climates. It produces 60- to 90-years natural oscillations that distort the warming trend suggesting it should be filtered before attempting to model the trend. Consequently, following EPM, we remove the low frequency natural component of the AMO from the NH and global temperature series in order to obtain a better measure of the low frequency trend, i.e., to isolate the trend in climate. The AMO series (1856-2010) was obtained from the National Oceanic and Atmospheric Administration (http://www.esrl.noaa.gov/). As discussed in EPM, applying standard unit root tests lead to a non-rejection of the unit root. This could be due to a genuine nonlinear trend, which biases the unit root tests towards non-rejections, or to a genuine I(1) noise component. Hence, it is important to allow for both I(0) and I(1) noise when testing for the presence of nonlinear components in the trend. We applied both the ASW-based and our testing procedures. We first used the sequential procedure described in Section 4.2 to determine the number of frequencies. The results are presented in Table 7. Our method selects the first three frequencies as being significant, while the ASW-based method fails to find any nonlinearities. The parameter estimates are presented in Table 8. Using the fitted nonlinear trend function from our procedure, the Enders and Lee (2012) unit root test (presented in Table 7) show that the remaining noise is deemed stationary at the 1% significance level.

The fitted trend functions are presented in Figure 7. The slowdown in the 40s-mid-50s and the marked increase in the growth rate after 1960 are clearly present in all series. However, the hiatus after the mid-90s is present only in the global and SH series. This is consistent with the argument in EPM that the reduction in the emissions of CFC was a major factor for the slowdown in global temperatures. As argued by Previdi and Polvani (2014), the ozone recovery (due to the reduction in the emissions of CFC) has been instrumental in driving SH climate by altering the tropospheric midlatitude jet. Hence, our fitted nonlinear trends are consistent with the main features of the climate trend since the early 20th century.

6 Conclusions

This paper proposes a new test for the presence of nonlinear deterministic trends approximated by Fourier expansions in a univariate time series without any prior knowledge as to whether the noise component is stationary or contains an autoregressive unit root. Our approach builds on the work of Perron and Yabu (2009a) and is based on a FGLS procedure that uses a super-efficient estimator of the sum of the autoregressive coefficients α when $\alpha = 1$. The resulting Wald test statistic asymptotically follows a chi-square limit distribution in both the $I(0)$ and $I(1)$ cases. To improve the finite sample properties of the tests, we use a bias corrected version of the OLS estimator of α proposed by Roy and Fuller (2001). We show that our procedure is substantially more powerful than currently available alternatives. An empirical application to global and hemispheric temperatures series shows the usefulness of our proposed method and offers additional insights into the differences in climate change in the Northern and Southern hemispheres.

References

Amemiya, T. (1985). *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts.

Andrews, D. W. K. (1991). 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica*, Vol. 59, pp. 817-858.

Astill, S., Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2015). 'Robust and powerful tests for nonlinear deterministic components', *Oxford Bulletin of Economics and Statistics*, Vol. 77, pp. 780-799.

Becker, R., Enders, W. and Hurn, S. (2004). 'A general test for time dependence in parameters', *Journal of Applied Econometrics*, Vol. 19, pp. 899–906.

Becker, R., Enders, W. and Lee, J. (2006). 'A stationarity test in the presence of an unknown number of smooth breaks', *Journal of Time Series Analysis*, Vol. 27, pp. 381–409.

Berk, K. N. (1974). 'Consistent autoregressive spectral estimates', *Annals of Statistics*, Vol. 2, pp. 489-502.

Campbell, J. Y. and Perron, P. (1991). 'Pitfalls and opportunities: what macroeconomists should know about unit roots', in *NBER Macroeconomics Annual 1991*, Vol. 6, pp. 141-220, MIT Press.

Canjels, E. and Watson, M. W. (1997). 'Estimating deterministic trends in the presence of serially correlated errors', *Review of Economics and Statistics*, Vol. 79, pp. 184–200.

Cochrane, D., and Orcutt, G. H. (1949). 'Application of least squares regression to relationships containing auto-correlated error terms', *Journal of the American Statistical Association*, Vol. 44, pp. 32-61.

Enders, W. and Lee, J. (2012). 'A unit root test using a Fourier series to approximate smooth breaks', *Oxford Bulletin of Economics and Statistics*, Vol. 74, pp. 574–599.

Estrada, F., Perron, P., Gay-García, C., and Martínez-López, B. (2013). 'A time-series analysis of the 20th century climate simulations produced for the IPCC's fourth assessment report', PloS one, Vol. 8(3), e60017.

Estrada, F., Perron, P. and Martínez-López, B. (2013). 'Statistically derived contributions of diverse human influences to twentieth-century temperature changes', *Nature Geoscience*, Vol. 6, pp. 1050-1055.

Gallant, A. R. (1981). 'On the bias in flexible functional forms and an essentially unbiased form: the Fourier flexible form', *Journal of Econometrics*, Vol. 15, pp. 211–245.

Gallant, A. R. and Souza, G. (1991). 'On the asymptotic normality of Fourier flexible form estimates', *Journal of Econometrics*, Vol. 50, pp. 329–353.

Harvey, D. I., Leybourne, S. J. and Xiao, L. (2010). 'Testing for nonlinear deterministic components when the order of integration is unknown', *Journal of Time Series Analysis*, Vol. 31, pp. 379–391.

Jones, P., and Enders, W. (2014). 'On the use of the flexible Fourier form in unit root tests, endogenous breaks, and parameter instability', in *Recent Advances in Estimating Nonlinear Models*, eds. J. Ma and M. Wohar. New York: Springer, pp 59-83.

Maher, N., Gupta, A. S. and England, M. H. (2014). 'Drivers of decadal hiatus periods in the 20th and 21st centuries', *Geophysical Research Letters*, Vol. 41, pp. 5978-5986.

Ng, S., and Perron, P. (1995). 'Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag', *Journal of the American Statistical Association*, Vol. 90, pp. 268-281.

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

Perron, P. (1988). 'Trends and random walks in macroeconomic time series: Further evidence from a new approach', *Journal of Economic Dynamics and Control*, Vol. 12, pp. 297-332.

Perron, P. (1989). 'The great crash, the oil price shock, and the unit root hypothesis', *Econometrica*, Vol. 57, pp. 1361-1401.

Perron, P. (1990). 'Testing for a unit root in a time series with a changing mean', *Journal of Business and Economic Statistics*, Vol. 8, pp. 153-162.

Perron, P. and Yabu, T. (2009a). 'Estimating deterministic trends with an integrated or stationary noise component', *Journal of Econometrics*, Vol. 151, pp. 56-69.

Perron, P. and Yabu, T. (2009b). 'Testing for shifts in trend with an integrated or stationary noise component', *Journal of Business and Economic Statistics*, Vol. 27, 369-396.

Perron, P. and Yabu, T. (2012). 'Testing for trend in the presence of autoregressive error: A comment', *Journal of the American Statistical Association*, Vol. 107, p. 844.

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

Previdi, M. and Polvani, L.M. (2014). 'Climate system response to stratospheric ozone depletion and recovery', *Quarterly Journal of the Royal Meteorological Society, Vol.* 140, pp. 2401-2419.

Rodrigues, P. M. M. and Taylor, A. M. R. (2012). 'The flexible Fourier form and local generalised least squares de-trended unit root tests', *Oxford Bulletin of Economics and Statistics*, Vol. 74, pp. 736–759.

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

Roy, A., and Fuller, W. A. (2001). 'Estimation for autoregressive time series with a root near 1', *Journal of Business and Economic Statistics*, Vol. 19, pp. 482-493.

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

Table 1: Values of $\tau_{.50}$ and $\tau_{.85}$.

	$p_d = 0$		$p_d = 1$	
	$\tau_{.50}$	$\tau_{.85}$	$\tau_{.50}$	$\tau_{.85}$
	Single Frequency			
$k = 1$	-2.39	-3.26	-3.09	-3.83
2	-1.71	-2.67	-2.56	-3.45
3	-1.63	-2.51	-2.33	-3.21
4	-1.60	-2.45	-2.27	-3.09
5	-1.59	-2.43	-2.23	-3.05
	Multiple Frequencies			
$n = 1$	-2.39	-3.26	-3.09	-3.83
2	-2.99	-3.93	-3.79	-4.50
3	-3.51	-4.49	-4.40	-5.10
4	-3.98	-4.99	-4.92	-5.64
5	-4.36	-5.44	-5.41	-6.11

Table 2.a: Finite Sample Null Rejection Probabilities of the ASW and FGLS Tests; $\mathbf{p}_d = 0$; 5% Nominal Size.

