A Wavelet Variance Ratio Test for Unit Roots in the Presence of GARCH(1,1) Errors

Ali, AA and Locking, H

Department of Economics and Statistics
Linnæus University, Sweden

Abstract

In this paper we test for autoregressive unit roots in univariate time series with generalized autoregressive conditionally heteroscedastic errors (GARCH(1,1) errors) using a spectral variance decomposition method based on the discrete wavelet transform. Results from using the wavelet variance ratio tests on filtered time series are compared to those from tests on the unfiltered series using four widely used unit root tests namely; the Augumented Dickey-Fuller (ADF) test, the Phillips-Perron (PP) test, and the two Elliot, Rothenberg and Stock (ERS) unit root tests i.e Piont-optimal and DF-GLS tests. Monte Carlo simulations show that, for finite sample sizes, the wavelet based methods show better performance than the four named tests in addressing the problem of over-sizing, and have size-adjusted power within the range of that of the other tests, in the presence of GARCH(1,1) errors.

Keywords: Time series, unit roots, variance ratio, wavelets, GARCH

1 Introduction

Many economic and financial time series have conditional variances that change over time. Unit root tests done on such time series are known to suffer from over-sizing and power deterioration. Kim and Schmidt (1993), Cook (2006) and Sjölander (2008) among others, show that the Dickey-Fuller tests (Dickey and Fuller; 1979) suffer from size distortion, particularly for finite sample sizes in the presence of autoregressive conditional heteroscedasticity (ARCH) Engle (1982) or Generalized ARCH (GARCH) Bollerslev (1986). While the asymptotic distributions of the Dickey-Fuller test statistics are unchanged by the presence of GARCH type errors (Pantula; 1988; Wang and Mao; 2008), over-sizing remains a problem in empirically relevant sample sizes, the extent of which depends on the strength of the (G)ARCH effects (Kim and Schmidt op. cit.). Thus, there is an opportunity to find methods that can reconcile the actual and nominal test sizes when using small sample sizes.

Several methods have been suggested to get around the problem of over-sizing in finite samples. However, most studies have not used the full range of sample sizes and GARCH processes that can realistically be encountered by the data analyst. Two exceptions are Cook (op. cit) who used 4 sample sizes and seven different GARCH(1,1) processes to compare the rejection frequencies of the original Dickey-Fuller test to 5 modified versions of the test, and Sjölander (op. cit.) who used 15 different GARCH (1, 1) processes and 5 sample sizes in a Monte Carlo study to compare his ADF-BEST test to 8 of the most commonly occurring unit root tests for size retention and power.

One approach to improve the size retention of the Dickey-Fuller test is to use the heteroscedisity consistent standard errors of White (1980), or one of its variations e.g. MacKinnon and White (1985), Davidson and MacKinnon (1993) and Cribari-Neto (2004). Heteroscedasicity consistent standard errors have been used by Andrews and Guggenberger (2011) to construct confidence intervals with asymptotically correct sizes for near unit root and unit root processes with GARCH errors. Demetrescu (2010) showed that while the small-sample distribution of the Augmented Dickey-Fuller and Dickey-Fuller-White (Dickey-Fuller with robust standard errors) react sensitively to the degree of persistence in the conditional variance, this is less the case with simple combinations of the two tests. Long and Ervin (2000) show that the Heteroscedasticity Consistent Covariance Matrix (HC3) estimator is superior to the commonly used HC0 when used for small sample sizes ($N < 250$) and Cribari-Neto (op. cit.) further modified the HC3 estimator and shows that the modification, the HC4 estimator, gives better results for small sample sizes.

An improvement on using robust standard errors alone has been to combine them with the Recursive Mean or trend adjustment of So and Shin (1999). Using the Dickey-Fuller test, the robust standard error induces a slight under-sizing while the recursive mean method tends to over-size. A wighted combination of the two Dickey-Fuller test statistics results in a test with empirical size closer to the nominal size can be obtained (Patterson; 2012, p. 516).

Unit root tests using the Maximum Likelihood (ML) estimation of the AR(p)-GARCH(p, q) model have been considered by Ling and Li (1998), Li et al. (2002), Ling and Li (2003) and Ling et al. (2003), among others. Seo

(1999) suggested a ML based test which estimates the autoregressive unit root and the GARCH parameters jointly. However, the ML based tests do not always perform better than the Dickey-Fuller tests. Charles and Darné (2008) show that under the conditions that $0.8 \leqslant (\alpha_1 + \gamma_1) < 1$ and $\gamma_1 > \alpha_1$ (α_1 and γ_1 are parameters of the GARCH(1,1) process and will be discussed later in the text), the empirical size and power properties of the ML test due to Seo (op. cit), for example, do not provide improvements on the standard Dickey-Fuller test statistics.

Sjölander (op. cit.) used a 3 step procedure to first estimate the GARCH risk under the uncertainty of mean equation stationarity in a reliable way, and then perform the unit root test using the Dickey-Fuller test. Monte Carlo simulations showed that the test introduced by Sjölander (op. cit.), called the ADF-BEST test, is unbiased for size and has better power than 8 of the most commonly used unit root tests in the presence of stationary GARCH(1,1) errors. The Monte Carlo study was extensive; using 15 GARCH(1,1) error processes, 12 different AR parameters for the mean equation, and 5 sample sizes ranging from 50 to 5000. The Design thus covered GARCH(1,1) processes that have very weak to very strong volatility and persistence in the conditional variance as well as near degenerate cases. The ADF-BEST unit root test, however, requires prior knowledge of the GARCH process i.e. it is a multi-step procedure that requires knowledge or estimation of GARCH risk.

Li and Shukur (2011) used the wavelet filter to remove the highest frequency components from the time series prior to applying the Dickey-Fuller test on the filtered data. For their method, the scaling coefficients from the wavelet analysis are tested for unit roots in place of the original series.

The asymptotic distribution of the Dickey-Fuller test statistic is shown to be the same as that of the Dickey-Fuller test on the original data with independently distributed errors. Monte Carlo simulations show that there is less size distortion using the wavelet improved method.

The test described in this paper filters the time series from the high frequency components using the wavelet low-pass filter in a way similar to that of Li and Shukur (op. cit.) i.e. the input series to the unit root test is the unit scale scaling coefficients obtained from the first level wavelet analysis. However, departing from the work cited earlier, which has mainly used the Dickey-Fuller test with modifications, the unit root test is done in the frequency domain using the wavelet variance ratio test introduced by Fan and Gençay (2010). Monte Carlo simulations are performed to compare the size retention and power of the variance ratio test on the filtered series to four commonly used unit root tests (see section A.1) performed when the GARCH behavior of the errors is neglected. The Monte Carlo study also extends the design used by Sjölander (op. cit.) by including two addition conditional distributions (fat-tailed and leptokurtic), both commonly encountered in financial economic data. The same 15 GARCH(1,1) error processes used by Sjölander (op. cit.) and 5 different dyadic sample sizes are used.

There are a multitude of candidate unit root tests that can be used for comparison with the wavelet ratio tests. The four alternative tests used here are well understood and extensively used by practitioners due to their accessibility. They also provide a good baseline for comparison in terms of statistical power, with the tests on efficiently detrended data known to have good power for finite samples, and the Augmented Dickey-Fuller test known to suffer from power deterioration.

It should be noted that the main goal of the method described here is not to model the GARCH(1,1) process with prediction or forecasting in mind, but to remove or minimize the effect of heteroscedasticity on the unit root test, and reconcile the actual and nominal test sizes for empirically interesting sample sizes.

The rest of this paper is organized as follows. Section 2 covers a short introduction to the wavelet transform, Section 3 covers the methodology of the wavelet variance ratio test for unit roots, Section 4 covers the Monte Carlo design and data generating processes (DGPs), Section 5 presents the main results in tabular, graphical and discussion format, and conclusions are given in Section 6. The appendix gives tables of the test sizes for all the data generatin processes (DGPs) considered, and a select set of figures corresponding to tables that have been discussed in the text.

2 The wavelet transform

A wavelet is an oscillatory function which starts at zero, achieves its maximum amplitude, and then reverts to zero. When considered as a function of time, as is the case in time series applications, it deviates from zero for a finite duration. It therefore has a location where it reaches its maximum, an oscillation period, and also a scale over which it grows and declines. The wavelet function, $\psi_{\lambda,t}(\cdot)$, which is real valued and defined over the real axis, must satisfy the following two conditions:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 \, dt = 1 \tag{2.1}$$

$$\int_{-\infty}^{\infty} \psi(t) \, dt = 0 \tag{2.2}$$

The first condition implies that the function is 'small' as compared to a sinusoidal function defined on $\mathbb{R}$, for example, and the second condition implies that it a 'wave', since it averages to zero.

Wavelets are indexed by the integer pair (λ, t). λ refers to the scale, which is the length of the wavelet, and t, refers to the location of the wavelet on the time axis.

The basic wavelet can be dilated (or compressed) and translated along the time axis and matched with signal segments in order to determine and localize the dominant or important frequencies. The wavelets which result from the translations and dilations of the basic wavelet are given as:

$$\psi_{\lambda,t}(u) = \frac{1}{\sqrt{\lambda}} \psi\left(\frac{u-t}{\lambda}\right) \tag{2.3}$$

For time series, which are sampled at discrete time points, the Discrete Wavelet Transform (DWT) is used. The DWT linearly transforms the vector $\boldsymbol{y} = [y_0, y_1, \ldots, y_{T-1}]^T$ of dimension $T = 2^J$ into a vector of dimension T containing the DWT coefficients. The transformation, given by $\boldsymbol{\mathcal{W}} \boldsymbol{y}$, outputs the vector of DWT coefficients $\boldsymbol{W} = [\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_J, \boldsymbol{V}_J]^T$. The j^{th} sub-vector from the partition, $\boldsymbol{W}_j$, contains $T/2^j$ wavelet coefficients which are differences between weighted, adjacent, and localized averages, and describe the changes at scale τ_j ($\tau_j \equiv 2^{j-1}$). $\boldsymbol{V}_J$ is a length $T/2^J$ vector of scaling coefficients associated with averages on scale λ_J ($\lambda_j \equiv 2^j$). $\boldsymbol{\mathcal{W}}$ is a $T \times T$ orthonormal transform matrix i.e. $\boldsymbol{\mathcal{W}}^T\boldsymbol{\mathcal{W}} = \boldsymbol{I}_T$. The transformation matrix $\boldsymbol{\mathcal{W}}$ can be similarly partitioned with $\boldsymbol{\mathcal{W}}_j$ being a $T/2^j \times T$ matrix related to changes at scale τ_j. Its rows are orthonormal vectors and are constructed from the high-pass wavelet filter. The last element of $\boldsymbol{\mathcal{W}}$ is $\boldsymbol{\mathcal{V}}_J$, a $1 \times T$ dimensional vector related to averaging at scale λ_J.

The scaling coefficients capture low frequency components of the time series (such as those due to unit root processes), while the wavelet coefficients capture the high frequency fluctuations. Because the power spectrum of a unit root process is characterized by a concentration of energy at the lower frequency bands, the ratio between the energy of the scaling coefficients to the total energy in a series forms the basis of the wavelet variance ratio test.

The Haar wavelet is a symmetric and compactly supported wavelet of length $L = 2$. Its wavelet (h_l) and scaling (g_l) filters are given by:

$$h_0 = \frac{1}{\sqrt{2}} \quad h_1 = -\frac{1}{\sqrt{2}} \quad g_0 = \frac{1}{\sqrt{2}} \quad g_1 = \frac{1}{\sqrt{2}} \tag{2.4}$$

The scaling filter $\{g_l\}$ is the 'quadrature mirror' filter corresponding to $\{h_l\}$ i.e $g_0 = -h_1$, and $g_1 = h_0$. The wavelet filter has the three properties of summation to zero, unit energy and orthogonal to its even shifts, while the scaling filter has the latter two properties.

Limitations of the Haar DWT in time series applications are the requirement that the length of input vector be dyadic (2^J), as well as not being invariant to circular shifts in the series. Also, the short filter may not provide as good band-pass properties as those of the longer and smoother filters. However, the Haar wavelet filter, because of its short length, is least affected by boundary effects, is symmetric, has closed form expressions, and is easy to understand and implement. For these reasons, we use the Haar DWT as one of two wavelet functions in the wavelet variance ratio tests. For comparison purposes, the lengths (sample sizes) of the simulated time series are dyadic, even when the unit root tests are performed using tests that do not have this restriction.

The Maximal Overlap DWT (MODWT[1]) is invariant to circular shifts in the original time series, meaning that shifts in the series result in equivalent shifts in the wavelet coefficients. It also has better resolution at lower scales because it does not down-sample as the DWT does. However, this comes at the cost of loss of orthogonality. The transform matrix $\widetilde{\boldsymbol{\mathcal{W}}}_j$ is not orthogonal, and as a result, smooths and details from the multiresolution analysis (MRA)[2] of the MODWT do not decompose the total variance on a scale-by-scale basis in the way that the wavelet coefficients do, hence only the wavelet and scaling coefficients will be used for the variance ratio test.

The MODWT wavelet and scaling filters are rescaled versions of the DWT filters (see equations 2.4). The wavelet filter $\{\tilde{h}_l\}$ is related to its DWT equivalent through $\tilde{h}_l \equiv h_l \big/ \sqrt{2}$ and the scaling filter $\tilde{g}_l \equiv g_l / \sqrt{2}$. It follows that for the scaling filter:

$$\sum_{l=0}^{L-1} \tilde{g}_l = 1 \quad \sum_{l=0}^{L-1} \tilde{g}_l^2 = \frac{1}{2} \quad \text{and} \quad \sum_{l=-\infty}^{\infty} \tilde{g}_l \tilde{g}_{l+2} = 0 \tag{2.5}$$

[1] also called “undecimated DWT”, “stationary DWT”, “time-invariant DWT” and “translation-invariant DWT”)

[2] The MRA of $\boldsymbol{y}$ expresses $\boldsymbol{y}$ as the sum of a constant vector $\boldsymbol{\mathcal{S}}_J$ (smooth) and $\boldsymbol{\mathcal{D}}_j$ (details, $j = 1 \ldots, J$) other vectors, each a time series related to variations of $\boldsymbol{y}$ at a given scale (Percival and Walden; 2000, p. 64–65)

Where L is the length of the filter.

The wavelet and scaling coefficients $\{\widetilde{W}_{j,t}\}$ and $\{\widetilde{V}_{j,t}\}$ are given as (Percival and Walden; 2000, Chapter 5):

$$\widetilde{W}_{j,t} = \sum\nolimits_{l=0}^{L-1} \tilde{h}_l^o y_{t-l \bmod T} \quad \tilde{V}_{j,t} = \sum\nolimits_{l=0}^{L-1} \tilde{g}_l^o y_{t-l \bmod T} \tag{2.6}$$

where $\{\tilde{h}_{j,l}^o\}$ and $\{\tilde{g}_{j,l}^o\}$ are filters obtained by periodizing $\{\tilde{h}_{j,l}\}$ and $\{\tilde{g}_{j,l}\}$ to length T. The wavelet transform is an energy preserving transform. The energy preservation (using the partial decomposition to level J_0) is given by:

$$\|\boldsymbol{y}\|^2 = \sum\nolimits_{j=1}^{J_0} \left\|\widetilde{\boldsymbol{W}}_j\right\|^2 + \left\|\widetilde{\boldsymbol{V}}_{J_0}\right\|^2 \tag{2.7}$$

with $J_0 < J$.

This is an important property of the wavelet transform for unit root testing because the wavelet ratio unit root test is based on the energy dis-balance between the wavelet and scaling coefficients.

In this paper, the LA(8) or Least Asymmetric wavelet of length 8 is used because of its desirable band-pass properties. Detailed expositions on wavelet filters and their band-pass properties can be found in the texts Percival and Walden (2000) by and Gençay et al. (2001).

