Unit Root Tests with Wavelets

Gencay, Ramazan and Fan, Yanqin

February 2007

Unit Root Tests with Wavelets*

Yanqin Fan† Ramazan Gençay‡

First version: February 2007, This version: April 2009
Forthcoming in Econometric Theory

Abstract

This paper develops a wavelet (spectral) approach to testing the presence of a unit root in a stochastic process. The wavelet approach is appealing, since it is based directly on the different behavior of the spectra of a unit root process and that of a short memory stationary process. By decomposing the variance (energy) of the underlying process into the variance of its low frequency components and that of its high frequency components via the discrete wavelet transformation (DWT), we design unit root tests against near unit root alternatives. Since DWT is an energy preserving transformation and able to disbalance energy across high and low frequency components of a series, it is possible to isolate the most persistent component of a series in a small number of scaling coefficients. We demonstrate the size and power properties of our tests through Monte Carlo simulations.

Keywords: Unit root tests, discrete wavelet transformation, maximum overlap wavelet transformation, energy decomposition.
JEL No: C1, C2, C12, C22, F31, G0, G1.

---

*This is a substantially shortened version of the paper: "Unit root and cointegration tests with wavelets." We are grateful to Pentti Saikkonen and two anonymous referees for detailed comments on the early paper which have helped improve the presentation of the results in the current paper. We also thank Stelios Bekiros, Buz Brock, Russell Davidson, Cees Diks, Cars Hommes, Benoit Perron, Hashem Pesaran, James MacKinnon, James Ramsey, Alessio Sancetta, Mototsugu Shintani and Zhijie Xiao for helpful discussions. All errors belong to the authors.

†Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, U.S.A. Part of the work in this paper was done when Fan visited the Department of Economics at Simon Fraser University whose hospitality and support are acknowledged. Yanqin Fan is grateful to the National Science Foundation for research support. `Email: yanqin.fan@vanderbilt.edu`

‡Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada. Ramo Gençay is grateful to the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada for research support. Ramo Gençay is grateful for the feedback of seminar participants at CESG, CIREQ, IRMACS, University of Amsterdam, University of British Columbia, Cambridge University, University College Dublin and New York University. `Email: rgencay@sfu.ca`. The software for this paper is available at `www.sfu.ca/∼rgencay`.

1 Introduction

As Granger (1966) pointed out, the vast majority of economic variables, after removal of any trend in mean and seasonal components, have similar shaped power spectra where the power of the spectrum peaks at the lowest frequency with exponential decline towards higher frequencies. Since Nelson and Plosser (1982) argued that this persistence was captured by modeling the series as having a unit autoregressive root, designing tests for unit root has attracted the attention of many researchers. The well-known Dickey and Fuller (1979) unit root tests have limited power to separate a unit root process from near unit root alternatives in small samples. Phillips (1986) and Phillips (1987a) pioneered the use of the functional central limit theorem to establish the asymptotic distribution of statistics constructed from unit root processes. To construct unit root tests with serially correlated errors, one approach is due to Phillips (1987a) and Phillips and Perron (1988) by adjusting the test statistic to take account for the serial correlation and heteroskedasticity in the disturbances. The other approach is due to Dickey and Fuller (1979) by adding lagged dependent variables as explanatory variables in the regression. Other important contributions are Chan and Wei (1987), Park and Phillips (1988), Park and Phillips (1989), Sims *et al.* (1990), Phillips and Solo (1992) and Park and Fuller (1995). In general, unit root tests cannot distinguish highly persistent stationary processes from nonstationary processes and the power of unit root tests diminish as deterministic terms are added to the test regressions. For maximum power against very persistent alternatives, Elliott *et al.* (1996) (ERS) use a framework similar to Dufour and King (1991) (DK) to derive the asymptotic power envelope for point-optimal tests of a unit root under various trend specifications. Ng and Perron (2001) exploits the finding of ERS and DK to develop modified tests with enhanced power subject to proper selection of a truncation lag.

Most existing unit root tests make direct use of time domain estimators of the coefficient of the lagged value of the variable in a regression with its current value as the dependent variable, except Choi and Phillips (1993), the Von Neumann variance ratio (VN) tests of Sargan and Bhargava (1983) and their extensions. Recently, Cai and Shintani (2006) provide alternative VN tests based on combinations of consistent and inconsistent long run variance estimators. Phillips and Xiao (1998) and Stock (1999) provide a helpful review of the main tests and an extensive list of references.

In this paper, we develop a general wavelet spectral approach to testing unit roots inspired by Granger (1966). The method of wavelets decomposes a stochastic process into its components, each of which is associated with a particular frequency band. The wavelet power spectrum measures the contribution of the variance at a particular frequency band relative to the overall variance of the process. If a particular band contributes substantially more to the overall variance relative to another frequency band, it is considered an important driver of this process. Recall that the spectrum of a unit root process is infinite at the origin, and hence the variance of a unit root process is largely contributed by low frequencies. By decomposing the variance[1] of the underlying process into the variance of its low frequency components and that of its high frequency components via the discrete wavelet transformation (DWT), we design wavelet-based unit root tests. Since DWT is an energy preserving transformation and able to disbalance energy across high and low frequency components of a series, it is possible to isolate the most persistent component of a series in a small number of coefficients referred to as the scaling coefficients. Our tests utilize the scaling coefficients of the unit scale. In particular, we construct test statistics from the ratio of the energy from the unit scale to the total energy (variance) of the time series. We establish asymptotic properties of our tests, including their asymptotic null distributions, consistency, and local power properties. Our tests are easy to implement, as their asymptotic null distributions are nuisance parameter free and the corresponding critical values can be tabulated. The Monte Carlo simulations are conducted to compare the empirical size and power of our tests to the Dickey and Fuller (1979) (ADF), Phillips and Perron (1988) (PP), Elliott *et al.* (1996) (ERS) and Ng and Perron (2001) (MPP) tests. Our tests have good size and comparable power against near unit root alternatives in finite samples.

Choi and Phillips (1993) developed unit root tests based on an alternative spectral approach to time series analysis, the Fourier spectral analysis, and demonstrated advantages of their tests over tests based on time domain approach. Unlike our tests, however, their tests make use of frequency domain estimators of the autoregressive coefficient. The DWT is an orthonormal transformation which may be relaxed through an oversampling approach termed as the maximum overlap DWT (MODWT), see, for example, Percival and Mofjeld (1997).[2] The VN tests of Sargan and Bhargava (1983) are based on the ratio of the sample variance of the first differences and the levels of the time series. These tests avoid the problem of redundant trend to gain efficiency. Sargan and Bhargava (1983) suggested using the VN statistic for testing the Gaussian random walk hypothesis, and Bhargava (1986) extended to the case of the time trend. Stock (1995) studied unit root tests with a linear time trend and Schmidt and Phillips (1992), working with polynomial trends, showed that the Lagrange multiplier principle leads to a VN test. Interestingly, we show that the VN tests are special

[1] In the signal processing literature, the variance of a process is referred to as the energy of the process. In this paper, we use the two terminologies interchangeably.

[2] The MODWT goes by several names in the literature, such as the stationary DWT by Nason and Silverman (1995) and the translation-invariant DWT by Coifman and Donoho (1995). A detailed treatment of MODWT can be found in Percival and Walden (2000) and Gençay *et al.* (2001).

cases of our wavelet tests when we use the Haar wavelet filter and unit scale MODWT. The Haar wavelet filter is the member of Daubechies compactly supported wavelet filter of the shortest length. By using Daubechies wavelet filter of longer length, our tests gain power over the VN tests in finite samples.

The Fourier approach is appealing when working with stationary time series. However, restricting ourselves to stationary time series is not appealing since most economic/financial time series exhibit quite complicated patterns over time (e.g., trends, abrupt changes, and volatility clustering). In fact, if the frequency components are not stationary such that they may appear, disappear, and then reappear over time, traditional spectral tools may miss such frequency components. Wavelet filters provide a natural platform to deal with the time-varying characteristics found in most real-world time series, and thus the assumption of stationarity may be avoided. The wavelet transform intelligently adapts itself to capture features across a wide range of frequencies and thus has the ability to capture events that are local in time. This makes the wavelet transform an ideal tool for studying nonstationary time series. Early applications of wavelets in economics and finance are by Ramsey and his coauthors (see Ramsey *et al.* (1995), Ramsey and Zhang (1997), Ramsey (1999), Ramsey (2002) for a review and references) who explore the use of DWT in decomposing various economic and financial data. Davidson *et al.* (1998) investigated U.S. commodity prices via wavelets. Gençay *et al.* (2003, 2005) propose a wavelet approach for estimating the systematic risk or the beta of an asset in a capital asset pricing model. The proposed method is based on a wavelet multiscaling approach where the wavelet variance of the market return and the wavelet covariance between the market return and a portfolio are calculated to obtain an estimate of the portfolio's systematic risk (beta) at each scale. In time series econometrics, one example of the successful application of wavelets is in the context of long memory processes where a number of estimation methods have been developed. These include wavelet-based OLS, the approximate wavelet-based maximum likelihood approach, and wavelet-based Bayesian approach. Fan (2003) and Fan and Whitcher (2003) provide an extensive list of references. The success of these methods relies on the so called 'approximate decorrelation' property of the DWT of a possibly nonstationary long memory process, see Fan (2003) for a rigorous proof of this result for a nonstationary fractionally differenced process. Fan and Whitcher (2003) propose overcoming the problem of spurious regression between fractionally differenced processes by applying the DWT to both processes and then estimating the regression in the wavelet domain. Other examples of applications of wavelets in econometrics include wavelet-based spectral density estimators and their applications in testing for serial correlation/conditional heteroscedasticity, see e.g., Hong (2000), Hong and Lee (2001), Lee and Hong (2001), Duchesne (2006a), Duchesne (2006b), and Hong and Kao

(2004).

This paper provides another context in which the use of the wavelet (spectral) approach may have advantages over the time domain approach or the Fourier approach. Unlike Hong (2000), Hong and Lee (2001), Lee and Hong (2001), Duchesne (2006a), Duchesne (2006b), and Hong and Kao (2004) who develop and/or make use of wavelet estimators of spectral density functions of the relevant processes, we employ directly the DWT of the observed time series. We contribute to the unit root literature on three different fronts. First, we propose a unified wavelet spectral approach to unit root testing; second, we provide a spectral interpretation of existing VN unit root tests; and finally, we propose higher order wavelet filters to capture low-frequency stochastic trends parsimoniously and gain power against near unit root alternatives.

In section two, we begin with a brief overview of wavelets, discrete wavelet filters and discrete wavelet transformation. In section three, we develop our wavelet-based unit root tests against purely stationary alternatives and trend stationary alternatives. Section four provides Monte Carlo simulations on the size and power properties of our tests. We conclude thereafter. An appendix contains technical proofs. Throughout this paper, we use $\Longrightarrow$ to denote weak convergence. All the limits are taken as the sample size approaches ∞.