		ASW			*FGLS*					
					Median Unbiased			Upper Biased		
ϕ	θ	$T = 150$	300	600	$T = 150$	300	600	$T = 150$	300	600
1.00	-0.80	0.139	0.104	0.074	0.197	0.197	0.186	0.079	0.071	0.068
	-0.40	0.067	0.054	0.048	0.149	0.111	0.083	0.066	0.051	0.049
	0.00	0.062	0.051	0.048	0.121	0.082	0.060	0.080	0.065	0.056
	0.40	0.057	0.048	0.045	0.124	0.080	0.065	0.103	0.076	0.064
	0.80	0.055	0.054	0.048	0.120	0.087	0.065	0.111	0.085	0.065
0.95	-0.80	0.139	0.073	0.036	0.165	0.110	0.076	0.080	0.059	0.053
	-0.40	0.063	0.022	0.008	0.139	0.087	0.058	0.057	0.042	0.037
	0.00	0.045	0.016	0.007	0.107	0.072	0.060	0.041	0.036	0.040
	0.40	0.042	0.016	0.007	0.073	0.048	0.039	0.024	0.023	0.021
	0.80	0.041	0.016	0.006	0.037	0.022	0.022	0.013	0.009	0.012
0.90	-0.80	0.082	0.059	0.042	0.095	0.075	0.061	0.053	0.050	0.052
	-0.40	0.036	0.015	0.009	0.091	0.062	0.041	0.043	0.038	0.033
	0.00	0.025	0.010	0.005	0.080	0.050	0.043	0.038	0.033	0.035
	0.40	0.024	0.009	0.005	0.066	0.056	0.040	0.022	0.033	0.032
	0.80	0.021	0.009	0.005	0.048	0.048	0.045	0.012	0.021	0.032
0.80	-0.80	0.054	0.052	0.054	0.056	0.059	0.056	0.032	0.044	0.050
	-0.40	0.029	0.021	0.018	0.062	0.044	0.047	0.037	0.031	0.041
	0.00	0.017	0.012	0.009	0.054	0.037	0.036	0.033	0.028	0.031
	0.40	0.013	0.009	0.008	0.031	0.030	0.034	0.010	0.020	0.030
	0.80	0.015	0.009	0.007	0.044	0.033	0.032	0.014	0.020	0.028

Note: *ASW* denotes the test of Astill et al. (2015); *FGLS* (Median Unbiased) is the $W_{\widehat{\gamma}}$ test with $\tau_{0.5}$; *FGLS* (Upper Biased) is the $W_{\widehat{\gamma}}$ test with $\tau_{0.85}$. The data are generated by: $y_t = u_t = \phi u_{t-1} + e_t + \theta e_{t-1}$.

Table 2.b: Finite Sample Null Rejection Probabilities of the ASW and FGLS Tests; $p_d = 1$; 5% Nominal Size.

		ASW			*FGLS*					
					Median Unbiased			Upper Biased		
ϕ	θ	$T = 150$	300	600	$T = 150$	300	600	$T = 150$	300	600
1.00	-0.80	0.173	0.137	0.097	0.191	0.162	0.168	0.115	0.063	0.061
	-0.40	0.073	0.054	0.051	0.161	0.137	0.107	0.074	0.058	0.052
	0.00	0.061	0.049	0.044	0.145	0.106	0.071	0.087	0.077	0.062
	0.40	0.055	0.041	0.040	0.130	0.082	0.064	0.100	0.073	0.063
	0.80	0.060	0.051	0.042	0.122	0.089	0.065	0.112	0.085	0.065
0.95	-0.80	0.078	0.038	0.015	0.117	0.072	0.051	0.078	0.039	0.030
	-0.40	0.025	0.006	0.001	0.095	0.064	0.042	0.041	0.030	0.025
	0.00	0.015	0.003	0.001	0.078	0.051	0.044	0.027	0.022	0.027
	0.40	0.015	0.003	0.001	0.051	0.039	0.029	0.016	0.015	0.017
	0.80	0.014	0.002	0.001	0.025	0.012	0.018	0.008	0.004	0.010
0.90	-0.80	0.050	0.032	0.019	0.078	0.049	0.050	0.064	0.034	0.037
	-0.40	0.011	0.005	0.002	0.066	0.041	0.035	0.036	0.026	0.027
	0.00	0.007	0.002	0.002	0.066	0.048	0.031	0.031	0.031	0.026
	0.40	0.004	0.001	0.000	0.033	0.045	0.035	0.008	0.023	0.027
	0.80	0.005	0.002	0.000	0.029	0.029	0.033	0.006	0.010	0.020
0.80	-0.80	0.026	0.031	0.031	0.036	0.040	0.042	0.024	0.030	0.038
	-0.40	0.009	0.004	0.006	0.053	0.031	0.035	0.036	0.024	0.030
	0.00	0.003	0.002	0.000	0.049	0.033	0.036	0.034	0.027	0.033
	0.40	0.002	0.000	0.001	0.014	0.024	0.028	0.005	0.017	0.025
	0.80	0.003	0.001	0.001	0.026	0.024	0.025	0.009	0.014	0.022

Note: *ASW* denotes the test of Astill et al. (2015); *FGLS* (Median Unbiased) is the $W_{\widehat{\gamma}}$ test with $\tau_{0.5}$; *FGLS* (Upper Biased) is the $W_{\widehat{\gamma}}$ test with $\tau_{0.85}$. The data are generated by: $y_t = u_t = \phi u_{t-1} + e_t + \theta e_{t-1}$.

Table 3: Number of Frequencies Selected by the ASW and FGLS tests; $p_d = 1$; T=150.

		ASW				*FGLS* (UB)							
						sig5				sig1			
α	γ	n=0	n=1	n=2	n=3	n=0	n=1	n=2	n=3	n=0	n=1	n=2	n=3
1.00	0	0.864	0.058	0.042	0.036	0.722	0.082	0.093	0.103	0.869	0.045	0.042	0.044
	1	0.870	0.043	0.050	0.037	0.630	0.076	0.181	0.113	0.826	0.034	0.092	0.048
	2	0.841	0.016	0.091	0.052	0.371	0.067	0.459	0.103	0.626	0.037	0.293	0.044
	3	0.780	0.006	0.141	0.074	0.118	0.039	0.742	0.102	0.302	0.033	0.620	0.045
	4	0.677	0.000	0.215	0.108	0.016	0.009	0.873	0.103	0.072	0.017	0.866	0.045
	5	0.555	0.000	0.313	0.132	0.000	0.001	0.895	0.103	0.005	0.004	0.948	0.044
0.95	0	0.970	0.015	0.008	0.006	0.837	0.029	0.054	0.081	0.918	0.020	0.028	0.035
	1	0.957	0.017	0.017	0.009	0.765	0.025	0.134	0.076	0.885	0.018	0.064	0.033
	2	0.936	0.008	0.037	0.020	0.485	0.018	0.419	0.079	0.726	0.007	0.236	0.031
	3	0.858	0.001	0.093	0.048	0.134	0.016	0.771	0.079	0.345	0.005	0.618	0.032
	4	0.745	0.000	0.175	0.081	0.012	0.006	0.901	0.082	0.071	0.004	0.893	0.032
	5	0.588	0.000	0.283	0.129	0.000	0.000	0.919	0.080	0.005	0.001	0.962	0.032
0.90	0	0.991	0.007	0.002	0.000	0.885	0.030	0.038	0.048	0.942	0.017	0.020	0.021
	1	0.984	0.008	0.006	0.003	0.811	0.024	0.109	0.056	0.899	0.016	0.060	0.025
	2	0.957	0.002	0.031	0.010	0.532	0.004	0.415	0.049	0.753	0.003	0.221	0.023
	3	0.864	0.000	0.097	0.039	0.113	0.001	0.832	0.054	0.342	0.000	0.634	0.024
	4	0.703	0.000	0.221	0.077	0.005	0.000	0.941	0.054	0.042	0.000	0.934	0.023
	5	0.512	0.000	0.368	0.120	0.000	0.000	0.944	0.056	0.001	0.000	0.973	0.026
0.80	0	0.994	0.006	0.000	0.000	0.890	0.030	0.035	0.044	0.952	0.011	0.016	0.021
	1	0.992	0.002	0.006	0.000	0.662	0.031	0.257	0.050	0.793	0.014	0.172	0.020
	2	0.928	0.000	0.061	0.011	0.350	0.001	0.602	0.047	0.502	0.001	0.478	0.021
	3	0.713	0.000	0.243	0.044	0.036	0.000	0.918	0.046	0.192	0.000	0.788	0.020
	4	0.424	0.000	0.490	0.087	0.000	0.000	0.955	0.045	0.004	0.000	0.977	0.019
	5	0.193	0.000	0.710	0.098	0.000	0.000	0.957	0.044	0.000	0.000	0.980	0.020

Note: *ASW* denotes the test of Astill et al. (2015); *FGLS* (UB) (sig5), resp. *FGLS* (UB) (sig1), are the $W_{\widehat{\gamma}}$ test with $\tau_{0.85}$ and a 5%, resp. 1%, test for the sequential procedure to select the number of frequencies.The data are generated by: $y_t = \gamma(\sum_{k=1}^{2}\sin(2\pi kt/T) + \sum_{k=1}^{2}\cos(2\pi kt/T)) + u_t$, $u_t = \alpha u_{t-1} + e_t$.

