Unit Root and Cointegration Tests with Wavelets*

Yanqin Fan† Ramazan Gençay‡

February 2007

Abstract

This paper develops a wavelet (spectral) approach to test 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 which have substantial power 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. Our tests utilize the wavelet coefficients of the coarsest scale. We demonstrate the size and power properties of our tests through Monte Carlo simulations, and apply them to financial time series. A generalization of our unit root tests to the maximum overlap DWT (MODWT) and to residual based tests for cointegration are also provided.

Keywords: Unit root tests, cointegration, discrete wavelet transformation, maximum overlap wavelet transformation, energy decomposition.

JEL No: C1, C2, C12, C22, F31, G0, G1.

*We thank Stelios Bekiros, Buz Brock, Cees Diks, Cars Hommes, Benoit Perron, Hashem Pesaran, James MacKinnon, 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 CIREQ, IRMACS, University of Amsterdam, Cambridge University and University College Dublin. `Email: gencay@sfu.ca`

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[1] has attracted the attention of many researchers. 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 the Von Neumann variance ratio (VN) tests of Sargan and Bhargava (1983) and their extensions.[2] 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 spectral (wavelet) approach to testing unit roots inspired by Granger (1966). As we demonstrate, the VN tests of Sargan and Bhargava (1983) and Bhargava (1986) are special cases of our wavelet unit root tests.

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 (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 which have substantial power 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 coefficients referred to as the scaling coefficients. Our tests utilize the scaling coefficients of the coarsest scale. In particular, we con-

[1]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), Phillips (1987) pioneered the use of the functional central limit theorem to calculate 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 (1987) 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.

[2]Cai and Shintani (2006) provide alternative VN tests.

struct test statistics from the ratio of the energy from the low-frequency scale to the total energy (variance) of the time series. We establish asymptotic properties of our tests and generalize our unit root tests to residual based tests for cointegration. 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).[3] Thus, orthogonality of the transform is lost but it has been shown that the wavelet variance utilizing MODWT coefficients is more efficient than the one obtained through the orthonormal DWT. Percival (1995) gives the asymptotic relative efficiencies for the wavelet variance estimator based on the orthonormal DWT compared to the estimator based on the MODWT. We generalize our tests to the MODWT setting to utilize these efficiency gains.

The VN tests 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. The VN tests are special 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 and also DWT/MODWT of higher scales, our tests gain power over the VN tests in finite samples.

An alternative spectral approach to time series analysis is that of the Fourier spectral analysis. 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. One example of the successful application of wavelets in nonstationary time series analysis 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

[3]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).

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.

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. We contribute to the unit root literature on three different fronts. First, we propose a unified 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. 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 remarkable power against near unit root alternatives. The application to financial time series provides a means to compare the power within these tests.

In section two, we begin with an 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 via DWT. Section four considers the empirically more relevant cases where the alternative processes against which the unit root process is tested may have deterministic trends. Section five develops similar tests via MODWT. Section six provides Monte Carlo simulations on the size and power properties of our tests. In section seven, an application to financial time series is carried out. In section eight, we consider extensions of the previously developed unit root tests to residual-based tests for cointegration. We conclude thereafter. An appendix contains technical proofs.

2 Wavelets

A wavelet is a small wave which grows and decays in a limited time period.[4] To formalize the notion of a wavelet, let $\psi(.)$ be a real valued function such that its integral is zero,

$$\int_{-\infty}^{\infty} \psi(t)\, dt = 0, \tag{1}$$

and its square integrates to unity,

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

[4]This section closely follows Gençay *et al.* (2001). The contrasting notion is a big wave such as the sine function which keeps oscillating indefinitely.

While Equation (2) indicates that $\psi(.)$ has to make some excursions away from zero, any excursions it makes above zero must cancel out excursions below zero due to Equation (1), and hence $\psi(.)$ is a wave, or a wavelet.

Fundamental properties of the continuous wavelet functions (filters), such as integration to zero and unit energy (Equations (1) and (2)), have discrete counterparts. Let $h = (h_0, \ldots, h_{L-1})$ be a finite length discrete wavelet filter such that it integrates (sums) to zero

$$\sum_{l=0}^{L-1} h_l = 0 \tag{3}$$

and has unit energy

$$\sum_{l=0}^{L-1} h_l^2 = 1. \tag{4}$$

In addition to Equations (3) and (4), the wavelet (or high-pass) filter h is orthogonal to its even shifts; that is,

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0, \quad \text{for all nonzero integers } n. \tag{5}$$

These conditions state that a wavelet filter should sum to zero, must have unit energy and must be orthogonal to its even shifts. We have already discussed the first two conditions earlier in Equations (1) and (2). Equations (4) and (5) are known as the orthonormality property of wavelet filters.

The natural object to complement a high-pass filter is a low-pass (scaling) filter g. By applying both h and g to an observed time series, we can separate high-frequency oscillations from low-frequency ones. We will denote a low-pass filter as $g = (g_0, \ldots, g_{L-1})$. The low-pass filter coefficients are determined by the *quadrature mirror relationship*[5]

$$g_l = (-1)^{l+1} h_{L-1-l} \quad \text{for} \quad l = 0, \ldots, L-1 \tag{6}$$

and the inverse relationship is given by $h_l = (-1)^l g_{L-1-l}$. The basic properties of the scaling filter are

$$\sum_{l=0}^{L-1} g_l = \sqrt{2} \tag{7}$$

$$\sum_{l=0}^{L-1} g_l^2 = 1 \tag{8}$$

[5]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.

$$\sum_{l=0}^{L-1} g_l g_{l+2n} = \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0, \tag{9}$$

for all nonzero integers n, and

$$\sum_{l=0}^{L-1} g_l h_{l+2n} = \sum_{l=-\infty}^{\infty} g_l h_{l+2n} = 0 \tag{10}$$

for all integers n. Equation (7) states that scaling coefficients are local weighted averages. Equations (8) and (9) indicate that scale coefficients satisfy the orthonormality property that they possess unit energy and are orthogonal to even shifts.

The filter sequences $\{h_l\}$ and $\{g_l\}$ are high-pass and low-pass filters, respectively. Let $H(f)$ be the transfer (or gain) function of $\{h_l\}$ defined via the discrete Fourier transform (DFT); i.e.,

$$H(f) = \sum_{l=0}^{L-1} h_l \exp(-i2\pi f l),$$

and let $G(f)$ be the discrete Fourier transform of $\{g_l\}$. Displaying the squared gain functions $\mathcal{H}(f)$ and $\mathcal{G}(f) = |G(f)|^2$ illustrates the frequency range captured by the wavelet and scaling filters. To construct the orthonormal matrix that defines the discrete wavelet transformation (DWT), wavelet coefficients cannot interact with one another. Equations 4 and 5 may be succinctly expressed in the frequency domain via the squared gain function

$$\mathcal{H}(f) + \mathcal{H}(f + \tfrac{1}{2}) = 2 \quad \text{for all } f. \tag{11}$$

This result can be motivated by Equation 5. The DFT of the left-hand side of Equation 5 is given by $[\mathcal{H}(f/2) + \mathcal{H}(f/2 + 1/2)]/2$. Hence, Equation 5 can be re-expressed via its inverse DFT

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = \int_{-\infty}^{\infty} \frac{1}{2} \left[\mathcal{H}(\tfrac{f}{2}) + \mathcal{H}(\tfrac{f}{2} + \tfrac{1}{2}) \right] e^{i2\pi f n} \, df.$$

By plugging Equation 11 into the above equation, we obtain

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = \int_{-\infty}^{\infty} e^{i2\pi f n} \, df = \begin{cases} 1 & n = 0 \\ 0 & \text{otherwise} \end{cases}$$

and, thus, Equations 4 and 5 are both satisfied. Figure 1a shows an ideal high-pass filter for $f \in [0, 1/4]$ and an approximation to this ideal filter by the Daubechies extremal phase wavelet filter of length 4.

Finally, a band-pass filter has a squared gain function that covers an interval of frequencies and then decays to zero as $f \to 0$ and $f \to 1/2$. We may construct a band-pass filter by recursively

applying a combination of low-pass and high-pass filters. Let $h_{J,l}$ be a band-pass filter produced by a combination of J filters. Starting with $h_{1,l}$, we obtain the filtered output $u_{J,t}$ via

$$\begin{aligned} u_{1,t} &= h_{1,l} * x_t \\ u_{2,t} &= h_{2,l} * u_{1,t} \\ \vdots &\ \vdots\ \vdots \\ u_{J,t} &= h_{J,l} * u_{J-1,t} \end{aligned} \tag{12}$$

for $t = 0, \pm 1, \pm 2, \ldots$ The *equivalent filter* h_l for the "cascade" of filters in Equation 12 has the transfer function

$$H(f) = \prod_{j=1}^{J} H_j(f). \tag{13}$$

This may be seen by viewing Equation 12 in the frequency domain via Fourier transforms; that is,

$$\begin{aligned} U_1(f) &= H_1(f)X(f) \\ U_2(f) &= H_2(f)U_1(f) = H_2(f)H_1(f)X(f) \\ \vdots &\ \vdots\ \vdots \\ U_J(f) &= H_J(f)U_{J-1}(f) = H_J(f)H_{J-1}(f)\cdots H_1(f)X(f), \end{aligned}$$

where $X(f)$ is the DFT of x_t. After applying a series of J filters, the DFT of the output series $u_{J,t}$ is given by the product of all previous Fourier transforms and may be depicted graphically through the following flowchart:

$$x_t \longrightarrow \boxed{H_1(f)} \longrightarrow \boxed{H_2(f)} \longrightarrow \cdots \longrightarrow \boxed{H_J(f)} \longrightarrow u_{J,t}.$$

Using Equation 13 and the convolution property of the Fourier transform, $U_J(f) = H(f)X(f)$ and, therefore, $u_{J,t}$ is simply the convolution of x_t with h_l; that is,

$$x_t \longrightarrow \boxed{H(f)} \longrightarrow u_{J,t}.$$

Figure 1c shows an ideal band-pass filter for $f \in [1/8, 1/4]$ and an approximation to this ideal filter by a filter cascade of the Daubechies extremal phase scaling and wavelet filters of length 4.

2.1 The Haar Wavelet and Daubechies Wavelet Filters

The Haar (1910) wavelet filter is an excellent benchmark to illustrate $\{h_{j,l}\}$ and $\{g_{j,l}\}$. The Haar wavelet remained in relative obscurity until the convergence of several disciplines to form what we

now know in a broad sense as wavelet methodology. It is a filter of length $L = 2$ that can be succinctly defined by its scaling (low-pass) filter coefficients

$$g_0 = g_1 = \frac{1}{\sqrt{2}},$$

or equivalently by its wavelet (high-pass) filter coefficients $h_0 = 1/\sqrt{2}$ and $h_1 = -1/\sqrt{2}$ through the inverse quadrature mirror relationship. The Haar wavelet is special since it is the only symmetric compactly supported orthonormal wavelet (Daubechies, 1992, Ch. 8). Figure 2 shows the unit scale wavelet filter coefficients $h_{1,l} = (1/\sqrt{2}, -1/\sqrt{2})$, where the first subscript denotes the scale of the filter, along with higher scale wavelet filter coefficients. The first row of filter coefficients illustrate a simple difference operation, where projecting $h_{1,l}$ onto a vector produces a difference between two adjacent observations. The next set of filter coefficients $h_{2,l}$ is a simple difference also, but two pairs of observations are averaged and those averages differenced. This theme is repeated in $h_{3,l}$ and $h_{4,l}$, where the averages are growing in length before being differenced.

Although the Haar wavelet filter is easy to visualize and implement, it is a poor approximation to an ideal band-pass filter. Figure 3 shows the squared gain functions for the scale 1–4 wavelet filter coefficients. An ideal band-pass filter is proportional to one inside the desired frequency interval and zero at all other frequencies. The dotted lines in each panel of Figure 3 indicate the frequency interval for the ideal (nominal) band-pass filter. The squared gain functions associated with the level j Haar wavelet filter do not decay rapidly outside this nominal frequency range, indicating the filter is a poor approximation to the ideal band-pass filter.

The Haar wavelet filter is a member of the so-called Daubechies (1992) wavelet filters. The Daubechies (1992) wavelet filters represent a collection of wavelets that improve on the frequency-domain characteristics of the Haar wavelet and may still be interpreted as generalized differences of adjacent averages. Daubechies derived these wavelets from the criterion of a compactly supported function with the maximum number of vanishing moments.[6] In general, there are no explicit time-domain formulae for this class of wavelet filters. The squared gain function of $\{h_l\}$ can be written explicitly as

$$\mathcal{H}(f) \equiv |H(f)|^2 = \mathcal{D}^{L/2}(f)\mathcal{C}(f), \quad \mathcal{C}(f) \equiv \frac{1}{2^{L-1}} \sum_{l=0}^{L/2-1} \binom{L/2-1+l}{l} \cos^{2l}(\pi f),$$

where $\mathcal{D}(f) \equiv 4\sin^2(\pi f)$ is the squared gain function of a first order backward difference filter and $\mathcal{C}(f)$ is the squared gain function of a low-pass or (weighted) averaging filter. We can thus regard $\{h_l\}$ as a two stage filter; the first stage is an $L/2$th order backward difference filter, and the second is a weighted average filter.

The squared gain function does not uniquely characterize a sequence of Daubechies wavelet filters. If we think of a sequence of wavelet filter coefficients h_l, it may be represented by using the

[6] A function $\psi(t)$ with P vanishing moments satisfies $\int t^p \psi(t) dt = 0$, where $p = 0, 1, \ldots, P-1$.

DFT yielding its transfer function $H(f)$. In polar notation the complex-valued transfer function may be written as

$$H(f) = |H(f)| \exp[i\theta(f)] = [\mathcal{H}(f)]^{1/2} \exp[i\theta(f)],$$

so that although two filters share the same squared gain function, they may differ in phase. The process of *spectral factorization* (Oppenheim and Schafer, 1989) may be used to obtain the different filters via finding the roots of $|H(f)|$. The number of possible factorizations increases as the length of the filter L increases. Daubechies first choose an *extremal phase* factorization,[7] whose resulting wavelets we denote by D(L) where L is the length of the filter. An alternative factorization leads to the *least asymmetric* class of wavelets, which we denote by LA(L).[8] The D(4) wavelets have a simple expression in the time domain via

$$h_0 = \frac{1-\sqrt{3}}{4\sqrt{2}},\ h_1 = \frac{-3+\sqrt{3}}{4\sqrt{2}},\ h_2 = \frac{3+\sqrt{3}}{4\sqrt{2}} \quad \text{and} \quad h_3 = \frac{-1-\sqrt{3}}{4\sqrt{2}}.$$

Figures 4 and 5 illustrate the LA(8) filter coefficients and the squared gain functions, respectively. Note, the D(8) will have an identical squared gain function since it only differs from the LA(8) with respect to its phase function. Contrasting the squared gain functions between Figures 3 and 5, it is obvious that the LA(8) is a much better approximation to an ideal band-pass filter than the Haar (less leakage). This improvement may lead to power gains in the unit root tests developed in this paper. Throughout the rest of this paper, we use the Daubechies compactly supported wavelet filters exclusively. Figure 6 illustrates wavelet filters associated with Daubechies families of wavelets with lengths $L \in \{2, 4, 8\}$.[9]

2.2 Discrete wavelet transformation

In principle, wavelet analysis can be carried out in all arbitrary time scales. This may not be necessary if only key features of the data are in question, and if so, discrete wavelet transformation (DWT) is an efficient and parsimonious route as compared to the continuous wavelet transformation. The DWT is a subsampling of $W(\lambda, t)$ with only dyadic scales, i.e., λ is of the form $2^{j-1},\ j = 1, 2, 3, \ldots$ and, within a given dyadic scale 2^{j-1}, t's are separated by multiples of 2^j.

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

$$\mathbf{w} = \mathcal{W}\mathbf{y},$$

[7] The term *extremal* (or minimum) *phase spectral factorization* is associated with a solution to the roots of $|H(f)|$ that are all inside the unit circle (Daubechies, 1992, Ch. 6).

[8] Symmetric filters are known as *linear phase* filters in the engineering literature. The degree of asymmetry for a filter may therefore be measured by the deviation from linearity of its phase. Least asymmetric filters are associated with a phase that is as close to linear as possible (Daubechies, 1992, Ch. 8).

[9] Tables 4.1 and 4.2 in Gençay *et al.* (2001) provide selected Daubechies wavelet filters.