2 Discrete Wavelet Transformation

A wavelet is a small wave which grows and decays in a limited time period.[3] To formalize the notion of a wavelet, let $\psi(.)$ be a real valued function such that its integral is zero, $\int_{-\infty}^{\infty} \psi(t)\, dt = 0$, and its square integrates to unity, $\int_{-\infty}^{\infty} \psi(t)^2\, dt = 1$. Thus, although $\psi(.)$ has to make some excursions away from zero, any excursions it makes above zero must cancel out excursions below zero, i.e., $\psi(.)$ is a small wave, or a wavelet.

Fundamental properties of the continuous wavelet functions (filters), such as integration to zero and unit energy, have discrete counterparts. Let $h = (h_0, \ldots, h_{L-1})$ be a finite length discrete wavelet (or high pass) filter such that it integrates (sums) to zero, $\sum_{l=0}^{L-1} h_l = 0$, and has unit energy, $\sum_{l=0}^{L-1} h_l^2 = 1$. In addition, the wavelet filter h is orthogonal to its even shifts; that is,

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0, \quad \text{for all nonzero integers } n. \tag{1}$$

The natural object to complement a high-pass filter is a low-pass (scaling) filter g. We will denote a low-pass filter as $g = (g_0, \ldots, g_{L-1})$. The low-pass filter coefficients are determined

[3] This section closely follows Gençay *et al.* (2001), see also Percival and Walden (2000). The contrasting notion is a big wave such as the sine function which keeps oscillating indefinitely.

by the *quadrature mirror relationship*[4]

$$g_l = (-1)^{l+1} h_{L-1-l} \quad \text{for} \quad l = 0, \ldots, L-1 \tag{2}$$

and the inverse relationship is given by $h_l = (-1)^l g_{L-1-l}$. The basic properties of the scaling filter are: $\sum_{l=0}^{L-1} g_l = \sqrt{2}$, $\sum_{l=0}^{L-1} g_l^2 = 1$,

$$\sum_{l=0}^{L-1} g_l g_{l+2n} = \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0, \tag{3}$$

for all nonzero integers n, and

$$\sum_{l=0}^{L-1} g_l h_{l+2n} = \sum_{l=-\infty}^{\infty} g_l h_{l+2n} = 0 \tag{4}$$

for all integers n. Thus, scaling filters are average filters and their coefficients satisfy the orthonormality property that they possess unit energy and are orthogonal to even shifts.

By applying both h and g to an observed time series, we can separate high-frequency oscillations from low-frequency ones. Let $\mathbf{y} = \{y_t\}_{t=1}^T$ be a dyadic length vector ($T = 2^M$) of observations where $M = \log_2(T)$. The length T vector of discrete wavelet coefficients $\mathbf{w}$ is obtained by $\mathbf{w} = \mathcal{W}\mathbf{y}$, where $\mathcal{W}$ is a $T \times T$ real-valued orthonormal matrix defining the DWT which satisfies $\mathcal{W}^T\mathcal{W} = I_T$ ($T \times T$ identity matrix). We refer the interested reader to Percival and Walden (2000) for a detailed discussion on the construction of $\mathcal{W}$ from the wavelet and scaling filters. The vector of wavelet coefficients may be organized into $M+1$ vectors,

$$\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M, \mathbf{v}_M]^T, \tag{5}$$

where $\mathbf{w}_j$ is a length $T/2^j$ vector of wavelet coefficients associated with changes on a scale of length $\lambda_j = 2^{j-1}$ and $\mathbf{v}_M$ is a length $T/2^M$ vector of scaling coefficients associated with averages on a scale of length $2^M = 2\lambda_M$.

In practice the DWT is implemented via a pyramid algorithm of Mallat (1989, 1998). The first iteration of the pyramid algorithm begins by filtering (convolving) the data with each filter to obtain the unit-scale wavelet and scaling coefficients:

$$W_{t,1} = \sum_{l=0}^{L-1} h_l y_{2t-l \bmod T} \quad \text{and} \quad V_{t,1} = \sum_{l=0}^{L-1} g_l y_{2t-l \bmod T},$$

[4]Quadrature mirror filters (QMFs) are often used in the engineering literature because of their ability for perfect reconstruction of a signal without aliasing effects. Aliasing occurs when a continuous signal is sampled to obtain a discrete time series.

where $t = 1, \dots, T/2$. Let $\mathbf{w}_1 = \left(W_{1,1}, ..., W_{T/2,1}\right)'$ and $\mathbf{v}_1 = \left(V_{1,1}, ..., V_{T/2,1}\right)'$ denote respectively the vectors of unit-scale wavelet and scaling coefficients. We obtain the level 1 partial DWT $\mathbf{w} = [\mathbf{w}_1,\ \mathbf{v}_1]^T$.

The second step of the pyramid algorithm starts by defining the "data" to be the scaling coefficients $\mathbf{v}_1$ from the first iteration and apply the filtering operations as above to obtain the second level of wavelet and scaling coefficients:

$$W_{t,2} = \sum_{l=0}^{L-1} h_l V_{2t-l,1 \bmod T/2} \quad \text{and} \quad V_{t,2} = \sum_{l=0}^{L-1} g_l V_{2t-l,1 \bmod T/2},$$

$t = 1, \dots, T/4$. Keeping all vectors of wavelet coefficients, and the final level of scaling coefficients, we have the following length T decomposition $\mathbf{w} = [\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{v}_2]^T$, where $\mathbf{w}_2,\ \mathbf{v}_2$ denote respectively the vectors of second scale wavelet and scaling coefficients. This procedure may be repeated up to M times where $M = \log_2(T)$ and gives the vector of wavelet coefficients in Equation (5).

The orthonormality of the matrix $\mathcal{W}$ implies that the DWT is a variance preserving transformation:

$$\|\mathbf{w}\|^2 = \sum_{t=1}^{T/2^M} V_{t,M}^2 + \sum_{j=1}^{M} \left(\sum_{t=1}^{T/2^j} W_{t,j}^2 \right) = \sum_{t=1}^{T} y_t^2 = \|\mathbf{y}\|^2 .$$

This can be easily proven through basic matrix manipulation via

$$\|\mathbf{y}\|^2 = \mathbf{y}^T\mathbf{y} = (\mathcal{W}\mathbf{w})^T \mathcal{W}\mathbf{w} = \mathbf{w}^T \mathcal{W}^T \mathcal{W}\mathbf{w} = \mathbf{w}^T\mathbf{w} = \|\mathbf{w}\|^2 .$$

Given the structure of the wavelet coefficients, $\|\mathbf{y}\|^2$ is decomposed on a scale-by-scale basis via

$$\|\mathbf{y}\|^2 = \sum_{j=1}^{M} \|\mathbf{w}_j\|^2 + \|\mathbf{v}_M\|^2 , \tag{6}$$

where $\|\mathbf{w}_j\|^2 = \sum_{t=1}^{T/2^j} W_{t,j}^2$ is the sum of squared variation of $\mathbf{y}$ due to changes at scale λ_j and $\|\mathbf{v}_M\|^2 = \sum_{t=1}^{T/2^M} V_{t,M}^2$ is the information due to changes at scales λ_M and higher.

The idea behind our wavelet unit root tests can be best understood through the energy (variance) decomposition of a white noise process and that of a unit root process. To illustrate, in Figure 1, the dot chart of a Gaussian white noise process is plotted for 1024 observations ($M = 2^{10} = 1024$). A six level ($J = 6$)[5] DWT is used. "Data" represents the total energy of the data which is normalized at one, wi, $i = 1, \dots, 6$ represents the percentage energy of wavelet coefficients, and v6 is the percentage energy of the scaling coefficients. The

[5] There is no specific reason for choosing $J = 6$. Any level $J < M$ could be used.

sum of the energies of the wavelet and the scaling coefficients is equal to the total energy of the data. The energy is the highest at the highest frequency wavelet coefficient (w1) and declines gradually towards the lowest frequency wavelet coefficient (w6). The percentage energy of the scaling coefficient (v6), i.e., $\|\mathbf{v}_J\|^2 / \|\mathbf{y}\|^2$, is close to zero. In Figure 2, the dot chart of a unit root process

$$y_t = y_{t-1} + u_t, \qquad u_t \sim \text{i.i.d. } N(0,1) \tag{7}$$

is plotted for $y_0 = 0$ and $t = 1, 2, \ldots, 1024$ observations. The energy is the highest for the scaling coefficients and almost zero at all wavelet coefficients. The percentage energy of the scaling coefficients (v6), i.e., $\|\mathbf{v}_J\|^2 / \|\mathbf{y}\|^2$, is almost equal to one. The number of coefficients needed equals 41 ($41/1024 = 4\%$) of the total number of coefficients to account for almost all energy of the data. Heuristically, when a white noise process is added up (say, as in a unit root process), the high frequencies are smoothed out (those spikes in the white noise disappear) and what is left is the long term stochastic trend. On the contrary, when we do differencing (e.g., first differencing to a unit root, then we are back to the white noise series), we get rid of the long term trend, and what is left is the high frequencies (spikes) with mean zero. Since a unit root process can be succinctly approximated by a few scaling coefficients and the energy of the scaling coefficients is almost equal to the total energy of the data, we develop our statistical tests for a unit root process based on this principle of energy decomposition.

3 New Unit Root Tests

Let $\{y_t\}_{t=1}^T$ be a univariate time series generated by

$$y_t = \rho y_{t-1} + u_t, \tag{8}$$

where $\{u_t\}$ is a weakly stationary zero-mean error with a strictly positive long run variance defined by $\omega^2 \equiv \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j$ where $\gamma_j = E(u_t u_{t-j})$. Throughout this paper, the initial condition is set to $y_0 = O_p(1)$ and the following assumption on the error term is maintained.

Assumption 1:

(a) $\{u_t\}$ is a linear process defined as $u_t = \psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$, $\psi(1) \neq 0$, and $\sum_{j=0}^{\infty} j|\psi_j| < \infty$;

(b) $\{\epsilon_t\}$ is i.i.d. with $E(\epsilon_t) = 0$, $Var(\epsilon_t) = \sigma^2$, and finite fourth cumulants, and $\epsilon_s = 0$ for $s \leq 0$.