Table 4: Exact Size of the Enders and Lee (2012) Unit Root Test with Sequential Frequency Selections; $p_d = 1$; 5% Nominal Size.

		Fixed n		*ASW*	*FGLS (UB)*	
	γ	n=0	n=2		sig5	sig1
T=150	0	0.048	0.049	0.093	0.118	0.103
	1	0.029	0.048	0.065	0.095	0.083
	2	0.006	0.048	0.044	0.071	0.061
	3	0.001	0.048	0.045	0.069	0.063
	4	0.000	0.052	0.058	0.073	0.068
	5	0.000	0.052	0.066	0.072	0.068
T=300	0	0.047	0.047	0.080	0.094	0.083
	1	0.041	0.048	0.071	0.089	0.077
	2	0.021	0.047	0.051	0.076	0.064
	3	0.004	0.042	0.033	0.058	0.047
	4	0.001	0.052	0.043	0.067	0.059
	5	0.000	0.051	0.048	0.068	0.061
T=600	0	0.048	0.053	0.078	0.082	0.072
	1	0.046	0.053	0.074	0.079	0.070
	2	0.026	0.049	0.052	0.061	0.049
	3	0.017	0.053	0.042	0.055	0.043
	4	0.007	0.055	0.036	0.052	0.039
	5	0.002	0.053	0.032	0.055	0.040

Note: *ASW* denotes the test of Astill et al. (2015); *FGLS* (*UB*) (sig5), resp. *FGLS* (*UB*) (sig1), are the $W_{\widehat{\gamma}}$ test with $\tau_{0.85}$ and a 5%, resp. 1%, test for the sequential procedure to select the number of frequencies. The data are generated by: $y_t = \gamma(\sum_{k=1}^{2} \sin(2\pi kt/T) + \sum_{k=1}^{2} \cos(2\pi kt/T)) + u_t$, $u_t = u_{t-1} + e_t$.

Table 5. Probability of Selecting a Nonlinear Model (n > 0) when the Data are Generated by Different Types of Nonlinear Models; $p_d = 1$; T=150.

		ASW				*FGLS* $(UB, sig5)$			
	η_0	$\alpha = 0.80$	0.90	0.95	1.0	$\alpha = 0.80$	0.90	0.95	1.0
Model 1	1	0.005	0.012	0.035	0.128	0.138	0.128	0.169	0.281
	3	0.035	0.038	0.089	0.179	0.320	0.187	0.261	0.417
	5	0.129	0.112	0.197	0.283	0.543	0.283	0.554	0.643
Model 2	1	0.016	0.021	0.058	0.147	0.215	0.153	0.184	0.316
	3	0.175	0.161	0.278	0.322	0.518	0.285	0.598	0.678
	5	0.438	0.430	0.584	0.558	0.516	0.822	0.988	0.961
Model 3	1	0.056	0.058	0.113	0.204	0.256	0.166	0.186	0.318
	3	0.599	0.526	0.674	0.608	0.576	0.255	0.573	0.657
	5	0.938	0.909	0.963	0.912	0.563	0.824	0.994	0.954
Model 4	1	0.005	0.011	0.034	0.129	0.147	0.127	0.186	0.308
	3	0.009	0.020	0.063	0.156	0.281	0.247	0.498	0.575
	5	0.041	0.057	0.140	0.211	0.435	0.588	0.898	0.901
Model 5	1	0.013	0.017	0.048	0.146	0.207	0.154	0.195	0.340
	3	0.174	0.142	0.250	0.326	0.609	0.411	0.789	0.792
	5	0.546	0.467	0.647	0.632	0.844	0.976	1.000	0.992

Note: *ASW* denotes the test of Astill et al. (2015); $FGLS(UB, sig5)$ is the W_γ test with $\tau_{0.85}$ and a 5% test for the sequential procedure to select the number of frequencies.The models are described in Section 4.4.

Table 6. Exact Size and Power Enders and Lee's (2012) Unit Root Test when the Data are Generated by Different Types of Nonlinear Models; $p_d = 1$; T = 150.

		ASW					*FGLS* $(UB, sig5)$				
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
$\alpha = 1.0$	$\eta_0 = 0$	0.090	0.091	0.084	0.080	0.086	0.115	0.113	0.114	0.107	0.109
	$\eta_0 = 1$	0.065	0.076	0.083	0.077	0.073	0.085	0.101	0.103	0.101	0.097
	$\eta_0 = 2$	0.054	0.064	0.078	0.059	0.069	0.063	0.087	0.087	0.082	0.091
	$\eta_0 = 3$	0.039	0.063	0.080	0.041	0.075	0.041	0.074	0.082	0.059	0.087
	$\eta_0 = 4$	0.035	0.059	0.071	0.037	0.087	0.037	0.064	0.074	0.051	0.094
	$\eta_0 = 5$	0.023	0.055	0.067	0.034	0.086	0.025	0.057	0.074	0.039	0.091
$\alpha = 0.8$	$\eta_0 = 0$	0.973	0.972	0.969	0.974	0.970	0.973	0.972	0.969	0.974	0.972
	$\eta_0 = 1$	0.765	0.874	0.902	0.956	0.809	0.822	0.904	0.916	0.958	0.855
	$\eta_0 = 2$	0.179	0.469	0.579	0.902	0.302	0.422	0.668	0.678	0.912	0.602
	$\eta_0 = 3$	0.134	0.221	0.554	0.736	0.174	0.305	0.515	0.555	0.788	0.582
	$\eta_0 = 4$	0.208	0.284	0.677	0.456	0.330	0.268	0.459	0.535	0.589	0.625
	$\eta_0 = 5$	0.256	0.377	0.690	0.204	0.509	0.244	0.441	0.514	0.425	0.721

Note: *ASW* denotes the test of Astill et al. (2015); $FGLS(UB, sig5)$ is the W_γ test with $\tau_{0.85}$ and a 5% test for the sequential procedure to select the number of frequencies.The models are described in Section 4.4.

Table 7: Empirical Applications to Temperature Series.

	ASW		*FGLS* (UB)			
			sig5		sig1	
	$\widehat{n}$	*LM*	$\widehat{n}$	*LM*	$\widehat{n}$	*LM*
Global	0	-2.039	3	-8.485***	3	-8.485***
Nothern Hemisphere	0	-2.271	3	-9.715***	3	-9.715***
Southern Hemisphere	0	-2.904*	3	-6.073***	3	-6.073***

Note: ***, **, and * denote a statistic significant at the 1%, 5%, and 10% level, respectively. LM is the unit root test of Enders and Lee (2012). $\widehat{n}$ is the number of frequencies estimated.

Table 8: Estimates of the Nonlinear Trend Functions.

	Global	Northern Hemisphere	Southern Hemisphere
	1856-2010	1856-2010	1850-2010
Constant	-0.436***	-0.509***	-0.577***
	(0.042)	(0.042)	(0.054)
Trend	0.006***	0.007***	0.005***
	(0.001)	(0.001)	(0.001)
$\sin(2\pi t/T)$	0.082***	0.101***	0.075**
	(0.030)	(0.029)	(0.038)
$\cos(2\pi t/T)$	0.105***	0.081***	0.138***
	(0.015)	(0.014)	(0.019)
$\sin(4\pi t/T)$	0.006	0.016	0.029
	(0.020)	(0.019)	(0.025)
$\cos(4\pi t/T)$	0.006	0.030***	-0.001
	(0.015)	(0.014)	(0.019)
$\sin(6\pi t/T)$	0.013	0.022	-0.055***
	(0.017)	(0.017)	(0.022)
$\cos(6\pi t/T)$	-0.056***	-0.042***	-0.046
	(0.015)	(0.014)	(0.019)

Note: ***, **, and * denote a statistic significant at the 1%, 5%, and 10% level, respectively.