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).[10] The vector of wavelet coefficients may be organized into $M+1$ vectors,

$$\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M, \mathbf{v}_M]^T, \tag{14}$$

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

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

$$W_{t,1} = \sum_{l=0}^{L-1} h_l y_{2t+1-l} \qquad V_{t,1} = \sum_{l=0}^{L-1} g_l y_{2t+1-l} \tag{15}$$

where $t = (L-2)/2, (L-2)/2+1, \ldots, T/2-1$. Similarly, the level j BI wavelet and scaling coefficients of $\{y_t\}$ are given by

$$W_{t,j} = \sum_{l=0}^{L^{(j)}-1} h_{j,l} y_{2^j t+1-l} \qquad V_{t,j} = \sum_{l=0}^{L^{(j)}-1} g_{j,l} y_{2^j t+1-l} \tag{16}$$

where $L^{(j)} = (2^j - 1)(L-1) + 1$ is the length of the level j wavelet filter $\{h_{j,l}\}$.

2.3 Analysis of variance

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

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

This can be easily proven through basic matrix manipulation via

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

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

$$\|\mathbf{y}\|^2 = \sum_{j=1}^{M} \|\mathbf{w}_j\|^2 + \|\mathbf{v}_M\|^2 , \tag{17}$$

where $\|\mathbf{w}_j\|^2$ is the sum of squared variation of $\mathbf{y}$ due to changes at scale λ_j and $\|\mathbf{v}_M\|^2$ is the information due to changes at scales λ_M and higher.

[10] Since DWT is an orthonormal transform, orthonormality implies that $\mathbf{y} = \mathcal{W}^T\mathbf{w}$ and $||\mathbf{w}||^2 = ||\mathbf{y}||^2$.

The motivation behind a wavelet based unit root test can be illustrated through the energy (variance) decomposition of the process at a given level of decomposition $J < M$. For a white noise process, $\|\mathbf{v}_J\|^2 / \|\mathbf{y}\|^2$ is close to zero whereas $\|\mathbf{v}_J\|^2 / \|\mathbf{y}\|^2$ is close to one for a unit root process. In Figure 7, the dot chart of a Gaussian white noise process is plotted for 1024 observations ($M = 2^{10} = 1024$). A six level ($J = 6$) discrete wavelet decomposition (DWT) is used. "Data" represents the total energy of the data which is normalized at one. w$i, i = 1, \ldots, 6$ represents the percentage energy of wavelet coefficients. v6 is the percentage energy of the scaling coefficients. The 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) is close to zero.

In Figure 8, the dot chart of a unit root process

$$y_t = y_{t-1} + u_t, \qquad u_t \sim iidN(0,1) \tag{18}$$

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) is almost equal to the energy of the data.[11] 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.

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, our statistical test for a unit root process is based on this principle of energy decomposition.

3 New Unit Root Tests — No Drift Case

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

$$y_t = \rho y_{t-1} + u_t, \tag{19}$$

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. Also, we use the symbols $\Longrightarrow$ and $\longrightarrow$ to denote convergence of the associated probability measures and convergence in probability respectively. All the limits are taken as the sample size approaches ∞.

[11] 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.

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 $\psi(L)$. Under Assumption 1, we have $\omega^2 = \psi(1)^2\sigma^2$ and $T^{-1/2}\sum_{t=1}^{[Tr]} u_t \implies \omega W(r)$ where $[Tr]$ denotes the integer part of Tr and $W(r)$ denotes a standard Brownian motion defined on $C[0,1]$. It is known that the weak convergence result: $T^{-1/2}\sum_{t=1}^{[Tr]} u_t \implies \omega W(r)$ holds for more general/other classes of processes than the class of linear processes specified in Assumption 1. 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 this section, we consider tests for $H_0 : \rho = 1$ against $H_1 : |\rho| < 1$. Therefore, under the alternative hypothesis, $\{y_t\}$ is a zero-mean stationary process with the long run variance $(1-\rho)^{-2}$ ω^2. As mentioned above, 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. Different tests can be constructed depending on (i) the wavelet filter being used to decompose the time series; and (ii) the level of DWT of the time series. To introduce the fundamental idea, we first develop a test based on the Haar wavelet filter and unit scale DWT and then extend it to tests based on any Daubechies (1992) compactly supported wavelet filter of finite length and the DWT of any finite level.

3.1 The first test — Haar wavelet filter and the DWT of unit scale

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

$$W_{t,1} = \frac{1}{\sqrt{2}}(y_{2t} - y_{2t-1}),\ t = 1, 2, \ldots, T/2, \tag{20}$$

$$V_{t,1} = \frac{1}{\sqrt{2}}(y_{2t} + y_{2t-1}),\ t = 1, 2, \ldots, T/2. \tag{21}$$

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:

$$\hat{S}_{T,1} = \frac{\sum_{t=1}^{T/2} V_{t,1}^2}{\sum_{t=1}^{T/2} V_{t,1}^2 + \sum_{t=1}^{T/2} W_{t,1}^2}. \tag{22}$$

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 much 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, $\hat{S}_{T,1} = \frac{E(y_{2t}+y_{2t-1})^2}{E(y_{2t}+y_{2t-1})^2+E(y_{2t}-y_{2t-1})^2}+o_p(1)$.*

Noting the inequality: $\frac{E(y_{2t}+y_{2t-1})^2}{E(y_{2t}+y_{2t-1})^2+E(y_{2t}-y_{2t-1})^2} < 1$, we expect our tests based on $\hat{S}_{T,1}$ to have power against H_1. Below, we provide the main steps of a proof of Lemma 3.1, because the fundamental idea of this paper is easily grasped from this proof. Suppose H_0 holds; $\rho = 1$ and hence $y_t = y_{t-1} + u_t$. Equations (20) and (21) imply:

$$W_{t,1} = \frac{1}{\sqrt{2}} u_{2t} \text{ and } V_{t,1} = \frac{1}{\sqrt{2}}(2y_{2t-1} + u_{2t}). \tag{23}$$

Using Equation (23), together with Equation (22), we obtain

$$\hat{S}_{T,1} = \frac{\sum_{t=1}^{T/2} V_{t,1}^2}{\sum_{t=1}^{T/2} V_{t,1}^2 + \frac{1}{2}\sum_{t=1}^{T/2} u_{2t}^2}, \tag{24}$$

where

$$\begin{aligned} \sum_{t=1}^{T/2} V_{t,1}^2 &= \frac{1}{2}\{4\sum_{t=1}^{T/2} y_{2t-1}^2 + 4\sum_{t=1}^{T/2} u_{2t}y_{2t-1} + \sum_{t=1}^{T/2} u_{2t}^2\} \\ &\equiv 2A_T + 2B_T + \frac{1}{2}C_T, \end{aligned} \tag{25}$$

in which

$$A_T = \sum_{t=1}^{T/2} x_t^2, \qquad B_T = \sum_{t=1}^{T/2} u_{2t}x_t, \quad C_T = \sum_{t=1}^{T/2} u_{2t}^2$$

with $x_t \equiv y_{2t-1}$ for $t = 1, 2, \ldots, T/2$. In the Appendix, we show that under H_0,

$$A_T = O_p(T^2),\ B_T = O_p(T),\ C_T = O_p(T). \tag{26}$$

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

$$\hat{S}_{T,1} = \frac{T^{-2}\sum_{t=1}^{T/2} V_{t,1}^2}{T^{-2}(\sum_{t=1}^{T/2} V_{t,1}^2 + \sum_{t=1}^{T/2} W_{t,1}^2)} = 1 + o_p(1). \tag{27}$$

The orders stated in (26) are easily understood, given that $\{u_{2t}\}$ is a linear, short memory stationary process and $\{x_t\}$ can be easily shown to be a unit root process, i.e., $x_t = x_{t-1} + v_t$, where

$$v_t = u_{2t-2} + u_{2t-1},\ t = 1, ..., \frac{T}{2}.$$

The detailed proofs of (26) are given in the Appendix.

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 and by the Law of Large Numbers, we get $\sum_{t=1}^{T/2} V_{t,1}^2 = O_p(T)$ and $\sum_{t=1}^{T/2} W_{t,1}^2 = O_p(T)$ leading to:

$$\begin{aligned}
\hat{S}_{T,1} &= \frac{\left(\frac{T}{2}\right)^{-1} \sum_{t=1}^{T/2} V_{t,1}^2}{\left(\frac{T}{2}\right)^{-1} \sum_{t=1}^{T/2} V_{t,1}^2 + \left(\frac{T}{2}\right)^{-1} \sum_{t=1}^{T/2} W_{t,1}^2} \\
&\longrightarrow \frac{E(V_{t,1}^2)}{E(V_{t,1}^2) + E(W_{t,1}^2)} \\
&= \frac{E(y_{2t} + y_{2t-1})^2}{E(y_{2t} + y_{2t-1})^2 + E(y_{2t} - y_{2t-1})^2}. \qquad (28)
\end{aligned}$$

To sum up, 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 $\hat{S}_{T,1}$. To establish the asymptotic distribution of $\hat{S}_{T,1}$ 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 $C_T = \sum_{t=1}^{T/2} u_{2t}^2$ and $E(C_T) = \frac{T}{2}E(u_{2t}^2) = \frac{T}{2}\gamma_0$, in which $\gamma_0 = \sigma^2 \sum_{s=0}^{\infty} \psi_s^2$. Hence,

$$\begin{aligned}
\left(\frac{T}{2}\right)(\hat{S}_{T,1} - 1) &= -\frac{(T/2)^{-1}\left(C_T/2 - \frac{T}{4}\gamma_0\right)}{2(T/2)^{-2}(A_T + B_T + C_T/2)} - \frac{\frac{1}{2}\gamma_0}{2(T/2)^{-2}(A_T + B_T + C_T/2)} \\
&= -\frac{o_p(1)}{\lambda_v^2 \int_0^1 [W(r)]^2 dr} - \frac{\frac{1}{2}\gamma_0}{\lambda_v^2 \int_0^1 [W(r)]^2 dr} \\
&= -\frac{\gamma_0}{2\lambda_v^2 \int_0^1 [W(r)]^2 dr} + o_p(1),
\end{aligned}$$

where $\lambda_v^2 = 4\omega^2$. This result is summarized in the following theorem.

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

The above proof 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). Andrews (1991) showed that this long run variance estimator is consistent 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_1 under H_0 is extremely easy to simulate.[12] Draw a large sample of i.i.d. random numbers from $N(0,1)$ denoted as $\{z_i\}_{i=1}^{N}$. Compute the following quantity:

$$\frac{-1}{N^{-2}\sum_{i=1}^{N}\left(\sum_{s=1}^{i} z_s\right)^2}.$$

Simulate the above quantity a large number of times and use the resulting empirical distribution to approximate the null limiting distribution of FG_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.

[12] Please see MacKinnon (2000) for a detailed treatment.

3.2 General filter case: unit scale decomposition

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

$$W_{t,1} = \sum_{l=0}^{L-1} h_l y_{2t+1-l},\ V_{t,1} = \sum_{l=0}^{L-1} g_l y_{2t+1-l}, \tag{29}$$

where $t = (L-2)/2, (L-2)/2+1, \ldots, T/2-1$. 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.

Let $L_1 = (L-2)/2$ denote the number of boundary dependent coefficients. Define

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

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

Theorem 3.3 *(i)* $\hat{S}_{T,1}^L = 1+o_p(1)$ *under the null hypothesis of unit root and* $\hat{S}_{T,1}^L = c_L+o_p(1)$ *under the alternative hypothesis with* $c_L = \frac{EV_{t,1}^2}{EV_{t,1}^2+EW_{t,1}^2} < 1$*; (ii)* $\left(\frac{T}{2}\right)(\hat{S}_{T,1}^L - 1) = -\frac{EW_{t,1}^2}{\lambda_v^2 \int_0^1 [W(r)]^2 dr} + o_p(1)$ *under the null hypothesis.*

The proof of Theorem 3.3(i) is given in the Appendix. It implies that a consistent test for unit root can be based on $\hat{S}_{T,1}^L$. Theorem 3.3(ii) extends Theorem 3.2 from the Haar filter to any Daubechies compactly supported wavelet filter of finite length. Since as the length of the filter increases, the approximation of the Daubechies wavelet filter to the ideal high-pass filter improves, we expect tests based on $\hat{S}_{T,1}^L$ to gain power over the test based on the Haar filter for $L \geq 4$.

The proof of Theorem 3.3(ii) is outlined below. Since under the null hypothesis,

$$\frac{1}{T^2}\sum_{t=L_1}^{T/2-1} V_{t,1}^2 = \frac{2}{T^2}\sum_{t=L_1}^{T/2-1} y_{2t+2-L}^2 + o_p(1),$$

the asymptotic distribution of $\frac{1}{T^2}\sum_{t=L_1}^{T/2-1} V_{t,1}^2$ is given by that of $2A_T^L \equiv \frac{2}{T^2}\sum_{t=L_1}^{T/2-1} y_{2t+2-L}^2$. Similar to the derivation of the asymptotic distribution of A_T, one can show that

$$\left(\frac{T}{2}\right)^{-2} A_T^L \implies \frac{1}{2}\lambda_v^2 \int_0^1 [W(r)]^2 dr.$$

On the other hand, by the LLN for a stationary process, we get

$$\left(\frac{T}{2}\right)^{-1}\left(\sum_{t=L_1}^{T/2-1}(W_{t,1}^2 - EW_{t,1}^2)\right) = o_p(1).$$

Hence under the null hypothesis,

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

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

$$\widehat{v}_{y,1}^2 = \frac{1}{(T/2 - L_1)}\sum_{t=L_1}^{T/2-1}W_{t,1}^2. \tag{30}$$

Define the test statistic:

$$FG_1^L = \left(\frac{T}{2}\right)\frac{\hat{\lambda}_v^2}{\widehat{v}_{y,1}^2}\left[\hat{S}_{T,1} - 1\right].$$

Under the null hypothesis, the limiting distribution of FG_1^L is the same as that of FG_1. Percival (1995) showed that the wavelet variance estimator based on MODWT is more efficient than that based on DWT which may lead to power gains in finite samples for our test.

An alternative test is based on a modification of $(\hat{S}_{T,1}^L - 1)$. Define $\widehat{D}_{T,1}^L$ as follows:

$$\widehat{D}_{T,1}^L = -\frac{\sum_{t=L_1}^{T/2-1}W_{t,1}^2}{\sum_{t=1}^{T}y_t^2}.$$

Then we can easily show that Theorem 3.3 holds for $\widehat{D}_{T,1}^L$. A further modification would be to include the boundary dependent coefficients in the numerator of the above expression. This will not affect the asymptotic properties of the tests, but may affect their finite sample performance in particular when the length of the wavelet filter is long.

3.3 General filter case: Higher scale decomposition

The above tests make use of the unit scale DWT and hence 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 provides a useful device. We consider DWT of a finite scale J. Let $\{W_{t,j}\}$ denote the scale j BI wavelet coefficients of the $\{y_t\}$, i.e.,

$$W_{t,j} = \sum_{l=0}^{L^{(j)}-1} h_{j,l} y_{2^j t+1-l},$$

where $L^{(j)} = (2^j - 1)(L - 1) + 1$ is the length of the level j wavelet filter $\{h_{j,l}\}$. Similarly we use $\{V_{t,j}\}$ to denote the level j BI scaling coefficients of the $\{y_t\}$, i.e.,

$$V_{t,j} = \sum_{l=0}^{L^{(j)}-1} g_{j,l} y_{2^j t+1-l}. \tag{31}$$

Generalizing the tests developed in the previous subsection, we will consider the following statistic:

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

where L_j denotes the number of boundary dependent coefficients at level j. The level J scaling coefficients $\{V_{t,J}\}$ summarize the behavior of $\{y_t\}$ in the frequency band $[0, 1/2^J]$, while the wavelet coefficients at levels $1, 2, ..., J$ together summarize the behavior of the process $\{y_t\}$ in the frequency band $[1/2^J, 1]$.

Similar to the proof for the unit scale case, we can show that under H_0 :

$$\begin{aligned}
\left(\frac{T}{2^J}\right)(\hat{S}^L_{T,J} - 1) &= -\frac{\left(\frac{T}{2^J}\right)^{-1} \sum_{j=1}^{J} \left(\sum_{t=L_j}^{T/2^j-1} \left[W^2_{t,j} - \upsilon^2_{y,j} \right] \right) + \left(\frac{T}{2^J}\right)^{-1} \sum_{j=1}^{J} \left(\sum_{t=L_j}^{T/2^j-1} \upsilon^2_{y,j} \right)}{\left(\frac{T}{2^J}\right)^{-2} \left\{ \sum_{t=L_J}^{T/2^J-1} V^2_{t,J} + \sum_{j=1}^{J} \left(\sum_{t=L_j}^{T/2^j-1} W^2_{t,j} \right) \right\}} \\
&= -\frac{\left(\frac{1}{2^J}\right)^{-1} \sum_{j=1}^{J} \frac{1}{2^j} \upsilon^2_{y,j}}{\left(\frac{T}{2^J}\right)^{-2} \left\{ \sum_{t=L_J}^{T/2^J-1} V^2_{t,J} \right\}} + o_p(1),
\end{aligned}$$

where $\upsilon^2_{y,j} = EW^2_{t,j}$. It remains to work out the limiting distribution of $\left(\frac{T}{2^J}\right)^{-2} \left\{ \sum_{t=L_J}^{T/2^J-1} V^2_{t,J} \right\}$.