The last condition in Assumption 1(a) is referred to as 1-summability of $\psi(L)$. The assumption $\epsilon_s = 0$ for $s \leq 0$ in Assumption 1(b) is made for convenience. Under Assumption 1, we have $\omega^2 = \psi(1)^2\sigma^2$ and $T^{-1/2}\sum_{t=1}^{[T\cdot]} u_t \implies \omega W(\cdot)$ where $[Tr]$ denotes the integer part of Tr and $W(\cdot)$ denotes a standard Brownian motion defined on $C[0,1]$, the space of continuous functions on $[0,1]$. It is known that the weak convergence result: $T^{-1/2}\sum_{t=1}^{[T\cdot]} u_t \implies \omega W(\cdot)$ holds for more general/other classes of processes than the class of linear processes specified in Assumption 1 including linear processes with martingale difference innovations $\{\epsilon_t\}$, see Phillips and Solo (1992). One may also extend the weak convergence result to linear processes with GARCH innovations by making use of the weak convergence result for GARCH processes, see Berkes *et al.* (2008). It is possible to extend the results to be developed in this paper to these other processes. For ease of exposition, we will stick to Assumption 1 in this paper.

In Subsections 3.1 and 3.2, we consider tests for $H_0 : \rho = 1$ against $H_1 : |\rho| < 1$ in (8). Under the alternative hypothesis, $\{y_t\}$ is a zero-mean stationary process with the long run variance $\omega^2/(1-\rho)^2$. As mentioned in Section 2, our tests for unit root are based on the different behavior of the energy decomposition of a unit root process and that of a short-memory such as a white noise process. To introduce the fundamental idea, we first develop a test based on the Haar wavelet filter and unit scale DWT in Subsection 3.1. In Subsection 3.2, we extend it to tests based on any Daubechies (1992) compactly supported wavelet filter of finite length. Finally, we extend the tests developed in Subsections 3.1 and 3.2 to trend stationary alternatives in Subsection 3.3.

3.1 The first test — Haar wavelet filter

Consider the unit scale Haar DWT of $\{y_t\}_{t=1}^T$ where T is assumed to be even. The wavelet and scaling coefficients are given by

$$W_{t,1} = \frac{1}{\sqrt{2}}(y_{2t} - y_{2t-1}),\ t = 1, 2, \ldots, T/2, \tag{9}$$

$$V_{t,1} = \frac{1}{\sqrt{2}}(y_{2t} + y_{2t-1}),\ t = 1, 2, \ldots, T/2. \tag{10}$$

The wavelet coefficients $\{W_{t,1}\}$ capture the behavior of $\{y_t\}$ in the high frequency band $[1/2, 1]$, while the scaling coefficients $\{V_{t,1}\}$ capture the behavior of $\{y_t\}$ in the low frequency band $[0, 1/2]$. The total energy of $\{y_t\}_{t=1}^T$ is given by the sum of the energies of $\{W_{t,1}\}$ and $\{V_{t,1}\}$. Since for a unit root process, the energy of the scaling coefficients $\{V_{t,1}\}$ dominates that of the wavelet coefficients $\{W_{t,1}\}$, we propose the following test statistic:

$$\hat{S}_{T,1} = \frac{\sum_{t=1}^{T/2} V_{t,1}^2}{\sum_{t=1}^{T/2} V_{t,1}^2 + \sum_{t=1}^{T/2} W_{t,1}^2}. \tag{11}$$

Heuristically, under H_0, $\hat{S}_{T,1}$ should be close to 1, since $\sum_{t=1}^{T/2} V_{t,1}^2$ dominates $\sum_{t=1}^{T/2} W_{t,1}^2$, while under H_1, $\hat{S}_{T,1}$ should be smaller than 1. We formalize these statements in the following lemma.

Lemma 3.1 *Under H_0, $\hat{S}_{T,1} = 1 + o_p(1)$, while under H_1, $\hat{S}_{T,1} = \frac{E(y_{2t}+y_{2t-1})^2}{E(y_{2t}+y_{2t-1})^2+E(y_{2t}-y_{2t-1})^2} + o_p(1)$.*

Note that:

$$\frac{E(y_{2t}+y_{2t-1})^2}{E(y_{2t}+y_{2t-1})^2+E(y_{2t}-y_{2t-1})^2} = \frac{E\left(V_{t,1}^2\right)}{E\left(V_{t,1}^2\right)+E\left(W_{t,1}^2\right)} < 1.$$

We conclude that it is the relative magnitude of the energy of the scaling coefficients to that of the wavelet coefficients that determines the power of the test based on $\hat{S}_{T,1}$ and we expect our test based on $\hat{S}_{T,1}$ to have power against H_1.

The asymptotic distribution of $\hat{S}_{T,1}$ under H_0 is summarized in the following theorem.

Theorem 3.2 *Under H_0, $T(\hat{S}_{T,1}-1) \Longrightarrow -\frac{\gamma_0}{\lambda_v^2 \int_0^1 [W(r)]^2 dr}$, where $\lambda_v^2 = 4\omega^2$.*

The proof of Theorem 3.2 in the Appendix makes it clear that it is the energy of the scaling coefficients that drives the asymptotic behavior of $\hat{S}_{T,1}$ under the null hypothesis. Alternatively, noting the energy decomposition: $\sum_{t=1}^{T/2} V_{t,1}^2 + \sum_{t=1}^{T/2} W_{t,1}^2 = \sum_{t=1}^{T} y_t^2$, we get immediately,

$$\begin{aligned} T(\hat{S}_{T,1}-1) &= -\frac{T^{-1}\sum_{t=1}^{T/2}\left(W_{t,1}^2 - EW_{t,1}^2\right)}{T^{-2}\sum_{t=1}^{T} y_t^2} - \frac{\frac{1}{2}EW_{t,1}^2}{T^{-2}\sum_{t=1}^{T} y_t^2} \\ &= -\frac{o_p(1)}{\omega^2\int_0^1 [W(r)]^2 dr} - \frac{\gamma_0}{4\omega^2\int_0^1 [W(r)]^2 dr} \\ &= -\frac{\gamma_0}{\lambda_v^2\int_0^1 [W(r)]^2 dr} + o_p(1) \text{ under } H_0. \end{aligned}$$

There are two unknown parameters in the asymptotic null distribution of $\hat{S}_{T,1}$: $\gamma_0 = E(u_{2t}^2)$ and λ_v^2 or ω^2. To estimate these parameters, we let $\hat{u}_t = y_t - \hat{\rho} y_{t-1}$ denote the OLS residual. Then $\hat{\gamma}_0 = T^{-1}\sum_{t=1}^{T} \hat{u}_t^2$ is a consistent estimator of γ_0. Being the long run variance of $\{u_t\}$, ω^2 can be consistently estimated by a nonparametric kernel estimator with the Bartlett kernel:

$$\widehat{\omega}^2 = 4\hat{\gamma}_0 + 2\sum_{j=1}^{q}[1 - j/(q+1)]\hat{\gamma}_j,$$

where q is the bandwidth/lag truncation parameter and $\hat{\gamma}_j = T^{-1}\sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}$, see Newey and West (1987).[6] Andrews (1991) showed that this long run variance estimator is consistent

[6] Newey and West (1987) suggest setting the bandwidth using the sample size dependent rule of $4(T/100)^{2/9}$. We use this rule with Bartlett kernel in this paper.

when the bandwidth q grows at a rate slower than $T^{1/2}$, with an optimal growth rate being $T^{1/3}$ under some moment conditions. Let $\hat{\lambda}_v^2 = 4\widehat{\omega}^2$ and define the test statistic as

$$FG_1 = \frac{T\hat{\lambda}_v^2}{\hat{\gamma}_0}\left[\hat{S}_{T,1} - 1\right].$$

Then under the null hypothesis, the limiting distribution of the test statistic FG_1 is given by the distribution of

$$-\frac{1}{\int_0^1 [W(r)]^2 dr}.$$

The limiting distribution of FG_1under H_0 is free from nuisance parameters and is extremely easy to simulate, see MacKinnon (2000) for a detailed treatment. The critical values of this test are tabulated in the first row of Table 1.

We note that an alternative way to estimate γ_0 is via the wavelet variance estimators. We will elaborate on this approach in the next subsection when we allow the use of a general filter. Also, ω^2 can be estimated by any existing long run variance estimators, including the wavelet-based estimator of Hong (2000).

3.2 A general test — Daubechies compactly supported wavelet filter

For a general Daubechies compactly supported wavelet filter $\{h_l\}_{l=0}^{L-1}$, the boundary-independent (BI) unit scale wavelet and scaling coefficients are given by

$$W_{t,1} = \sum_{l=0}^{L-1} h_l y_{2t-l}, \;\; V_{t,1} = \sum_{l=0}^{L-1} g_l y_{2t-l}, \tag{12}$$

where $t = L_1, L_1 + 1, \ldots, T/2$ with $L_1 = L/2$. Again the wavelet coefficients $\{W_{t,1}\}$ extract the high frequency information in $\{y_t\}$. Since any Daubechies wavelet filter has a difference filter embedded in it, $\{W_{t,1}\}$ is stationary under both H_0 and H_1. However the sequence of scaling coefficients $\{V_{t,1}\}$, extracting the low frequency information in $\{y_t\}$, is nonstationary under H_0 and stationary under H_1. Reflected in their respective energies, this implies that the energy of the scaling coefficients dominates that of the wavelet coefficients under H_0, which forms the basis for our tests.

Define[7]

$$\hat{S}_{T,1}^L = \frac{\sum_{t=L_1}^{T/2} V_{t,1}^2}{\sum_{t=L_1}^{T/2} V_{t,1}^2 + \sum_{t=L_1}^{T/2} W_{t,1}^2}.$$

We will construct a test for unit root based on the following asymptotic properties of $\hat{S}_{T,1}^L$.

[7] Instead of using the BI wavelet and scaling coefficients only, one could use all the wavelet and scaling coefficients. This would not change the asymptotic results, as there is only a finite number of boundary dependent wavelet and scaling coefficients.

Theorem 3.3 *(i)* $\hat{S}^L_{T,1} = 1 + o_p(1)$ *under* H_0 *and* $\hat{S}^L_{T,1} = c_L + o_p(1)$ *under* H_1 *with* $c_L = \frac{E\left(V^2_{t,1}\right)}{E\left(V^2_{t,1}\right)+E\left(W^2_{t,1}\right)} < 1$*; (ii)* $\left(\frac{T}{2}\right)\left(\hat{S}^L_{T,1} - 1\right) \Longrightarrow -\frac{E\left(W^2_{t,1}\right)}{\lambda^2_v \int_0^1 [W(r)]^2 dr}$ *under* H_0.

Theorem 3.3(i) implies that a consistent test for unit root can be based on $\hat{S}^L_{T,1}$. Theorem 3.3(ii) extends Theorem 3.2 from the Haar filter ($L = 2$) to any Daubechies compactly supported wavelet filter of finite length L. Since as the length of the filter L increases, the approximation of the Daubechies wavelet filter to the ideal high-pass filter improves[8], we expect tests based on $\hat{S}^L_{T,1}$ to gain power as L increases. On the other hand, as L increases, the number of BI wavelet and scaling coefficients will decrease which would have an adverse effect on the power of our tests. It might be possible to choose L based on some power criterion function, but this is beyond the scope of this paper. In other applications of DWT with Daubechies compactly supported wavelet filter, $L = 2$ or 4 are often used.