Figure 1. Local Asymptotic Power (p_d=0)

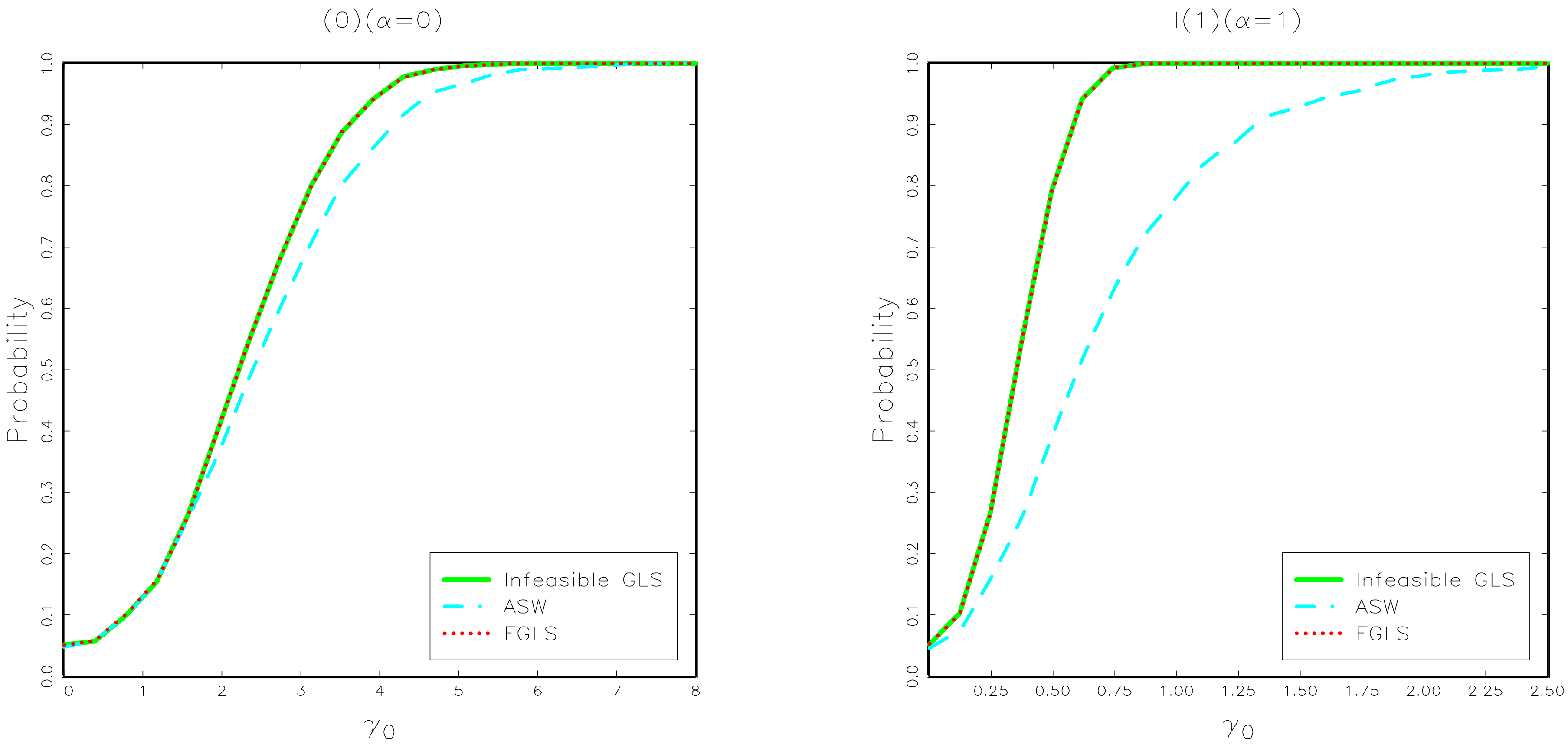

Figure 2a. Finite Sample Size of FGLS tests (p_d=0)

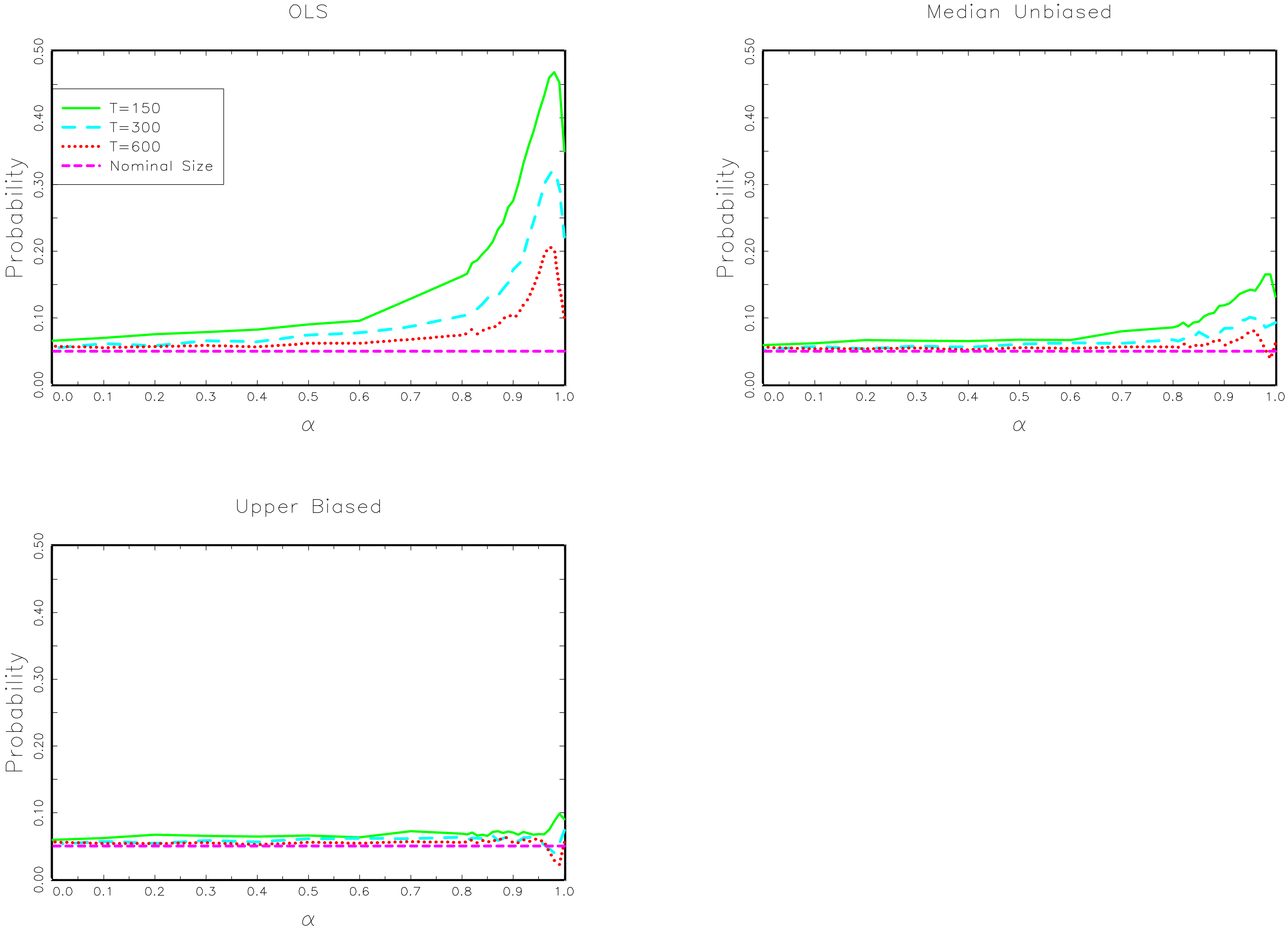

Figure 2b. Finite Sample Size of FGLS tests(p_d=1)