Suppose $\rho = 1$. Then

$$V_{t,J} = Y_{2^J t+2-L^{(J)}} \left(\sum_{l=0}^{L^{(J)}-1} g_{J,l} \right) + \sum_{l=0}^{L^{(J)}-2} g_{J,l} \left\{ \sum_{j=0}^{L^{(J)}-2-l} u_{2^{(J)}t+1-j-l} \right\}.$$

Since $\sum_{l=0}^{L^{(J)}-1} g_{J,l} = 2^{J/2}$, we obtain

$$\begin{aligned}
\frac{1}{T^2}\sum_{t=L_J}^{T/2^J-1} V_{t,J}^2 &= \frac{1}{T^2}\sum_{t=L_J}^{T/2^J-1}\left[2^{J/2}y_{2^Jt+2-L^{(J)}} + \sum_{l=0}^{L^{(J)}-2} g_{J,l}\left\{\sum_{j=0}^{L^{(J)}-2-l} u_{2^{(J)}t+1-j-l}\right\}\right]^2 \\
&= \frac{2^J}{T^2}\sum_{t=L_J}^{T/2^J-1} y_{2^Jt+2-L^{(J)}}^2 + \frac{1}{T^2}\sum_{t=L_J}^{T/2^J-1}\left[\sum_{l=0}^{L^{(J)}-2} g_{J,l}\left\{\sum_{j=0}^{L^{(J)}-2-l} u_{2^{(J)}t+1-j-l}\right\}\right]^2 \\
&\quad + \frac{2^{J/2+1}}{T^2}\sum_{t=L_J}^{T/2^J-1}\left[y_{2^Jt+2-L^{(J)}}\sum_{l=0}^{L^{(J)}-2} g_{J,l}\left\{\sum_{j=0}^{L^{(J)}-2-l} u_{2^{(J)}t+1-j-l}\right\}\right] \\
&= \frac{2^J}{T^2}\sum_{t=L_J}^{T/2^J-1} y_{2^Jt+2-L^{(J)}}^2 + o_p(1).
\end{aligned}$$

Hence,

$$\begin{aligned}
\left(\frac{T}{2^J}\right)^{-2}\left\{\sum_{t=L_J}^{T/2^J-1} V_{t,J}^2\right\} &= 2^J\left(\frac{T}{2^J}\right)^{-2}\sum_{t=L_J}^{T/2^J-1} y_{2^Jt+2-L^{(J)}}^2 + o_p(1) \\
&= \lambda_{v,J}^2\int_0^1 [W(r)]^2 dr + o_p(1),
\end{aligned}$$

where $\lambda_{v,J}^2 = 4^J\omega^2$. Consequently,

$$\begin{aligned}
\left(\frac{T}{2^J}\right)(\hat{S}_{T,J}^L - 1) &= -\frac{\left(\frac{1}{2^J}\right)^{-1}\sum_{j=1}^J \frac{1}{2^j}\upsilon_{y,j}^2}{\left(\frac{T}{2^J}\right)^{-2}\left\{\sum_{t=L_J}^{T/2^J-1} V_{t,J}^2\right\}} + o_p(1) \\
&= -\frac{\left(\frac{1}{2^J}\right)^{-1}\sum_{j=1}^J \frac{1}{2^j}\upsilon_{y,j}^2}{\lambda_{v,J}^2\int_0^1 [W(r)]^2 dr} + o_p(1).
\end{aligned}$$

Let $\widehat{\lambda}_{v,J}^2 = 4^J\widehat{\omega}^2$ and $\widehat{\upsilon}_{y,j}^2$ denote a consistent estimator of $\upsilon_{y,j}^2$, $j = 1, ..., J$. The test statistic is defined as

$$FG_J^L = \frac{T\widehat{\lambda}_{v,J}^2}{4^J\sum_{j=1}^J \frac{1}{2^j}\widehat{\upsilon}_{y,j}^2}(\hat{S}_{T,J}^L - 1)$$

whose limiting distribution under the null hypothesis is the same as that of FG_1^L. Obviously, if $J = 1$, then $FG_J^L = FG_1^L$ defined earlier.

An alternative test can be constructed based on a modification of $(\hat{S}_{T,J}^L - 1)$, i.e.,

$$\widehat{D}_{T,J}^L = -\frac{\sum_{j=1}^J\left(\sum_{t=L_j}^{T/2^j-1} W_{t,j}^2\right)}{\sum_{t=1}^T y_t^2}.$$

It is easy to show that under H_0,

$$T\widehat{D}^L_{T,J} \implies -\frac{4^J \sum_{j=1}^J \frac{1}{2^j} \upsilon^2_{y,j}}{\lambda^2_{v,J} \int_0^1 [W(r)]^2 dr}.$$

3.4 Power of the tests

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

$$\rho = \exp\left(\frac{c}{T}\right) \sim 1 + \frac{c}{T} \tag{32}$$

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

$$T^{-2} \sum_{t=1}^T y_t^2 \implies \omega^2 \int_0^1 [J_c(r)]^2 \, dr,$$

where

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

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 of local alternatives, the asymptotic distributions of the test statistics developed in the previous subsections 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 (32) and their asymptotic power is the same as that of Sargan and Bhargava test. As a result, any power difference among these tests must be determined via alternatives routes such as higher order asymptotic power functions, local alternatives of a different form, or Monte-Carlo simulation. We will explore the last approach in Section 6.

4 Incorporating a Drift Term

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

$$y_t = \alpha + \rho y_{t-1} + u_t, \tag{33}$$

where $\{u_t\}$ satisfies Assumption 1. In this section, we first construct tests for $H_0 : \rho = 1$ against $H_1 : |\rho| < 1$ for any given $\alpha \neq 0$. Therefore, under the alternative hypothesis, $\{y_t\}$ is a stationary process with a non-zero mean and the long run variance $\omega_y^2 = (1-\rho)^{-2} \, \omega^2$. We then develop tests that have power against trend stationary alternatives.

4.1 Haar filter case: unit scale decomposition

As in the no-intercept case, we use $\hat{S}_{T,1}$ defined in (24) as the basis for our test. Because of the presence of the intercept term, the asymptotic distribution of $\hat{S}_{T,1}$ under H_0 will be different from that when there is no intercept, leading to a completely different test. To be specific, note that under H_0, $y_t = \alpha + y_{t-1} + u_t$ implying $y_t = y_0 + \alpha t + \sum_{j=1}^{t} u_j$. As a result, the process $\{x_t\}$ is characterized by:

$$x_t = y_{2t-1} = y_0 + \alpha(2t-1) + \sum_{j=1}^{2t-1} u_j.$$

Using the above expression, one can show that

$$\begin{aligned} A_T &= \sum_{t=1}^{T_1} x_t^2 = \alpha^2 \sum_{t=1}^{T_1} (2t-1)^2 + o_p(T^3) \\ &= \frac{4}{3}\alpha^2 T_1^3 + o_p(T^3). \end{aligned}$$

Recall the following expression:

$$\begin{aligned} \hat{S}_{T,1} - 1 &= -\frac{C_T/2}{2A_T + 2B_T + C_T} \\ &= -\frac{C_T/2 - \frac{T}{4}E(\alpha + u_{2t})^2}{2A_T + 2B_T + C_T} - \frac{\frac{T}{4}E(\alpha + u_{2t})^2}{2A_T + 2B_T + C_T} \\ &\equiv -\frac{1}{2}F_T - \frac{1}{2}S_T. \end{aligned} \tag{34}$$

It is easy to show that

$$T_1^{5/2} F_T = \frac{T_1^{-1/2} \sum_{t=1}^{T_1} \left[(\alpha + u_{2t})^2 - E(\alpha + u_{2t})^2\right]}{T_1^3 \left(2A_T + 2B_T + C_T\right)} = \frac{N(0, \sigma_*^2)}{8\alpha^2/3} + o_p(1), \tag{35}$$

where σ_*^2 is the long run variance of $\left\{(\alpha + u_{2t})^2\right\}$. The second term satisfies

$$T_1^2 S_T = \frac{E(\alpha + u_{2t})^2}{T_1^{-3}\left[2A_T + 2B_T + C_T\right]} = \frac{E(\alpha + u_{2t})^2}{8\alpha^2/3} + o_p(1). \tag{36}$$

The following lemma follows immediately from Equations (34), (35), and (36).

Lemma 4.1 *Under H_0, (i) $T_1^2\left(\hat{S}_{T,1} - 1\right) = -\frac{3E(\alpha+u_{2t})^2}{16\alpha^2} + o_p(1)$; (ii) $T_1^{5/2}\left(\hat{S}_{T,1} - 1 + \frac{1}{2}S_T\right) = -\frac{3N(0,\sigma_*^2)}{16\alpha^2} + o_p(1)$.*

The first result in Lemma 4.1 is not useful for the construction of a proper test, because the limiting distribution is degenerate. The second result gives a non-degenerate limiting distribution,

but involves an unknown term S_T in the center. We need to find an estimator of S_T with a convergence rate faster than $T_1^{-5/2}$. To achieve this, we suggest the following procedure. Given the sample size T, choose T^* such that $T^*/T = o(1)$. Construct $\hat{S}_{T^*,1}$ using the first T^* observations. Then Lemma 4.1(ii) implies:

$$(T_1^*)^{5/2}\left(\hat{S}_{T^*,1} - 1 + \frac{1}{2}S_{T^*}\right) \implies -\frac{3N(0,\sigma_*^2)}{16\alpha^2},$$

where

$$S_{T^*} = \frac{2T_1^* EW_{t,1}^2}{\sum_{t=1}^{T^*/2} V_{t,1}^2 + \sum_{t=1}^{T^*/2} W_{t,1}^2} = \frac{2T_1^* EW_{t,1}^2}{\sum_{t=1}^{T^*} y_t^2}.$$

We need to estimate S_{T^*} with a convergence rate faster than $(T_1^*)^{-5/2}$. For this purpose, we use the whole sample $\{y_t\}_{t=1}^T$ to construct the unit scale wavelet coefficients: $\{W_{t,1}\}_{t=1}^{T_1}$ and estimate S_{T^*} by

$$\widehat{S}_{T^*} = \frac{2T_1^* \sum_{t=1}^{T_1} W_{t,1}^2}{T_1 \sum_{t=1}^{T^*} y_t^2}.$$

Since

$$\begin{aligned}
(T_1^*)^{5/2}\left(\widehat{S}_{T^*} - S_{T^*}\right) &= (T_1^*)^{5/2}\left[\frac{2T_1^* \sum_{t=1}^{T_1}(W_{t,1}^2 - EW_{t,1}^2)}{T_1 \sum_{t=1}^{T^*} y_t^2}\right] \\
&= \left[\frac{2T_1^{*1/2} \sum_{t=1}^{T_1}(W_{t,1}^2 - EW_{t,1}^2)}{T_1 (T_1^*)^{-3} \sum_{t=1}^{T^*} y_t^2}\right] \\
&= O_p\left(\frac{T_1^{*1/2}}{T_1^{1/2}}\right) \\
&= o_p(1),
\end{aligned}$$

we get

$$(T_1^*)^{5/2}\left(\hat{S}_{T^*,1} - 1 + \frac{1}{2}\widehat{S}_{T^*}\right) \implies -\frac{3N(0,\sigma_*^2)}{16\alpha^2}.$$

Now, let $\widehat{\alpha}$ be an estimator of α and $\widehat{\sigma}_*^2$ be an estimator of σ_*^2. Then the test statistic is given by

$$FG_M^L(drift) = -\frac{16\widehat{\alpha}^2 (T_1^*)^{5/2}\left(\hat{S}_{T^*,1} - 1 + \frac{1}{2}\widehat{S}_{T^*}\right)}{3\widehat{\sigma}_*}$$

and the limiting distribution of $FG_M^L(drift)$ under H_0 is the standard normal. Compared with tests in the no-drift case developed above, implementation of the tests in the drift case is complicated by the fact that one has to choose T^* and the only requirement for the asymptotic theory to hold is $T^*/T = o(1)$. Of course, the choice of T^* in finite samples would affect the power of the tests. An appropriate data-driven method for choosing T^* is definitely an important issue, although it is beyond the scope of the current paper.

4.2 General filter case: unit scale decomposition

Similar to the Haar case, using $y_t = y_0 + \alpha t + \sum_{j=1}^{t} u_j$ under H_0, we can establish the order of the energy of the scaling coefficients:

$$\sum_{t=L_1}^{T_1-1} V_{t,1}^2 = 2\alpha^2 \sum_{t=L_1}^{T_1-1} (2t+1)^2 + o_p(T^3) = \frac{8\alpha^2}{3} T_1^3 + o_p(T^3).$$

Decomposing $\hat{S}_{T,1}^L$ as follows:

$$\begin{aligned} \hat{S}_{T,1}^L - 1 &= -\frac{\sum_{t=L_1}^{T/2-1}(W_{t,1}^2 - EW_{t,1}^2)}{\sum_{t=L_1}^{T/2-1} V_{t,1}^2 + \sum_{t=L_1}^{T/2-1} W_{t,1}^2} - \frac{(T/2 - L_1)EW_{t,1}^2}{\sum_{t=L_1}^{T/2-1} V_{t,1}^2 + \sum_{t=L_1}^{T/2-1} W_{t,1}^2} \\ &= -F_T^L - S_T^L, \end{aligned}$$

we immediately get the following results:

$$T_1^{5/2} F_T^L = \frac{T_1^{-1/2} \sum_{t=L_1}^{T/2-1}(W_{t,1}^2 - EW_{t,1}^2)}{T_1^{-3} \sum_{t=L_1}^{T/2-1} V_{t,1}^2 + T_1^{-3} \sum_{t=L_1}^{T/2-1} W_{t,1}^2} = \frac{N(0, \sigma_{*W}^2)}{8\alpha^2/3} + o_p(1),$$

$$T_1^2 S_T^L = \frac{T_1^{-1}(T/2 - L_1)EW_{t,1}^2}{T_1^{-3} \sum_{t=L_1}^{T/2-1} V_{t,1}^2 + T_1^{-3} \sum_{t=L_1}^{T/2-1} W_{t,1}^2} = \frac{EW_{t,1}^2}{8\alpha^2/3} + o_p(1),$$

where σ_{*W}^2 is the long run variance of the process $\left\{W_{t,1}^2\right\}$.

Summarizing these results, we obtain the Lemma below, extending Lemma 4.1 to any Daubechies compactly supported wavelet filter of finite length. Following exactly the same procedure as in the Haar filter case, the Lemma below provides the basis for testing H_0.

Lemma 4.2 *Under H_0, (i)* $T_1^2 \left(\hat{S}_{T,1}^L - 1\right) = -\frac{3EW_{t,1}^2}{8\alpha^2} + o_p(1)$; *(ii)* $T_1^{5/2} \left(\hat{S}_{T,1}^L - 1 + S_T^L\right) = -\frac{3N(0,\sigma_{*W}^2)}{8\alpha^2} + o_p(1)$.

4.3 Tests against trend stationarity

Note that under H_0, model (33) implies that $y_t = y_0 + \alpha t + \sum_{j=1}^{t} u_j$. Thus y_t has a linear deterministic trend and a stochastic trend. Under the alternative, however, model (33) implies that the process $\{y_t\}$ is a stationary process with a non-zero mean. If one tests H_0 against the alternative hypothesis of a (linear) trend stationary process, then the above tests may not have power. To deal with trend stationary alternatives, components representation of a time series is often used and detrending performed, see Schmidt and Phillips (1992), Phillips and Xiao (1998), and Stock (1999). Phillips and Xiao (1998) also have a detailed discussion on efficient detrending for general trends. For ease of exposition, we restrict ourselves to non-zero mean and linear trend cases only.