Note that $E\left(W^2_{t,1}\right)$ equals twice of the so-called wavelet variance at the unit scale. As a result, existing wavelet variance estimators can be used to estimate $E\left(W^2_{t,1}\right)$, see Percival (1995) for a detailed comparison of the wavelet variance estimators based on DWT and MODWT respectively. Based on DWT, $2\widehat{\upsilon}^2_{y,1}$ is a consistent estimator of the wavelet variance, where

$$\widehat{\upsilon}^2_{y,1} = \frac{1}{(T/2 - L_1 + 1)} \sum_{t=L_1}^{T/2} W^2_{t,1}. \tag{13}$$

Define the test statistic:

$$FG^L_1 = \left(\frac{T}{2}\right) \frac{\hat{\lambda}^2_v}{\widehat{\upsilon}^2_{y,1}} \left[\hat{S}^L_{T,1} - 1\right].$$

Under the null hypothesis, the limiting distribution of FG^L_1 is the same as that of FG_1.

We now develop asymptotic power functions for our unit root tests by considering the sequence of local alternatives given by

$$\rho = \exp\left(\frac{c}{T}\right) \sim 1 + \frac{c}{T} \tag{14}$$

for a particular value of $c < 0$. Under this sequence of local alternatives, it is well known that

$$T^{-2} \sum_{t=1}^{T} y^2_t \Longrightarrow \omega^2 \int_0^1 \left[J_c(r)\right]^2 dr,$$

where

$$J_c(r) = \int_0^r \exp\left\{(r-u)c\right\} dW(u)$$

is the Ornstein-Uhlenbeck process generated in continuous time by the stochastic differential equation $dJ_c(r) = cJ_c(r)dr + dW(r)$. Using this, one can easily show that under this sequence

[8] Percival and Walden (2000) provides an excellent discussion on this.

of local alternatives, the asymptotic distributions of the test statistics FG_1^L, FG_1 are of the same form as those under the null hypothesis except that the Brownian motion $W(\cdot)$ is replaced with the Ornstein-Uhlenbeck process $J_c(\cdot)$, i.e., $-1/\int_0^1 [J_c(r)]^2 dr$. In particular, this leads to the conclusion that all these tests have the same asymptotic power (to the first order) against the sequence of local alternatives of the form (14). The following theorem states consistency and local power properties of our tests.

Theorem 3.4 *(i) Under H_1, $\Pr\left(FG_1^L < -C\right) \to 1$ for any fixed positive constant C; (ii) Under (14), we obtain:*

$$FG_1^L \implies -\frac{1}{\int_0^1 [J_c(r)]^2 dr}.$$

3.3 Tests against trend stationarity

Tests developed in the previous subsections can be extended to deal with trend stationary alternatives. We adopt the components representation of a time series and work with the detrended series, see Schmidt and Phillips (1992), Phillips and Xiao (1998), and Stock (1999). For ease of exposition, we restrict ourselves to non-zero mean and linear trend cases only. Phillips and Xiao (1998) also have a detailed discussion on efficient detrending for general trends.

The process $\{y_t\}$ is assumed to be of the form:

$$y_t = \mu + \alpha t + y_t^s, \tag{15}$$

where $\{y_t^s\}$ is generated by model (8). Under $H_0 : \rho = 1$, $\{y_t^s\}$ is a unit root process while under $H_0 : |\rho| < 1$, $\{y_t^s\}$ is a zero mean stationary process. If $\alpha = 0$, we consider the demeaned series $\{y_t - \overline{y}\}$, where $\overline{y} = T^{-1}\sum_{t=1}^T y_t$ is the sample mean of $\{y_t\}$. If $\alpha \neq 0$, we work with the detrended series $\left\{\widetilde{y}_t - \overline{\widetilde{y}}\right\}$, where $\widetilde{y}_t = \sum_{j=1}^t \left(\Delta y_j - \overline{\Delta y}\right)$ and $\overline{\widetilde{y}}$ is the sample mean of $\{\widetilde{y}_t\}$, in which $\Delta y_t = y_t - y_{t-1}$ and $\overline{\Delta y}$ is the sample mean of $\{\Delta y_t\}$. Alternative expressions for the detrended series $\left\{\widetilde{y}_t - \overline{\widetilde{y}}\right\}$ can be found in Schmidt and Phillips (1992).

Let $\left\{W_{t,1}^M\right\}$ and $\left\{V_{t,1}^M\right\}$ denote respectively the unit scale DWT wavelet and scaling coefficients of the demeaned series $\{y_t - \overline{y}\}$. We will construct our tests based on

$$\widehat{S}_{T,1}^{LM} = \frac{\sum_{t=1}^{T/2}(V_{t,1}^M)^2}{\sum_{t=1}^T (y_t - \overline{y})^2}.$$

Similarly, let $\left\{W_{t,1}^d\right\}$ and $\left\{V_{t,1}^d\right\}$ denote respectively the unit scale DWT wavelet and scaling coefficients of the detrended series $\left\{\widetilde{y}_t - \overline{\widetilde{y}}\right\}$. We will construct our tests based on

$$\widehat{S}_{T,1}^{Ld} = -\frac{\sum_{t=1}^{T/2}(V_{t,1}^d)^2}{\sum_{t=1}^T (\widetilde{y}_t - \overline{\widetilde{y}})^2}.$$

Under H_0, it is known that $T^{-2}\sum_{t=1}^{T}(y_t-\overline{y})^2 \implies \omega^2\int_0^1 [W_\mu(r)]^2\,dr$ and $T^{-2}\sum_{t=1}^{T}(\widetilde{y}_t-\overline{\widetilde{y}})^2 \implies \omega^2\int_0^1 [V_\mu(r)]^2\,dr$, where $W_\mu(r)=W(r)-\int_0^1 W(u)du$ and $V_\mu(r)=V(r)-\int_0^1 V(u)du$ in which $V(r)=W(r)-rW(1)$.

Theorem 3.5 *Under H_0, we have: (i)* $T\left(\widehat{S}_{T,1}^{LM}-1\right) \implies -\frac{E(W_{t,1}^{M})^2}{2\omega^2\int_0^1[W_\mu(r)]^2dr}$; *(ii)* $T\left(\widehat{S}_{T,1}^{Ld}-1\right) \implies -\frac{E(W_{t,1}^{d})^2}{2\omega^2\int_0^1[V_\mu(r)]^2dr}$. *Under (14), we have: (i)* $T\left(\widehat{S}_{T,1}^{LM}-1\right) \implies -\frac{E(W_{t,1}^{M})^2}{2\omega^2\int_0^1[J_c^M(r)]^2dr}$; *(ii)* $T\left(\widehat{S}_{T,1}^{Ld}-1\right) \implies -\frac{E(W_{t,1}^{d})^2}{2\omega^2\int_0^1\left[J_c^d(r)\right]^2dr}$, *where* $J_c^M(r)=\int_0^r \exp\{(r-u)c\}\,dW_\mu(u)$ *and* $J_c^d(r)=\int_0^r \exp\{(r-u)c\}\,dV_\mu(u)$.

To estimate ω^2, we take the OLS residuals from a regression of y_t on a linear trend and y_{t-1} and then apply a nonparametric kernel estimator with the Bartlett kernel to the residuals.

Remark 3.1. It is interesting to note that when the Haar wavelet filter is used,

$$\widehat{S}_{T,1}^{LM}=1-\frac{\sum_{t=1}^{T/2}(y_{2t}-y_{2t-1})^2/2}{\sum_{t=1}^{T}(y_t-\overline{y})^2}.$$

This expression resembles that of the Sargan and Bhargava (1983) and Bhargava (1986) test. In fact, we can obtain the Sargan and Bhargava (1983) and Bhargava (1986) test from an extension of $\widehat{S}_{T,1}^{LM}$ by using MODWT instead of DWT. To see this, we recall that apart from a factor of $\sqrt{2}$, the unit scale MODWT wavelet and scaling coefficients of $\{y_t-\overline{y}\}$ are given by

$$\widetilde{W}_{t,1}=\sum_{l=0}^{L-1}h_l y_{t-l \bmod T},\ \widetilde{V}_{t,1}=\sum_{l=0}^{L-1}g_l\left(y_{t-l \bmod T}-\overline{y}\right), \tag{16}$$

where $t=1,\ldots,T$. It is easy to see that the DWT coefficients are obtained from the corresponding MODWT coefficients via downsampling by 2. At each scale, there are T MODWT wavelet coefficients and T MODWT scaling coefficients. Let

$$\widetilde{S}_{T,1}^{LM}=\frac{\sum_{t=1}^{T}\widetilde{V}_{t,1}^2}{\sum_{t=1}^{T}\widetilde{V}_{t,1}^2+\sum_{t=1}^{T}\widetilde{W}_{t,1}^2}.$$

With the Haar wavelet filter, apart from one coefficient $\widetilde{V}_{1,1}^2$ in the numerator, $\widetilde{S}_{T,1}^{LM}$ reduces to

$$\widetilde{S}_{T,1}^{LM}=1-\frac{\sum_{t=2}^{T}(y_t-y_{t-1})^2}{\sum_{t=1}^{T}(y_t-\overline{y})^2},$$

so that $\left(1-\widetilde{S}_{T,1}^{LM}\right)$ with the Haar wavelet filter is the VN ratio used in Sargan and Bhargava (1983).

Remark 3.2. Generalizing the local power properties of FG_1^L in Theorem 3.4 (ii) to trend stationary alternatives, Theorem 3.5 shows that under the sequence of local alternatives (14), the asymptotic distributions of the test statistics developed in this subsection are of the same form as those under the null hypothesis except that the Brownian motion is replaced with the Ornstein-Uhlenbeck process $J_c^M(\cdot)$ when $\alpha = 0$ and with $J_c^d(\cdot)$ when $\alpha \neq 0$. This implies that their asymptotic power is the same as that of Sargan-Bhargava test. Hence, the local power analysis provided in Elliott *et al.* (1996) (ERS) applies to our tests.

4 Monte Carlo Simulations

In this section, we investigate the finite sample performance of the new unit root tests against trend stationary alternatives and compare them against the Elliott *et al.* (1996) (ERS) and Ng and Perron (2001) (MPP) tests. To save space, we restrict ourselves to non-zero mean and linear trend cases only.[9]

The asymptotic critical values of tests based on $\widehat{S}_{T,1}^{LM}$ and $\widehat{S}_{T,1}^{Ld}$ are tabulated in Table 1. These critical values are calculated from one million replications. The simulations are carried out for a sample size of 1,000 observations and 5,000 replications. Under the alternative, we discard the first 1,000 observations as transients. We have purposely chosen ρ values of 0.99 and 0.98 to seek the power of the tests for very near unit root alternatives.