3 Methodology

3.1 Wavelet variance ratio unit root test

The wavelet variance ratio test has found use in unit root testing because the wavelet variance decomposition of a time series makes it possible to determine the contribution to the total variation corresponding to changes within each scale. When the contribution to the total variation from the higher scales dominates that from the lower scales, the data generating process is more likely to be a unit root process than a stationary process. Also, because the wavelet transform can be formulated in terms of filters, there is a direct relationship between scale and frequency. The DWT decomposes the frequency interval $[0, 1/2]$ such that, at scale τ_j, the wavelet filter is a band-pass filter for the frequency interval $[1/2^{j+1}, 1/2^j]$. In the presence of a unit root, changes in the higher scales (lower frequencies) will be associated with the smoother movements due to the unit root, and the energy balance will be such that most of the energy will be in the low frequency bands. The wavelet variance ratio test is therefore constructed as a ratio of the energy of the scaling coefficients to the total energy of the data. Fan and Gençay (2010) introduced the wavelet variance ratio test using the unit level wavelet and scaling coefficients and base their test statistic on

$$\hat{S}_{T,1} = \frac{\sum_{t=1}^{T/2} V^2{}_{1,t}}{\sum_{t=1}^{T/2} V^2{}_{1,t} + \sum_{t=1}^{T/2} W^2{}_{1,t}} \tag{3.1}$$

For a unit root process, the scaling coefficients will be non-stationary and their energy at the finest level $\sum_{t=1}^{T/2} V^2{}_{1,t}$ dominates that of the stationary wavelet coefficients $\sum_{t=1}^{T/2} W^2{}_{1,t}$. The ratio of the variance of the scaling coefficients to that of the series two will be close to unity. The asymptotic distribution of the test statistic based on $\hat{S}_{T,1}$ has been derived by Fan and Gençay (2010) who also show that under the null hypothesis of a unit root

$$\hat{S}_{T,1} = \frac{\sum_{t=1}^{T/2} V^2{}_{1,t}}{\sum_{t=1}^{T/2} V^2{}_{1,t} + \sum_{t=1}^{T/2} W^2{}_{1,t}} = 1 + o_p(1) \tag{3.2}$$

and under the alternative hypothesis

$$\hat{S}_{T,1} = \frac{E(y_{2t} + y_{2t+1})^2}{E(y_{2t} + y_{2t+1})^2 + E(y_{2t} - y_{2t+1})^2} < 1 \tag{3.3}$$

This result can be generalized to higher level decompositions (Fan and Gençay; 2006), with the wavelet and scaling DWT coefficients computed respectively from

$$W_{j,t} = \sum\nolimits_{l=0}^{L^{(j)}-1} h_{j,l} y_{2^j t+1-l} \quad V_{j,t} = \sum\nolimits_{l=0}^{L^{(j)}-1} g_{j,l} y_{2^j t+1-l} \tag{3.4}$$

where $L^{(j)}$ is the length of the level j wavelet filter.

The corresponding energy ratio on which a test statistic can based is therefore

$$\hat{S}_{T,J} = \frac{\sum_{t=L_J}^{T/2^J-1} V^2{}_{j,t}}{\sum_{t=L_J}^{T/2^J-1} V^2{}_{j,t} + \sum_{j=1}^{J} \left(\sum_{t=L_j}^{T/2^J-1} W^2{}_{j,t} \right)} \tag{3.5}$$

In this paper we make use of the variance ratio unit root test but instead of the data $\{y_t\}$, we consider using the unit level DWT (and MODWT) scaling coefficients as the input vector; (cf. Li and Shukur; 2011) who apply the Dickey-Fuller unit root test to the scaling coefficients. This does not change the asymptotic properties of the test from that of $\hat{S}_{T,1}$ because under the null hypothesis, $\sum_{t=L_J}^{T/2^J-1} V^2{}_{j,t} = O_p(T^2)$ and each term of the inner summation of the squared wavelet coefficients, $\sum_{t=L_J}^{T/2^J-1} W^2{}_{j,t} = O_p(T)$. As a result, $\sum_{j=1}^{J} \left(\sum_{t=L_j}^{T/2^J-1} W^2{}_{j,t} \right) = O_p(T)$ holds even when (for the purposes of filtering the high frequency components) the unit level wavelet coefficients are set to zero i.e $\sum_{t=L_1}^{T/2} W^2{}_{1,t} = 0$. L_1 is the number of boundary level dependent coefficients at the unit level.

It therefore holds that, under the null hypothesis,

$$\frac{\sum_{t=L_J}^{T/2^J-1} V^2{}_{j,t}}{\sum_{t=L_J}^{T/2^J-1} V^2{}_{j,t} + \sum_{j=2}^{J}\left(\sum_{t=L_j}^{T/2^J-1} W^2{}_{j,t}\right)} = 1 + o_p(1) \tag{3.6}$$

and under the alternative,

$$\frac{\sum_{t=L_J}^{T/2^J-1} V^2{}_{j,t}}{\sum_{t=L_J}^{T/2^J-1} V^2{}_{j,t} + \sum_{j=2}^{J}\left(\sum_{t=L_j}^{T/2^J-1} W^2{}_{j,t}\right)} < 1 \tag{3.7}$$

Intuitively, because the power spectrum of a unit root process is concentrated at the lowest frequencies (near zero frequencies), a test based on the higher scale wavelet variance decompositions will not change the energy ratio for large sample sizes. The energy of the scaling coefficients will always dominate that of the wavelet coefficients, and the ratio will be close to 1 under the null hypothesis of unit root.

$\widetilde{S}$, $\widetilde{V}$ and $\widetilde{W}$ replace S, V and W in the case of the variance ratio test using the Maximal Overlap DWT, and the test proceeds in a similar way to the DWT case.

Using the scaling coefficients at first level as the input vector for unit root testing restricts the frequency decomposition of $\{y_t\}$ to the bands $f \in [0, 1/4]$ because the unit level wavelet coefficients corresponding to $f \in [1/4, 1/2]$ are omitted. This has two consequences; firstly the test retains power against alternatives that have considerable energy in frequencies closer to zero, such as near integrated processes, and secondly it modifies the behavior of $\{y_t\}$ with respect to the highest frequency band by dampening the effects of volatility in the series. Setting the unit level wavelet coefficients to zero can be considered the simplest wavelet thresholding method. Other thresholding methods have been used to smooth volatility effects e.g. (Gençay et al.; 2001, p. 516) where five thresholding methods are applied to the classic dataset on IBM daily returns studied in Box and Jenkins (1976). More on thresholding procedures can be found in Nason (2008) and Gençay et al. (2001).

4 The Monte Carlo experiment

Monte Carlo simulation was used to compare the size retention and power of the wavelet variance ratio unit root test to the ADF test (Dickey and Fuller; 1981; Said and Dickey; 1984), the PP test (Phillips and Perron; 1988), and the DF-GLS and Point Optimal tests of Eliott, Rothenburg and Stock (Elliot et al.; 1996). The ADF and PP tests differ in the how they handle serial correlation in the error terms. In order to compare each test with the wavelet ratio test, a DGP having a higher order autoregressive process than AR(1) has to be used to generate the simulated series. For this reason, the DGP is an AR(2) process with a single unit root. Also, the GARCH (1,1) model is used to generate the errors because it provides a more parsimonious representation of possible higher order ARCH models.

4.1 The Data Generating Process (DGP)

The DGP for the AR(2)-GARCH(1,1) is specified as follows: $y_t = \mu_t + z_t$ with the deterministic components specified as $\mu_t = \mu + \beta t$. For the purposes of this study, an allowance is not made for a time trend and it follows that $\mu_t = \mu$. The unit root dynamics are captured in the error term $\phi(L) z_t = \varepsilon_t$ with $\phi(L) = \left(1 - \phi_1 L - \phi_2 L^2\right)$ for the AR(2) process. Written with the isolated unit root, $(1 - L)(1 - \varphi L) z_t = \varepsilon_t$ and the DGP is therefore given by $(1-\varphi L)\,\Delta y_t = \varepsilon_t$. The coefficients of the lag polynomial of the AR(2) processes that are simulated are $\phi_1 = 1.25$ and $\phi_2 = -0.25$ so that the DGP is $(1-0.25L)\,\Delta y_t = \varepsilon_t$, and ε_t follows a stationary GARCH(1,1) process as shown below

$$\varepsilon_t = u_t \sigma_t \tag{4.1}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 \sigma_{t-1}^2 \tag{4.2}$$

$$\sigma^2 = \frac{\alpha_0}{1-\alpha_1-\gamma_1} \tag{4.3}$$

The scaled conditional error $u_t \sim$ iid(0,1); σ_t^2 and σ^2 are the conditional and unconditional variances respectively. Also, for σ_t^2 to be positive, $\alpha_0 > 0$, $\alpha_1 \geqslant 0$, and $\gamma_1 \geqslant 0$. For the simulation we set α_0 to be equal to $(1-\alpha_1-\gamma_1)$ so that the unconditional variance is fixed at 1. The condition $\alpha_0 > 0$ is required to avoid degeneracy of the unconditional variance and $\alpha_1 + \gamma_1 < 1$ is required for finite unconditional variance.

A burn-in of 100 was used and the initial value for the unconditional variance was set to 1. y_0 was set to zero by subtracting y_0 from each y_t. For this study, an allowance is made for a constant term but no time trend is included in the specification of the deterministic terms. Testing using the variance ratio test will therefore have to be done using the demeaned series $\{\tilde{y}_t\}$ with $\tilde{y}_t \equiv (y_t - \mu)$ where μ is estimated by the sample mean $\hat{\mu} = T^{-1}\sum_{t=1}^{T} y_t$. This can be extended to allow for a constant term and time trend, but in that case, the data will have to be demeaned and detrended using a prior regression on a constant term and time trend, and the residuals tested for the unit root.

4.2 The Monte Carlo Design

Table 4.1 shows the factors that are varied together with their respective ranges. The nominal size is 5% as per convention and its acceptance range is the 95% confidence interval for the estimated proportion as used by Edgerton and Shukur (1999).

$$CL = \hat{\pi} \pm \Phi^{-1}(1-\hat{\pi})\sqrt{\frac{\hat{\pi}\,(1-\hat{\pi})}{N}} \tag{4.4}$$

where $\hat{\pi}$ is the proportion of rejections of the null hypothesis and N is the number of repetitions of the test.

For this study, N=10,000 and $\hat{\pi}$=0.05. $CL = [0.0457, 0.0543]$.

It is therefore expected that 1 in 20 simulations, each using 10,000 repetitions, will result in spurious under- or over-sizing. 75 series are simulated for each experimental setting so 1 in 15 results are expected to fall outside the interval for an unbiased test. Three standardized conditional distributions are used; the standard Normal, the standardized t distribution with 4 degrees of freedom, and the Generalized Error Distribution (GED) Nelson (1991) with the tail-thickness parameter ν equal to 1.5. The standardized t and GED distributions used here have fat tails and are more typical of financial applications than the normal distribution.

The ranges for α_0, α_1 and γ_1 are replicated from the study by Sjölander (2008) with justification given therein. Fifteen combinations of α_0, α_1 and γ_1 meet the constrains for positive conditional variance and unconditional variance equal to 1. The generated series will have GARCH(1,1) errors that can be considered to be representative of processes that are encountered in practice.

Table 4.1: Factors that vary for the different DGPs

Factor	Symbol	Design
Nominal size	π_0	0.05
Number of repetitions	N	10 000
Number of observations	T	64, 128, 256, 512, 1024
Dominant root[3]	α	0.8, 0.9, 0.95, 0.98, 0.99, 1
GARCH parameters	α_0	$\alpha_0 = (1 - \alpha_1 - \gamma_1)$
	α_1	0.005, 0.25, 0.50, 0.75, 0.99
	γ_1	0.005, 0.24, 0.49, 0.74, 0.99
Conditional distributions[4]	u_t	iid N(0,1)
		iid GED($\nu = 1.5$)
		iid $t^{std}(4)$
Wavelet function	DWT Haar	
	MODWT LA(8)	

The interpretation of the effects of the GARCH(1,1) process is that α_1 represents the degree of volatility and $(\alpha_1 + \gamma_1)$ the degree of autoregressive persistence of the conditional variance. Table 4.2 shows the combinations of α_1 and $(\alpha_1 + \gamma_1)$ that are used by the 15 different DGPs.

Critical values for the wavelet variance ratio unit root tests were determined using simulation as follows:

For each wavelet function and sample size

1. A series was generated from the unit root AR(2) process with iid N(0,1)

[3] An AR(2) process is generated by $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$. The solution to this difference equation depends on the 2 starting values of $\{y_t\}$, and on $\{\varepsilon_t\}$. The difference equation will be stable (hence AR(2) series stationary) if the larger (dominant) characteristic root, z_d=max(z_1, z_2), of the characteristic equation, $z^2 = \phi_1 z + \phi_2$, satisfies $|z_d| < 1$, i.e. is inside the unit circle (see Enders; 2010, pp. 22-23). The larger root determines the impact of past shocks i.e. persistence of the shocks. When the root equals unity, the process is said to have a unit root, and shocks have permanent effects.

[4] a t distributed random variable with with n d.f has variance = (n-2)/n. The draws from such a distribution have to be scaled by $\left(\sqrt{(n-2)/n}\right)^{-1}$ to generate pseudo-random numbers with variance equal to 1.

Table 4.2: Volatility and persistence used for the DGPs

Model	**Volatility** (α_1)	**Persistance** ($\alpha_1 + \gamma_1$)
1	0.005	0.010
2		0.245
3		0.495
4		0.745
5		0.995
6	0.250	0.255
7		0.490
8		0.740
9		0.990
10	0.500	0.505
11		0.740
12		0.990
13	0.750	0.755
14		0.990
15	0.990	0.995

errors, as given in section 4.1

2. The unit level wavelet decomposition was performed on the series
3. The scaling coefficients were extracted
4. The wavelet ratio unit root test was done on the scaling coefficients
5. The process was repeated 10,000 times
6. The critical values were determined as the 5$^{\text{th}}$ percentile of the distribution of the 10,000 values

The actual size for each unit root test was determined as follows:

1. The (MO)DWT variance ratio (MODWT-VR and DWT-VR) tests:

For each experimental setting (combination of sample size, wavelet function, GARCH(1,1) process and conditional distribution), 10,000 variance ratios are calculated, and the proportion of values that are more extreme than the critical value provide the empirical size under the Monte Carlo experimental conditions. Results that are outside of the interval $[0.0457, 0.0543]$ are considered to be from biased tests i.e under- or over-sized tests

2. The Augmented Dickey-Fuller (ADF) test: Because the DGP is an AR process of known order (2), the Dickey-Fuller regression is augmented by one lagged difference, Δy_{t-1}, so that the maintained regression is

$$\Delta y_t = \mu + \gamma y_{t-1} + \alpha_1 \Delta y_{t-1} + \varepsilon_t \tag{4.5}$$

$$\gamma = \phi_1 + \phi_2 - 1 \quad \text{and} \quad \alpha_1 = -\phi_2$$

ϕ_1 and ϕ_2 are the AR(2) coefficients.

The citical values for the ADF test are known to depend on the sample size as well as the lag order (see Cheung and Lai; 1995). The approximate finite sample critical values that are used for the ADF test are obtained from to the response surface method of Cheung and Lai (op. cit.) The response surface function is given as:

$$cv(T,p) = \kappa_0 + \kappa_1 \,/\, T + \kappa_2 \,/\, T^2 \omega_1[(p) \,/\, T] + \omega_2[(p) \,/\, T]^2 \tag{4.6}$$

κ_1 and κ_2 relate the critical values to the sample size and ω_1 and ω_2

relate the critical values to the dependence of the lag order of the fitted ADF regression. $p = 1$ and therefore ADF(1) used here.

10,000 series were generated for each experimental condition and tested for a unit root using the ADF test. The values of the test statistic are compared to the critical values derived using the method described above.