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

$$y_t = \mu + \alpha t + y_t^s, \tag{37}$$

where $\{y_t^s\}$ is generated by model (19). 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

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

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

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

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 4.3 *Under H_0, we have: (i) $T\left(\widehat{D}_{T,1}^{LM}\right) \implies -\frac{E(W_{t,1}^M)^2}{2\omega^2 \int_0^1 [W_\mu(r)]^2 dr}$ and (ii) $T\left(\widehat{D}_{T,1}^{Ld}\right) \implies -\frac{E(W_{t,1}^d)^2}{2\omega^2 \int_0^1 [V_\mu(r)]^2 dr}$.*

The above theorem can be easily extended to (i) DWT of any finite scale J and (ii) a general trend in (39) using the detrending procedure discussed in Phillips and Xiao (1998). 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.

5 Maximum Overlap DWT

MODWT has been demonstrated to have advantages over DWT in several situations including the estimation of wavelet variance.[13] Apart from a factor of $\sqrt{2}$, the BI unit scale MODWT wavelet

[13] See Allan (1966), Howe and Percival (1995), Percival (1983), Percival and Guttorp (1994) and Percival (1995).

and scaling coefficients are given by

$$\widetilde{W}_{t,1} = \sum_{l=0}^{L-1} h_l y_{t+1-l},\ \widetilde{V}_{t,1} = \sum_{l=0}^{L-1} g_l y_{t+1-l}, \tag{38}$$

where $t = (L-2), (L-2)+1, \ldots, T-1$. 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. In this section, we will exploit these unit scale MODWT coefficients to construct tests for unit root. Extensions to higher scale MODWT are straightforward and are omitted for space considerations.

Let

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

We first consider the no-intercept case, then the intercept case, and finally trend-stationary alternatives. For the *no-intercept* case, following exactly the same derivation as in Subsection 3.2, we obtain

$$\frac{1}{T^2} \sum_{t=L-2}^{T-1} \widetilde{V}_{t,1}^2 = \frac{2}{T^2} \sum_{t=L-2}^{T-1} y_{t+2-L}^2 + o_p(1) \implies \frac{\lambda_v^2}{2} \int_0^1 [W(r)]^2 dr$$

so that under H_0, we have

$$T(\widetilde{S}_{T,1}^{L} - 1) = -\frac{E\widetilde{W}_{t,1}^2}{T^{-2} \sum_{t=L-2}^{T-1} \widetilde{V}_{t,1}^2} + o_p(1) \implies -\frac{2E\widetilde{W}_{t,1}^2}{\lambda_v^2 \int_0^1 [W(r)]^2 dr}.$$

Since $\left\{\widetilde{W}_{t,1}\right\}$ is a stationary process, $E\widetilde{W}_{t,1}^2 = EW_{t,1}^2$, implying that under the null hypothesis, the following holds:

$$\widetilde{FG}_1^L \equiv \frac{T\widehat{\lambda}_v^2(\widetilde{S}_{T,1}^{L} - 1)}{2\widetilde{v}_{y,1}^2} \implies -\frac{1}{\int_0^1 [W(r)]^2 dr},$$

where

$$\widetilde{v}_{y,1}^2 = \frac{1}{(T-L)} \sum_{t=L-1}^{T-1} \widetilde{W}_{t,1}^2.$$

It is interesting to note that the test statistic $\widetilde{FG}_1^L$ is of exactly the same form as that of FG_1^L except that the former uses MODWT while the latter uses DWT and both have the same limiting distribution under the null hypothesis.

Alternatively, noting that $\sum_{t=1}^{T-1} \widetilde{V}_{t,1}^2 + \sum_{t=1}^{T-1} \widetilde{W}_{t,1}^2 = 2\sum_{t=1}^{T} y_t^2$, we can modify $(\widetilde{S}_{T,1}^{L} - 1)$ and employ the following test statistic:

$$\widetilde{D}_{T,1}^{L} = \frac{\sum_{t=L-2}^{T-1} \widetilde{W}_{t,1}^2}{2\sum_{t=1}^{T} y_t^2}.$$

Under H_0, $T\widetilde{D}^L_{T,1} \implies -\frac{2E\widetilde{W}^2_{t,1}}{\lambda_v^2 \int_0^1 [W(r)]^2 dr}$.

Now consider the case with an *intercept*. By following the same arguments as in Section 4, we get $\sum_{t=L-2}^{T-1} \widetilde{V}^2_{t,1} = \frac{2\alpha^2 T^3}{3} + o_p(T^3)$. Decomposing $\widetilde{S}^L_{T,1}$ as below,

$$\begin{aligned}\widetilde{S}^L_{T,1} - 1 &= -\frac{\sum_{t=L-2}^{T-1}[\widetilde{W}^2_{t,1} - E\widetilde{W}^2_{t,1}]}{\sum_{t=L-2}^{T-1} \widetilde{V}^2_{t,1} + \sum_{t=L-2}^{T-1} \widetilde{W}^2_{t,1}} - \frac{TE\widetilde{W}^2_{t,1}}{\sum_{t=L-2}^{T-1} \widetilde{V}^2_{t,1} + \sum_{t=L-2}^{T-1} \widetilde{W}^2_{t,1}} \\ &= -\widetilde{F}^L_T - \widetilde{S}^L_T,\end{aligned}$$

we get the following lemma, the basis of our tests for H_0.

Lemma 5.1 *Under H_0, (i)* $T^2\left(\widetilde{S}^L_{T,1} - 1\right) = -\frac{3E\widetilde{W}^2_{t,1}}{2\alpha^2} + o_p(1)$; *(ii)* $T^{5/2}\left(\widetilde{S}^L_{T,1} - 1 + \widetilde{S}^L_T\right) = -\frac{3N(0,\widetilde{\sigma}^2_{*W})}{2\alpha^2} + o_p(1)$, *where $\widetilde{\sigma}^2_{*W}$ is the long run variance of the process $\left\{\widetilde{W}^2_{t,1}\right\}$.*

Similar to the DWT case, we can also construct tests against non-zero mean stationary processes and linear *trend stationary* processes based on MODWT of the demeaned and detrended series. Let $\left\{\widetilde{W}^M_{t,1}\right\}$ and $\left\{\widetilde{V}^M_{t,1}\right\}$ 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

$$\widetilde{D}^{LM}_{T,1} = -\frac{\sum_{t=1}^{T/2}(\widetilde{W}^M_{t,1})^2}{\sum_{t=1}^{T}(y_t - \overline{y})^2}.$$

Similarly, let $\left\{\widetilde{W}^d_{t,1}\right\}$ and $\left\{\widetilde{V}^d_{t,1}\right\}$ 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

$$\widetilde{D}^{Ld}_{T,1} = -\frac{\sum_{t=1}^{T/2}(\widetilde{W}^d_{t,1})^2}{\sum_{t=1}^{T}(\widetilde{y}_t - \overline{\widetilde{y}})^2}.$$

One can easily show that under H_0, $T\widetilde{D}^{LM}_{T,1} \implies -\frac{E(\widetilde{W}^M_{t,1})^2}{2\omega^2 \int_0^1 [W_\mu(r)]^2 dr}$ and $T\widetilde{D}^{Ld}_{T,1} \implies -\frac{E(\widetilde{W}^d_{t,1})^2}{2\omega^2 \int_0^1 [V_\mu(r)]^2 dr}$. It is interesting to note that our tests based on the Haar wavelet filter are identical to the Sargan and Bhargava (1983) and Bhargava (1986) tests. In view of the improved performance of the wavelet filter as an approximation to the ideal band-pass filter when the length of the filter increases, we expect power gains of our tests when a Daubechies filter other than the Haar is used.

6 Monte Carlo Simulations

In this section, we investigate the sampling performance of the new unit root tests and compare them against the Augmented Dickey and Fuller (1979) (ADF), Phillips and Perron (1988) (PP), Elliott *et al.* (1996) (ERS) and Ng and Perron (2001) (MPP) tests. We will study the no drift, the drift and tests against trend stationarity with detrending.

6.1 No Drift

The true process (data generating mechanism) is an AR(1) process without a drift

$$y_t = \rho y_{t-1} + u_t, \qquad u_t \sim iidN(0, \sigma^2)$$

when $\rho = 1$, the null hypothesis holds and the alternative hypothesis holds when $|\rho| < 1$. Since most unit root tests are sensitive to the starting value of the unit root process, we draw the starting values randomly from $y_0 \sim N(0, 1)$ which permits a large range of starting values. To keep a balance between the signal and noise components, we set $\sigma^2 = |y_0|$. Under the alternative, the data is generated from $y_t = \rho y_{t-1} + u_t$ where $u_t \sim iidN(0, 1)$ and $y_0 \sim N(0, 1/(1-\rho^2))$.

The size and the power of the test are calculated for 100, 500 and 1,000 observations at the 1, 5 and 10 percent levels.[14] For size calculations, no initial values are discarded from the simulations. For power calculations, five times of sample size are discarded as transients.

Since $y_0 \neq 0$, the ADF (Augmented Dickey-Fuller) test requires a constant term in the test regression. Accordingly, the asymptotic critical values of the ADF test with $y_0 \neq 0$ are -3.43, -2.86, and -2.57 at the 1, 5, and 10 percent levels, respectively. In the calculation of the ADF test, the lag length is set to one. The asymptotic critical values of our test, FG_M^L, are -29.04, -17.75 and -13.09 at the 1, 5, and 10 percent levels, respectively. These critical values are calculated from one million replications. The asymptotic and finite sample distributions of the FG_M^L test are illustrated in Figure 9 where our test statistic has desirable size in small samples.

Table 1 compares the size and power of the FG_M^L against the ADF and Phillips and Perron (1988) (PP) tests. All simulations are with discrete wavelet transformation (DWT) at scale-1 ($M = 1$) and with the Haar filter. The actual size of our test is reasonably close to the nominal size at 1%, 5% and 10% levels. The power gain of our test relative to the ADF and PP tests are 89% and 75% at the 5% level for $\rho = 0.90$ and T=100. For $T = 500$, we calculate the power for alternatives much closer to the null than traditionally reported. Namely, the power calculations are carried out for $\rho = 0.99, 0.98$ and 0.97. For $\rho = 0.99$ and $\rho = 0.97$, the powers of our test at the 5% level are 21.9% and 93.8%, respectively. These are 84% and 55% power gains relative to the ADF test, and 71% and 51% relative to the PP test. For $T = 1000$, our test has a power of 68.4% for $\rho = 0.99$ and reaches 99.7% power for $\rho = 0.98$. For $\rho = 0.99$, these are 111% and 103% power improvements over the ADF and PP tests, respectively.

Table 2 contains results with the Haar, LA(10) and LA(20)[15] filters. The LA(10) and LA(20) are better band-pass filters than the Haar filter and therefore we expect these filters to have higher power. The results indicate that both filters perform slightly better than the Haar filter for a sample size of 100 observations. $FG_J^L(J = 4)$ test does slightly better than the $FG_J^L(J = 1)$ tests

[14] The limiting distribution of $-1/\int_0^1 [W(r)]^2 dr$ is calculated from 1 million replications through the method proposed in MacKinnon (2000). The simulated data for the null distribution is generated from $y_t = y_{t-1} + u_t$, $u_t \sim i.i.d.N(0, \sigma^2)$ where $y_0 \neq 0$.

[15] Least asymmetric filters of length 10 and 20.

as reported in Table 3. Although a higher scale decomposition is preferable, there should always be a balance between the scale of decomposition and the number of observations due to the boundary coefficients at the higher scales.[16]

6.2 With Drift

The true process is an AR(1) process with a drift

$$y_t = \alpha + \rho y_{t-1} + u_t$$

where the error term follows either an AR(1) or an MA(1) process

$$u_t = \begin{cases} \rho u_{t-1} + \epsilon_t \\ \\ \epsilon_t + \theta \epsilon_{t-1} \end{cases}$$

where $\epsilon_t \sim iidN(0, \sigma^2)$. For $\rho = 1$, the null hypothesis holds and the alternative hypothesis holds when $|\rho| < 1$. The starting value of the unit root process is drawn from $y_0 \sim N(0, 1)$, $\alpha = 1$ and $\sigma^2 = 1$. Our test requires a choice for T^*. For a given T, $T_1 = T/2$, $T_1^* = T^*/2$ and the choice for T^* requires that the ratio T^*/T should go to zero asymptotically. We set $T^* = T^{0.95}$ which provides reasonable empirical size at the 1, 5, and 10 percent levels.

Since $y_0 \neq 0$, the ADF test requires a constant term and a linear trend component in the test regression. Accordingly, the asymptotic critical values of the ADF test with $y_0 \neq 0$ are -3.96, -3.41, and -3.12 at the 1, 5, and 10 percent levels, respectively. The asymptotic critical values of the Elliott *et al.* (1996) (ERS) are 1.99, 3.26 and 4.48 at the 1, 5, and 10 percent levels. The asymptotic critical values of the Ng and Perron (2001) (MPP) test are -2.58, -1.98, and -1.62 at the 1, 5, and 10 percent levels, respectively.

The asymptotic distribution of our test is $N(0, 1)$. The asymptotic and finite sample distributions of the $FG_J^L(drift)$ test are illustrated in Figure 10 where our test statistic has desirable size in small samples. The size and the power of the tests are calculated for 100 observations at the 1, 5 and 10 percent levels. For size calculations no initial values of y_0 are discarded from the simulations. For power calculations, five times of sample size are discarded as transients. The lag length of the ADF, ERS and MPP tests are determined by choosing the modified Akaike Information Criteria (AIC) with a maximum length of 12.[17]

All simulations for the $FG_J^L(drift)$ test are with discrete wavelet transformation (DWT) at scale-1 ($J = 1$) and with Haar filter. Table 4 studies the size and power with AR(1) errors for $T = 100$. The power of our test is largest when the parameter of the AR(1) error term is negative. This is because of the fact that there is minimal energy in the smooth component of the DWT

[16] We also studied $D_{T,1}^L$ and $D_{T,4}^L$ tests and their performances are similar to that of FG tests. Tables for these tests are not reported for brevity.

[17] All calculations are carried out in SPlus using the FinMetrics and Wavelet modules.

transformation when γ is negative. For $\gamma = -0.50$ and $\rho = 0.95$, the power of our test is 67.3% whereas ADF, ERS, and MPP have 0.4%, 2.4% and 2% power at the 1% level. At the 5% level, our test has 70.8% power whereas ADF, ERS, and MPP have 3.3%, 13.4% and 13.6% power.

For $\gamma = 0$ and $\rho = 0.95$, the power of our test is 32% (ADF, ERS and MPP are 0.4%, 3.4%, 3.3%) at the 1% level. At the 5% level, ours is 37.1% and ADF, ERS and MPP are 4.7%, 16%, 17.1% respectively. For positive values of γ, the power performance of our test is not as striking. This is due to the fact that positive γ adds additional energy to the smooth component of the DWT transformation which in turn leads to a reduction in power. For $\gamma = 0.50$ and $\rho = 0.95$, the power of our test is 10.7% and ADF, ERS and MPP are 4.5%, 18.4%, 19.8% at the 5% level.

The reported power performance of our tests are intuitive and the following pattern emerges. The power of our tests is largest for AR(1) errors with negative coefficients, followed by i.i.d. errors, then AR(1) errors with positive coefficients, because in terms of energy, AR(1) with negative coefficients has the least energy relative to that of a random walk, while AR(1) with positive coefficients has the largest energy relative to that of a random walk. A further advantage of our test is that it does not involve lag selection which is an issue for our comparison tests. The size of our test remains stable across the range of AR(1) error parameter whereas the comparison tests suffer size problems.

Table 5 studies the size and power with MA(1) errors for $T = 100$. The empirical size of our test is reasonable across different values of the MA(1) parameter. When $\theta = 0.50$, the size of our test is 1.3%, 4.2% and 7.4% at the 1, 5, and 10 percent levels, respectively. The ERS and MPP tests severely underreject for $\theta = 0.50$. For $\theta = 0$, our test has good empirical size with 1.2%, 4.4% and 8.2% at the 1, 5, and 10 percent levels. ERS and MPP underreject at the 10 percent level while overrejecting at the 1 percent level. For $\theta = -0.50$, our test underrejects with empirical size of 0.4%, 2.4% and 5.8% at the 1, 5, and 10 percent levels. ERS and MPP tests grossly overreject for $\theta = 0.5$. For instance, the size of ERS are 13.6%, 17.4%, and 0.19.6% at the 1, 5, and 10 percent levels.