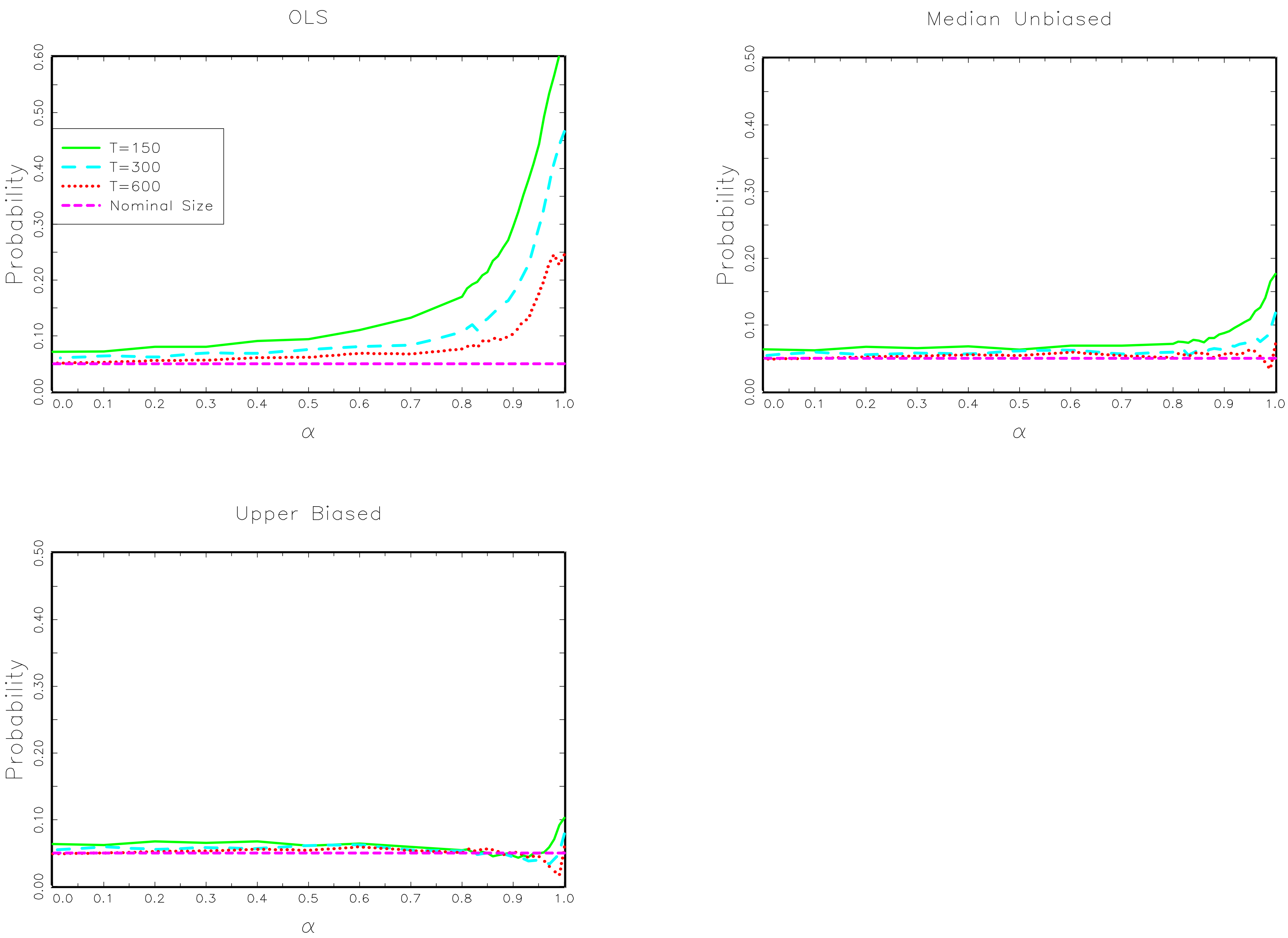

Figure 3a. Finite Sample Power Comparisons ($p_d=0$): T=150

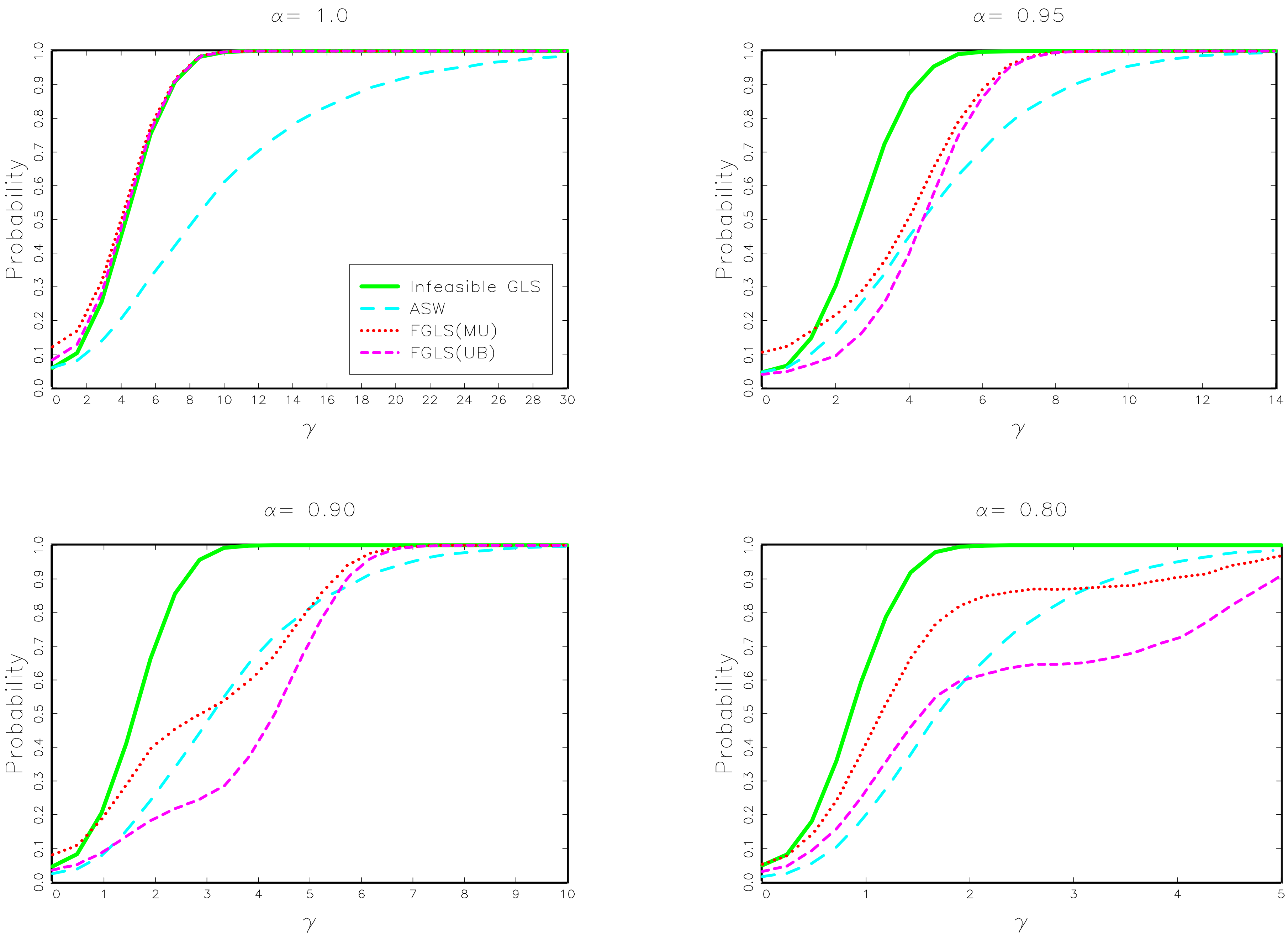

Figure 3b. Finite Sample Power Comparisons (p_d=0): T = 300

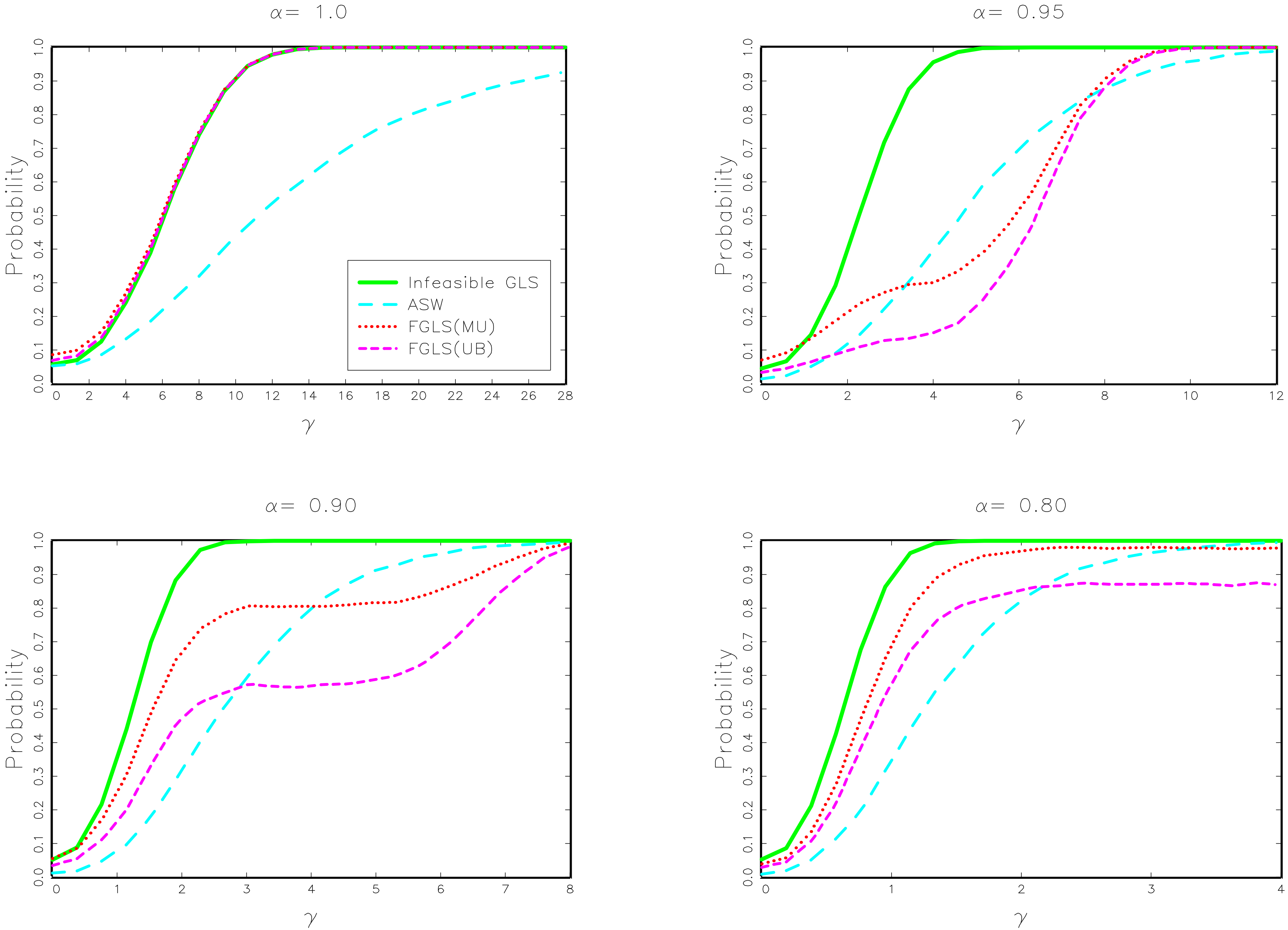

Figure 3c. Finite Sample Power Comparisons $(p_d = 0)$: T = 600