Values that are more extreme than the critical value indicate false rejection of the null hypothesis of unit root. The proportion of rejections provides the empirical size under the Monte Carlo experimental conditions. Results that are outside of the interval $[0.0457, 0.0543]$ are considered to be from biased tests

3. The ERS Tests - Dickey-Fuller GLS (DF-GLS) and ERS Point Optimal (ERS-PT) tests:

   The DF-GLS test uses an efficient version of the ADF t-statistic i.e the test is a ADF test applied to efficiently detrended data without an intercept. Because DGP used is known (an AR(2) process), the number of lags used in the DF-GLS test is set to 1. The critical values for both the ERS tests are found by interpolation of the simulation results in Table 1. of Elliot et al. (1996, p. 825).

   10,000 series were generated for each experimental condition and each test was performed on the series. The proportion of values that are more extreme than the critical value provide the empirical size under the Monte Carlo experimental conditions. Results that are outside of the interval $[0.0457, 0.0543]$ are considered to be from biased tests i.e under- or over-sized tests

4. The Phillips-Perron test statistic (Phillips and Perron; 1988) is a modification of the basic Dickey-Fuller test statistic by a non-parametric adjustment. Autocorrelation of the error terms that might result from the underlying GDP not being AR(1) are addressed by using the Newey-West heteroscedasticity and autocorrelation (HAC) estimate of the long run variance (Newey and West; 1987). The truncation lag used for the autocovariance in for this paper is set to $\mathrm{trunc}(4(T/100)^{0.25})$ [5]. The critical values are found by interpolation of values in Table 4.2 Banerjee et al. (1993, p.103)

   10,000 series were generated for each experimental condition and each test was performed on the series. The proportion of values that are more extreme than the critical value provide the empirical size of the test.

5 Results

5.1 Comparison of size retention for unit root tests

5.1.1 GARCH(1,1) errors with N(0,1) conditional distribution

The test sizes for all DGPs are reported in Tables A.1, A.2 and A.3 in the appendix. All sizes which fall within the acceptance interval i.e $[0.0457, 0.0543]$ are shaded in gray. The wavelet ratio and ADF tests are shown to be unbiased for series generated by models 1-5 of Table 4.2 i.e $\alpha_1 = 0.005$ and $(\alpha_1 + \gamma_1) = (0.10,\ 0.245,\ 0.495,\ 0.745,\ 0.995)$. The ERS tests (ERS-PT

[5] see Said and Dickey (1985) and Schwert (1987) for how to choose the lag length when approximating an ARMA(p, q) process with an AR(l) process. The method used here is the fixed rule (l_4) due to Schwert (1987)

and DF-GLS) are also shown to be unbiased when used on the largest sample sizes. This result is expected because the volatility of the conditional variance in these series is low.

Models 6-9 of Table 4.2 have moderate volatility ($\alpha_1 = 0.25$). The wavelet ratio and ADF test is unbiased for models 6, 7, and 8. These models have a moderate to high persistence in the conditional variance of GARCH errors i.e 0.255, 0.49 and 0.74 respectively. Model number 9 has high persistence in the conditional variance $(\alpha_1 + \gamma_1) = 0.99$. For this model, the wavelet variance ratio tests are shown to be unbiased but the ADF, the two ERS tests and the PP tests are all shown to be biased.

Model 10 of Table 4.2 has moderately high volatility ($\alpha_1 = 0.5$) as well as moderately high persistence in the conditional variance $(\alpha_1 + \gamma_1) = 0.505$. The wavelet ratio- and the ADF tests are unbiased for all sample sizes, and the ERS and PP tests are unbiased for the largest sample size.

Models 12-15 of Table 4.2 have high persistence in the conditional variance of the GARCH errors (as a result of a large volatility parameter). Only the wavelet ratio tests are unbiased for these models, except model 13 where the PP test has good performance.

5.1.2 GARCH(1,1) errors with GED($\nu = 1.5$) conditional distribution

For models 1-8 and model 10 of Table 4.2, the wavelet ratio and ADF tests are unbiased. The DF-GLS and ERS-PT also perform well for the larger sample sizes (T=512 and T=1024). Models 12, 14 and 15 have high persistence in the conditional variance of the GARCH(1,1) error (0.99, 0.99

and 0.995 respectively). The only unbiased tests are the wavelet ratio tests.

5.1.3 GARCH(1,1) errors with t(4 df) conditional distribution

The results for models 1-7 and 10-13 of Table 4.2 follow a similar pattern to those where the standardized normal and GED conditional distributions are used. All tests perform poorly for models 8, 14 and 15.

In general, over-sizing in the unit root tests results mainly from high volatility in the conditional variance of the GARCH error. The DWT-VR and MODWT-VR tests perform well in comparison to the alternative unit root tests, except when the conditional distribution of the GARCH error is the fat-tailed t(4 df) distribution, where all the tests perform poorly. Tables 5.1, 5.2 and 5.3 show the actual test sizes for unit root tests performed on series with GARCH errors for which $\alpha_1 = (0.25, 0.5, 0.75, 0.99)$ and $(\alpha_1 + \gamma_1) = (0.99, 0.995)$, using the three conditional distributions.

Table 5.1: Test sizes for series with GARCH(1,1) errors with N(0,1) conditional distribution

GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.990$						
T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
64	0.0535	0.0489	0.0752	0.0997	0.1092	0.0523
128	0.0469	0.0534	0.0828	0.0878	0.0905	0.0681
256	0.0493	0.0479	0.0891	0.0792	0.0831	0.0735
512	0.0544	0.0507	0.0896	0.0787	0.0810	0.0724
1024	0.0505	0.0468	0.0902	0.0711	0.0723	0.0789

Table 5.1: Test sizes for series with GARCH(1,1) errors with N(0,1) conditional distribution

GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.990$						
T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.500$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0525	0.0515	0.0796	0.1039	0.1091	0.0715
128	0.0469	0.0487	0.0890	0.0891	0.0914	0.0692
256	0.0465	0.0482	0.0920	0.0850	0.0843	0.0775
512	0.0534	0.0488	0.0957	0.0785	0.0770	0.0743
1024	0.0532	0.0537	0.0882	0.0751	0.0738	0.0787
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.750$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0537	0.0521	0.0758	0.0909	0.1034	0.0703
128	0.0467	0.0483	0.0841	0.0821	0.0762	0.0712
256	0.0469	0.0502	0.0834	0.0736	0.0714	0.0689
512	0.0494	0.0512	0.0815	0.0748	0.0664	0.0667
1024	0.0525	0.0500	0.0773	0.0597	0.0656	0.0717
GARCH parameters: $\alpha_0 = 0.005, \alpha_1 = 0.990$, and $(\alpha_1 + \gamma_1) = 0.995$						
64	0.0467	0.0557	0.0660	0.0818	0.0926	0.0605
128	0.0519	0.0436	0.0693	0.0698	0.0723	0.0639
256	0.0501	0.0510	0.0696	0.0604	0.0670	0.0621
512	0.0512	0.0470	0.0707	0.0598	0.0613	0.0603
1024	0.0484	0.0466	0.0688	0.0572	0.0578	0.0628

Table 5.2: Test sizes for series with GARCH(1,1) errors with GED (ν = 1.5) conditional distribution

GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.990$						
T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
64	0.0486	0.0554	0.0788	0.0975	0.1060	0.0551
128	0.0473	0.0567	0.0866	0.0877	0.0837	0.0642
256	0.0547	0.0500	0.0928	0.0878	0.0809	0.0732
512	0.0462	0.0536	0.0945	0.0837	0.0795	0.0803
1024	0.0532	0.0426	0.0967	0.0790	0.0764	0.0776
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.500$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0486	0.0474	0.0853	0.1016	0.1077	0.0646
128	0.0533	0.0472	0.0846	0.0899	0.0866	0.0685
256	0.0509	0.0538	0.0930	0.0824	0.0829	0.0708
512	0.0479	0.0480	0.0921	0.0784	0.0770	0.0745
1024	0.0524	0.0460	0.0916	0.0725	0.0768	0.0734
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.750$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0476	0.0468	0.0804	0.0986	0.1036	0.0667
128	0.0492	0.0488	0.0815	0.0826	0.0832	0.0690
256	0.0514	0.0423	0.0784	0.0720	0.0710	0.0713
512	0.0521	0.0531	0.0812	0.0658	0.0672	0.0694
1024	0.0487	0.0492	0.0826	0.0647	0.0689	0.0640
GARCH parameters: $\alpha_0 = 0.005, \alpha_1 = 0.990$, and $(\alpha_1 + \gamma_1) = 0.995$						
64	0.0500	0.0501	0.0669	0.0810	0.0925	0.0634
128	0.0495	0.0482	0.0698	0.0699	0.0722	0.0653

Table 5.2: Test sizes for series with GARCH(1,1) errors with GED ($\nu = 1.5$) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.990$						
256	0.0505	0.0537	0.0714	0.0580	0.0663	0.0593
512	0.0551	0.0527	0.0697	0.0592	0.0598	0.0620
1024	0.0467	0.0490	0.0702	0.0567	0.0573	0.0610

Table 5.3: Test sizes for series with GARCH(1,1) errors from a standardized t conditional distribution with 4 degrees of freedom

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0532	0.0538	0.0769	0.1015	0.1058	0.0615
128	0.0569	0.0586	0.0897	0.0857	0.0869	0.0657
256	0.0545	0.0586	0.0938	0.0826	0.0868	0.0705
512	0.0516	0.0582	0.0955	0.0818	0.0771	0.0721
1024	0.0571	0.0546	0.0931	0.0779	0.0723	0.0740
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.500$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0542	0.0531	0.0782	0.0957	0.1031	0.0639
128	0.0468	0.0474	0.0851	0.0877	0.0829	0.0667
256	0.0506	0.0497	0.0881	0.0774	0.0778	0.0766
512	0.0494	0.0499	0.0909	0.0718	0.0737	0.0718
1024	0.0422	0.0492	0.0850	0.0664	0.0690	0.0702
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.750$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0525	0.0451	0.0756	0.0919	0.1014	0.0624

Table 5.3: Test sizes for series with GARCH(1,1) errors from a standardized t conditional distribution with 4 degrees of freedom

GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.990$						
T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
128	0.0481	0.0546	0.0721	0.0771	0.0766	0.0655
256	0.0437	0.0439	0.0776	0.0689	0.0688	0.0681
512	0.0415	0.0434	0.0730	0.0616	0.0583	0.0644
1024	0.0443	0.0441	0.0738	0.0600	0.0621	0.0622
GARCH parameters: $\alpha_0 = 0.005, \alpha_1 = 0.990$, and $(\alpha_1 + \gamma_1) = 0.995$						
64	0.0452	0.0441	0.0658	0.0733	0.0872	0.0593
128	0.0435	0.0452	0.0645	0.0661	0.0671	0.0564
256	0.0432	0.0453	0.0654	0.0581	0.0603	0.0620
512	0.0456	0.0400	0.0661	0.0542	0.0530	0.0556
1024	0.0441	0.0475	0.0626	0.0523	0.0495	0.0581

5.2 Comparison of power functions for the unit root tests

The Monte Carlo design for the power comparisons is similar to that for size calculations. However, the data are generated under the alternative hypothesis with the dominant roots of the characteristic equations of the AR(2) process equal being (0.99, 0.98, 0.95, 0.90 and 0.80); which correspond to the following autoregressive coefficients respectively,

$\phi_1 = 1.25$ and $\phi_2 =$((0.2574), (0.2646), (0.285), (0.315), (0.36))

The scaled conditional error, u_t (from equation 4.3), follows the three conditional distributions used before. Over-sized tests, when used for small samples, will have more power than unbiased tests. To avoid this, size-adjusted power is reported for all the unit root tests. The tests are correctly sized by simulating critical values for each sample size.

The size-adjusted power curves for the simulated series with the GARCH(1,1) parameters $\alpha_0 = 0.01$, $\alpha_1 = 0.25$ and $(\alpha_1 + \gamma_1) = 0.99$ are given in Figures 5.1, 5.2 and 5.3 for each of the three conditional distributions. The corresponding power functions for the tests reported in Tables 5.1, 5.2 and 5.3 are presented in the appendix. These are the power curves for DGPs with GARCH(1,1) errors for which the persistence in the conditional variance is $(\alpha_1 + \gamma_1) \geq 0.99$.

For the smaller sample sizes (T=64 and 128), the tests have comparable power in the near integrated case (dominant root = 0.99). The difference in power becomes bigger as dominant root gets smaller. When the dominant root is 0.8, the DF-GLS and ERS-PT have the highest power, followed by the two variance ratio tests, then the Dickey-Fuller test. The Phillips-Perron test has the lowest power.

For the largest sample size (T=1024) there is a considerable difference in power between the tests that perform best and worst.

Figure 5.1: Power curves for tests using a DGP with Normal (0,1) conditional distribution, $\alpha_0 = 0.01$ $\alpha_1 = 0.25$ $(\alpha_1 + \gamma_1) = 0.99$

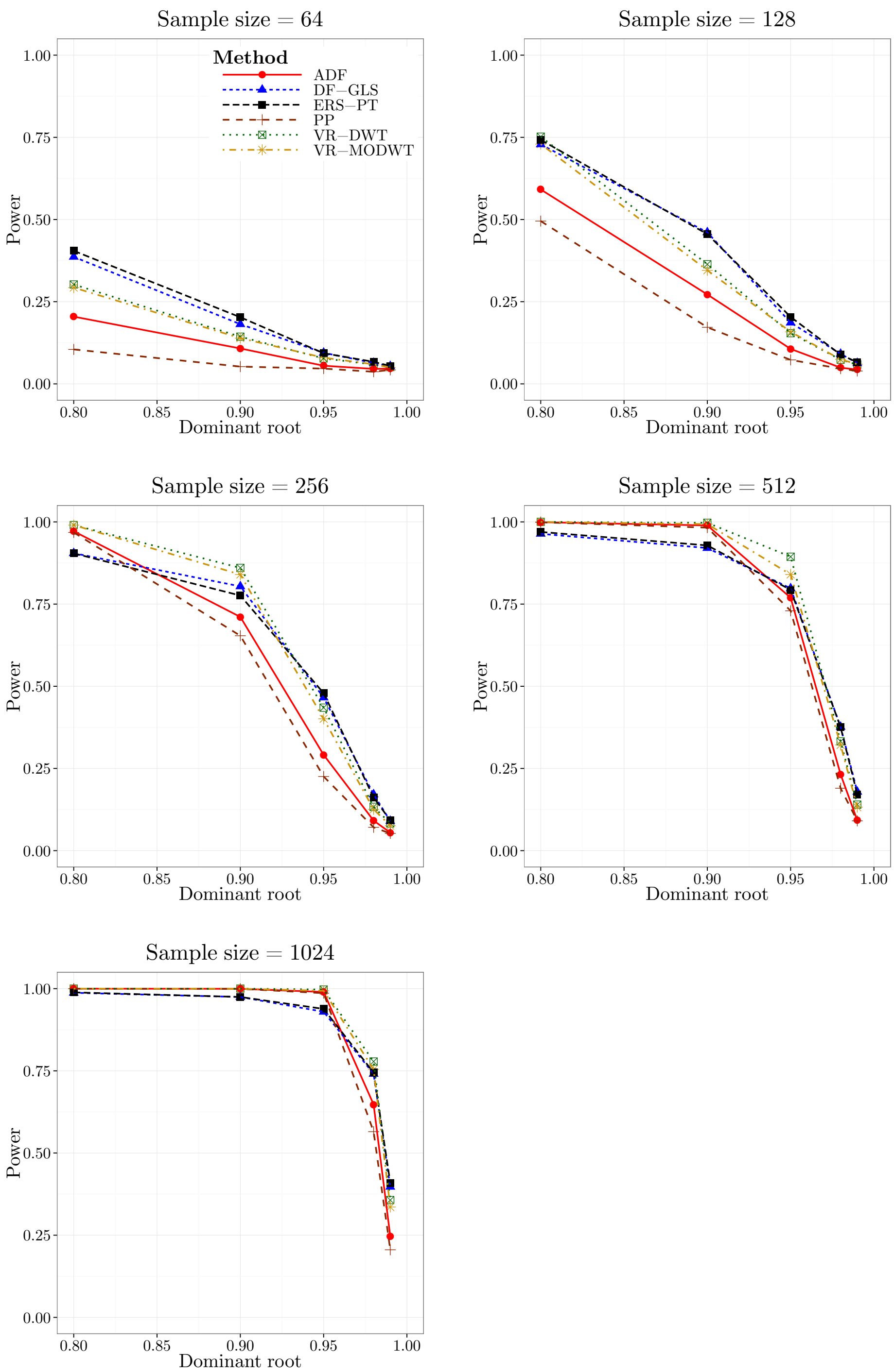


Figure 5.2: Power curves for tests using a DGP with GED($\nu = 1.5$) conditional distribution, $\alpha_0 = 0.01$, $\alpha_1 = 0.25$ and $(\alpha_1 + \gamma_1) = 0.99$