When $\theta = 0.50$ and $\rho = 0.95$, the power of our test is 20% whereas the power of ADF, ERS and MPP are 3%, 10.6%, 10.8% respectively. This is 85% power improvement over the MPP test. For $\theta = 0$ and $\rho = 0.95$, the power of our test is 36.9% whereas the power of ADF, ERS and MPP are 2.9%, 13%, 15.4% respectively. This is 140% power improvement over the MPP test. For $\theta = -0.50$ and $\rho = 0.95$, the power of our test is 81.4% whereas the power of ADF, ERS and MPP remain at 5.3%, 14.5% and 15.2%.

Similar to that of the AR(1) process, the following pattern emerges for MA(1) errors. The power of our tests is largest for MA(1) errors with negative coefficients, followed by i.i.d. errors, then MA(1) errors with positive coefficients, because in terms of energy, MA(1) with negative coefficients has the least energy relative to that of a random walk, while MA(1) with positive coefficients has the largest energy relative to that of a random walk. For most ranges of θ, our test provides reasonable size and substantially higher power relative to the ERS and MPP tests.

6.3 Tests against trend stationarity

To deal with trend stationary alternatives, components representation of a time series is often used and detrending performed, see Schmidt and Phillips (1992), Phillips and Xiao (1998), and Stock (1999). Phillips and Xiao (1998) also have a detailed discussion on efficient detrending for general trends. For ease of exposition, we restrict ourselves to non-zero mean and linear trend cases only. The process $\{y_t\}$ is of the form:

$$y_t = \mu + \alpha t + y_t^s, \tag{39}$$

where $\{y_t^s\}$ is generated by model (19). 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 $\{y_t\}$.

The asymptotic critical values of our test, $\widehat{D}_{T,1}^{LM}$, are -40.38, -27.38 and -21.75 at the 1, 5, and 10 percent levels, respectively. These critical values are calculated from one million replications. The asymptotic and finite sample distributions of the $\widehat{D}_{T,1}^{LM}$ test are illustrated in Figure 11 where our test statistic has desirable size in small samples. Table 6 studies the size and power of the demeaned series across sample sizes of 100, 200 and 1,000 observations with i.i.d. errors. All four tests have empirical sizes close to their nominal counterparts. Ours, $\widehat{D}_{T,1}^{LM}$, has comparable power for $T = 100$ and provides consistently more power for $T = 200$ and $T = 1000$.

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 $\{\Delta y_t\}$. The asymptotic critical values of our test, $\widehat{D}_{T,1}^{Lt}$, are -50.77, -36.54 and -30.23 at the 1, 5, and 10 percent levels, respectively. These critical values are calculated from one million replications. The asymptotic and finite sample distributions of the $\widehat{D}_{T,1}^{Lt}$ test are illustrated in Figure 12 where our test statistic has desirable size in small samples. Table 7 studies the size and power of the detrended series across sample sizes of 100, 200 and 1,000 observations with i.i.d. errors. All four tests have empirical sizes close to their nominal counterparts. ERS and MPP does slightly better than ours for $T = 100$ and $T = 200$. Ours has slightly more power for $T = 1000$. The advantage of ours is that it is free of lag selection issues which other tests require.

An extension to GARCH(1,1) errors is carried out in Table 9. Ours tests provide higher power against ERS and MPP tests and have reasonable empirical size. For a sample size of $T = 1000$ and $\rho = 0.98$, $\widehat{D}_{T,1}^{LM}$ has a power of 93.4 percent whereas ERS and MPP tests remain at 77.4 and 77.9 percent levels. For $\widehat{D}_{T,1}^{Lt}$ and $\rho = 0.98$, the power is 72.6 and the powers of ERS and MPP tests are 70.7 and 68.5 percent. In addition to these power improvements, we wish to reiterate the fact that our tests are free of lag selection issues which other tests require.

7 Applications

We apply and compare our tests with the Elliott *et al.* (1996) (ERS) and Ng and Perron (2001) (MPP) tests. We use data at 30-minute, daily and monthly frequency from FX, interest rate, equity markets and inflation rate. For all series, we use the natural logarithm of the levels for all tests. The test regression contains a drift and trend component for index and individual equity series and only a drift term for the FX and interest rate series. The lag lengths of the ERS and MPP test regressions are determined by minimizing the modified Akaike Information Criteria with a maximum length of 12.

In the 80 years of daily DJIA series, spanning from 1928 to 2007, all three tests fail to reject the null of unit root with $\widehat{D}_{T,1}^{Lt}$ providing the smallest p-value of 55%. ERS and MPP p-values are 89% and 90%. As a major individual stock, we studied daily Microsoft shares adjusted for dividends and splits from 1986 until 2007, a total of 5,305 observations. The p-values are 90, 98 and 99 percent for the $\widehat{D}_{T,1}^{Lt}$, ERS and MPP tests. The Canadian inflation, from 1995 to 2007, a total of 146 observations, lead to p-values of 55, 62 and 57 percent for the $\widehat{D}_{T,1}^{Lt}$, ERS and MPP tests.

We also study high-frequency FX series. The data is one year of 30-min series for 1996, a total of 17,568 observations for the USD-JPY rate.[18] With high-frequency series, the p-values are 43%, 80% and 90%, for $\widehat{D}_{T,1}^{LM}$, ERS and MPP, respectively. The short-term U.S. Treasury Bills, from 1897 to 1996, a total of 27,567 observations lead to p-values of 1e-008, 0.003 and 0.003 for the $\widehat{D}_{T,1}^{LM}$, ERS and MPP tests, respectively. The null hypothesis of unit root is rejected at the 1 percent level for all three tests. The strongest rejection lies with the $\widehat{D}_{T,1}^{LM}$ test.

8 Testing for Cointegration

The unit root tests developed in the previous sections 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). In this section, we provide such an extension for the no-drift case using unit scale DWT. Extensions for other cases are straightforward.

Our notation and formulation here are similar to those in Phillips and Ouliaris (1990). Let $\{z_t\}$ be an $(m+1)$-dimensional multivariate time series generated by an integrated process of the form:

$$z_t = z_{t-1} + \xi_t,$$

where $\{\xi_t\}$ is a linear process satisfying:

[18] This is the HFDF-96 data provided by the Olsen Group (www.olsen.ch). More details regarding the properties of high-frequency FX series can be found in Dacorogna *et al.* (2001).

Assumption 2:

(a) $\{\xi_t\}$ is a linear process defined as $\xi_t = \Psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}$, $\Psi(1) \neq 0$, and $\Psi(L)$ is one-summable;

(b) $\{\varepsilon_t\}$ is i.i.d. with $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \Sigma > 0$, and finite fourth cumulants, and $\varepsilon_s = 0$ for $s \leq 0$.

Let Ω denote the long run variance-covariance matrix of $\{\xi_t\}$. Under Assumption 2, it is known that $T^{-1/2} \sum_{t=1}^{[Tr]} \xi_t \implies B(r)$, where $B(r)$ is $(m+1)$-vector Brownian motion with covariance matrix Ω. We now partition $z_t = (y_{1t}, y_{2t}')'$ into the scalar variable y_{1t} and the m-dimensional vector y_{2t}. Consider the linear cointegrating regressions:

$$y_{1t} = \widehat{\beta}' y_{2t} + \widehat{u}_t,$$

where $\widehat{\beta}$ is the OLS estimator of β in the regression of y_{1t} on y_{2t}. We now extend our tests for unit root based on unit scale DWT developed in Subsection 3.2 to the corresponding tests for no-cointegration. In particular, we use:

$$\widehat{CD}_{T,1}^{L} = -\frac{\sum_{t=L_1}^{T/2-1} \widehat{W}_{t,1}^2}{\sum_{t=1}^{T} \widehat{u}_t^2},$$

where $\left\{\widehat{W}_{t,1}\right\}$ is the unit scale wavelet coefficients of $\{\widehat{u}_t\}$. To state the asymptotic distribution of $\widehat{CD}_{T,1}^{L}$ under the null hypothesis of no-cointegration, we need to introduce some notation. Partition Ω conformably with that of z_t so that

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{bmatrix}.$$

Let $\omega_{11.2} = \omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21}$. Further let $W_{m+1}(r) = (W_1(r), W_2(r)')'$ be an $(M+1)$-dimensional standard Brownian motion. Phillips and Ouliaris (1990) showed that under the null of no-cointegration, $T^{-2} \sum_{t=1}^{T} \widehat{u}_t^2 \implies \omega_{11.2} \int_0^1 Q^2(r) dr$, where

$$Q(r) = W_1(r) - \left(\int_0^1 W_1(r) W_2'(r) dr \right) \left(\int_0^1 W_2(r) W_2'(r) dr \right)^{-1} W_2(r).$$

Let

$$A = \int_0^1 B(r) B(r)' dr = \begin{bmatrix} a_{11} & a_{21}' \\ a_{21} & A_{22} \end{bmatrix}.$$

Theorem 8.1 *Under the null hypothesis of no-cointegration,*

$$T\left(\widehat{CD}_{T,1}^{L}\right) \implies -\frac{\eta' Var(W_{t,1}^{z})\eta}{\omega_{11.2}\int_0^1 Q^2(r)dr},$$

where $\eta' = (1, -a_{21}' A_{22}^{-1})$ *and* $\left\{W_{t,1}^{z}\right\}$ *is the unit scale wavelet coefficient of* $\{z_t\}$.

Both η and $\omega_{11.2}$ depend on the long run covariance matrix Ω. We now discuss its estimation. Let $\widehat{\xi}_t$ denote the OLS residual in the regression: $z_t = \widehat{\Pi} z_{t-1} + \widehat{\xi}_t$. Then similar to the estimation of the long run variance ω^2, we can use a nonparametric kernel estimator with the Bartlett kernel to estimate Ω.

9 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 spectral approach to unit root testing; second, we provide a 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 our tests, the intuitive construction and simplicity are worth emphasizing. The simulation studies demonstrate the superior power of our tests with reasonable empirical sizes. A generalization of our unit root tests to the maximum overlap DWT (MODWT) and to residual based tests for cointegration are also provided.

Appendix: Technical Proofs

Proof of (26): Let $T_1 = \frac{T}{2}$. By Proposition 17.2 in Hamilton (1994), we have

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

Define the partial sum process associated with $\{v_t\}$ as

$$X_{T_1}(r) = \frac{1}{T_1} \sum_{t=1}^{[T_1 r]} v_t, \ 0 \leq r \leq 1.$$

Then it follows that

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

By the functional CLT, we obtain

$$\sqrt{T} X_{T_1}(\cdot) \stackrel{L}{\to} 2\psi(1)\sigma W(\cdot),$$

where $W(\cdot)$ is the standard Brownian motion. 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 CMT,

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

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

$$\left(\frac{T}{2} \right)^{-2} A_T \stackrel{L}{\to} \frac{1}{2} \lambda_v^2 \int_0^1 [W(r)]^2 dr.$$

We now look at B_T. Recall that $B_T = \sum_{t=1}^{T/2} u_{2t} y_{2t-1}$. Simple algebra shows that $B_T = \frac{1}{2} \sum_{s=1}^{T-1} (T-s-1)\gamma_s = O(T)$ 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 LLN.

Proof of Theorem 3.3(i): Since $\{W_{t,1}\}$ is stationary, $\sum_{t=L_1}^{T/2-1} W_{t,1}^2 = O_p(T)$. Now consider the order of $\sum_{t=L_1}^{T/2-1} V_{t,1}^2$. Suppose $\rho = 1$. Noting that

$$\begin{aligned} V_{t,1} &= Y_{2t+2-L} \sum_{l=0}^{L-1} g_l + \sum_{l=0}^{L-2} g_l \left\{ \sum_{j=0}^{L-2-l} u_{2t+1-j-l} \right\} \\ &= \sqrt{2} Y_{2t+2-L} + \sum_{l=0}^{L-2} g_l \left\{ \sum_{j=0}^{L-2-l} u_{2t+1-j-l} \right\}, \end{aligned}$$

we obtain

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

Hence when $\rho = 1$, we obtain

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

If $|\rho| < 1$, then both $\{W_{t,1}\}$ and $\{V_{t,1}\}$ are stationary so that

$$\hat{S}_{T,1}^L = \frac{EV_{t,1}^2}{EV_{t,1}^2 + EW_{t,1}^2} + o_p(1).$$

				T=100					
ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%
	FG_J^L (Haar)			ADF			PP		
1.00	0.007	0.040	0.088	0.010	0.048	0.098	0.011	0.057	0.106
0.95	0.036	0.212	0.389	0.026	0.124	0.233	0.033	0.143	0.253
0.90	0.186	0.625	0.834	0.094	0.331	0.531	0.105	0.358	0.559
0.85	0.484	0.904	0.979	0.256	0.644	0.824	0.280	0.675	0.841
0.80	0.776	0.986	0.998	0.518	0.873	0.965	0.558	0.888	0.966
				T=500					
ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%
	FG_J^L (Haar)			ADF			PP		
1.00	0.010	0.047	0.096	0.011	0.048	0.097	0.012	0.051	0.100
0.99	0.048	0.219	0.405	0.031	0.119	0.228	0.035	0.128	0.242
0.98	0.231	0.657	0.867	0.090	0.324	0.511	0.100	0.339	0.525
0.97	0.569	0.938	0.994	0.236	0.607	0.805	0.256	0.621	0.809
				T=1,000					
ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%
	FG_J^L (Haar)			ADF			PP		
1.00	0.006	0.048	0.086	0.005	0.044	0.070	0.007	0.043	0.086
0.99	0.253	0.684	0.882	0.094	0.324	0.509	0.101	0.337	0.521
0.98	0.879	0.997	1	0.488	0.859	0.958	0.507	0.858	0.956

Table 1: Size and Power of the FG_J^L (J=1) - No Drift

The test statistic and its variance are calculated with discrete wavelet transformation. The data generating process is $y_t = y_{t-1} + u_t$, $u_t \sim iidN(0, \sigma^2)$ where $y_0 \sim N(0, 1)$ and $\sigma^2 = |y_0|$. Under the alternative, $y_t = \rho y_{t-1} + u_t$, $u_t \sim iidN(0, 1)$ where $y_0 \sim N(0, 1/(1-\rho^2))$. The asymptotic critical values are -29.04, -17.75 and -13.09 at the 1, 5, and 10 percent levels, respectively. These critical values are calculated from one million replications. The asymptotic critical values of the ADF/PP test with $y_0 \neq 0$ are -3.43, -2.86, and -2.57 at the 1, 5, and 10 percent levels, respectively. In the calculation of the ADF/PP tests, the lag length is set to one. All simulations are with 5,000 replications. Under the alternative, the five times of the sample size are discarded as transients.

				T=100					
ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%
	FG_J^L (Haar)			FG_J^L (LA(10))			FG_J^L (LA(20))		
1.00	0.007	0.040	0.088	0.007	0.044	0.093	0.012	0.054	0.109
0.95	0.036	0.212	0.389	0.045	0.221	0.400	0.055	0.244	0.415
0.90	0.186	0.625	0.834	0.198	0.626	0.829	0.220	0.627	0.860
0.85	0.484	0.904	0.979	0.486	0.904	0.979	0.492	0.905	0.980
0.80	0.776	0.986	0.998	0.774	0.985	0.998	0.776	0.985	0.998
				T=500					
ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%
	FG_J^L (Haar)			FG_J^L (LA(10))			FG_J^L (LA(20))		
1.00	0.010	0.047	0.096	0.009	0.047	0.095	0.009	0.047	0.095
0.99	0.048	0.219	0.405	0.049	0.221	0.406	0.052	0.224	0.407
0.98	0.231	0.657	0.867	0.234	0.655	0.866	0.236	0.657	0.866
0.97	0.569	0.938	0.994	0.570	0.934	0.994	0.571	0.936	0.994
				T=1,000					
ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%
	FG_J^L (Haar)			FG_J^L (LA(10))			FG_J^L (LA(20))		
1.00	0.006	0.048	0.086	0.007	0.047	0.093	0.007	0.047	0.093
0.99	0.253	0.684	0.882	0.256	0.683	0.882	0.259	0.683	0.884
0.98	0.879	0.997	1	0.879	0.997	1	0.879	0.997	1

Table 2: Size and Power of the FG_J^L (J=1) - No Drift

The test statistic and its variance are calculated with discrete wavelet transformation. The data generating process is $y_t = y_{t-1} + u_t$, $u_t \sim iidN(0, \sigma^2)$ where $y_0 \sim N(0, 1)$ and $\sigma^2 = |y_0|$. Under the alternative, $y_t = \rho y_{t-1} + u_t$, $u_t \sim iidN(0, 1)$ where $y_0 \sim N(0, 1/(1-\rho^2))$. The asymptotic critical values are -29.04, -17.75 and -13.09 at the 1, 5, and 10 percent levels, respectively. These critical values are calculated from one million replications. The asymptotic critical values of the ADF/PP test with $y_0 \neq 0$ are -3.43, -2.86, and -2.57 at the 1, 5, and 10 percent levels, respectively. In the calculation of the ADF/PP tests, the lag length is set to one. All simulations are with 5,000 replications. Under the alternative, the five times of the sample size are discarded as transients.