In Tables 2 and 3, we examine the size and power properties of the wavelet tests for $\widehat{S}_{T,1}^{LM}$ and $\widehat{S}_{T,1}^{Ld}$ with serially correlated errors. The error process is a stationary AR(1) with a parameter (γ) in the range of -0.8, -0.5, 0, 0.5, and 0.8. We set the bandwidth for the long-run variance to 20 with the Bartlett kernel for the $\widehat{S}_{T,1}^{LM}$ and $\widehat{S}_{T,1}^{Ld}$ tests to ensure that the empirical sizes are close to their nominal ones across these ranges of serially correlated errors.[10] As illustrated in Table 2, the wavelet test with demeaned series has higher power relative to ERS and MPP tests when $\gamma < 0$. At the 5% level, $\gamma = -0.8$, and $\rho = 0.99$, the $\widehat{S}_{T,1}^{LM}$, ERS and MPP tests have powers of 99.7%, 40.1% and 40.2%, respectively. For $\gamma = -0.5$, $\rho = 0.99$, and at the 5% level, the $\widehat{S}_{T,1}^{LM}$, ERS and MPP tests have powers of 87.1%, 39.6% and 39.3%, respectively. When $\gamma < 0$, the contribution of the error persistence concentrates in higher frequencies and it becomes easier for the wavelet test to separate such persistence from low frequency ones. For $\gamma > 0$, the simulation results between three tests are in the same order of magnitude for up to $\gamma = 0.5$. For $\gamma = 0.8$, the ERS and MPP tests perform better for certain critical levels. This is partly due to the fact that the wavelet test under-rejects for our choice of the bandwith for $\gamma = 0.8$.

A similar type of performance comparison is observed in Table 3 for $\widehat{S}_{T,1}^{Ld}$. The wavelet

[9] In the following tables, we report empirical size and power and do not adjust the empirical power for slight variations in empirical size.

[10] It might be possible to choose the bandwidth according to a criterion, but this is beyond the scope of this paper. We intend to investigate this issue in future work.

test does much better for $\gamma < 0$. For $\gamma = -0.8$, $\rho = 0.99$, and at the 5% level, the $\widehat{S}_{T,1}^{Ld}$, ERS and MPP tests have powers of 96.8%, 18.5% and 16.3%, respectively. On the other hand, ERS and MPP tests do better than the wavelet test for $\gamma = 0.8$. For $\gamma = 0.8$, $\rho = 0.99$, and at the 5% level, the $\widehat{S}_{T,1}^{Ld}$, ERS and MPP tests have powers of 6.9%, 26.0% and 24.5%, respectively. This is partly due to the fact that the wavelet test under-rejects for our choice of the bandwith for $\gamma = 0.8$.

5 Conclusions

Our unit root tests provide a novel approach in disbalancing the energy in the data by constructing test statistics from its lower frequency dynamics. We contribute to the unit root literature on three different fronts. First, we propose a unified wavelet spectral approach to unit root testing; second, we provide a wavelet spectral interpretation of existing Von Neumann variance ratio tests, and finally, we propose higher order wavelet filters to capture low-frequency stochastic trends parsimoniously and gain power against near unit root alternatives in finite samples. In our tests, the intuitive construction and simplicity are worth emphasizing. The simulation studies demonstrate the comparable power of our tests with reasonable empirical sizes.

Several extensions of our tests are possible. First, our tests make use of the unit scale DWT only ($J = 1$) and hence of the energy decomposition of $\{y_t\}$ into frequency bands $[0, 1/2]$ and $[1/2, 1]$. Heuristically, these tests are suitable for testing a unit root process against alternatives that have most energy concentrated in the frequency band $[1/2, 1]$. To distinguish between a unit root process and a 'strongly' dependent process that has substantial energy in frequencies close to zero, we need to further decompose the low frequency band $[0, 1/2]$. DWT of higher scales ($J > 1$) provides a useful device. The choice of J thus depends on the energy concentration of the alternative process against which the unit root hypothesis is being tested. It is possbile to extend our tests to make use of DWT of higher scales J, although the technical analysis would be more complicated. This and the choice of J are currently being pursued by the authors. Second, we show in the paper that the Sargan and Bhargava test is a special case of wavelet based tests with MODWT using unit scale Haar wavelet filter. MODWT has proven to have advantages over DWT in various situations including wavelet variance estimation. It would be interesting to see if it also has advantages in the context of testing unit root. Thirdly, the unit root tests developed in this paper can be extended to residual-based tests for cointegration in the same way that other unit root tests have been extended, see e.g., Phillips and Ouliaris (1990) and Stock (1999). This is also being investigated by the authors.

Appendix: Technical Proofs

Proof of Lemma 3.1. Suppose H_0 holds. Then $y_t = y_{t-1} + u_t$. Equations (9) and (10) imply:

$$W_{t,1} = \frac{1}{\sqrt{2}} u_{2t} \text{ and } V_{t,1} = \frac{1}{\sqrt{2}}(2y_{2t-1} + u_{2t}). \tag{17}$$

Using Equation (17), together with Equation (11), we obtain

$$\hat{S}_{T,1} = \frac{\sum_{t=1}^{T/2} V_{t,1}^2}{\sum_{t=1}^{T/2} V_{t,1}^2 + \frac{1}{2}\sum_{t=1}^{T/2} u_{2t}^2}, \tag{18}$$

where

$$\sum_{t=1}^{T/2} V_{t,1}^2 = \frac{1}{2}\{4\sum_{t=1}^{T/2} y_{2t-1}^2 + 4\sum_{t=1}^{T/2} u_{2t} y_{2t-1} + \sum_{t=1}^{T/2} u_{2t}^2\} \equiv 2A_T + 2B_T + \frac{1}{2}C_T, \tag{19}$$

in which $A_T = \sum_{t=1}^{T/2} x_t^2$, $B_T = \sum_{t=1}^{T/2} u_{2t} x_t$, and $C_T = \sum_{t=1}^{T/2} u_{2t}^2$ with $x_t \equiv y_{2t-1}$ for $t = 1, 2, \ldots, T/2$.

Below, we show that under H_0,

$$A_T = O_p(T^2),\ B_T = O_p(T),\ C_T = O_p(T). \tag{20}$$

Let $T_1 = \frac{T}{2}$. By Proposition 17.2 in Hamilton (1994), we have

$$x_t = x_0 + \sum_{j=1}^{t} v_t = x_0 + \sum_{j=0}^{2t-1} u_j = x_0 + \left\{ u_0 + \psi(1)\sum_{j=1}^{2t-1} \epsilon_j + \eta_{2t-1} - \eta_0 \right\}.$$

Define the partial sum process associated with $\{v_t\}$ as $X_{T_1}(r) = \frac{1}{T_1}\sum_{t=1}^{[T_1 r]} v_t$, $0 \le r \le 1$.Then it follows that

$$X_{T_1}(r) \overset{L}{=} \frac{1}{T_1}\psi(1) \sum_{j=1}^{2[T_1 r]-1} \epsilon_j = 2\psi(1)\frac{1}{T}\sum_{j=1}^{[Tr]-1} \epsilon_j.$$

By the functional Central Limit Theorem (CLT), we obtain $\sqrt{T} X_{T_1}(\cdot) \Longrightarrow 2\psi(1)\sigma W(\cdot)$. Observing that $\sum_{t=1}^{T_1} x_t^2 = \frac{T_1^2}{2}\int_0^1 \left\{T X_{T_1}^2(r)\right\} dr$, we obtain by the Continuous Mapping Theorem (CMT),

$$\frac{1}{T_1^2}\sum_{t=1}^{T_1} x_t^2 \Longrightarrow \frac{1}{2}\lambda_v^2 \int_0^1 W^2(r) dr,$$

where $\lambda_v = 2\psi(1)\sigma$. As a result, we get

$$T_1^{-2} A_T \Longrightarrow \frac{1}{2}\,\lambda_v^2 \int_0^1 [W(r)]^2 dr. \tag{21}$$

We now look at B_T. Recall that $B_T = \sum_{t=1}^{T_1} u_{2t}y_{2t-1}$. Simple algebra shows that $E(B_T) = \frac{1}{2}\sum_{s=1}^{T-1}(T-s-1)\gamma_s = O(T)$ and $Var(T^{-1}B_T) = o(1)$, where $\gamma_j = \sigma^2 \sum_{s=0}^{\infty} \psi_s\psi_{s+j}$, for $j = 0, 1, 2, \ldots$. Hence $B_T = O_p(T)$. The order of C_T follows from the Law of Large Numbers (LLN).

Hence under H_0, we get $\sum_{t=1}^{T_1} V_{t,1}^2 = O_p(T^2)$ and $\sum_{t=1}^{T_1} W_{t,1}^2 = O_p(T)$, implying that the energy of the scaling coefficients dominates that of the wavelet coefficients as mentioned above. Consequently,

$$\hat{S}_{T,1} = \frac{T^{-2}\sum_{t=1}^{T_1} V_{t,1}^2}{T^{-2}(\sum_{t=1}^{T_1} V_{t,1}^2 + \sum_{t=1}^{T_1} W_{t,1}^2)} = 1 + o_p(1). \tag{22}$$

Now consider what happens under H_1. In this case, $|\rho| < 1$ so that $y_t = \rho y_{t-1} + u_t$ and $\{y_t\}$ is a stationary short memory process. Thus, under H_1, both $\{V_{t,1}\}$ and $\{W_{t,1}\}$ are stationary, short memory processes. Moreover,

$$\frac{2}{T_1}\sum_{t=1}^{T_1} W_{t,1}^2 = \frac{1}{T_1}\sum_{t=1}^{T_1} y_{2t}^2 + \frac{1}{T_1}\sum_{t=1}^{T_1} y_{2t-1}^2 - \frac{2}{T_1}\sum_{t=1}^{T_1} y_{2t}y_{2t-1} = \frac{2\gamma_0}{1+\rho} + o_p(1),$$

implying $\sum_{t=1}^{T_1} W_{t,1}^2 = O_p(T)$. Similarly, we obtain $\sum_{t=1}^{T_1} V_{t,1}^2 = O_p(T)$, since $\frac{2}{T_1}\sum_{t=1}^{T_1} V_{t,1}^2 = \frac{2\gamma_0}{1-\rho} + o_p(1)$. As a result, we obtain

$$\begin{aligned}
\hat{S}_{T,1} &= \frac{T_1^{-1}\sum_{t=1}^{T_1} V_{t,1}^2}{T_1^{-1}\sum_{t=1}^{T_1} V_{t,1}^2 + T_1^{-1}\sum_{t=1}^{T_1} W_{t,1}^2} = \frac{E(V_{t,1}^2)}{E(V_{t,1}^2) + E(W_{t,1}^2)} + o_p(1) \\
&= \frac{E(y_{2t} + y_{2t-1})^2}{E(y_{2t} + y_{2t-1})^2 + E(y_{2t} - y_{2t-1})^2} + o_p(1). \qquad (23)
\end{aligned}$$