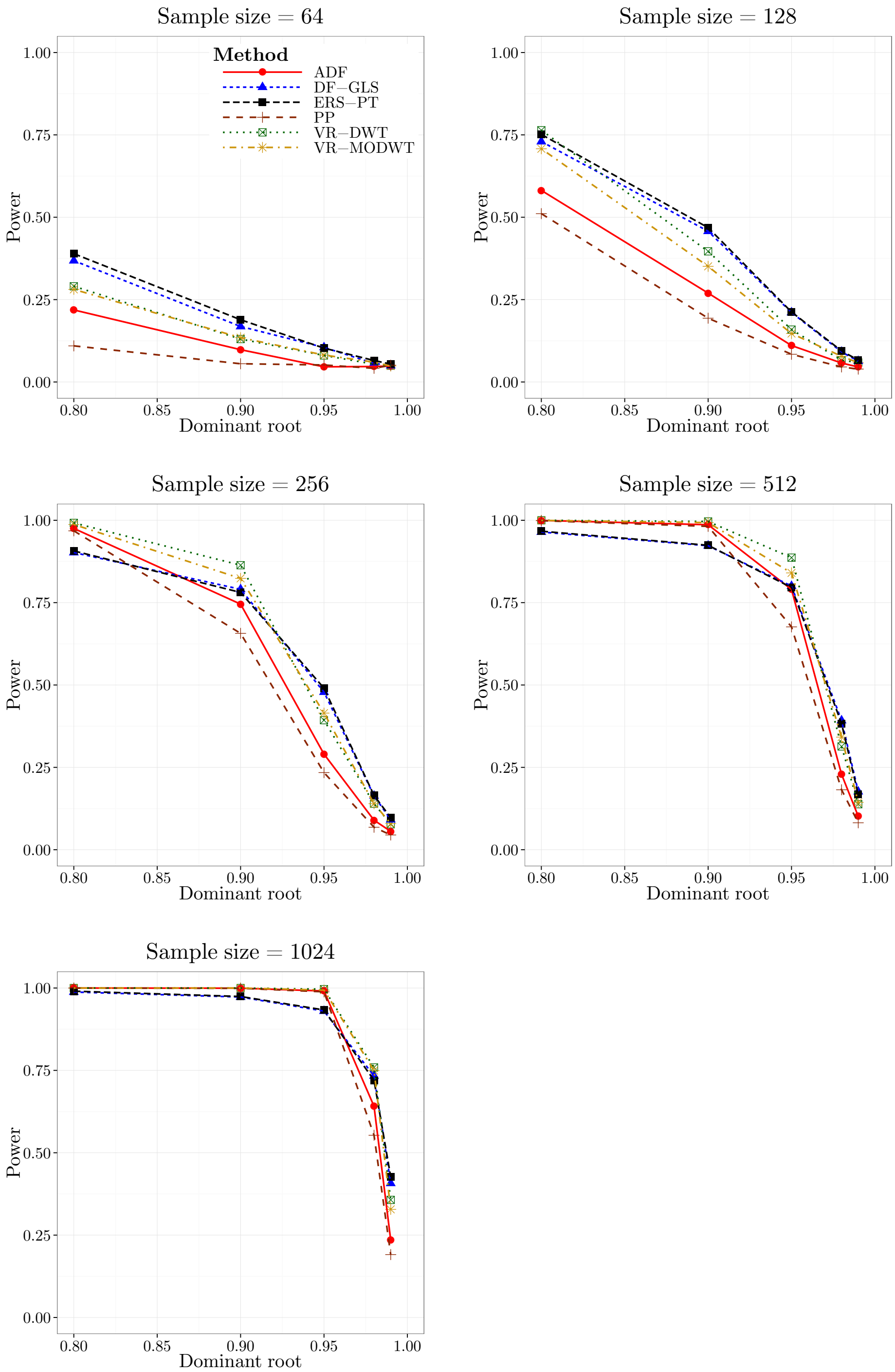
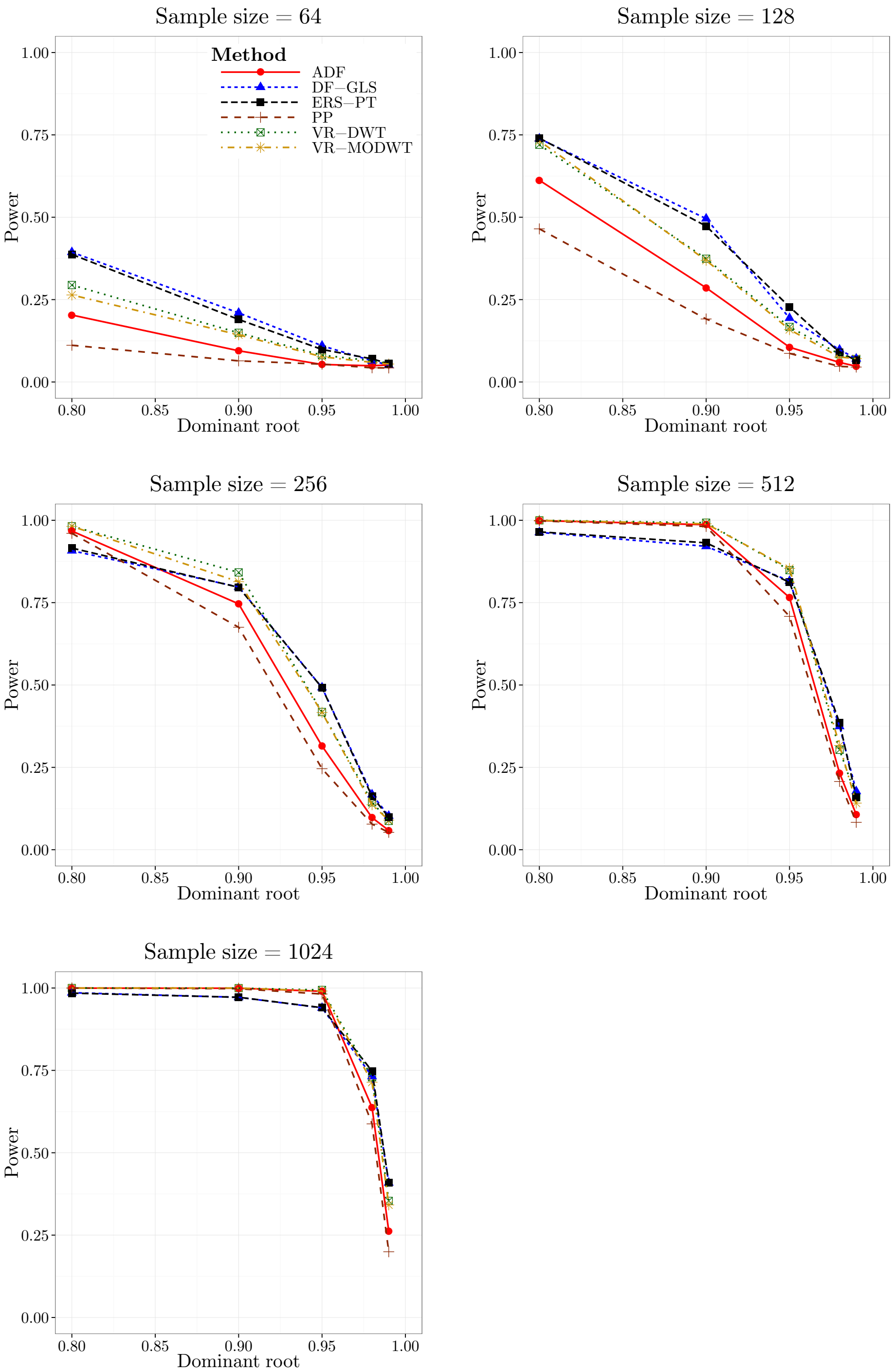


Figure 5.3: Power curves for tests using a DGP with $t(4)$ conditional distribution, $\alpha_0 = 0.01$, $\alpha_1 = 0.25$ and $(\alpha_1 + \gamma_1) = 0.99$



Depending on the conditional distribution, the DF-GLS and ERS-PT tests have between 39% and 40% power when the dominant root is 0.99. The ADF test has power between 23% and 26%, and the PP test has power between 19% and 21%. The DWT-VR and MODWT-VR have power of between 35% and 36%. These results are expected because the tests based on efficient detrending are designed to be powerful against near integrated alternatives compared to the Dickey-Fuller tests.

The three conditional distributions have similar power functions. This suggests that the six unit root tests considered here are robust to the leptokurtic and fat-tailed conditional distributions used in the study.

6 Conclusions

The problem of over-rejection of the unit root null hypothesis in the presence of GARCH(1,1) errors has been revisited in this paper. A unit root test based on the wavelet variance ratio has been used on data that is filtered from the high frequency components using the discrete wavelet transform. This is a combination of the methods used by Li and Shukur (2011) and Fan and Gençay (2010). The wavelet based unit root test is shown to be an improvement on 4 widely used unit root tests (the Augmented Dickey-Fuller, DF-GLS, ERS Point Optimal, and the Phillip-Perron tests) for sample sizes between 64 and 1024. The wavelet ratio tests perform as well as, or better than the other unit root tests when the GARCH(1,1) errors have fat-tailed and leptokurtic conditional distributions. Only in the case where the conditional distribution is a t distribution with 4 degrees, combined with moderate to high volatility and persistence in conditional variance of the GARCH error ($\alpha_1 + \gamma_1 \geq 0.99$), do the two wavelet variance ratio tests become slightly biased.

In addition, size-adjusted power comparisons were made between the unit root tests. For the smallest sample sizes T=(64, 128), all tests had similar power for the near unit root alternatives. As the alternatives moved further from the unit root, the DF-GLS and ERS-PT tests were most powerful, followed by the wavelet ratio tests. The ADF and PP test had least power. The power functions were similar for the three conditional distributions for all tests using the same sample sizes.

References

Andrews, D. W. and Guggenberger, P. (2011). A conditional-heteroskedasticity-robust confidence interval for the autoregressive parameter, *Review of Economics and Statistics* (0).

Banerjee, A., Dolado, J. J., Galbraith, J. W. and Hendry, D. (1993). *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of econometrics* **31**(3): 307–327.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control, Revised Edition*, Holden-Day.

Charles, A. and Darné, O. (2008). A note on unit root tests and GARCH errors a simulation experiment, *Communications in Statistics-Simulation and Computation* **37**(2): 314–319.

Cheung, Y.-W. and Lai, K. S. (1995). Lag order and critical values of the Augmented Dickey–Fuller test, *Journal of Business & Economic Statistics* **13**(3): 277–280.

Cook, S. (2006). The robustness of modified unit root tests in the presence of GARCH, *Quantitative Finance* **6**(4): 359–363.

Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form, *Computational Statistics & Data Analysis* **45**(2): 215–233.

Davidson, R. and MacKinnon, J. G. (1993). *Estimation and inference in econometrics*, Oxford University Press.

Demetrescu, M. (2010). On the dickey–fuller test with white standard errors, *Statistical Papers* **51**(1): 11–25.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American statistical association* **74**(366a): 427–431.

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica: Journal of the Econometric Society* pp. 1057–1072.

Edgerton, D. and Shukur, G. (1999). Testing autocorrelation in a system perspective testing autocorrelation, *Econometric Reviews* **18**(4): 343–386.

Elliot, G., Rothenberg, T. and Stock, J. (1996). Efficient tests for an autoregressive unit roo, *Econometrica* **64**(4): 813–836.

Enders, W. (2010). *Applied econometric time series*, John Wiley & Sons.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica: Journal of the Econometric Society* pp. 987–1007.

Fan, Y. and Gençay, R. (2006). Unit root and cointegration tests with wavelets, *Canadian Econometric Study Group Meeting, McGill University, Montreal, September*, pp. 29–30.

Fan, Y. and Gençay, R. (2010). Unit root tests with wavelets, *Econometric Theory* **26**(05): 1305–1331.

Gençay, R., Selçuk, F. and Whitcher, B. J. (2001). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, Elsevier.

Kim, K. and Schmidt, P. (1993). Unit root tests with conditional heteroskedasticity, *Journal of Econometrics* **59**(3): 287–300.

Li, W., Ling, S. and McAleer, M. (2002). Recent theoretical results for time series models with GARCH errors, *Journal of Economic Surveys* **16**(3): 245–269.

Li, Y. and Shukur, G. (2011). Wavelet improvement of the over-rejection of unit root test under GARCH errors: An application to swedish immigration data, *Communications in Statistics-Theory and Methods* **40**(13): 2385–2396.

Ling, S. and Li, W. (1998). Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscedastic errors, *The Annals of Statistics* **26**(1): 84–125.

Ling, S. and Li, W. (2003). Asymptotic inference for unit root processes with GARCH (1, 1) errors, *Econometric Theory* **19**(4): 541–564.

Ling, S., Li, W. and McAleer, M. (2003). Estimation and testing for unit root processes with GARCH (1, 1) errors: theory and monte carlo evidence, *Econometric Reviews* **22**(2): 179–202.

Long, J. S. and Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model, *The American Statistician* **54**(3): 217–224.

MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties, *Journal of Econometrics* **29**(3): 305–325.

Nason, G. P. (2008). *Wavelet Methods in Statistics with R*, Springer.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica: Journal of the Econometric Society* pp. 347–370.

Newey, W. and West, K. (1987). A simple positive semi-definite heteroscedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**(3): 703–708.

Pantula, S. G. (1988). Estimation of autoregressive models with ARCH errors, *Sankhyā: The Indian Journal of Statistics, Series B* pp. 119–138.

Patterson, K. (2009). *A Primer for Unit Root Testing*, Palgrave Macmillan.

Patterson, K. (2012). *Unit Root Tests in Time Series Volume 2: Extensions and Developments*, Vol. 2, Palgrave Macmillan.

Patterson, K. D. (2000). *An introduction to applied econometrics: a time series approach*, Macmillan.

Percival, D. B. and Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis (Cambridge Series in Statistical and Probabilistic Mathematics)*, Cambridge University Press.

Phillips, P. C. and Perron, P. (1988). Testing for a unit root in time series regression, *Biometrika* **75**(2): 335–346.

Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika* **71**(3): 599–607.

Said, S. E. and Dickey, D. A. (1985). Hypothesis testing in arima (p, 1, q) models, *Journal of the American Statistical Association* **80**(390): 369–374.

Schwert, G. W. (1987). Effects of model specification on tests for unit roots in macroeconomic data, *Journal of Monetary Economics* **20**(1): 73–103.

Seo, B. (1999). Distribution theory for unit root tests with conditional heteroskedasticity, *Journal of Econometrics* **91**(1): 113–144.

Sjölander, P. (2008). A new test for simultaneous estimation of unit roots and GARCH risk in the presence of stationary conditional heteroscedasticity disturbances, *Applied Financial Economics* **18**(7): 527–558.

So, B. S. and Shin, D. W. (1999). Recursive mean adjustment in time-series inferences, *Statistics & Probability Letters* **43**(1): 65–73.

Wang, G. and Mao, W.-L. (2008). Unit root testing in the presence of heavy-tailed garch errors, *Australian & New Zealand Journal of Statistics* **50**(3): 273–292.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica: Journal of the Econometric Society* pp. 817–838.