				T=100					
ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%
	FG_J^L (Haar)			ADF			PP		
1.00	0.010	0.047	0.098	0.010	0.048	0.098	0.011	0.057	0.106
0.95	0.036	0.217	0.391	0.026	0.124	0.233	0.033	0.143	0.253
0.90	0.187	0.629	0.834	0.094	0.331	0.531	0.105	0.358	0.559
0.85	0.489	0.906	0.979	0.256	0.644	0.824	0.280	0.675	0.841
0.80	0.776	0.988	0.999	0.518	0.873	0.965	0.558	0.888	0.966
				T=500					
ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%
	FG_J^L (Haar)			ADF			PP		
1.00	0.010	0.053	0.106	0.011	0.048	0.097	0.012	0.051	0.100
0.99	0.048	0.221	0.406	0.031	0.119	0.228	0.035	0.128	0.242
0.98	0.234	0.657	0.868	0.090	0.324	0.511	0.100	0.339	0.525
0.97	0.571	0.939	0.994	0.236	0.607	0.805	0.256	0.621	0.809
				T=1,000					
ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%
	FG_J^L (Haar)			ADF			PP		
1.00	0.009	0.053	0.107	0.005	0.044	0.070	0.007	0.043	0.086
0.99	0.258	0.685	0.884	0.094	0.324	0.509	0.101	0.337	0.521
0.98	0.880	0.997	1	0.488	0.859	0.958	0.507	0.858	0.956

Table 3: Size and Power of the FG_J^L - (J=4) No Drift

The test statistic and its variance are calculated with discrete wavelet transformation. The data generating process is $y_t = y_{t-1} + u_t$, $u_t \sim iidN(0, \sigma^2)$ where $y_0 \sim N(0, 1)$ and $\sigma^2 = |y_0|$. Under the alternative, $y_t = \rho y_{t-1} + u_t$, $u_t \sim iidN(0, 1)$ where $y_0 \sim N(0, 1/(1-\rho^2))$. The asymptotic critical values are -29.04, -17.75 and -13.09 at the 1, 5, and 10 percent levels, respectively. These critical values are calculated from one million replications. The asymptotic critical values of the ADF/PP test with $y_0 \neq 0$ are -3.43, -2.86, and -2.57 at the 1, 5, and 10 percent levels, respectively. In the calculation of the ADF/PP tests, the lag length is set to one. All simulations are with 5,000 replications. Under the alternative, the five times of the sample size are discarded as transients.

						T=100						
ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
		$FG_J^L(drift)$			ADF			ERS			MPP	
						$\gamma = -0.50$						
1.00	0.011	0.046	0.093	0.012	0.044	0.093	0.075	0.095	0.112	0.089	0.109	0.130
0.95	0.673	0.708	0.731	0.004	0.033	0.100	0.024	0.134	0.278	0.020	0.136	0.328
0.90	0.823	0.848	0.862	0.013	0.078	0.168	0.076	0.297	0.480	0.059	0.313	0.546
						$\gamma = 0$						
1.00	0.011	0.044	0.087	0.005	0.033	0.068	0.049	0.071	0.081	0.067	0.079	0.096
0.95	0.320	0.371	0.403	0.004	0.047	0.110	0.034	0.160	0.312	0.033	0.171	0.359
0.90	0.437	0.488	0.516	0.017	0.111	0.210	0.105	0.361	0.555	0.097	0.350	0.631
						$\gamma = 0.50$						
1.00	0.034	0.071	0.117	0.008	0.041	0.083	0.018	0.027	0.032	0.022	0.032	0.037
0.95	0.079	0.107	0.131	0.004	0.045	0.099	0.044	0.184	0.320	0.045	0.198	0.373
0.90	0.091	0.125	0.152	0.015	0.078	0.152	0.102	0.360	0.529	0.103	0.390	0.591

Table 4: Size and Power of the FG_J^L - With Drift - AR(1) Errors

The test statistic and its variance are calculated with discrete wavelet transformation where J=1 and with Haar filter. The data generating process is $y_t = \alpha + y_{t-1} + u_t$, $u_t = \gamma u_{t-1} + \epsilon_t$, $\epsilon_t \sim iidN(0, \sigma^2)$ where $y_0 \sim N(0, 1)$. Under the alternative, $y_t = \alpha + \rho y_{t-1} + u_t$, $u_t = \gamma u_{t-1} + \epsilon_t$, $\epsilon_t \sim iidN(0, \sigma^2)$. For both models, $\sigma^2 = 1$ and $\alpha = 1$. The asymptotic null distribution of $FG_J^L(drift)$ is standard normal. The asymptotic critical values of the ADF test with $y_0 \neq 0$ are -3.96, -3.41, and -3.12 at the 1, 5, and 10 percent levels, respectively. The asymptotic critical values of the ERS test are 1.99, 3.26, and 4.48 at the 1, 5, and 10 percent levels, respectively. The asymptotic critical values of the MPP test are -2.58, -1.98, and -1.62 at the 1, 5, and 10 percent levels, respectively. The lag length of the test regressions is determined by minimizing the modified AIC with the maximum lag length of 12. All simulations are with 5,000 replications. Under the alternative, the five times of the sample size are discarded as transients. All calculations are with $T^* = T^{0.95}$.

						T=100						
ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
		$FG_J^L(drift)$			ADF			ERS			MPP	
						$\theta = 0.50$						
1.00	0.013	0.042	0.074	0.004	0.030	0.064	0.013	0.015	0.017	0.015	0.016	0.018
0.95	0.154	0.200	0.230	0.003	0.030	0.076	0.024	0.106	0.244	0.016	0.108	0.302
0.90	0.202	0.245	0.279	0.012	0.049	0.104	0.090	0.333	0.548	0.094	0.333	0.548
						$\theta = 0$						
1.00	0.012	0.044	0.082	0.006	0.033	0.067	0.026	0.035	0.045	0.030	0.045	0.057
0.95	0.319	0.369	0.401	0.003	0.029	0.073	0.017	0.130	0.291	0.02	0.154	0.355
0.90	0.438	0.489	0.521	0.016	0.100	0.209	0.095	0.317	0.555	0.080	0.354	0.584
						$\theta = -0.50$						
1.00	0.004	0.024	0.058	0.012	0.056	0.094	0.098	0.144	0.176	0.136	0.174	0.196
0.95	0.793	0.814	0.831	0.012	0.053	0.111	0.034	0.145	0.284	0.033	0.152	0.330
0.90	0.927	0.937	0.944	0.028	0.113	0.208	0.088	0.265	0.429	0.085	0.278	0.468

Table 5: Size and Power of the FG_J^L - With Drift - MA(1) Errors

The test statistic and its variance are calculated with discrete wavelet transformation where J=1 and with Haar filter. The data generating process is $y_t = \alpha + y_{t-1} + u_t$, $u_t = \epsilon_t + \theta\epsilon_{t-1}$, $\epsilon_t \sim iidN(0, \sigma^2)$ where $y_0 \sim N(0, 1)$. Under the alternative, $y_t = \alpha + \rho y_{t-1} + u_t$, $u_t = \epsilon_t + \theta\epsilon_{t-1}$, $\epsilon_t \sim iidN(0, \sigma^2)$. For both models, $\sigma^2 = 1$ and $\alpha = 1$. The asymptotic null distribution of $FG_J^L(drift)$ is standard normal. The asymptotic critical values of the ADF test with $y_0 \neq 0$ are -3.96, -3.41, and -3.12 at the 1, 5, and 10 percent levels, respectively. The asymptotic critical values of the ERS test are 1.99, 3.26, and 4.48 at the 1, 5, and 10 percent levels, respectively. The asymptotic critical values of the MPP test are -2.58, -1.98, and -1.62 at the 1, 5, and 10 percent levels, respectively. The lag length of the test regressions is determined by minimizing the modified AIC with the maximum lag length of 12. All simulations are with 5,000 replications. Under the alternative, the five times of the sample size are discarded as transients. All calculations are with $T^* = T^{0.95}$.

ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
		$\widehat{D}^{Ld}_{T,1}$			ADF			ERS			MPP	
						$T = 100$						
1.00	0.013	0.056	0.104	0.008	0.042	0.094	0.011	0.048	0.104	0.010	0.054	0.126
0.95	0.047	0.199	0.357	0.019	0.077	0.145	0.035	0.199	0.347	0.031	0.212	0.413
0.90	0.190	0.527	0.707	0.039	0.199	0.336	0.155	0.464	0.663	0.164	0.508	0.745
						$T = 200$						
1.00	0.012	0.053	0.102	0.011	0.054	0.104	0.016	0.059	0.103	0.015	0.057	0.111
0.95	0.232	0.554	0.729	0.051	0.203	0.351	0.155	0.441	0.655	0.152	0.473	0.713
0.90	0.799	0.974	0.995	0.286	0.657	0.828	0.550	0.806	0.912	0.577	0.846	0.938
						$T = 1000$						
1.00	0.010	0.052	0.101	0.010	0.052	0.103	0.014	0.057	0.102	0.014	0.055	0.110
0.99	0.277	0.615	0.773	0.060	0.216	0.356	0.156	0.438	0.623	0.153	0.440	0.641
0.98	0.805	0.972	0.993	0.253	0.592	0.796	0.497	0.774	0.862	0.495	0.776	0.871

Table 6: Size and Power of the $\widehat{D}^{LM}_{T,1}$ - Demeaned Series

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 $y_t = \mu + y^s_t$, where $y^s_t = \rho y^s_{t-1} + u_t$, $u_t \sim iidN(0, \sigma^2)$ and $y_0 \sim N(0, \sigma^2)$. Under the null $\rho = 1$ and under the alternative $\rho < 1$. The asymptotic critical values of the $\widehat{D}^{LM}_{T,1}$ test are -40.38, -27.38, and -21.75 at the 1, 5, and 10 percent levels, respectively. The asymptotic critical values of the ADF test with $y_0 \neq 0$ are -3.96, -3.41, and -3.12 at the 1, 5, and 10 percent levels, respectively. The asymptotic critical values of the ERS test are 1.99, 3.26, and 4.48 at the 1, 5, and 10 percent levels, respectively. The asymptotic critical values of the MPP test are -2.58, -1.98, and -1.62 at the 1, 5, and 10 percent levels, respectively. The lag length of the test regressions is determined by minimizing the modified AIC with the maximum lag length of 12. All simulations are with 1,000 replications.

ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
		$\widehat{D}^{Ld}_{T,1}$			ADF			ERS			MPP	
						$T = 100$						
1.00	0.004	0.035	0.079	0.010	0.048	0.096	0.003	0.032	0.085	0.003	0.032	0.082
0.95	0.016	0.059	0.148	0.014	0.080	0.154	0.060	0.066	0.154	0.016	0.058	0.151
0.90	0.023	0.171	0.346	0.055	0.190	0.322	0.029	0.192	0.376	0.031	0.184	0.372
						$T = 200$						
1.00	0.011	0.046	0.089	0.013	0.058	0.104	0.008	0.041	0.083	0.008	0.042	0.082
0.95	0.044	0.205	0.366	0.044	0.184	0.315	0.047	0.206	0.374	0.052	0.196	0.363
0.90	0.258	0.667	0.832	0.257	0.635	0.811	0.267	0.663	0.830	0.283	0.663	0.833
						$T = 1000$						
1.00	0.013	0.054	0.103	0.008	0.052	0.099	0.010	0.056	0.099	0.011	0.054	0.094
0.99	0.059	0.221	0.363	0.048	0.213	0.328	0.054	0.207	0.369	0.058	0.196	0.339
0.98	0.327	0.690	0.834	0.266	0.637	0.795	0.297	0.663	0.830	0.304	0.650	0.821

Table 7: Size and Power of the $\widehat{D}^{Ld}_{T,1}$ - GLS Detrended Series

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 $y_t = \mu + \alpha t + y^s_t$, where $y^s_t = \rho y^s_{t-1} + u_t$, $u_t \sim iidN(0, \sigma^2)$ and $y_0 \sim N(0, \sigma^2)$. Under the null $\rho = 1$ and under the alternative $\rho < 1$. The asymptotic critical values of the $\widehat{D}^{Lt}_{T,1}$ test are -50.77, -36.54, and -30.23 at the 1, 5, and 10 percent levels, respectively. The asymptotic critical values of the ADF test with $y_0 \neq 0$ are -3.96, -3.41, and -3.12 at the 1, 5, and 10 percent levels, respectively. The asymptotic critical values of the ERS test are 3.96, 5.62, and 6.89 at the 1, 5, and 10 percent levels, respectively. The asymptotic critical values of the MPP test are -3.42, -2.91, and -2.62 at the 1, 5, and 10 percent levels, respectively. The lag length of the test regressions is determined by minimizing the modified AIC with the maximum lag length of 12. All simulations are with 1,000 replications.

ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
		$\widehat{D}^{Ld}_{T,1}$			ADF			ERS			MPP	
					$\widehat{D}^{LM}_{T,1}$ - Demeaned Series							
1.00	0.012	0.054	0.112	0.025	0.076	0.121	0.014	0.062	0.115	0.012	0.058	0.124
0.99	0.183	0.501	0.692	0.052	0.184	0.325	0.176	0.448	0.605	0.170	0.447	0.634
0.98	0.664	0.934	0.986	0.257	0.591	0.775	0.540	0.774	0.856	0.547	0.779	0.866
					$\widehat{D}^{Ld}_{T,1}$ - GLS Detrended Series							
1.00	0.016	0.058	0.105	0.019	0.061	0.114	0.016	0.061	0.111	0.015	0.059	0.103
0.99	0.095	0.294	0.427	0.065	0.208	0.356	0.087	0.250	0.393	0.090	0.241	0.377
0.98	0.372	0.726	0.855	0.235	0.597	0.774	0.353	0.707	0.840	0.361	0.685	0.823

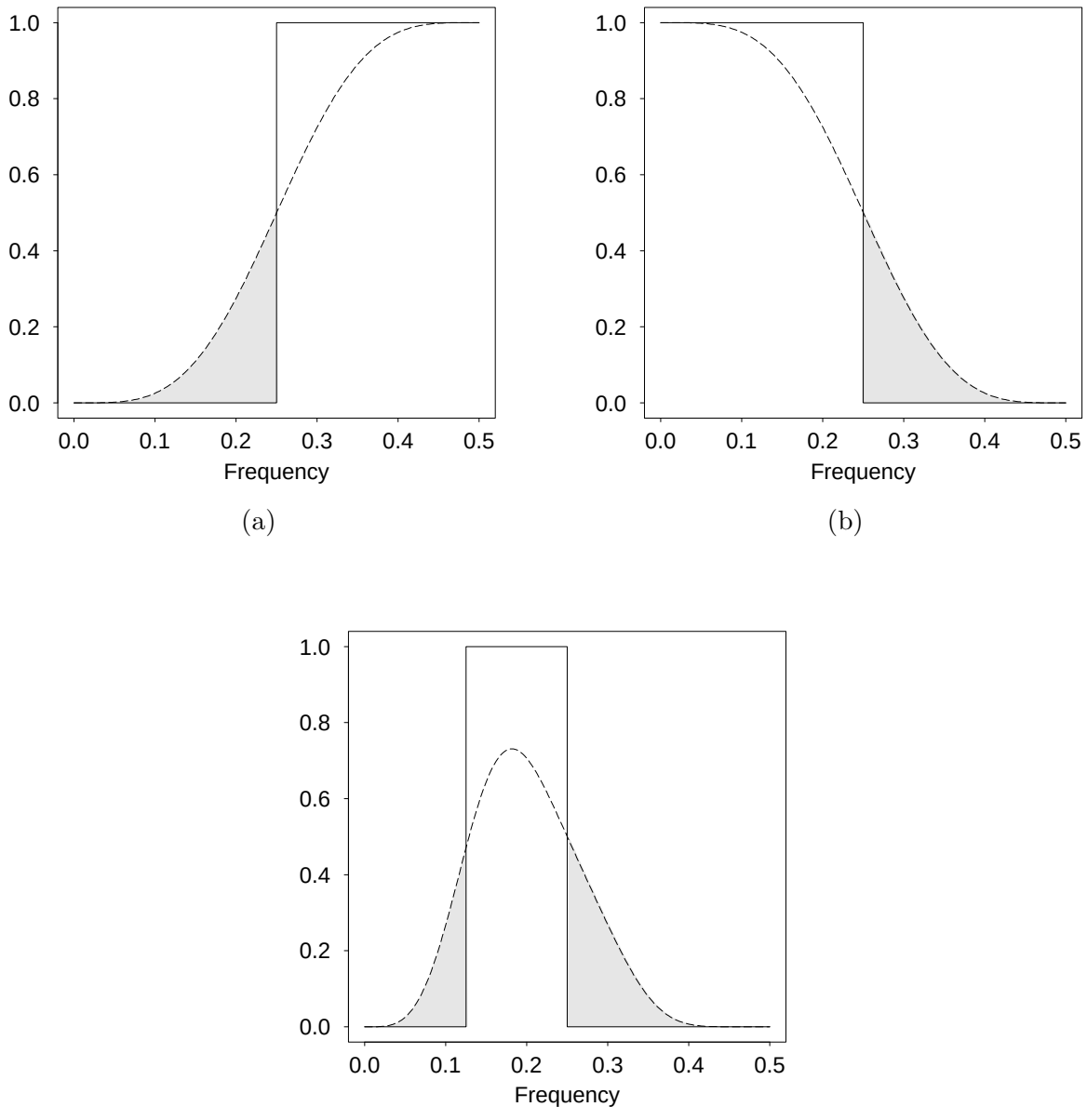
Table 8: Size and Power of the $\widehat{D}^{LM}_{T,1}$ and $\widehat{D}^{Ld}_{T,1}$ - GARCH(1,1) Errors

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 $y_t = \mu + \alpha t + y^s_t$, where $y^s_t = \rho y^s_{t-1} + \epsilon_t$, $\epsilon_t = u_t\sigma_t$ where $u_t \sim N(0,1)$, $\sigma^2_t = 1 + 0.25\epsilon^2_{t-1} + 0.65\sigma^2_{t-1}$ and $y_0 \sim N(0, \sigma^2)$. The lag length of the test regressions is determined by minimizing the modified AIC with the maximum lag length of 12. All simulations are with 1,000 replications.