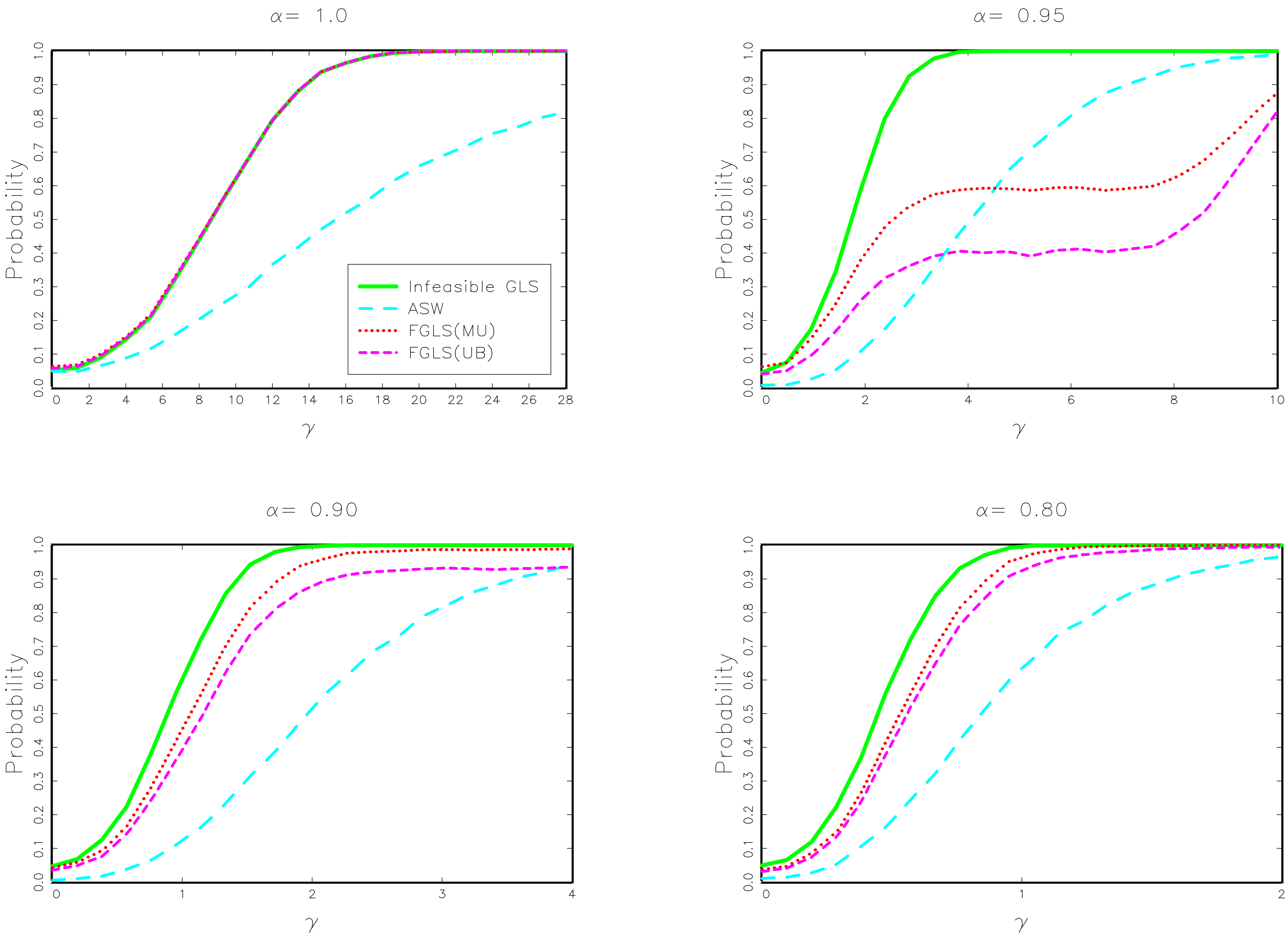

Figure 4a. Finite Sample Power Comparisons (p_d=1): T = 150

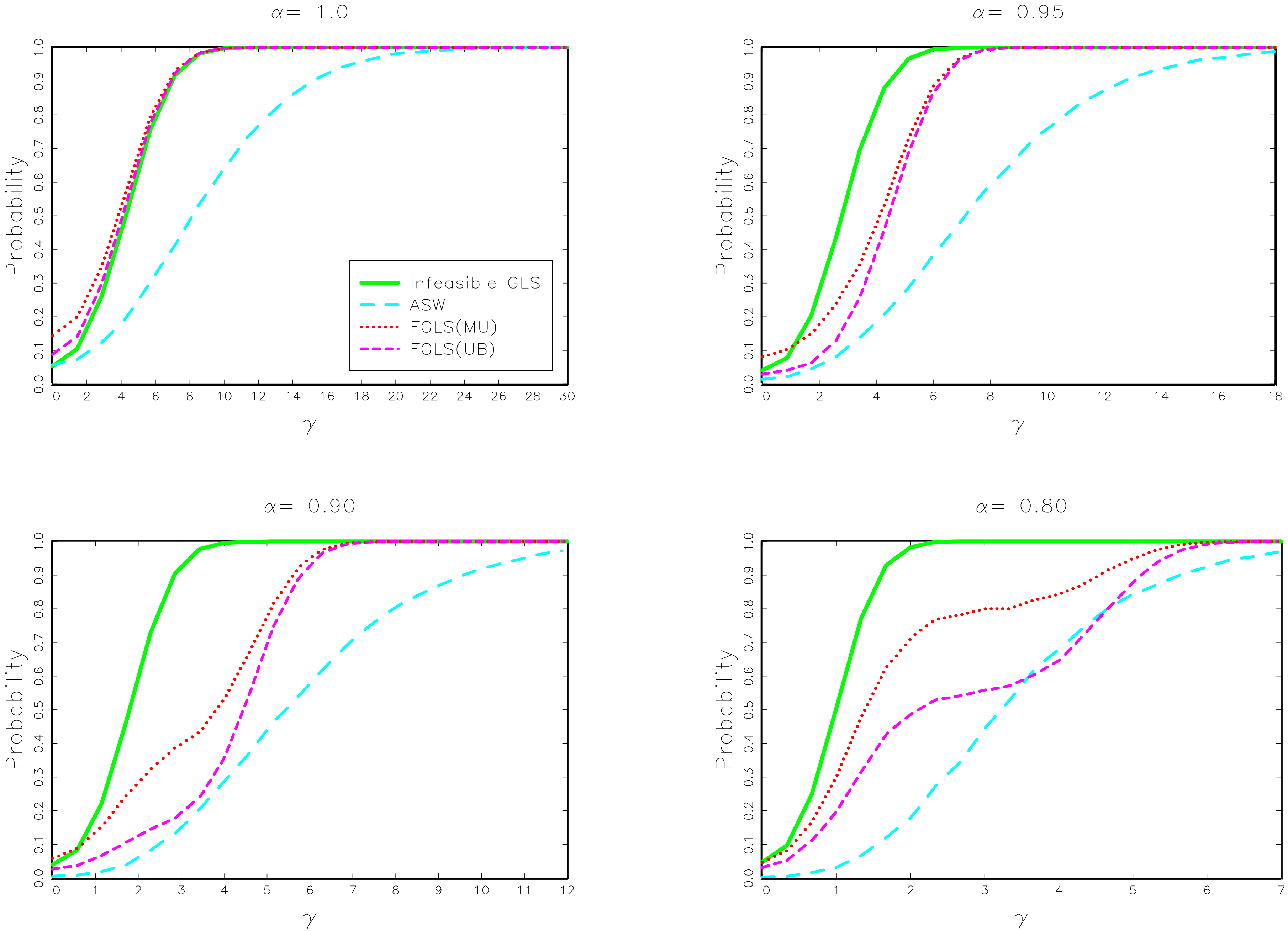

Figure 4b. Finite Sample Power Comparisons (p_d=1): T = 300

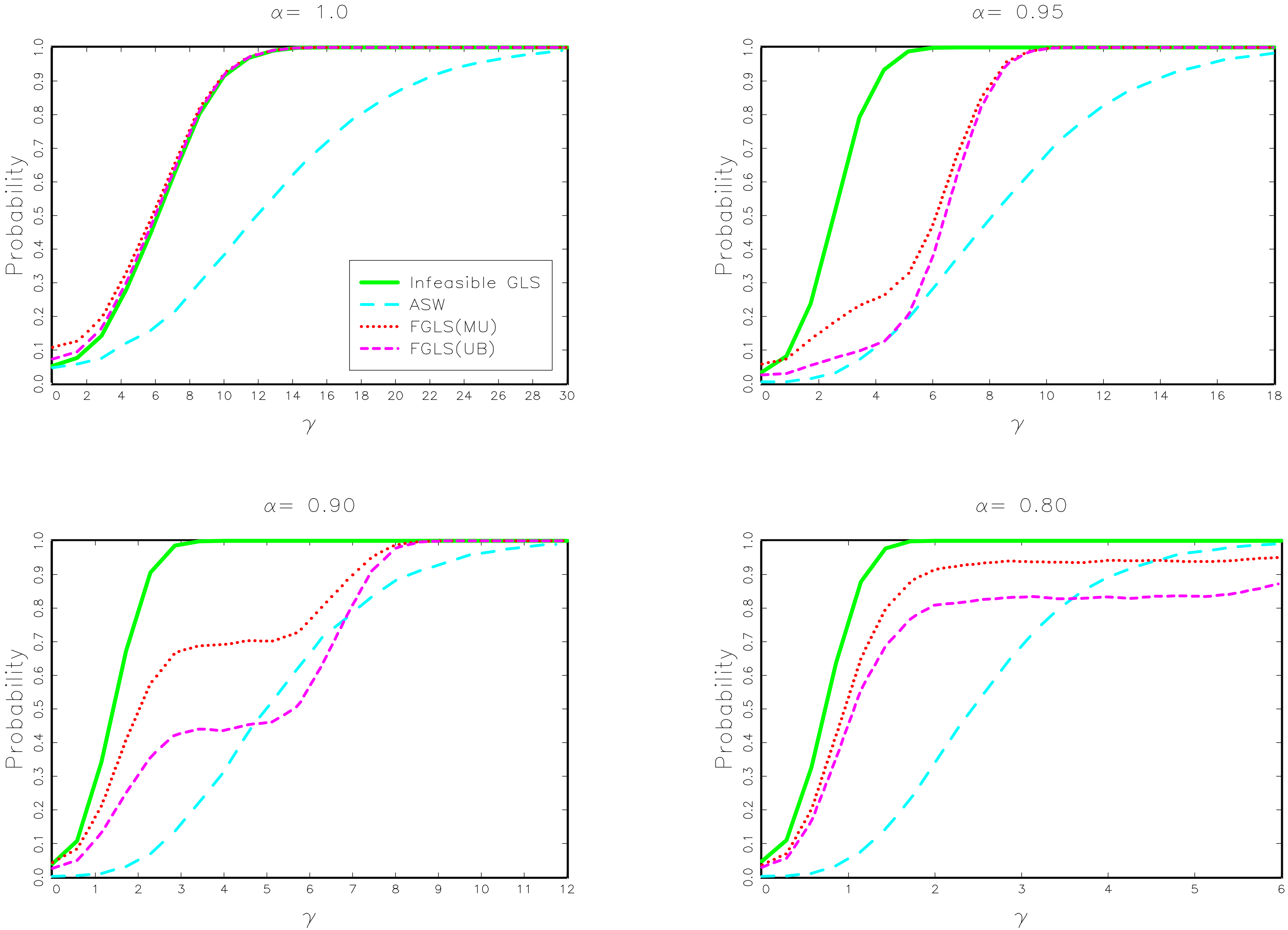