Zivot, E. and Wang, J. (2007). *Modeling Financial Time Series with S-PLUS®*, Vol. 191.

A Appendix

A.1 Alternative tests

A brief overview of the four alternative unit root tests is given in this appendix. The four tests are the Augmented DF test (Dickey and Fuller; 1981; Said and Dickey; 1984), The ERS Point Optimal and DF-GLS tests (Elliot, Rothenberg and Stock; 1996) and the PP test (Phillips and Perron; 1988).

A.1.1 The Augmented Dickey-Fuller test (ADF)

The ADF (Dickey and Fuller; 1981; Said and Dickey; 1984) augments the basic Dickey-Fuller tests with lagged differences in order to accommodate ARMA(p,q) dynamics, the order of which is usually unknown As a result, p in equation (A.1) is chosen to be sufficiently large so that ε_t is white noise. The ADF tests the null hypothesis that a time series is ARIMA$(p, 1, 0)$ against the alternative that it is a stationary AR$(p + 1)$ process. The tests regression is

$$\Delta y_t = \mu_t + \gamma y_{t-1} + \sum_{j=1}^{p} \alpha_j \Delta y_{t-j} + \varepsilon_t \tag{A.1}$$

Under the null hypothesis $\gamma = 0$ and y_t is $I(1)$. The test statistic (*pseudo-t* statistic) is the t statistic used for testing the significance of the coefficient of the lagged dependent variable y_{t-1}. The test statistic follows the DF (Dickey-Fuller) distribution and critical values are obtained through simulation.

A.1.2 Eliott, Rothenberg and Stock's Point Optimal Test (ERS-PT)

The ERS test uses efficient detrending to retain power when the dominant root is close to unity. For this test, under H_A, the dominant root (ρ) takes a value that is local to unity i.e.

$$H_{\mathrm{A}} : \rho = \rho_c = 1 + c/T < 1 \tag{A.2}$$

with $c < 0$ so that for $c = -10$ and $T = 100$, for example, $\rho_c = 0.90$. The test statistic is then asymptotically optimal for this alternative. Eliott, Rothenberg and Stock (op. cit) recommend to choose ($c = \bar{c}$) corresponding to the point on the power envelope that has 50% power.

The test is constructed as follows: Consider the DGP in common factor form as in Patterson (2009, p. 230-231)

$$y_t = \mu_t + u_t \tag{A.3}$$

$$(1 - \rho L)u_t = v_t \tag{A.4}$$

$$v_t = \varepsilon_t \tag{A.5}$$

so that,

$$y_1 = \mu_1 + u_1 \tag{A.6}$$

$$(1 - \rho L)y_t = (1 - \rho L)\mu_t + \varepsilon_t \qquad t = 2, \ldots, T \tag{A.7}$$

with u_1 assumed to be drawn from the unconditional distribution of u_t.

The dependent variable in equation (A.7) is a quasi-difference of y_t and the independent variable is a quasi-differenced trend. When ρ is evaluated at ρ_c, as in the alternative hypothesis ($\rho < 1$), then the dependent and independent (quasi-differenced) variables are $\boldsymbol{y}_c = \{y_t - \rho_c y_{t-1}\}$ and $\boldsymbol{X}_c = \{\mu_t - \rho_c \mu_{t-1}\}$, respectively.

Let $S(\rho_c)$ be the residual sum of squares from a regression of $\boldsymbol{y}_c$ on $\boldsymbol{X}_c$. Then the point optimal test against the alternative ρ_c is

$$P_T = [S(\rho_c) - \rho_c S(1)]/\hat{\lambda}^2 \tag{A.8}$$

where $\hat{\lambda}^2$ is a consistent estimate of the long-run variance of u_t. The critical values for this test are given by Eliott, Rothenberg and Stock (op. cit) for sample sizes $T = (50, 100, 200, \infty)$.

A.1.3 Dickey-Fuller Generalized Least Squares (DF-GLS)

Consider equation (A.3) and let the deterministic part be given as

$$\mu_t = \beta_0 + \beta_1 t \tag{A.9}$$

$$= [1 \;\; t]\boldsymbol{\beta} \tag{A.10}$$

for the linear trend case, with $\boldsymbol{\beta} = [\beta_0 \;\; \beta_1]^T$.

The quasi-differenced data for this linear trend case can be obtained in a way similar to that used in A.1.2 and the following detrending regression fitted

$$\boldsymbol{y}_c = \boldsymbol{X}_c \boldsymbol{\beta}_c + \boldsymbol{\nu}_c \tag{A.11}$$

The detrending parameters can $\boldsymbol{\beta}_c$, that are estimated under the alternative can then be used to GLS-detrend the data

$$\boldsymbol{y}_t^d = \boldsymbol{y}_t - \boldsymbol{X}_c \boldsymbol{\beta}_c \tag{A.12}$$

The ADF regression is then estimated on the GLS-detrended data

$$\Delta {y_t}^d = \gamma y_{t-1}^d + \sum_{j=1}^{p} \alpha_j \Delta y_{t-j}^d + \varepsilon \tag{A.13}$$

The test regression does not include the deterministic terms because these have been accounted for by the detrending. The test for the unit root proceeds using the ADF test statistic as explained in section A.1.1.

A.1.4 Phillips-Perron (PP) unit root test

The Phillips and Perron unit root test addresses the problems of serial correlation when the underlying DGP may not be AR(1). The test is a direct modification the ADF test statistic. Instead of augmenting the basic Dickey-Fuller regression with lagged differences of the dependent variable to whiten the residuals, the Phillips-Perron test uses the Newey-West heteroscedasticity and autocorrelation consistent (HAC) estimator of the long-run variance to make a correction to the basic Dickey-Fuller test statistic.

The test regression for the PP test is given as follows

$$\Delta y_t = \mu_t + \gamma y_{t-1} + u_t \tag{A.14}$$

where u_t may be autocorrelated and possibly heteroscedastic.

The PP test corrects for the autocorrelation and heteroscedasticity in a non-parametric way using the modified DF statistics Z_μ and Z_τ which correspond to the DF normalized bias and t-statistic respectively. Phillips and Perron (op. cit) show that, under the null hypothesis, each of the Z statistics has a limiting distribution that is the same as its corresponding DF statistic.

The modified test statistics are given as (Patterson; 2000; Zivot and Wang; 2007, p. 264; p. 127)

$$Z_\mu = T\hat{\gamma} - \frac{1}{2}\left(\frac{T^2 \times SE(\hat{\gamma}^2)}{\hat{\sigma}^2}\right)(\hat{\lambda}^2 - \hat{\sigma}^2) \tag{A.15}$$

$$Z_\tau = \frac{\hat{\sigma}}{\hat{\lambda}} t_{\gamma=0} - \frac{1}{2}\left(\frac{\hat{\lambda}^2 - \hat{\sigma}^2}{\hat{\lambda}^2}\right)\left(\frac{T \times SE(\hat{\gamma}^2)}{\hat{\sigma}^2}\right) \tag{A.16}$$

where

$$\begin{aligned}
\hat{\sigma}^2 &= T^{-1}\sum_{t=1}^{T}\hat{u}^2 \\
\hat{\lambda}^2 &= \hat{\sigma}^2 + 2\sum_{s=1}^{l} w_{sl} \sum_{t=s+1}^{T} \hat{\varepsilon}_t\hat{\varepsilon}_{t-s}/T \\
w_{sl} &= 1 - s/(l+1)
\end{aligned}$$

$\hat{\sigma}^2$ (sample variance of the least squares residuals) and $\hat{\lambda}^2$ (the Newey-West long-run variance estimate) are consistent estimates of the short- and long-run variances of u_t respectively. The truncation lag for the covariances, l, can be chosen using the autocorrelation function.

The critical values for the PP test are the same as those of the ADF test for large sample sizes.

A.2 Test sizes for all the considered DGPs

Table A.1: Test sizes for series with GARCH(1,1) errors with N(0,1) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0490	0.0479	0.0479	0.0808	0.0889	0.0348
128	0.0462	0.0462	0.0471	0.0671	0.0660	0.0357
256	0.0522	0.0535	0.0484	0.0600	0.0582	0.0393
512	0.0540	0.0502	0.0483	0.0524	0.0515	0.0397
1024	0.0529	0.0456	0.0510	0.0560	0.0495	0.0405
GARCH parameters: $\alpha_0 = 0.775, \alpha_1 = 0.005$, and $(\alpha_1 + \gamma_1) = 0.245$						
64	0.0504	0.0480	0.0524	0.0810	0.0902	0.0318
128	0.0545	0.0523	0.0477	0.0680	0.0717	0.0330
256	0.0507	0.0456	0.0511	0.0580	0.0612	0.0388
512	0.0481	0.0523	0.0456	0.0570	0.0576	0.0411
1024	0.0477	0.0500	0.0484	0.0564	0.0493	0.0409
GARCH parameters: $\alpha_0 = 0.505, \alpha_1 = 0.005$, and $(\alpha_1 + \gamma_1) = 0.495$						
64	0.0482	0.0491	0.0495	0.0838	0.0935	0.0288
128	0.0507	0.0505	0.0510	0.0708	0.0679	0.0326
256	0.0496	0.0452	0.0509	0.0571	0.0614	0.0361
512	0.0520	0.0500	0.0474	0.0576	0.0558	0.0414
1024	0.0510	0.0553	0.0477	0.0498	0.0522	0.0384
GARCH parameters: $\alpha_0 = 0.255, \alpha_1 = 0.005$, and $(\alpha_1 + \gamma_1) = 0.745$						
64	0.0508	0.0529	0.0485	0.0814	0.0906	0.0325

Table A.1: Test sizes for series with GARCH(1,1) errors with N(0,1) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
128	0.0447	0.0483	0.0459	0.0688	0.0653	0.0316
256	0.0526	0.0531	0.0509	0.0605	0.0645	0.0381
512	0.0500	0.0515	0.0478	0.0553	0.0501	0.0424
1024	0.0451	0.0441	0.0553	0.0492	0.0475	0.0396
GARCH parameters: $\alpha_0 = 0.005, \alpha_1 = 0.005$, and $(\alpha_1 + \gamma_1) = 0.995$						
64	0.0484	0.0490	0.0485	0.0843	0.0904	0.0300
128	0.0489	0.0477	0.0528	0.0652	0.0659	0.0370
256	0.0487	0.0517	0.0502	0.0602	0.0588	0.0370
512	0.0509	0.0570	0.0561	0.0597	0.0574	0.0461
1024	0.0485	0.0496	0.0483	0.0528	0.0505	0.0464
GARCH parameters: $\alpha_0 = 0.745, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.255$						
64	0.0455	0.0537	0.0440	0.0784	0.0915	0.0349
128	0.0548	0.0436	0.0470	0.0699	0.0635	0.0339
256	0.0496	0.0483	0.0546	0.0586	0.0583	0.0388
512	0.0440	0.0513	0.0516	0.0574	0.0542	0.0416
1024	0.0499	0.0471	0.0516	0.0482	0.0525	0.0429
GARCH parameters: $\alpha_0 = 0.051, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.490$						
64	0.0476	0.0540	0.0524	0.0789	0.0903	0.0384
128	0.0512	0.0462	0.0536	0.0642	0.0657	0.0385
256	0.0501	0.0529	0.0539	0.0597	0.0591	0.0409
512	0.0548	0.0502	0.0512	0.0606	0.0557	0.0388

Table A.1: Test sizes for series with GARCH(1,1) errors with N(0,1) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
1024	0.0494	0.0499	0.0503	0.0526	0.0508	0.0446
GARCH parameters: $\alpha_0 = 0.260, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.740$						
64	0.0454	0.0503	0.0529	0.0841	0.0945	0.0413
128	0.0475	0.0503	0.0510	0.0700	0.0698	0.0406
256	0.0494	0.0515	0.0548	0.0638	0.0617	0.0431
512	0.0499	0.0496	0.0504	0.0609	0.0534	0.0404
1024	0.0525	0.0538	0.0495	0.0540	0.0538	0.0424
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0535	0.0489	0.0752	0.0997	0.1092	0.0523
128	0.0469	0.0534	0.0828	0.0878	0.0905	0.0681
256	0.0493	0.0479	0.0891	0.0792	0.0831	0.0735
512	0.0544	0.0507	0.0896	0.0787	0.0810	0.0724
1024	0.0505	0.0468	0.0902	0.0711	0.0723	0.0789
GARCH parameters: $\alpha_0 = 0.495, \alpha_1 = 0.500$, and $(\alpha_1 + \gamma_1) = 0.505$						
64	0.0486	0.0510	0.0519	0.0811	0.0922	0.0420
128	0.0488	0.0485	0.0577	0.0686	0.0699	0.0368
256	0.0486	0.0482	0.0471	0.0603	0.0613	0.0377
512	0.0448	0.0522	0.0523	0.0552	0.0532	0.0434
1024	0.0520	0.0502	0.0500	0.0490	0.0453	0.0460
GARCH parameters: $\alpha_0 = 0.260, \alpha_1 = 0.500$, and $(\alpha_1 + \gamma_1) = 0.740$						
64	0.0500	0.0487	0.0572	0.0792	0.0943	0.0471

Table A.1: Test sizes for series with GARCH(1,1) errors with N(0,1) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
128	0.0532	0.0497	0.0582	0.0698	0.0700	0.0406
256	0.0482	0.0508	0.0589	0.0623	0.0619	0.0459
512	0.0459	0.0448	0.0576	0.0556	0.0580	0.0478
1024	0.0504	0.0500	0.0544	0.0531	0.0537	0.0459
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.500$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0525	0.0515	0.0796	0.1039	0.1091	0.0715
128	0.0469	0.0487	0.0890	0.0891	0.0914	0.0692
256	0.0465	0.0482	0.0920	0.0850	0.0843	0.0775
512	0.0534	0.0488	0.0957	0.0785	0.0770	0.0743
1024	0.0532	0.0537	0.0882	0.0751	0.0738	0.0787
GARCH parameters: $\alpha_0 = 0.245, \alpha_1 = 0.750$, and $(\alpha_1 + \gamma_1) = 0.755$						
64	0.0460	0.0458	0.0559	0.0763	0.0895	0.0497
128	0.0461	0.0448	0.0584	0.0661	0.0695	0.0491
256	0.0517	0.0465	0.0575	0.0611	0.0550	0.0501
512	0.0492	0.0466	0.0556	0.0565	0.0541	0.0503
1024	0.0505	0.0463	0.0540	0.0538	0.0507	0.0466
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.750$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0537	0.0521	0.0758	0.0909	0.1034	0.0703
128	0.0467	0.0483	0.0841	0.0821	0.0762	0.0712
256	0.0469	0.0502	0.0834	0.0736	0.0714	0.0689
512	0.0494	0.0512	0.0815	0.0748	0.0664	0.0667

Table A.1: Test sizes for series with GARCH(1,1) errors with N(0,1) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
1024	0.0525	0.0500	0.0773	0.0597	0.0656	0.0717
GARCH parameters: $\alpha_0 = 0.005, \alpha_1 = 0.990$, and $(\alpha_1 + \gamma_1) = 0.995$						
64	0.0467	0.0557	0.0660	0.0818	0.0926	0.0605
128	0.0519	0.0436	0.0693	0.0698	0.0723	0.0639
256	0.0501	0.0510	0.0696	0.0604	0.0670	0.0621
512	0.0512	0.0470	0.0707	0.0598	0.0613	0.0603
1024	0.0484	0.0466	0.0688	0.0572	0.0578	0.0628