Proof of Theorem 3.2. Under H_0, we note that

$$\hat{S}_{T,1} - 1 = -\frac{C_T/2 - \frac{T}{4}\gamma_0}{2A_T + 2B_T + C_T} - \frac{\frac{T}{4}\gamma_0}{2A_T + 2B_T + C_T},$$

where A_T, B_T, C_T are defined in (19). Note that $C_T = \sum_{t=1}^{T_1} u_{2t}^2$ and $E(C_T) = T_1 E(u_{2t}^2) = T_1\gamma_0$, in which $\gamma_0 = \sigma^2 \sum_{s=0}^{\infty} \psi_s^2$. We obtain $T_1^{-1}C_T - \gamma_0 = o_p(1)$. This, (21), and the fact that $B_T = O_p(T)$ imply:

$$\begin{aligned}
T_1(\hat{S}_{T,1} - 1) &= -\frac{T_1^{-1}\left(C_T/2 - \frac{T_1}{2}\gamma_0\right)}{2T_1^{-2}(A_T + B_T + C_T/2)} - \frac{\frac{1}{2}\gamma_0}{2T_1^{-2}(A_T + B_T + C_T/2)} \\
&= -\frac{o_p(1)}{\lambda_v^2\int_0^1 [W(r)]^2 dr} - \frac{\frac{1}{2}\gamma_0}{\lambda_v^2\int_0^1 [W(r)]^2 dr} \\
&= -\frac{\gamma_0}{2\lambda_v^2\int_0^1 [W(r)]^2 dr} + o_p(1),
\end{aligned}$$

where $\lambda_v^2 = 4\omega^2$.

Proof of Theorem 3.3. **(i)** Under $H_0 : \rho = 1$. We now show that $T_1^{-1} \sum_{t=L_1}^{T_1} W_{t,1}^2 = E\left(W_{t,1}^2\right) + o_p(1)$ and $T_1^{-2} \sum_{t=L_1}^{T_1} V_{t,1}^2 = O_p(1)$. Hence, under H_0, we obtain

$$\hat{S}_{T,1}^L = \frac{1}{1 + \frac{\sum_{t=L_1}^{T_1} W_{t,1}^2}{\sum_{t=L_1}^{T_1} V_{t,1}^2}} = \frac{1}{1 + \frac{O_p(T)}{O_p(T^2)}} = 1 + o_p(1).$$

To show $T_1^{-1} \sum_{t=L_1}^{T_1} W_{t,1}^2 = E\left(W_{t,1}^2\right) + o_p(1)$, we note:

$$W_{t,1} = y_{2t+1-L} \sum_{l=0}^{L-1} h_l + \sum_{l=0}^{L-2} h_l \left\{ \sum_{j=0}^{L-2-l} u_{2t-j-l} \right\} = \sum_{l=0}^{L-2} h_l \left\{ \sum_{j=0}^{L-2-l} u_{2t-j-l} \right\},$$

implying that $W_{t,1}$ is a finite linear combination of $\{u_t\}$. The claim follows immediately from Assumptions 1 and 2.

Now we consider the order of $\sum_{t=L_1}^{T_1} V_{t,1}^2$. Noting that

$$V_{t,1} = y_{2t+1-L} \sum_{l=0}^{L-1} g_l + \sum_{l=0}^{L-2} g_l \left\{ \sum_{j=0}^{L-2-l} u_{2t-j-l} \right\} = \sqrt{2} y_{2t+1-L} + \sum_{l=0}^{L-2} g_l \left\{ \sum_{j=0}^{L-2-l} u_{2t-j-l} \right\},$$

we obtain

$$\begin{aligned}
\frac{1}{T_1^2} \sum_{t=L_1}^{T_1} V_{t,1}^2 &= \frac{1}{T_1^2} \sum_{t=L_1}^{T_1} \left[\sqrt{2} y_{2t+1-L} + \sum_{l=0}^{L-2} g_l \left\{ \sum_{j=0}^{L-2-l} u_{2t-j-l} \right\} \right]^2 \\
&= \frac{2}{T_1^2} \sum_{t=L_1}^{T_1} y_{2t+1-L}^2 + \frac{1}{T_1^2} \sum_{t=L_1}^{T_1} \left[\sum_{l=0}^{L-2} g_l \left\{ \sum_{j=0}^{L-2-l} u_{2t-j-l} \right\} \right]^2 \\
&\quad + \frac{2\sqrt{2}}{T_1^2} \sum_{t=L_1}^{T_1} y_{2t+1-L} \left[\sum_{l=0}^{L-2} g_l \left\{ \sum_{j=0}^{L-2-l} u_{2t-j-l} \right\} \right] \\
&= \frac{2}{T_1^2} \sum_{t=L_1}^{T_1} y_{2t+1-L}^2 + o_p(1) \\
&= O_p(1).
\end{aligned}$$

If $|\rho| < 1$, then $\{y_t\}$ is a stationary short memory process. Since both $\{W_{t,1}\}$ and $\{V_{t,1}\}$ are obtained from finite linear combinations of $\{y_t\}$, we can show that $T_1^{-1} \sum_{t=L_1}^{T_1} W_{t,1}^2 = E\left(W_{t,1}^2\right) + o_p(1)$ and $T_1^{-1} \sum_{t=L_1}^{T_1} V_{t,1}^2 = E\left(V_{t,1}^2\right) + o_p(1)$, implying $\hat{S}_{T,1}^L = \frac{E\left(V_{t,1}^2\right)}{E\left(V_{t,1}^2\right) + E\left(W_{t,1}^2\right)} + o_p(1)$.

(ii) Since under the null hypothesis, $\frac{1}{T_1^2} \sum_{t=L_1}^{T_1} V_{t,1}^2 = \frac{2}{T_1^2} \sum_{t=L_1}^{T_1} y_{2t+1-L}^2 + o_p(1)$, the asymptotic distribution of $\frac{1}{T_1^2} \sum_{t=L_1}^{T_1} V_{t,1}^2$ is given by that of $2A_T^L \equiv \frac{2}{T_1^2} \sum_{t=L_1}^{T_1} y_{2t+1-L}^2$. Similar to the derivation of the asymptotic distribution of A_T in the proof of Lemma 3.1, one can show that $T_1^{-2} A_T^L \implies \frac{1}{2}\lambda_v^2 \int_0^1 [W(r)]^2 dr$. On the other hand, extending the proof of Lemma 3.1, we can show that $T_1^{-1} \sum_{t=L_1}^{T_1} W_{t,1}^2 - EW_{t,1}^2 = o_p(1)$. Hence under the null hypothesis,

$$\begin{aligned}
T_1(\hat{S}_{T,1}^L - 1) &= -\frac{T_1^{-1} \sum_{t=L_1}^{T_1} (W_{t,1}^2 - EW_{t,1}^2)}{T_1^{-2}\left(\sum_{t=L_1}^{T_1} V_{t,1}^2 + \sum_{t=L_1}^{T_1-1} W_{t,1}^2\right)} - \frac{T_1^{-1}(T_1 - L_1) EW_{t,1}^2}{T_1^{-2}\left(\sum_{t=L_1}^{T_1} V_{t,1}^2 + \sum_{t=L_1}^{T_1} W_{t,1}^2\right)} \\
&= -\frac{o_p(1)}{\lambda_v^2 \int_0^1 [W(r)]^2 dr} - \frac{EW_{t,1}^2}{\lambda_v^2 \int_0^1 [W(r)]^2 dr} \\
&= -\frac{EW_{t,1}^2}{\lambda_v^2 \int_0^1 [W(r)]^2 dr} + o_p(1).
\end{aligned}$$

Proof of Theorem 3.4. **(i)** From Theorem 3.3 (i), we know: $\hat{S}_{T,1}^L - 1 = (c_L - 1) + o_p(1)$, where

$$c_L - 1 = -\frac{E\left(W_{t,1}^2\right)}{E\left(W_{t,1}^2\right) + E\left(V_{t,1}^2\right)} < 0.$$

This, together with the consistency of $\hat{\lambda}_v^2$ and $\widehat{v}_{y,1}^2$, imply:

$$T_1^{-1}\left(FG_1^L\right) = \frac{\hat{\lambda}_v^2}{\widehat{v}_{y,1}^2}\left[\hat{S}_{T,1}^L - 1\right] = \frac{\lambda_\nu}{E\left(W_{t,1}^2\right)}\left(c_L - 1\right) + o_p(1).$$

The conclusion follows from this and the fact that $\frac{\lambda_\nu}{E\left(W_{t,1}^2\right)}(c_L - 1) < 0$.

(ii) For notational simplicity, we present a detailed proof for $L = 2$, i.e., for FG_1. The general case follows the same arguments with more tedious notation just as Theorem 3.3 (ii) extends Theorem 3.2. Under (14), $y_t = \exp\left(\frac{c}{T}\right) y_{t-1} + u_t$. Using the same arguments as in the proof of Lemma 3.1 and Lemma 1 in Phillips (1987b), we can show:

$$\frac{1}{T_1^2} \sum_{t=1}^{T/2} y_{2t-1}^2 \implies \frac{1}{2}\lambda_v^2 \int_0^1 [J_c(r)]^2 dr \text{ and } \sum_{t=1}^{T/2} y_{2t-1} u_{2t} = O_p(T) \tag{24}$$

where $T_1 = T/2$. Equations (9) and (10) imply:

$$W_{t,1} = \frac{1}{\sqrt{2}} u_{2t} - \frac{1}{\sqrt{2}}\left[1 - \exp\left(\frac{c}{T}\right)\right] y_{2t-1} \text{ and } V_{t,1} = \frac{1}{\sqrt{2}}\left[1 + \exp\left(\frac{c}{T}\right)\right] y_{2t-1} + \frac{1}{\sqrt{2}} u_{2t}.$$