Figure 4c. Finite Sample Power Comparisons (p_d=1): T = 600

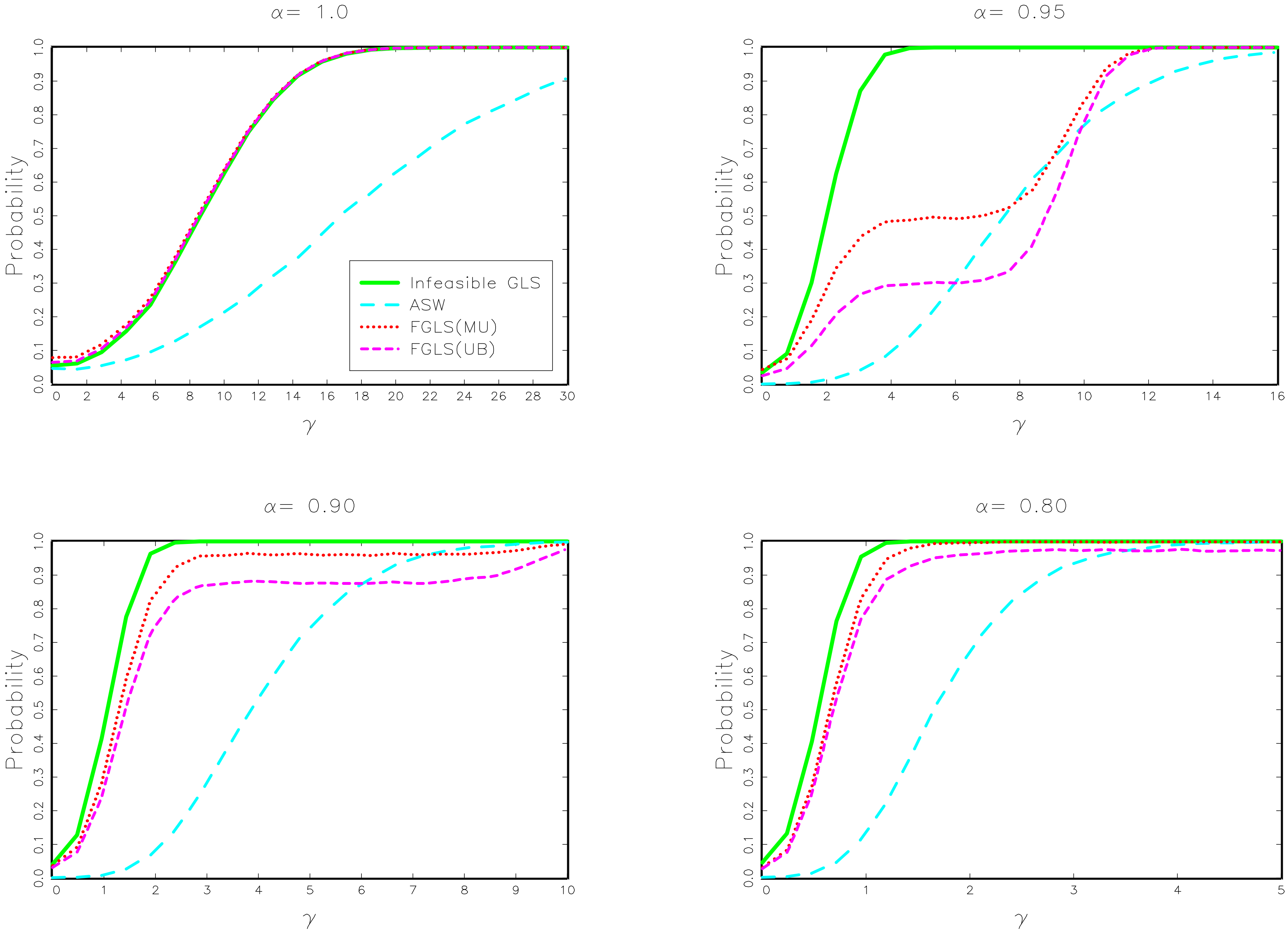

Figure 5a. Sequential Enders-Lee Unit Root Tests (p_d=1): α = 0.9, T=150

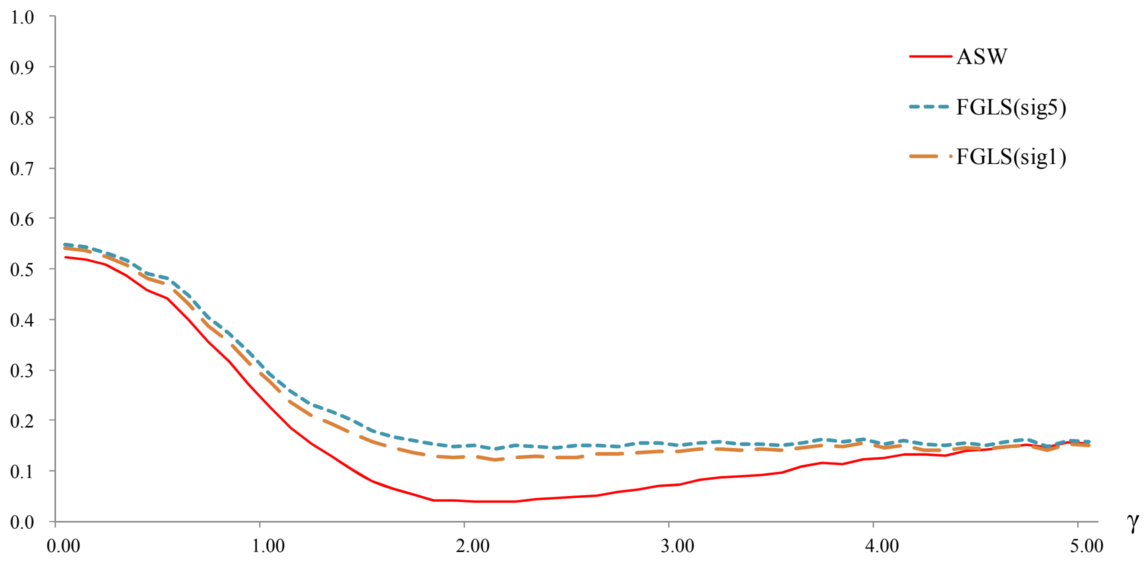

Figure 5b. Sequential Enders-Lee Unit Root Tests (p_d=1): α = 0.8, T=150

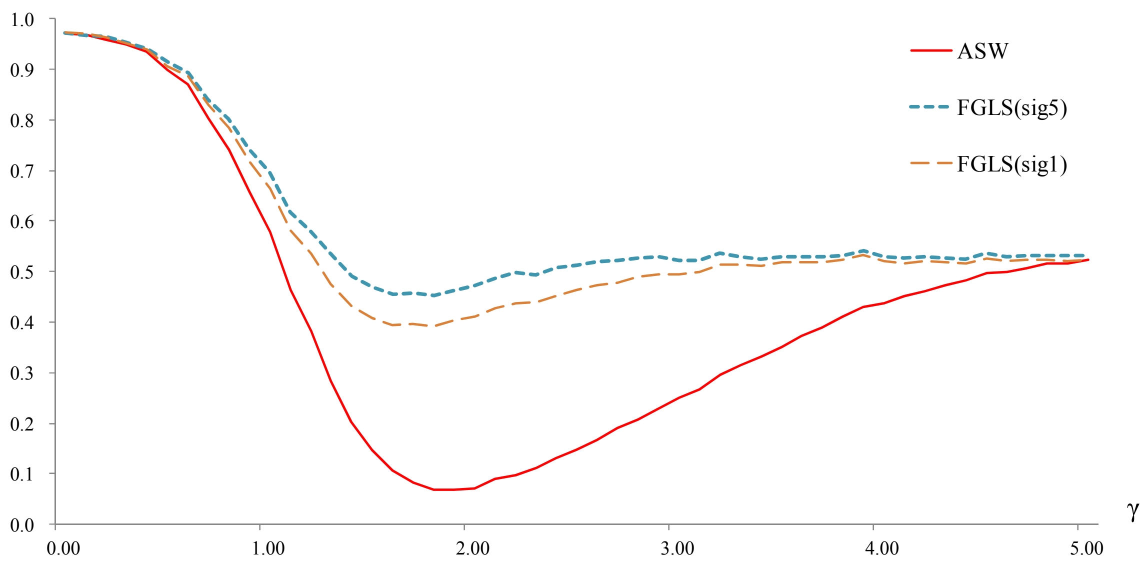