Table A.2: Test sizes for series with GARCH(1,1) errors with GED(ν = 1.5) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0557	0.0498	0.0478	0.0887	0.0918	0.0329
128	0.0443	0.0545	0.0503	0.0702	0.0688	0.0323
256	0.0482	0.0542	0.0502	0.0622	0.0570	0.0395
512	0.0495	0.0479	0.0503	0.0567	0.0509	0.0397
1024	0.0505	0.0438	0.0494	0.0495	0.0529	0.0413
GARCH parameters: $\alpha_0 = 0.775, \alpha_1 = 0.005$, and $(\alpha_1 + \gamma_1) = 0.245$						
64	0.0528	0.0519	0.0485	0.0827	0.0904	0.0332
128	0.0580	0.0514	0.0485	0.0722	0.0635	0.0398
256	0.0517	0.0486	0.0494	0.0563	0.0595	0.0375
512	0.0470	0.0547	0.0430	0.0539	0.0528	0.0412
1024	0.0519	0.0516	0.0483	0.0537	0.0462	0.0449
GARCH parameters: $\alpha_0 = 0.505, \alpha_1 = 0.005$, and $(\alpha_1 + \gamma_1) = 0.495$						
64	0.0527	0.0480	0.0499	0.0840	0.0962	0.0315
128	0.0469	0.0567	0.0478	0.0729	0.0703	0.0362
256	0.0472	0.0474	0.0499	0.0581	0.0593	0.0374
512	0.0512	0.0531	0.0493	0.0517	0.0536	0.0441
1024	0.0484	0.0467	0.0504	0.0503	0.0468	0.0435
GARCH parameters: $\alpha_0 = 0.255, \alpha_1 = 0.005$, and $(\alpha_1 + \gamma_1) = 0.745$						
64	0.0507	0.0466	0.0495	0.0859	0.0983	0.0330
128	0.0500	0.0486	0.0503	0.0686	0.0679	0.0359

Table A.2: Test sizes for series with GARCH(1,1) errors with GED($\nu = 1.5$) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
256	0.0448	0.0511	0.0510	0.0566	0.0593	0.0347
512	0.0489	0.0499	0.0506	0.0526	0.0548	0.0364
1024	0.0562	0.0549	0.0506	0.0516	0.0500	0.0418
GARCH parameters: $\alpha_0 = 0.005, \alpha_1 = 0.005$, and $(\alpha_1 + \gamma_1) = 0.995$						
64	0.0529	0.0503	0.0421	0.0851	0.0930	0.0315
128	0.0530	0.0426	0.0496	0.0676	0.0666	0.0364
256	0.0522	0.0530	0.0507	0.0626	0.0620	0.0376
512	0.0526	0.0500	0.0494	0.0520	0.0509	0.0409
1024	0.0486	0.0492	0.0530	0.0475	0.0519	0.0439
GARCH parameters: $\alpha_0 = 0.745, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.255$						
64	0.0467	0.0499	0.0509	0.0812	0.0928	0.0335
128	0.0543	0.0453	0.0481	0.0673	0.0674	0.0412
256	0.0519	0.0552	0.0467	0.0531	0.0557	0.0398
512	0.0496	0.0467	0.0503	0.0501	0.0517	0.0395
1024	0.0528	0.0536	0.0513	0.0505	0.0531	0.0402
GARCH parameters: $\alpha_0 = 0.051, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.490$						
64	0.0491	0.0492	0.0524	0.0849	0.0904	0.0374
128	0.0520	0.0485	0.0475	0.0673	0.0640	0.0375
256	0.0498	0.0491	0.0583	0.0543	0.0591	0.0406
512	0.0445	0.0514	0.0539	0.0569	0.0559	0.0378
1024	0.0521	0.0505	0.0517	0.0535	0.0507	0.0430

Table A.2: Test sizes for series with GARCH(1,1) errors with GED($\nu =$ 1.5) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
GARCH parameters: $\alpha_0 = 0.260, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.740$						
64	0.0489	0.0457	0.0550	0.0831	0.0929	0.0441
128	0.0491	0.0493	0.0538	0.0702	0.0671	0.0454
256	0.0478	0.0493	0.0532	0.0598	0.0624	0.0412
512	0.0544	0.0463	0.0510	0.0520	0.0540	0.0425
1024	0.0506	0.0473	0.0507	0.0518	0.0513	0.0412
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0486	0.0554	0.0788	0.0975	0.1060	0.0551
128	0.0473	0.0567	0.0866	0.0877	0.0837	0.0642
256	0.0547	0.0500	0.0928	0.0878	0.0809	0.0732
512	0.0462	0.0536	0.0945	0.0837	0.0795	0.0803
1024	0.0532	0.0426	0.0967	0.0790	0.0764	0.0776
GARCH parameters: $\alpha_0 = 0.495, \alpha_1 = 0.500$, and $(\alpha_1 + \gamma_1) = 0.055$						
64	0.0467	0.0566	0.0543	0.0750	0.0877	0.0378
128	0.0499	0.0462	0.0515	0.0645	0.0652	0.0429
256	0.0496	0.0506	0.0530	0.0583	0.0604	0.0420
512	0.0493	0.0497	0.0487	0.0530	0.0544	0.0399
1024	0.0453	0.0520	0.0538	0.0511	0.0468	0.0453
GARCH parameters: $\alpha_0 = 0.260, \alpha_1 = 0.500$, and $(\alpha_1 + \gamma_1) = 0.740$						
64	0.0486	0.0498	0.0586	0.0786	0.0982	0.0480
128	0.0478	0.0478	0.0563	0.0727	0.0718	0.0436

Table A.2: Test sizes for series with GARCH(1,1) errors with GED($\nu = 1.5$) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
256	0.0481	0.0463	0.0558	0.0602	0.0621	0.0492
512	0.0501	0.0452	0.0586	0.0529	0.0555	0.0460
1024	0.0576	0.0497	0.0545	0.0533	0.0516	0.0444
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.500$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0486	0.0474	0.0853	0.1016	0.1077	0.0646
128	0.0533	0.0472	0.0846	0.0899	0.0866	0.0685
256	0.0509	0.0538	0.0930	0.0824	0.0829	0.0708
512	0.0479	0.0480	0.0921	0.0784	0.0770	0.0745
1024	0.0524	0.0460	0.0916	0.0725	0.0768	0.0734
GARCH parameters: $\alpha_0 = 0.245, \alpha_1 = 0.750$, and $(\alpha_1 + \gamma_1) = 0.755$						
64	0.0457	0.0514	0.0584	0.0790	0.0883	0.0506
128	0.0467	0.0510	0.0577	0.0715	0.0656	0.0483
256	0.0522	0.0477	0.0579	0.0592	0.0615	0.0466
512	0.0473	0.0429	0.0550	0.0610	0.0589	0.0473
1024	0.0459	0.0511	0.0567	0.0563	0.0523	0.0477
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.750$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0476	0.0468	0.0804	0.0986	0.1036	0.0667
128	0.0492	0.0488	0.0815	0.0826	0.0832	0.0690
256	0.0514	0.0423	0.0784	0.0720	0.0710	0.0713
512	0.0521	0.0531	0.0812	0.0658	0.0672	0.0694
1024	0.0487	0.0492	0.0826	0.0647	0.0689	0.0640

Table A.2: Test sizes for series with GARCH(1,1) errors with GED(ν = 1.5) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
GARCH parameters: $\alpha_0 = 0.005, \alpha_1 = 0.990$, and $(\alpha_1 + \gamma_1) = 0.995$						
64	0.0500	0.0501	0.0669	0.0810	0.0925	0.0634
128	0.0495	0.0482	0.0698	0.0699	0.0722	0.0653
256	0.0505	0.0537	0.0714	0.0580	0.0663	0.0593
512	0.0551	0.0527	0.0697	0.0592	0.0598	0.0620
1024	0.0467	0.0490	0.0702	0.0567	0.0573	0.0610

Table A.3: Test sizes for series with GARCH(1,1) errors with GED($\nu=1.5$) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.250$, and $(\alpha_1+\gamma_1) = 0.990$						
64	0.0483	0.0429	0.0523	0.0843	0.0879	0.0329
128	0.0527	0.051	0.0507	0.0655	0.0653	0.0311
256	0.0487	0.0487	0.048	0.0548	0.0575	0.0349
512	0.0506	0.0451	0.0483	0.0554	0.0527	0.0349
1024	0.0539	0.0514	0.0469	0.0481	0.0526	0.0322
GARCH parameters: $\alpha_0 = 0.775, \alpha_1 = 0.005$, and $(\alpha_1+\gamma_1) = 0.245$						
64	0.0551	0.0514	0.0549	0.083	0.0881	0.0353
128	0.05	0.0564	0.0459	0.0662	0.0602	0.0337
256	0.0499	0.0518	0.0504	0.0557	0.0589	0.0340
512	0.0483	0.0438	0.0511	0.0594	0.0499	0.0327
1024	0.0498	0.0483	0.0468	0.0475	0.0519	0.0350
GARCH parameters: $\alpha_0 = 0.505, \alpha_1 = 0.005$, and $(\alpha_1+\gamma_1) = 0.495$						
64	0.0572	0.047	0.0486	0.0783	0.0874	0.0354
128	0.0563	0.0479	0.0462	0.0634	0.064	0.0351
256	0.0488	0.0455	0.053	0.0584	0.0609	0.0313
512	0.0537	0.048	0.0544	0.057	0.0516	0.0317
1024	0.0507	0.0533	0.046	0.0515	0.0481	0.0330
GARCH parameters: $\alpha_0 = 0.255, \alpha_1 = 0.005$, and $(\alpha_1+\gamma_1) = 0.745$						
64	0.0476	0.0505	0.0523	0.0841	0.0864	0.0320
128	0.0516	0.0499	0.0506	0.0664	0.0672	0.0350

Table A.3: Test sizes for series with GARCH(1,1) errors with GED(ν = 1.5) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
256	0.0574	0.0544	0.047	0.0606	0.0574	0.0335
512	0.0525	0.0531	0.049	0.0513	0.0563	0.0361
1024	0.0542	0.0569	0.0483	0.0536	0.0514	0.0361
GARCH parameters: $\alpha_0 = 0.005, \alpha_1 = 0.005$, and $(\alpha_1 + \gamma_1) = 0.995$						
64	0.0499	0.0524	0.0479	0.0827	0.0897	0.0344
128	0.0534	0.0519	0.0503	0.0694	0.0647	0.0333
256	0.0522	0.0505	0.0552	0.0563	0.0546	0.0354
512	0.0469	0.0508	0.0521	0.0549	0.0563	0.0381
1024	0.0531	0.0509	0.0521	0.0524	0.0506	0.0374
GARCH parameters: $\alpha_0 = 0.745, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.255$						
64	0.054	0.0506	0.0511	0.0811	0.0891	0.0398
128	0.0526	0.0593	0.0500	0.0632	0.0618	0.0341
256	0.0539	0.0505	0.0509	0.0622	0.0589	0.0315
512	0.0509	0.0503	0.0492	0.0548	0.0492	0.0354
1024	0.0481	0.0508	0.0506	0.0482	0.0512	0.0366
GARCH parameters: $\alpha_0 = 0.051, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.490$						
64	0.0551	0.0557	0.0550	0.0807	0.0859	0.0396
128	0.0526	0.0572	0.0512	0.0713	0.0642	0.0406
256	0.0556	0.0551	0.0539	0.0587	0.059	0.0371
512	0.0553	0.0506	0.0514	0.0557	0.0554	0.0383
1024	0.0527	0.0484	0.0517	0.0522	0.0532	0.0332

Table A.3: Test sizes for series with GARCH(1,1) errors with GED(ν = 1.5) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
GARCH parameters: $\alpha_0 = 0.260, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.740$						
64	0.0561	0.0577	0.0626	0.0878	0.0970	0.0411
128	0.0552	0.0584	0.0578	0.0728	0.0707	0.0444
256	0.0560	0.0603	0.0598	0.0610	0.0619	0.0415
512	0.0563	0.0599	0.0555	0.0553	0.0579	0.0390
1024	0.0567	0.0544	0.0557	0.0529	0.0532	0.0359
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.250$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0532	0.0538	0.0769	0.1015	0.1058	0.0615
128	0.0569	0.0586	0.0897	0.0857	0.0869	0.0657
256	0.0545	0.0586	0.0938	0.0826	0.0868	0.0705
512	0.0516	0.0582	0.0955	0.0818	0.0771	0.0721
1024	0.0571	0.0546	0.0931	0.0779	0.0723	0.0740
GARCH parameters: $\alpha_0 = 0.495, \alpha_1 = 0.500$, and $(\alpha_1 + \gamma_1) = 0.505$						
64	0.0576	0.0534	0.0556	0.0723	0.0792	0.0442
128	0.0507	0.0556	0.0531	0.0638	0.0603	0.0401
256	0.0532	0.0507	0.0497	0.0537	0.0571	0.0366
512	0.0489	0.0520	0.0511	0.0552	0.0505	0.0391
1024	0.0525	0.0614	0.0474	0.0507	0.0461	0.0352
GARCH parameters: $\alpha_0 = 0.260, \alpha_1 = 0.500$, and $(\alpha_1 + \gamma_1) = 0.740$						
64	0.0560	0.0510	0.0603	0.0855	0.0920	0.0513
128	0.0557	0.0539	0.0589	0.0706	0.0710	0.0488

Table A.3: Test sizes for series with GARCH(1,1) errors with GED($\nu = 1.5$) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
256	0.0524	0.0508	0.0599	0.0623	0.0612	0.0456
512	0.0589	0.0572	0.0569	0.0500	0.0540	0.0422
1024	0.0522	0.0643	0.0561	0.0541	0.0498	0.0417
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.500$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0542	0.0531	0.0782	0.0957	0.1031	0.0639
128	0.0468	0.0474	0.0851	0.0877	0.0829	0.0667
256	0.0506	0.0497	0.0881	0.0774	0.0778	0.0766
512	0.0494	0.0499	0.0909	0.0718	0.0737	0.0718
1024	0.0422	0.0492	0.0850	0.0664	0.0690	0.0702
GARCH parameters: $\alpha_0 = 0.245, \alpha_1 = 0.750$, and $(\alpha_1 + \gamma_1) = 0.755$						
64	0.0512	0.0463	0.0563	0.0720	0.0824	0.0537
128	0.0539	0.0530	0.0555	0.0675	0.0657	0.0505
256	0.0493	0.0534	0.0559	0.0562	0.0601	0.0462
512	0.0543	0.0474	0.0555	0.0525	0.0532	0.0478
1024	0.0483	0.0522	0.0558	0.0493	0.0501	0.0429
GARCH parameters: $\alpha_0 = 0.010, \alpha_1 = 0.750$, and $(\alpha_1 + \gamma_1) = 0.990$						
64	0.0525	0.0451	0.0756	0.0919	0.1014	0.0624
128	0.0481	0.0546	0.0721	0.0771	0.0766	0.0655
256	0.0437	0.0439	0.0776	0.0689	0.0688	0.0681
512	0.0415	0.0434	0.0730	0.0616	0.0583	0.0644
1024	0.0443	0.0441	0.0738	0.0600	0.0621	0.0622

Table A.3: Test sizes for series with GARCH(1,1) errors with GED(ν = 1.5) conditional distribution

T	VR-DWT	VR-MODWT	ADF	DF-GLS	ERS-PT	PP
GARCH parameters: $\alpha_0 = 0.005, \alpha_1 = 0.990$, and $(\alpha_1 + \gamma_1) = 0.995$						
64	0.0452	0.0441	0.0658	0.0733	0.0872	0.0593
128	0.0435	0.0452	0.0645	0.0661	0.0671	0.0564
256	0.0432	0.0453	0.0654	0.0581	0.0603	0.0620
512	0.0456	0.0400	0.0661	0.0542	0.0530	0.0556
1024	0.0441	0.0475	0.0626	0.0523	0.0495	0.0581

A.3 Power Curves for selected DGPs

Figure A.1: N(0,1); $\alpha_0 = 0.01$; $\alpha_1 = 0.5$; $(\alpha_1 + \gamma_1) = 0.99$