Thus,

$$\begin{aligned}
2\sum_{t=1}^{T/2} W_{t,1}^2 &= \sum_{t=1}^{T/2} u_{2t}^2 + \sum_{t=1}^{T/2}\left[1 - \exp\left(\frac{c}{T}\right)\right]^2 y_{2t-1}^2 - 2\sum_{t=1}^{T/2} u_{2t}\left[1 - \exp\left(\frac{c}{T}\right)\right] y_{2t-1} \\
&\sim \sum_{t=1}^{T/2} u_{2t}^2 + \frac{c^2}{T^2}\sum_{t=1}^{T/2} y_{2t-1}^2 + 2\frac{c}{T}\sum_{t=1}^{T/2} u_{2t} y_{2t-1} \\
&= \sum_{t=1}^{T/2} u_{2t}^2 + O_p(1),
\end{aligned}$$

where we have used: $\exp\left(\frac{c}{T}\right) \sim 1 + \frac{c}{T}$ and (24). Similarly, we obtain:

$$
\begin{aligned}
2\sum_{t=1}^{T/2} V_{t,1}^2 &= \sum_{t=1}^{T/2} u_{2t}^2 + \sum_{t=1}^{T/2} \left[1+\exp\left(\frac{c}{T}\right)\right]^2 y_{2t-1}^2 - 2\sum_{t=1}^{T/2} u_{2t}\left[1+\exp\left(\frac{c}{T}\right)\right] y_{2t-1} \\
&\sim \sum_{t=1}^{T/2} u_{2t}^2 + \sum_{t=1}^{T/2} \left[2+\frac{c}{T}\right]^2 y_{2t-1}^2 - 2\sum_{t=1}^{T/2} u_{2t}\left[2+\frac{c}{T}\right] y_{2t-1} \\
&= \sum_{t=1}^{T/2} u_{2t}^2 + 4\sum_{t=1}^{T/2} y_{2t-1}^2 + O_p(T).
\end{aligned}
$$

So, under (14), we have:

$$
\hat{S}_{T,1} - 1 = -\frac{\sum_{t=1}^{T/2} W_{t,1}^2}{\sum_{t=1}^{T/2} V_{t,1}^2 + \sum_{t=1}^{T/2} W_{t,1}^2} = \frac{\sum_{t=1}^{T/2} u_{2t}^2 + O_p(1)}{4\sum_{t=1}^{T/2} y_{2t-1}^2 + O_p(T) + 2\sum_{t=1}^{T/2} u_{2t}^2 + O_p(1)},
$$

implying:

$$
T_1(\hat{S}_{T,1} - 1) = -\frac{T_1^{-1}\sum_{t=1}^{T/2} u_{2t}^2 + o_p(1)}{4T_1^{-2}\sum_{t=1}^{T/2} y_{2t-1}^2 + o_p(1)} = -\frac{\gamma_0}{2\lambda_v^2\int_0^1 [J_c(r)]^2 dr} + o_p(1),
$$

where we have again used (24).

Proof of Theorem 3.5. The proofs are similar to those of Theorem 3.3 (ii) and Theorem 3.4 and are thus omitted.

Level		
1%	5%	10%
	FG_1^L	
-29.04	-17.75	-13.09
	$\widehat{S}_{T,1}^{LM}$	
-40.38	-27.38	-21.75
	$\widehat{S}_{T,1}^{Ld}$	
-50.77	-36.54	-30.23

FG_1^L is the wavelet test for no drift. $\widehat{S}_{T,1}^{LM}$ and $\widehat{S}_{T,1}^{Ld}$ are the wavelet tests for trend stationary alternatives without and with linear trends, respectively. Entries are based on one million Monte Carlo replications.

Table 1: Critical Values of Wavelet Tests

ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%
		$\widehat{S}_{T,1}^{LM}$			ERS			MPP	
				$\gamma = -0.80$					
1.00	0.009	0.068	0.119	0.014	0.047	0.097	0.011	0.045	0.099
0.99	0.982	0.997	0.998	0.156	0.401	0.587	0.144	0.402	0.610
0.98	1.000	1.000	1.000	0.451	0.702	0.827	0.443	0.704	0.840
				$\gamma = -0.50$					
1.00	0.006	0.045	0.103	0.011	0.051	0.102	0.011	0.049	0.108
0.99	0.668	0.871	0.937	0.148	0.396	0.569	0.141	0.393	0.592
0.98	0.984	1.000	1.000	0.487	0.746	0.846	0.479	0.748	0.863
				$\gamma = 0.00$					
1.00	0.006	0.046	0.087	0.013	0.052	0.099	0.011	0.052	0.106
0.99	0.153	0.486	0.687	0.163	0.423	0.596	0.156	0.416	0.611
0.98	0.683	0.954	0.991	0.488	0.741	0.846	0.495	0.743	0.855
				$\gamma = 0.50$					
1.00	0.006	0.038	0.085	0.015	0.055	0.112	0.013	0.053	0.118
0.99	0.069	0.316	0.543	0.168	0.422	0.605	0.162	0.417	0.619
0.98	0.374	0.845	0.953	0.475	0.715	0.844	0.473	0.721	0.856
				$\gamma = 0.80$					
1.00	0.007	0.031	0.056	0.013	0.048	0.098	0.011	0.048	0.097
0.99	0.021	0.189	0.386	0.155	0.405	0.585	0.148	0.402	0.601
0.98	0.198	0.668	0.883	0.460	0.708	0.821	0.454	0.712	0.833

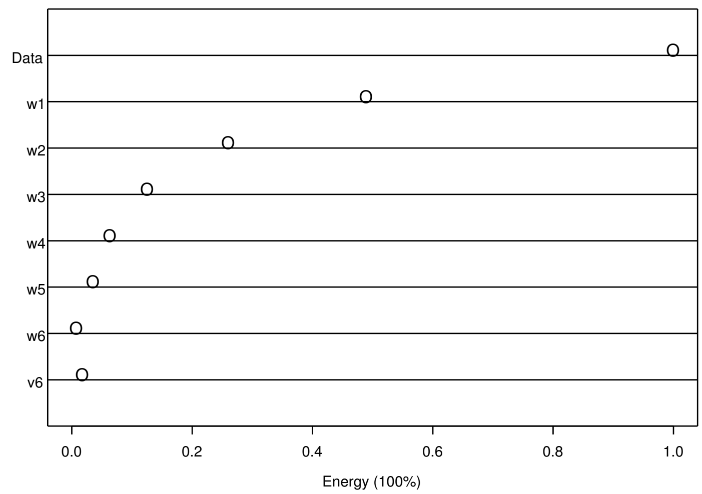
Table 2: Size and Power of the $\widehat{S}_{T,1}^{LM}$ - Demeaned Series with Serially Correlated Errors

The wavelet test statistic is calculated with a unit scale ($J = 1$) discrete wavelet transformation and with the Haar filter. The data generating process is $y_t = \mu + y_t^s$, where $y_t^s = \rho y_{t-1}^s + u_t$, $u_t = \gamma u_{t-1} + \epsilon_t$ $\epsilon_t \sim iidN(0, 1)$, $\mu = 1$ and $y_0 = 0$. Under the null $\rho = 1$ and under the alternative $\rho < 1$. The asymptotic critical values of the $\widehat{S}_{T,1}^{LM}$ test are tabulated in Table 1. The bandwidth is set to 20 with the Bartlett kernel in the calculation of the long-run variance of the wavelet test. The lag length of the ERS and MPP test regressions are determined by minimizing the modified AIC with the maximum lag length of 12. All simulations are with 1,000 observations and 5,000 replications.

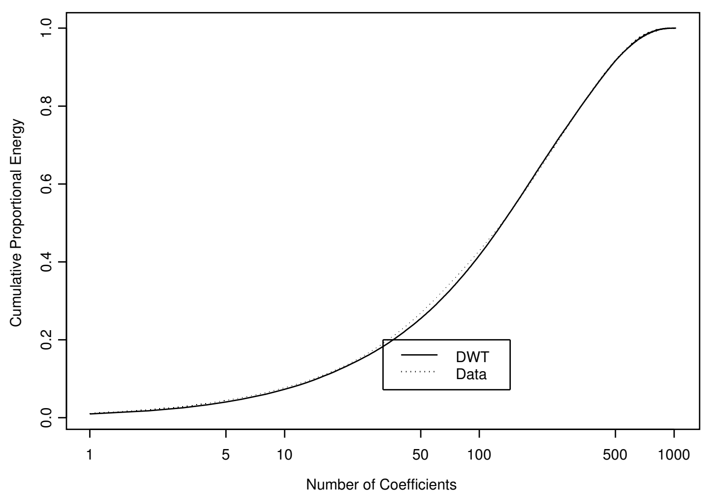
ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%
		$\widehat{S}_{T,1}^{Ld}$			ERS			MPP	
					$\gamma = -0.80$				
1.00	0.006	0.058	0.121	0.008	0.043	0.093	0.007	0.043	0.084
0.99	0.878	0.968	0.989	0.041	0.185	0.341	0.039	0.163	0.324
0.98	0.996	1.000	1.000	0.224	0.581	0.753	0.232	0.551	0.732
					$\gamma = -0.50$				
1.00	0.052	0.044	0.096	0.013	0.485	0.102	0.012	0.046	0.096
0.99	0.745	0.926	0.971	0.056	0.222	0.383	0.059	0.208	0.363
0.98	0.976	0.998	0.999	0.335	0.672	0.824	0.345	0.649	0.813
					$\gamma = 0.00$				
1.00	0.002	0.041	0.072	0.011	0.049	0.104	0.011	0.048	0.097
0.99	0.332	0.643	0.795	0.069	0.242	0.399	0.073	0.225	0.377
0.98	0.783	0.946	0.978	0.338	0.666	0.815	0.349	0.651	0.801
					$\gamma = 0.50$				
1.00	0.001	0.039	0.056	0.008	0.052	0.103	0.011	0.049	0.096
0.99	0.055	0.245	0.415	0.076	0.262	0.419	0.081	0.246	0.394
0.98	0.267	0.649	0.825	0.312	0.647	0.789	0.320	0.626	0.773
					$\gamma = 0.80$				
1.00	0.005	0.040	0.052	0.014	0.055	0.104	0.015	0.052	0.096
0.99	0.007	0.069	0.175	0.074	0.260	0.389	0.076	0.245	0.367
0.98	0.053	0.303	0.521	0.288	0.594	0.761	0.298	0.582	0.752

Table 3: SIZE AND POWER OF THE $\widehat{S}_{T,1}^{Ld}$ - GLS DETRENDED SERIES WITH SERIALLY CORRELATED ERRORS
The wavelet test statistic is calculated with a unit scale ($J = 1$) discrete wavelet transformation and with the Haar filter. The data generating process is $y_t = \mu + \alpha t + y_t^s$, where $y_t^s = \rho y_{t-1}^s + u_t$, $u_t = \gamma u_{t-1} + \epsilon_t \;\; \epsilon_t \sim iidN(0,1)$, $\mu = 1$, $\alpha = 1$ and $y_0 = 0$. Under the null $\rho = 1$ and under the alternative $\rho < 1$. The asymptotic critical values of the $\widehat{S}_{T,1}^{Ld}$ test are tabulated in Table 1. The bandwidth is set to 20 with the Bartlett kernel in the calculation of the long-run variance of the wavelet test. The lag length of the ERS and MPP test regressions are determined by minimizing the modified AIC with the maximum lag length of 12. All simulations are with 1,000 observations and 5,000 replications.