	Wavelet	ERS	MPP	Observations	Dates	Frequency
DJIA ($\widehat{D}^{Ld}_{T,1}$)	0.55	0.89	0.90	19,702	1928/10/01 - 2007/3/21	Daily
Microsoft ($\widehat{D}^{Ld}_{T,1}$)	0.90	0.98	0.99	5,305	1986/03/13 - 2007/3/21	Daily
Inflation (Canada) ($\widehat{D}^{Ld}_{T,1}$)	0.55	0.62	0.57	146	1995/01 - 2007/2	Monthly
USD-JPY ($\widehat{D}^{LM}_{T,1}$)	0.43	0.80	0.90	17,568	1996/01/01 - 1996/31/12	30-Minutes
T-BILL ($\widehat{D}^{LM}_{T,1}$)	1e-008	0.003	0.003	27,567	1897/01/02 - 1996/12/31	Daily

Table 9: p-values of Unit Root Tests with Equity, Inflation, FX and Interest Rates
The lag length of the ERS and MPP test regressions is determined by minimizing the modified AIC with the maximum lag length of 12.

Figure 1: Squared Gain Functions for Ideal Filters and Their Wavelet Approximations



Squared gain functions for ideal filters (solid line) and their wavelet approximations (dotted line). The shaded regions represent *leakage*, meaning frequencies outside the nominal pass-band persist in the filtered output. (a) An ideal high-pass filter (solid line) over the frequency interval $f \in [1/4, 1/2]$ and its approximation via the D(4) wavelet filter (dotted line). (b) An ideal low-pass filter over $f \in [0, 1/4]$ and its approximation via the D(4) scaling filter. (c) An ideal band-pass filter over $f \in [1/8, 1/4]$ and its approximation via the second scale D(4) wavelet filter.

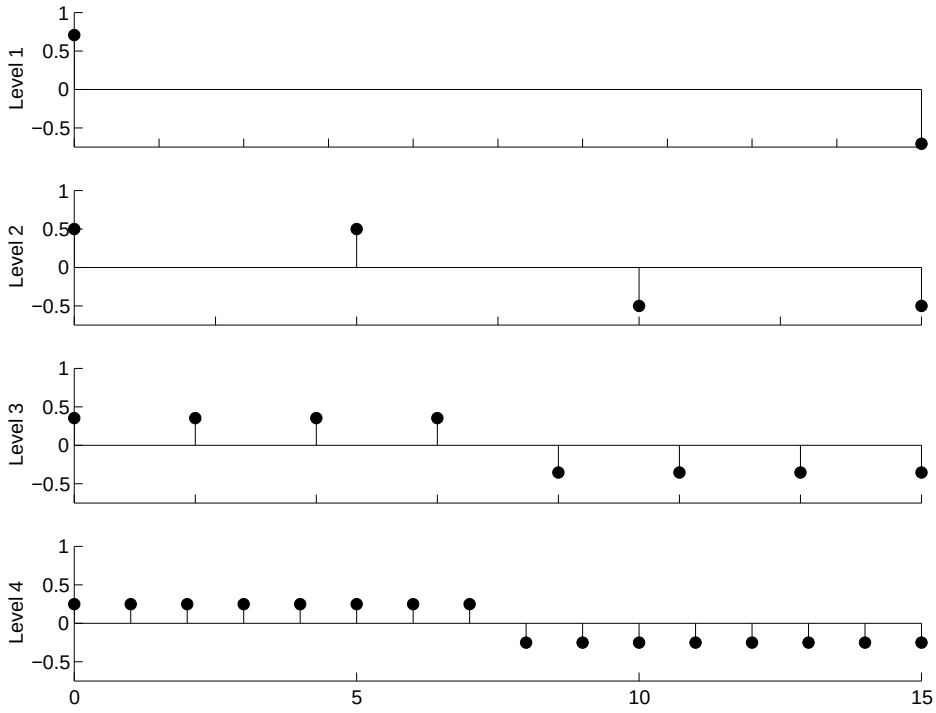


Figure 2: Haar Wavelet Filter Coefficients

Haar wavelet filter coefficients for the first four scales. The nonzero coefficients indicated by vertical lines attached to the solid circles—from top to bottom—are given by $h_{1,l} = (1, -1)/\sqrt{2}$, $h_{2,l} = (1, 1, -1, -1)/2$, $h_{3,l} = (1/\sqrt{8} \cdot \mathbf{1}_4, -1/\sqrt{8} \cdot \mathbf{1}_4)$, and $h_{4,l} = (1/4 \cdot \mathbf{1}_8, -1/4 \cdot \mathbf{1}_8)$, where $\mathbf{1}_N$ is a length N vector of ones.

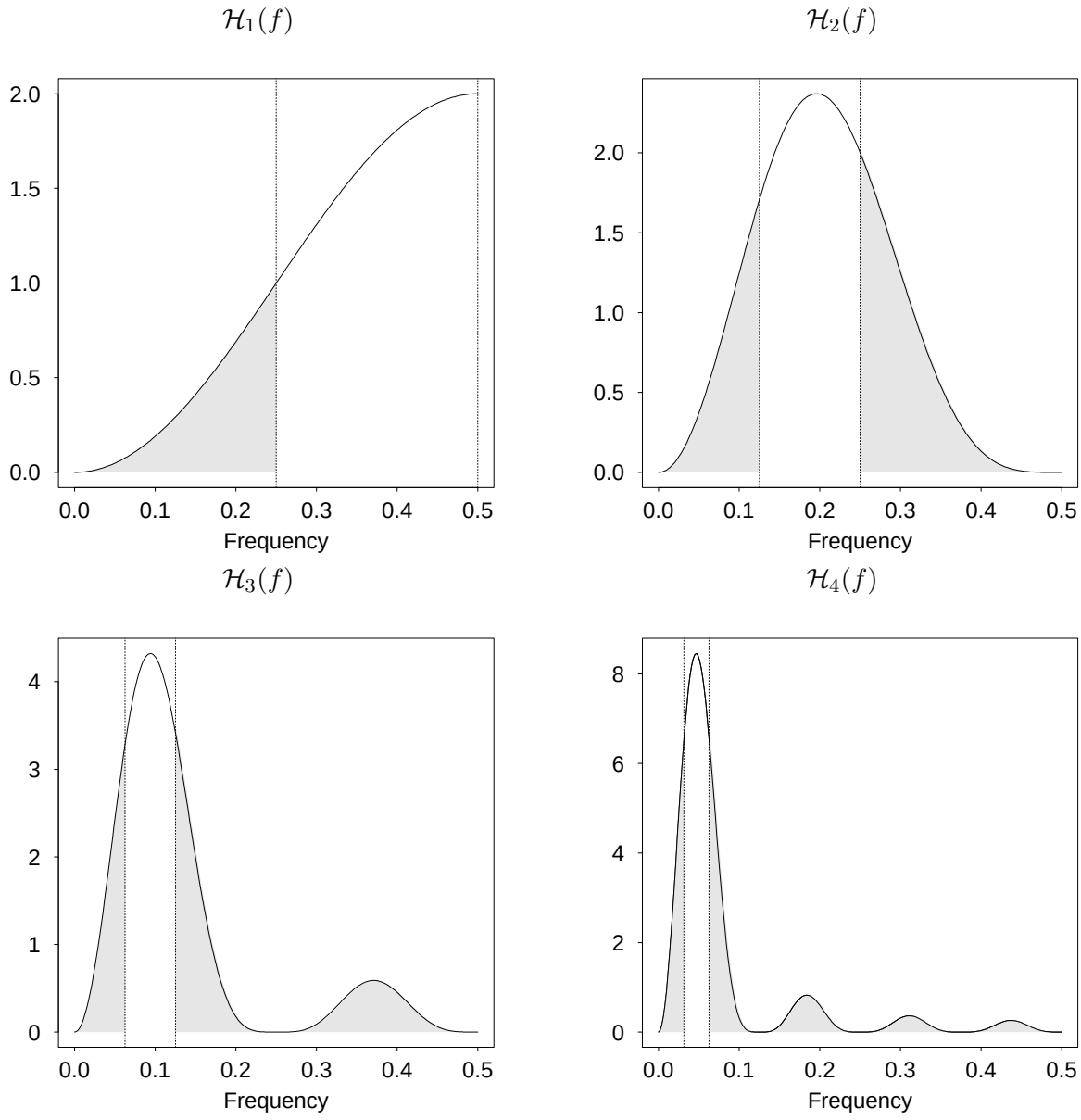


Figure 3: The Haar Wavelet Filter in Frequency Domain

Frequency-domain representation of the Haar wavelet filter. Each plot shows the squared gain function corresponding to the wavelet coefficient vectors in Figure 2. An ideal band-pass filter would only exhibit positive values on the frequencies between the dotted lines. Frequencies with positive weight $\mathcal{H}(f) > 0$ outside of the dotted lines (shaded regions) indicate poor approximation of the Haar wavelet filter to an ideal band-pass filter. This is also known as *leakage*.

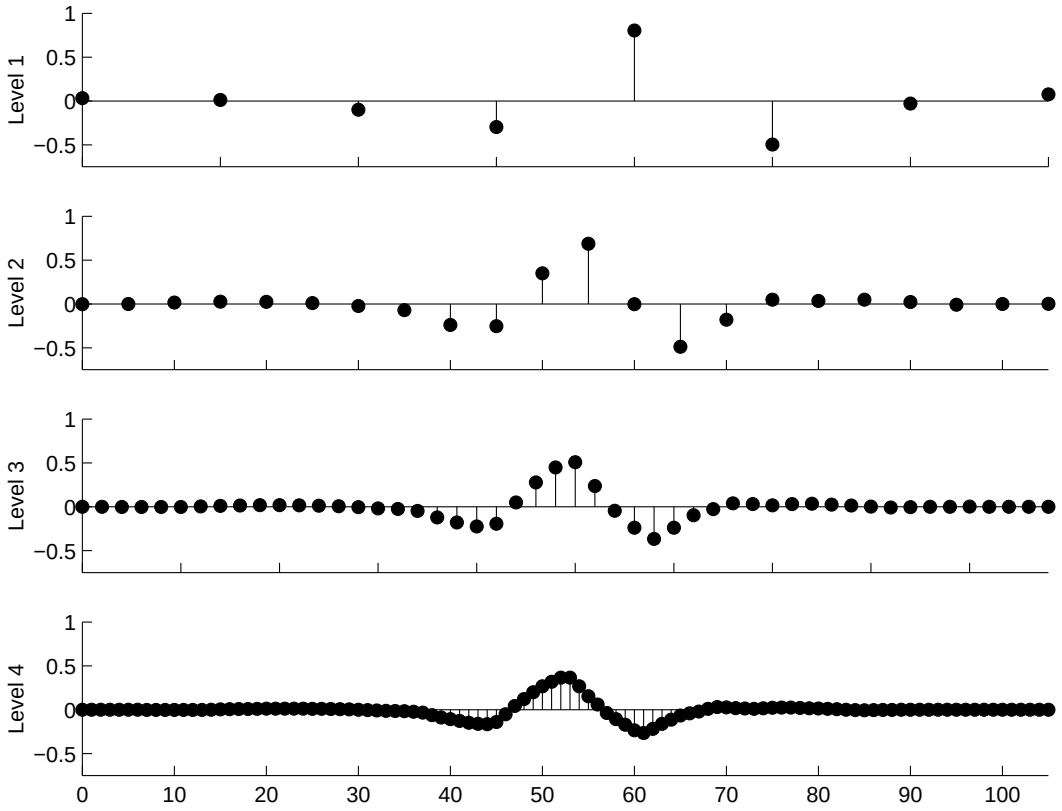


Figure 4: LA(8) Wavelet Filter Coefficients

LA(8) wavelet filter coefficients for the first four scales $h_{1,l}, \ldots, h_{4,l}$. Starting with eight nonzero coefficients at the first scale, the LA(8) wavelet filter is smoother than the Haar and is nearly symmetric with a positive peak and a negative dip on either side. Unlike the Haar wavelet filter, the LA(8) coefficients do not have a convenient closed-form expression.

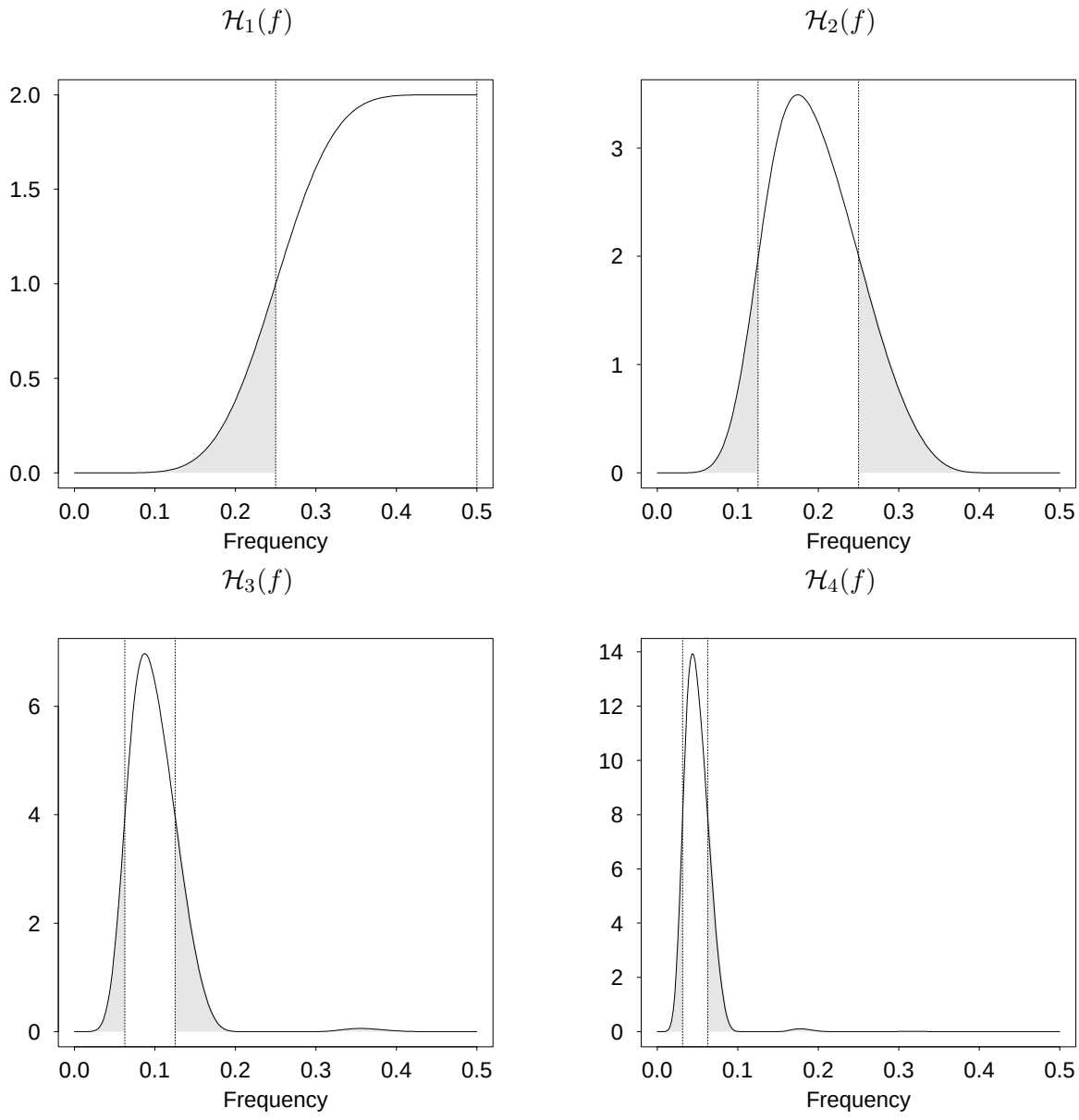


Figure 5: The LA(8) Wavelet Filter in Frequency Domain

Frequency-domain representations of the LA(8) wavelet filter. Each plot shows the squared gain function corresponding to the wavelet coefficient vectors in Figure 4. Frequencies with positive weight $\mathcal{H}(f) > 0$ outside of the dotted lines (shaded regions) correspond to the leakage associated with this approximation to an ideal band-pass filter. The filters associated with these squared gain functions suffer from much less leakage than the Haar wavelet filters.