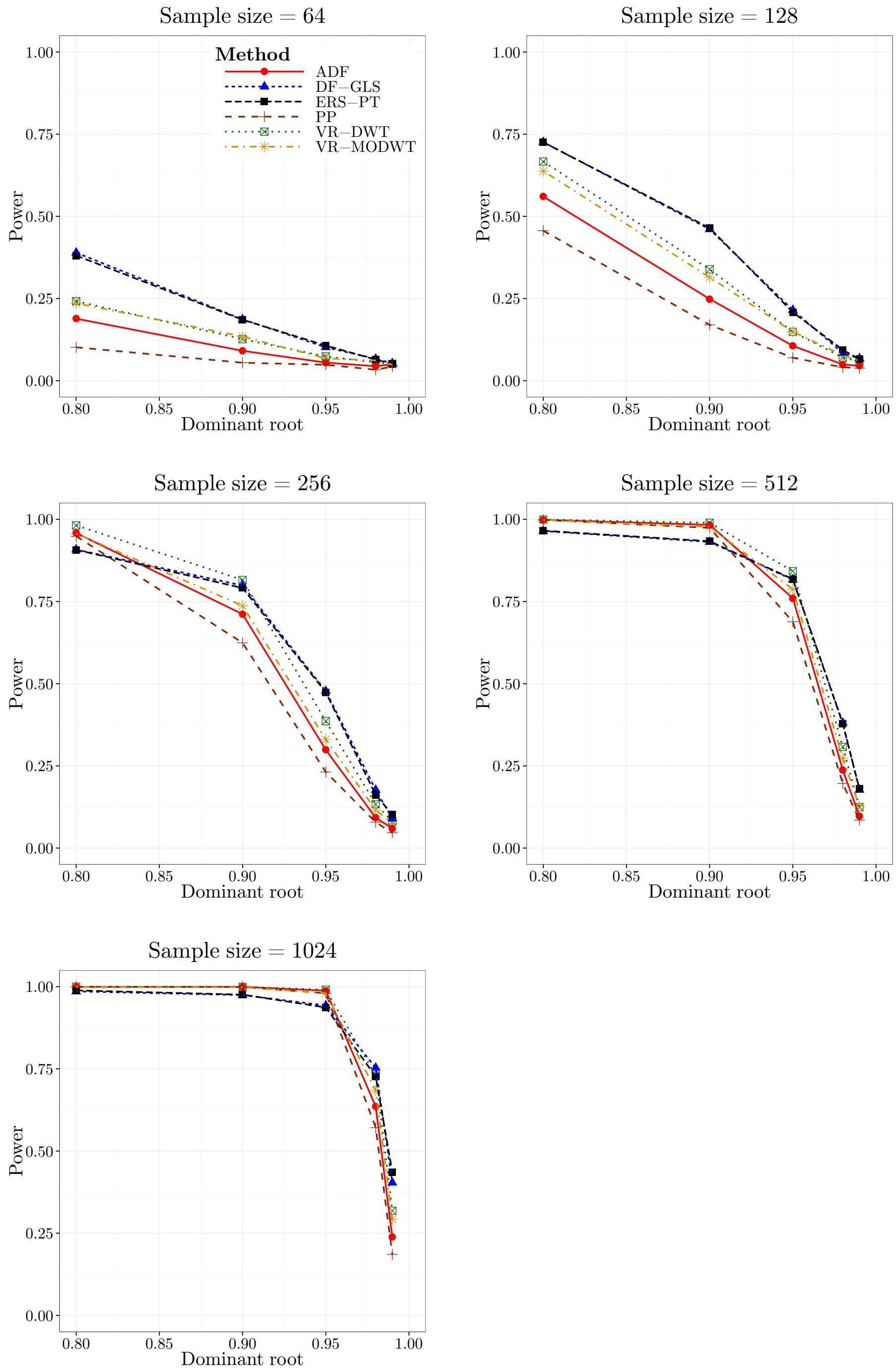


Figure A.2: GED ($\nu = 1.5$); $\alpha_0 = 0.01$; $\alpha_1 = 0.5$; $(\alpha_1 + \gamma_1) = 0.99$

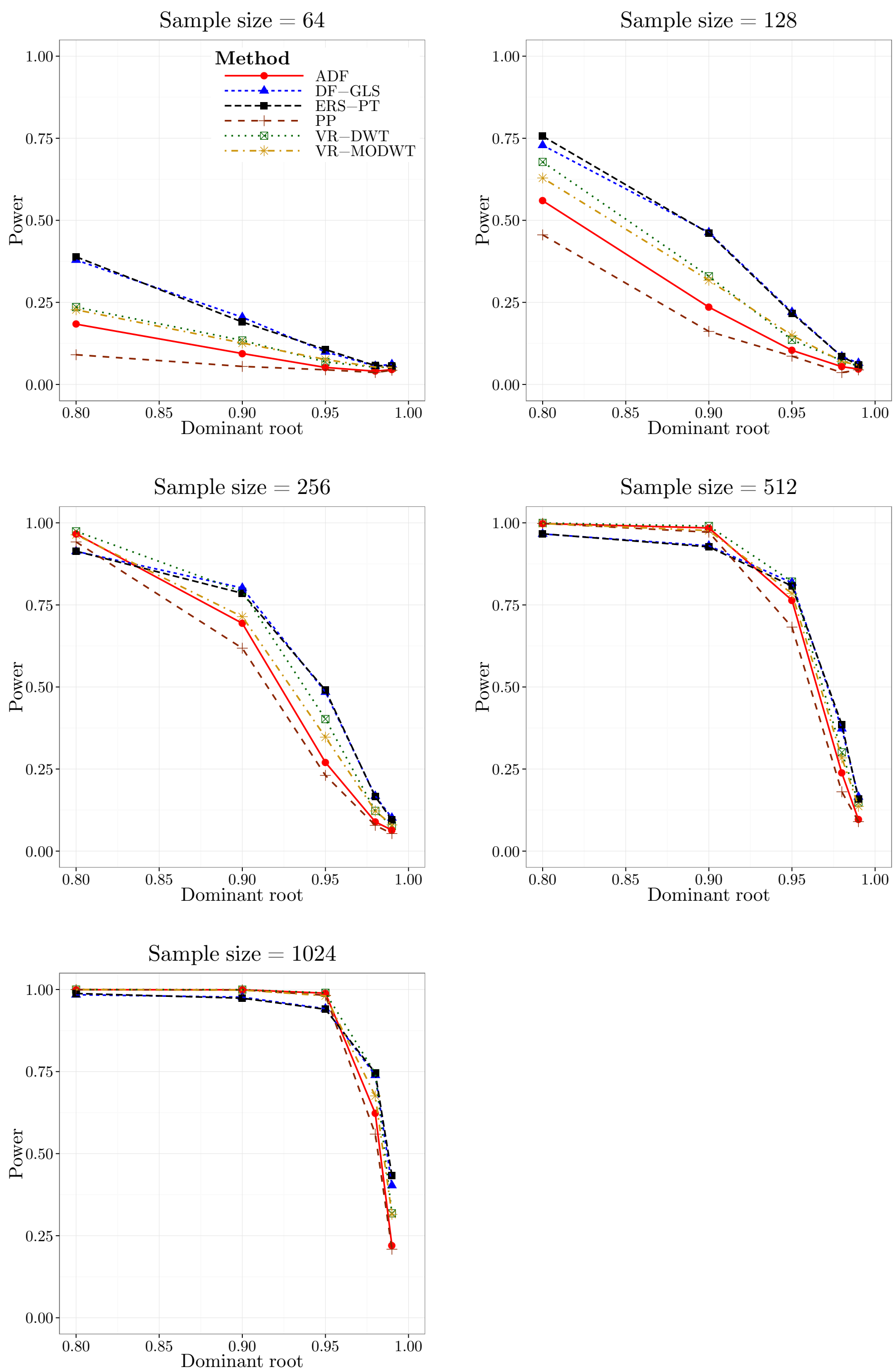


Figure A.3: $t(4)$; $\alpha_0 = 0.01$; $\alpha_1 = 0.5$; $(\alpha_1 + \gamma_1) = 0.99$

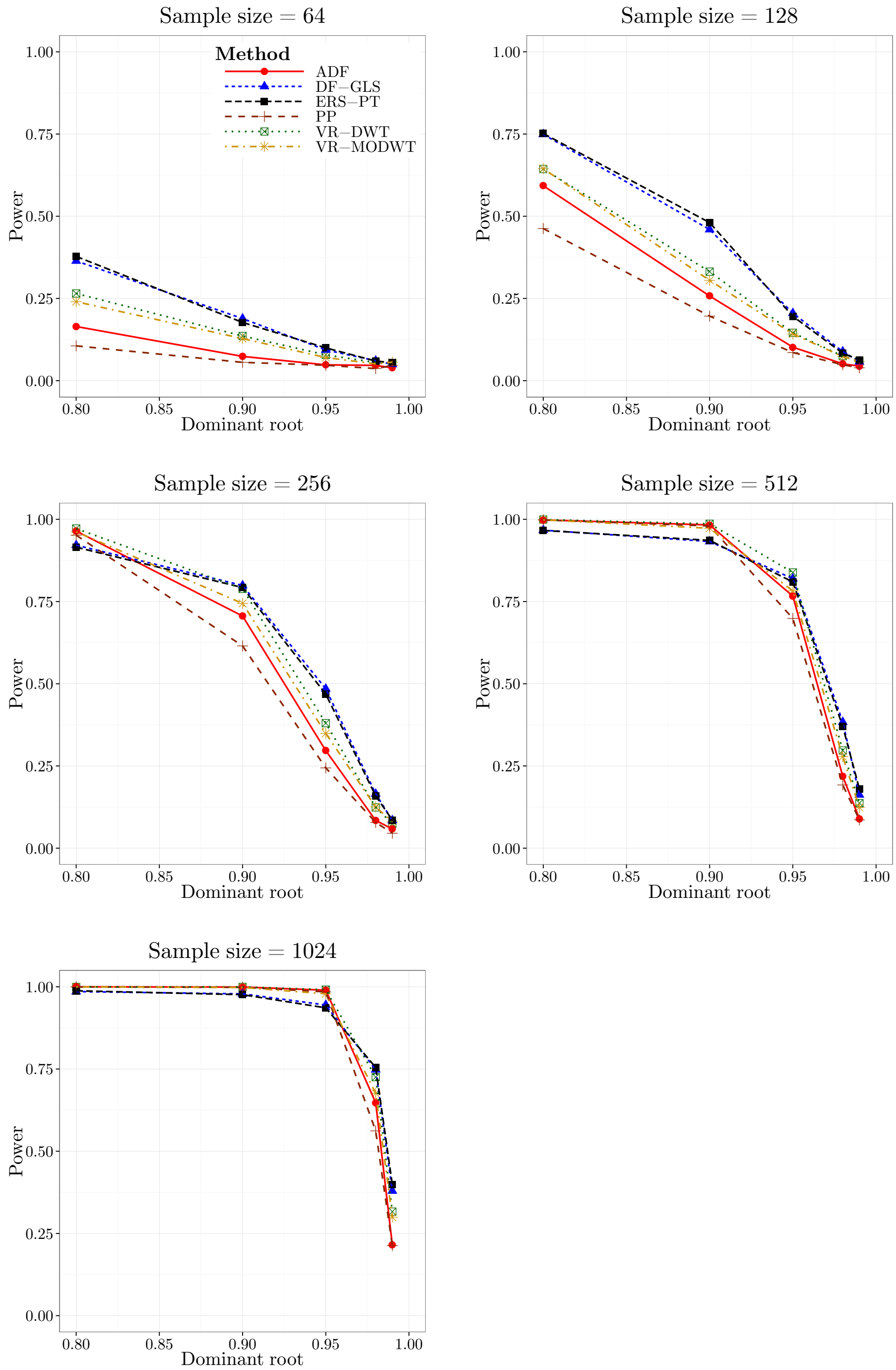


Figure A.4: N(0,1); $\alpha_0 = 0.01$; $\alpha_1 = 0.75$; $(\alpha_1 + \gamma_1) = 0.99$

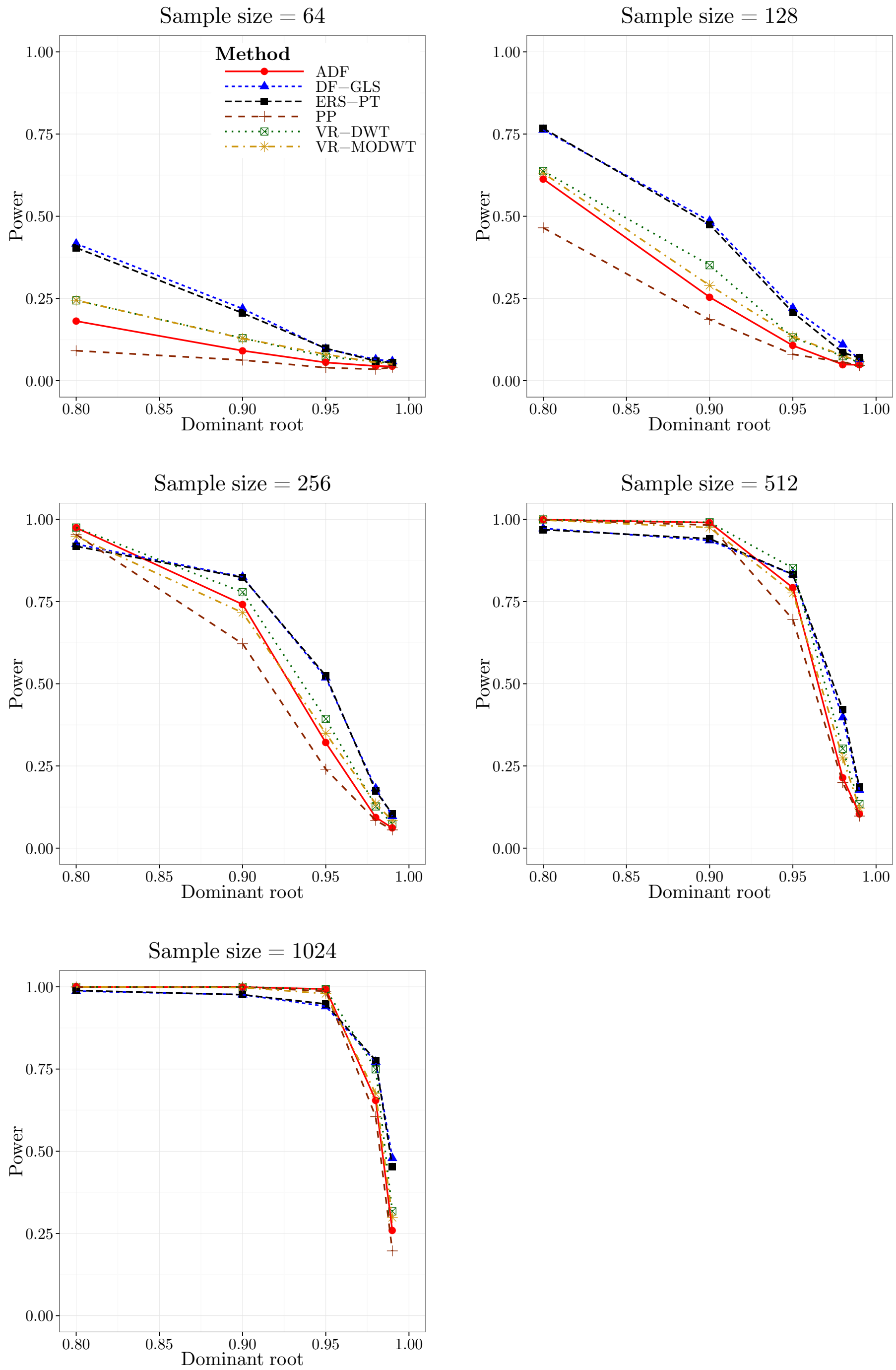


Figure A.5: GED ($\nu = 1.5$); $\alpha_0 = 0.01$; $\alpha_1 = 0.75$; $(\alpha_1 + \gamma_1) = 0.99$