Figure 1: Wavelet Variance Decomposition of a White Noise Process



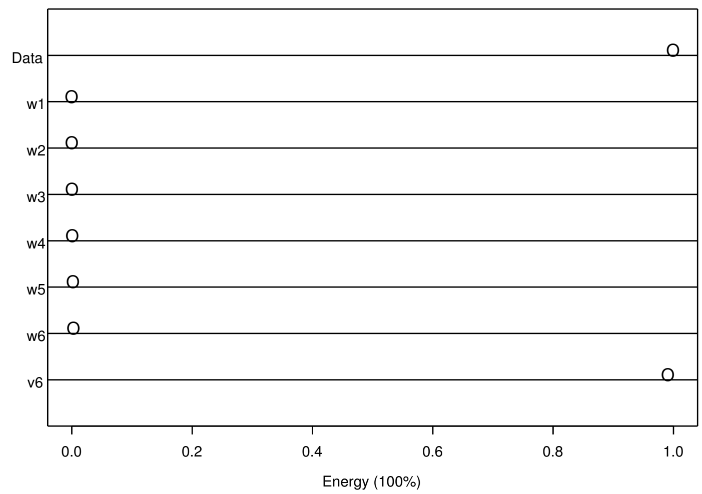
(a)



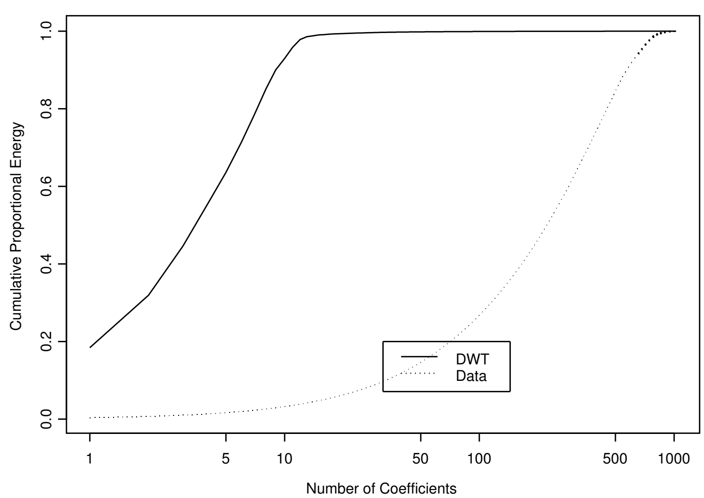
(b)

The energy decomposition of a white noise process through a six level discrete wavelet decomposition (DWT) with 1024 observations. **(a)** "Data" represents the total energy of the data which is normalized at one. w$i, i = 1, \ldots, 6$ represents the percentage energy of the wavelet coefficients. v6 is the percentage energy of the scale coefficients. The energies of the wavelet and scaling coefficients are equal to the sum of the energy of the data. The energy is the highest at the highest frequency wavelet coefficient (w1) and declines gradually towards the lowest frequency wavelet coefficient (w6). The percentage energy of the scaling coefficient (v6) is zero. **(b)** This figure compares the proportional energy of the data to the proportional energy of all coefficients. The number of coefficients needed is equal to the number of data points to account for the total energy of the data. The horizontal axis is on natural logarithmic scale.

Figure 2: Wavelet Variance Decomposition of a Unit Root Process



(a)



(b)

The energy decomposition of a unit root process through a six level discrete wavelet decomposition (DWT) with 1024 observations. **(a)** "Data" represents the total energy of the data which is normalized at one. w$i, i = 1, \dots, 6$ represents the percentage energy of wavelet coefficients. v6 is the percentage energy of the scaling coefficients. The energies of the wavelet and scaling coefficients are equal to the sum of the energy of the data. The energy is the highest for the scaling coefficients and close to zero for wavelet coefficients. The percentage energy of the scaling coefficients (v6) is close to the energy of the data. **(b)** This figure compares the proportional energy of the data to the proportional energy of all coefficients. The number of coefficients needed equals 41 ($41/1024 = 4\%$) of the total number of coefficients to account for the total energy of the data. The horizontal axis is on natural logarithmic scale.

References

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

Berkes, I., Hormann, S., and Horvath, L. (2008). The functional central limit theorem for a family of garch observations with applications. *Statistics & Probability Letters*, **78**, 2725–2730.

Bhargava, A. (1986). On the theory of testing for unit roots in observed time series. *Review of Economic Studies*, **53**, 369–384.

Cai, Y. and Shintani, M. (2006). On the alternative long-run variance ratio test for a unit root. *Econometric Theory*, **22**, 347–372.

Chan, N. H. and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics*, **15**, 1050–1063.

Choi, I. and Phillips, P. C. B. (1993). Testing for a unit root by frequency domain regression. *Journal of Econometrics*, **59**, 263–286.

Coifman, R. R. and Donoho, D. L. (1995). Translation invariant denoising. *Wavelets and Statistics, ed. A. Antoniadis and G. Oppenheim, Vol. 103, New York, Springer-Verlag*, pages 125–150.

Daubechies, I. (1992). *Ten Lectures on Wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia.

Davidson, R., Labys, W. C., and Lesourd, J.-B. (1998). Walvelet analysis of commodity price behavior. *Computational Economics*, **11**, 103–128.

Dickey, D. A. and Fuller, W. A. (1979). Distributions of the estimators for autoregressive time series with a unit root. *Journal of American Statistical Association*, **74**, 427–431.

Duchesne, P. (2006a). On testing for serial correlation with a wavelet-based spectral density estimator in multivariate time series. *Econometric Theory*, **22**, 633–676.

Duchesne, P. (2006b). Testing for multivariate autoregressive conditional heteroskedasticity using wavelets. *Computational Statistics & Data Analysis*, **51**, 2142–2163.

Dufour, J. M. and King, M. (1991). Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary and nonstationary errors. *Journal of Econometrics*, **47**, 115–143.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, **64**, 813–836.

Fan, Y. (2003). On the approximate decorrelation property of the discrete wavelet transform for fractionally differenced processes. *IEEE Transactions on Information Theory*, **49**, 516–521.

Fan, Y. and Whitcher, B. (2003). A wavelet solution to the spurious regression of fractionally differenced processes. *Applied Stochastic Models in Business and Industry*, **19**, 171–183.

Gençay, R., Selçuk, F., and Whitcher, B. (2001). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press, San Diego.

Gençay, R., Selçuk, F., and Whitcher, B. (2003). Systematic risk and time scales. *Quantitative Finance*, **3**, 108–116.

Gençay, R., Selçuk, F., and Whitcher, B. (2005). Multiscale systematic risk. *Journal of International Money and Finance*, **24**, 55–70.

Granger, C. W. J. (1966). The typical spectral shape of an economic variable. *Econometrica*, **34**, 150–161.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

Hong, Y. (2000). Wavelet-based estimation for heteroskedasticity and autocorrelation consistent variance-covariance matrices. Technical report, Department of Economics and Department of Statistical Science, Cornell University.

Hong, Y. and Kao, C. (2004). Wavelet-based testing for serial correlation of unknown form in panel models. *Econometrica*, **72**, 1519–1563.

Hong, Y. and Lee, J. (2001). One-sided testing for ARCH effects using wavelets. *Econometric Theory*, **17**, 1051–1081.

Lee, J. and Hong, Y. (2001). Testing for serial correlation of unknown form using wavelet methods. *Econometric Theory*, **17**, 386–423.

MacKinnon, J. G. (2000). Computing numerical distribution functions in econometrics. *High Performance Computing Systems and Applications, ed. A. Pollard, D. Mewhort, and D. Weaver, Amsterdam, Kluwer*, pages 455–470.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674–693.

Mallat, S. (1998). *A Wavelet Tour of Signal Processing*. Academic Press, San Diego.

Nason, G. P. and Silverman, B. W. (1995). The stationary wavelet transform and some statistical applications. *Wavelets and Statistics, Volume 103 of Lecture Notes in Statistics, ed. A. Antoniadis and G. Oppenheim, Springer Verlag, New York*, pages 281–300.

Nelson, C. R. and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, **10**, 139–162.

Newey, W. K. and West, K. D. (1987). A simple positive semidefinite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703–708.

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

Park, H. and Fuller, W. (1995). Alternative estimators and unit root tests for the autoregressive process. *Journal of Time Series Analysis*, **16**, 415–429.

Park, J. Y. and Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, **4**, 468–497.

Park, J. Y. and Phillips, P. C. B. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, **5**, 95–131.

Percival, D. B. (1995). On estimation of the wavelet variance. *Biometrica*, **82**, 619–631.

Percival, D. B. and Mofjeld, H. O. (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association*, **92**, 868–880.

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

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, **33**, 311–340.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica*, **55**, 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrica*, **74**, 535–547.

Phillips, P. C. B. and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica*, **58**, 165–193.

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

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

Phillips, P. C. B. and Xiao, Z. (1998). A primer on unit root testing. *Journal of Economic Surveys*, **12**, 423–469.

Ramsey, J. B. (1999). The contribution of wavelets to the anlaysis of economic and financial data. *Philosophical Transactions of the Royal Society of London A*, **357**, 2593–2606.

Ramsey, J. B. (2002). Wavelets in economics and finance: Past and future. *Studies in Nonlinear Dynamics & Econometrics*, **3**, 1–29.

Ramsey, J. B. and Zhang, Z. (1997). The analysis of foreign exchange data using waveform dictionaries. *Journal of Empirical Finance*, **4**, 341–372.

Ramsey, J. B., Zaslavsky, G., and Usikov, D. (1995). An analysis of U. S. stock price behavior using wavelets. *Fractals*, **3**, 377–389.

Sargan, J. D. and Bhargava, A. (1983). Testing residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica*, **51**, 153–174.

Schmidt, P. and Phillips, P. C. B. (1992). LM tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics*, **54**, 257–288.

Sims, C. A., Stock, J. H., and Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica*, **58**, 113–144.

Stock, J. H. (1995). Unit roots, structural breaks and trends. *Handbook of Econometrics, ed. R. F. Engle and D. McFadden, Amsterdam, North-Holland*, pages 2739–2841.

Stock, J. H. (1999). A class of tests for integration and cointegration. *Cointegration, Causality, and Forecasting Festschrift in Honour of Clive W. J. Granger, ed. R. F. Engle and H. White, Oxford, Oxford University Press, Chapter 6.*