Figure 6a. Sequential Enders-Lee Unit Root Tests (p_d=1): α = 0.9; T=150; Fixed n

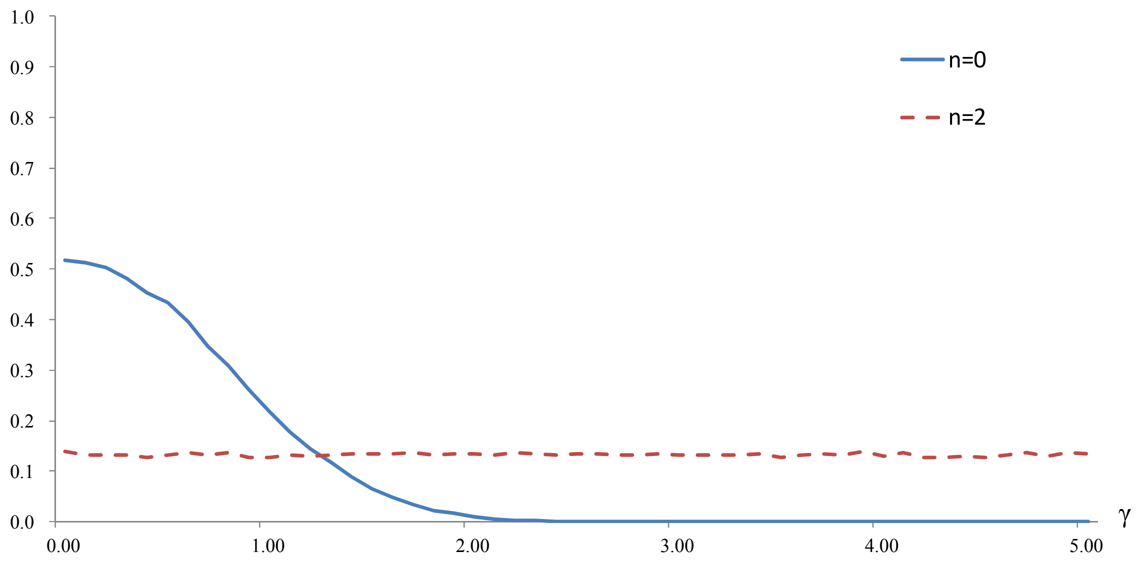

Figure 6b. Sequential Enders-Lee Unit Root Tests (p_d=1): α = 0.8; T=150; Fixed n

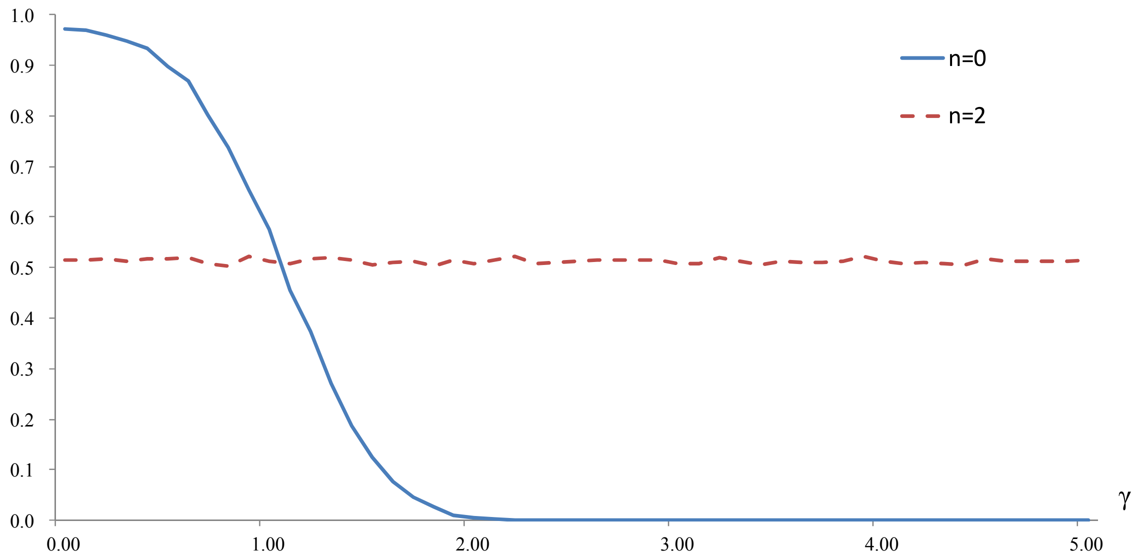

Figure 7. Temperature Series