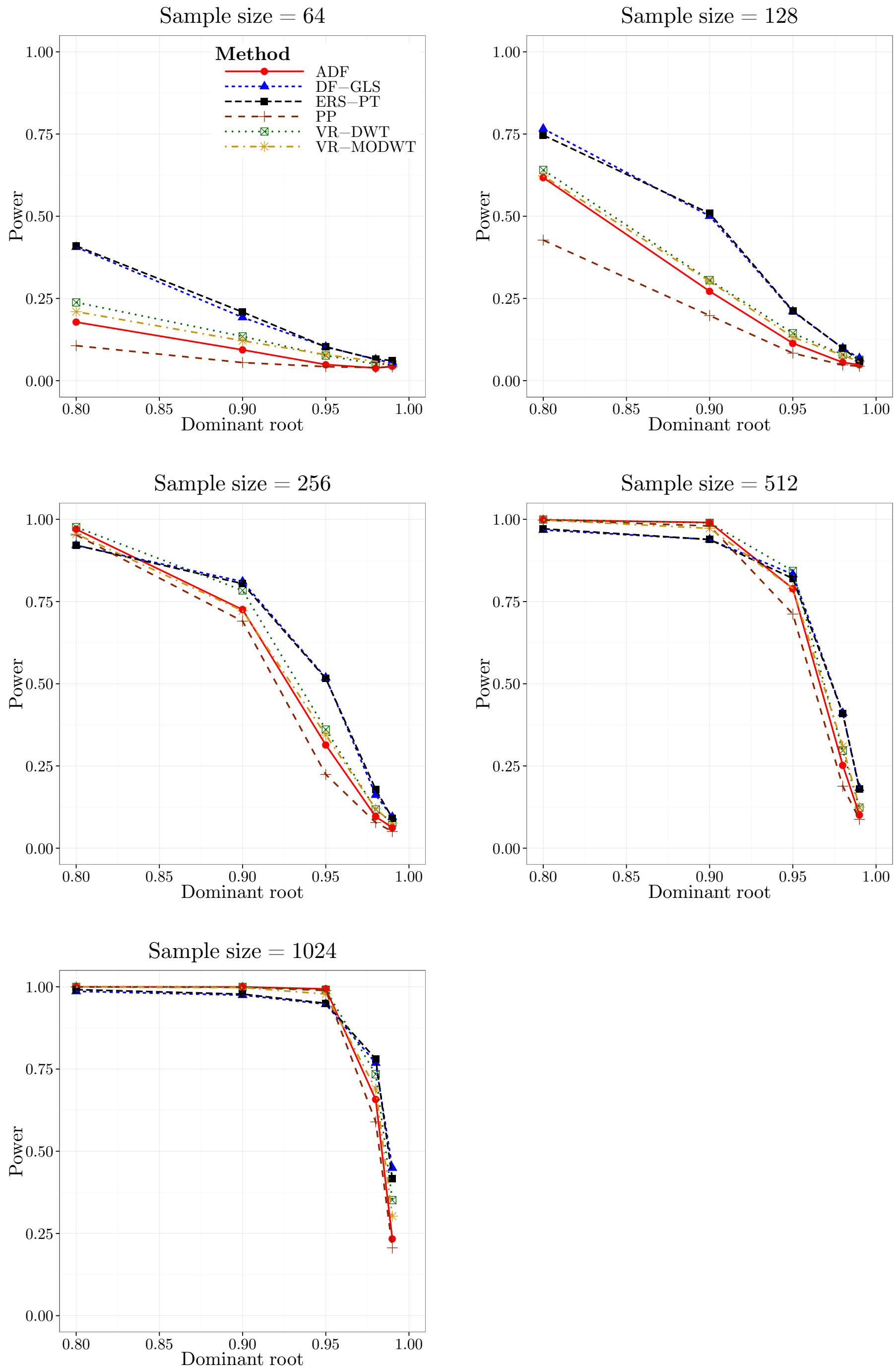


Figure A.6: $t(4)$; $\alpha_0 = 0.01$; $\alpha_1 = 0.75$; $(\alpha_1 + \gamma_1) = 0.99$

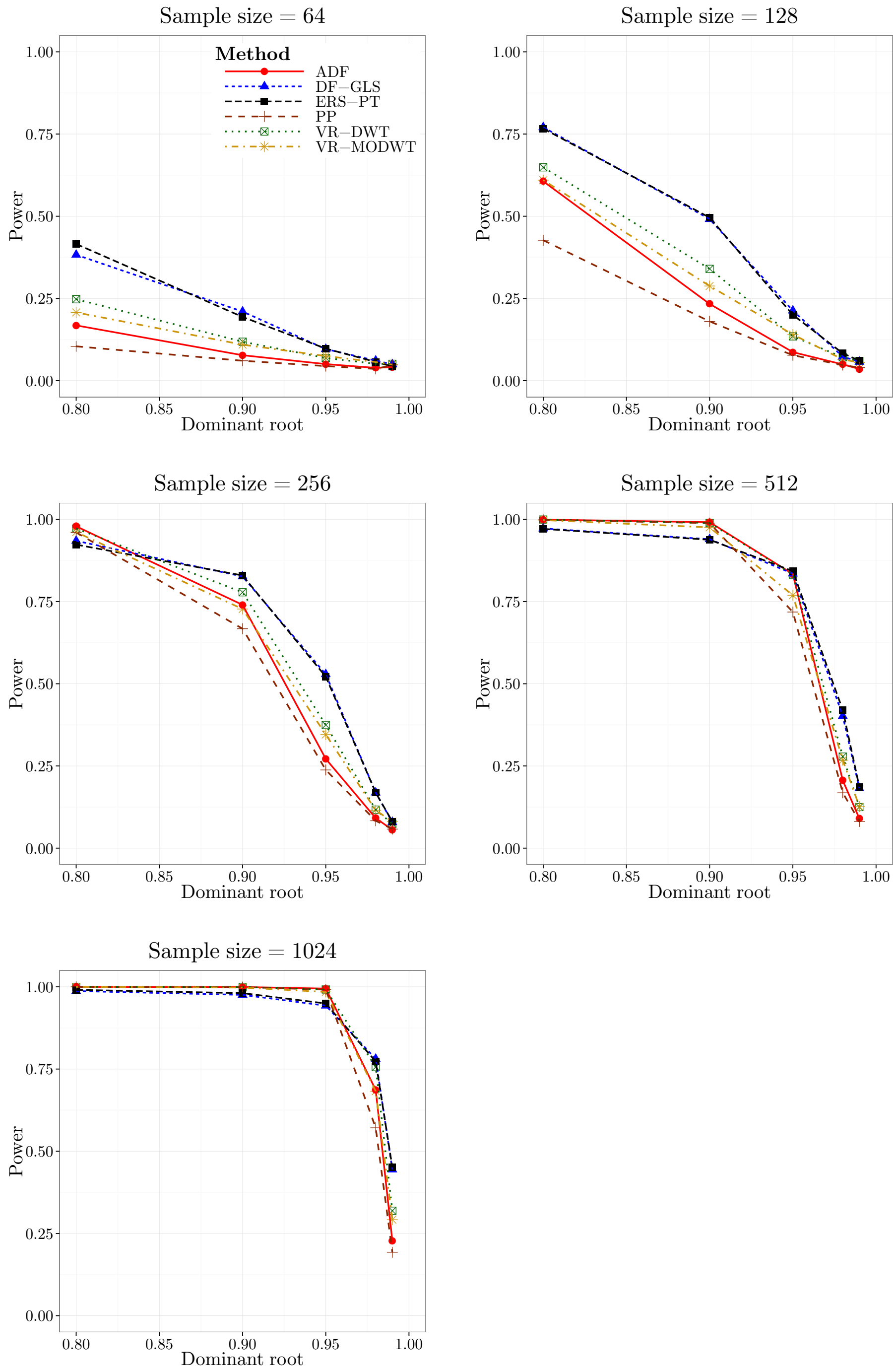


Figure A.7: N(0,1); $\alpha_0 = 0.005$; $\alpha_1 = 0.99$; $(\alpha_1 + \gamma_1) = 0.995$

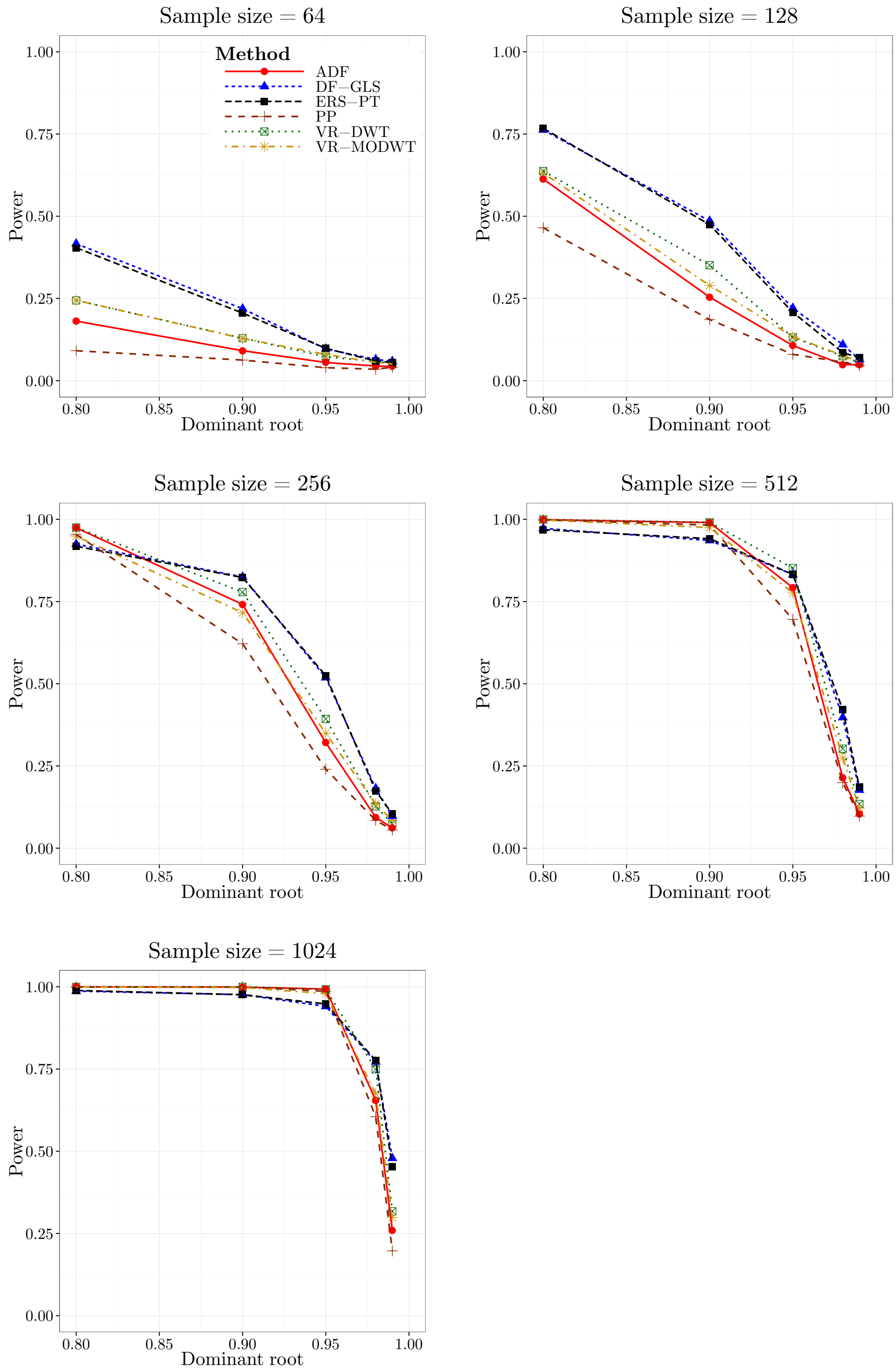


Figure A.8: GED ($\nu = 1.5$); $\alpha_0 = 0.005$; $\alpha_1 = 0.99$ $(\alpha_1 + \gamma_1) = 0.995$

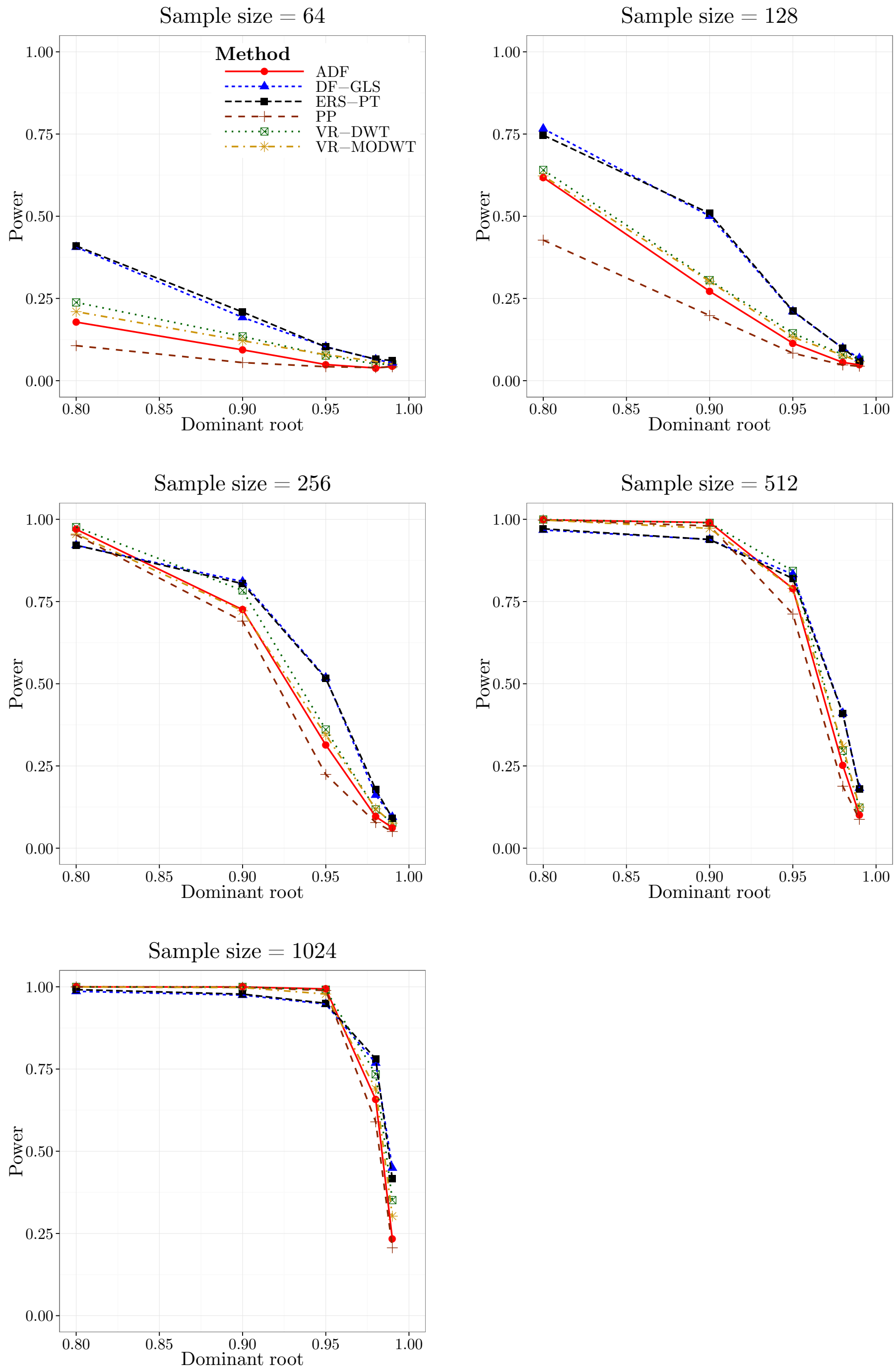


Figure A.9: $t(4)$; $\alpha_0 = 0.005$; $\alpha_1 = 0.99$; $(\alpha_1 + \gamma_1) = 0.995$