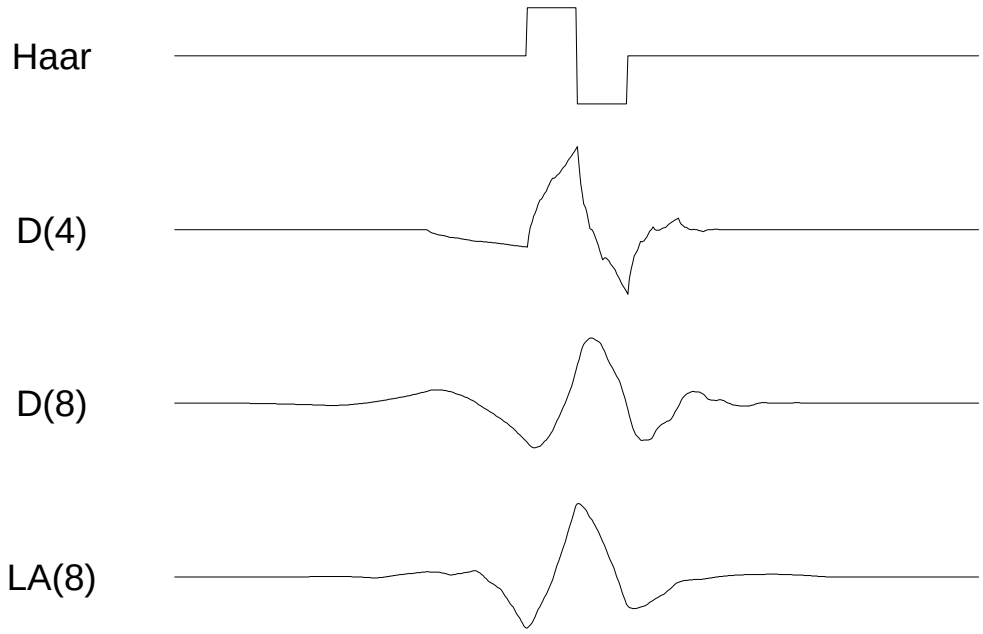
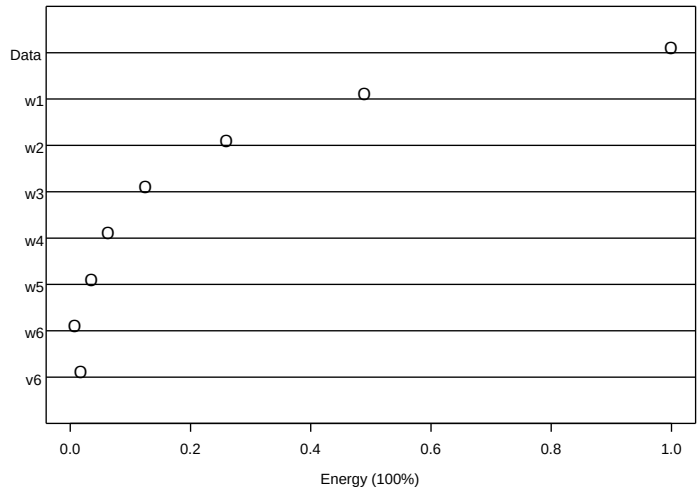
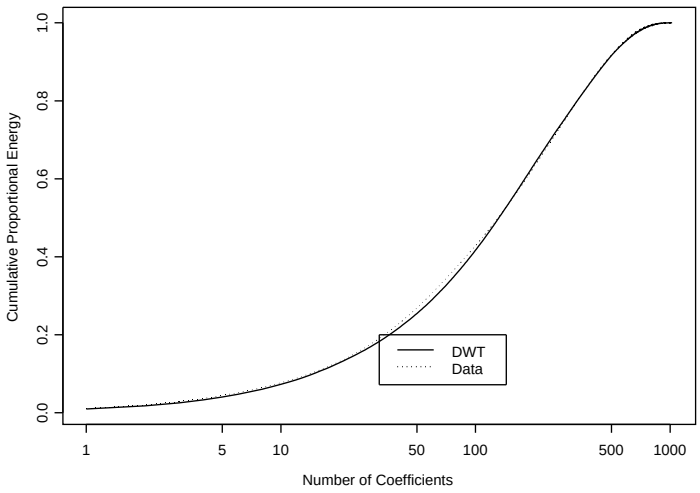


Figure 6: Daubechies Wavelet Filters of Lengths $L \in \{2, 4, 8\}$

Daubechies wavelet filters of lengths $L \in \{2, 4, 8\}$ for level $j = 6$. From top to bottom, the first three rows are extremal phase Daubechies compactly supported wavelets (the Haar wavelet is equivalent to the D(2)), while the last row is a least asymmetric Daubechies compactly supported wavelet.



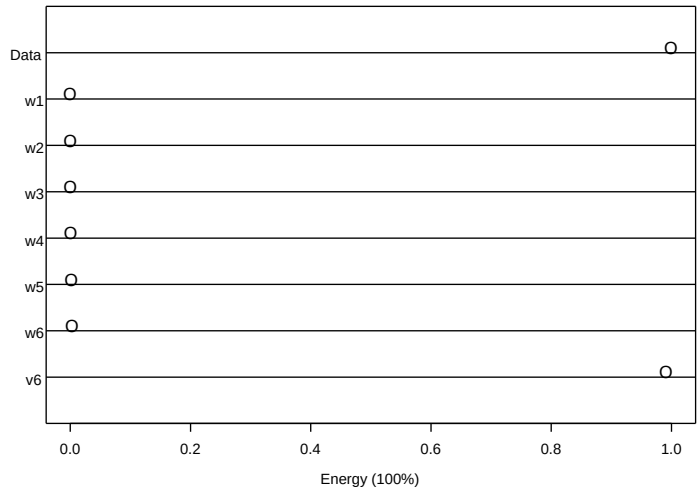
(a)



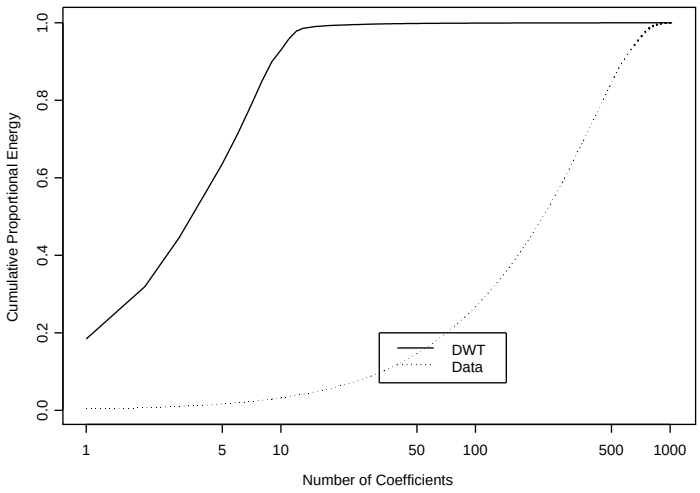
(b)

Figure 7: Wavelet Variance Decomposition of a White Noise Process

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.



(a)



(b)

Figure 8: Wavelet Variance Decomposition of a Unit Root Process

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$i, i = 1, \ldots, 6$ 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.

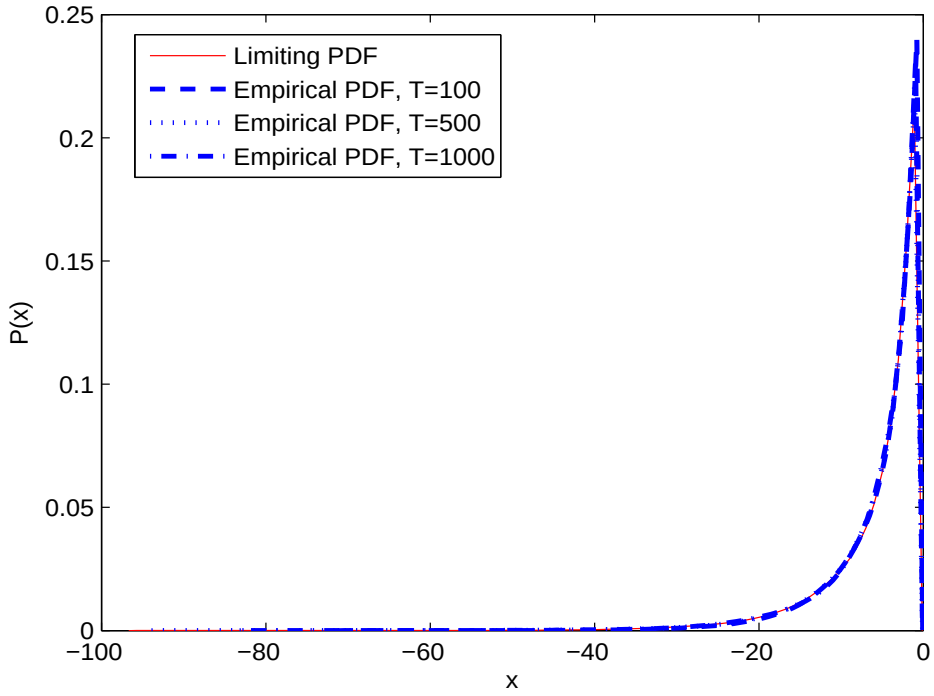


Figure 9: Limiting and Empirical Distributions of $FG_L(Haar)$

The limiting distribution of $-\frac{1}{\int_0^1 [W(r)]^2 dr}$ for 1 million replications. The empirical distribution of $FG_L(Haar)$ is with $T =$ 100 & 500 and 1,000 observations and for 20,000 replications. The simulated data for the null distribution is generated from $y_t = y_{t-1} + u_t$ where $u_t \sim i.i.d.N(0, \sigma^2)$ where $y_0 \sim N(0,1)$.

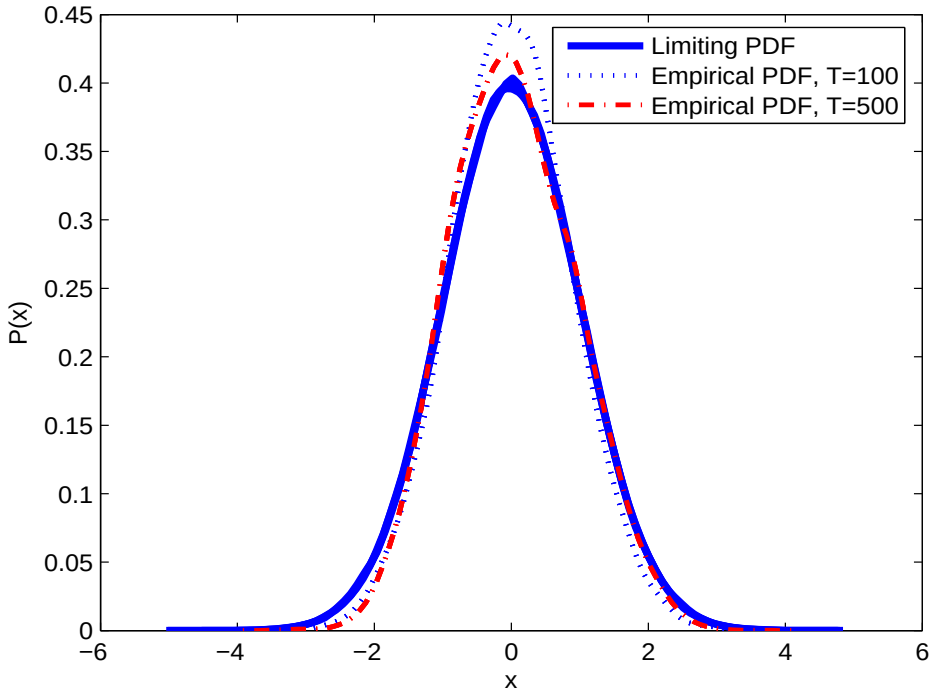


Figure 10: Limiting and Empirical Distributions of $FG_L(drift)(Haar)$

The limiting distribution is standard normal. The empirical distribution of $FG_L(drift)(Haar)$ is with $T = 100$ and $T = 500$ for 10,000 replications. The simulated data for the null distribution is generated from $y_t = \alpha + y_{t-1} + u_t$, $\alpha = 5$, $u_t \sim i.i.d.N(0, \sigma^2)$ where $y_0 \sim N(0, 1)$.

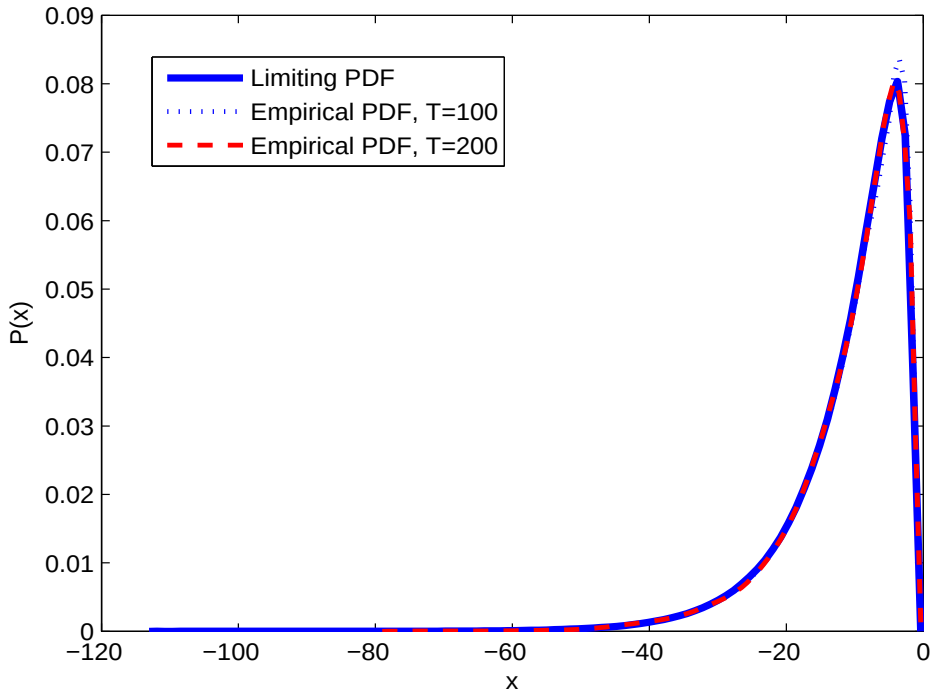


Figure 11: Limiting and Empirical Distributions of $\widehat{D}_{T,1}^{LM}$

The limiting distribution of $-\frac{1}{\int_0^1 [W_\mu(r)]^2 dr}$ for 1 million replications. The empirical distribution of $\widehat{D}_{T,1}^{LM}$ is with $T = 100$ and 200 observations and for 5,000 replications. The simulated data for the null distribution is generated from $y_t = \mu + y_t^s$, where $y_t^s = y_{t-1}^s + u_t$, $u_t \sim iidN(0, \sigma^2)$ and $y_0 \sim N(0,1)$.

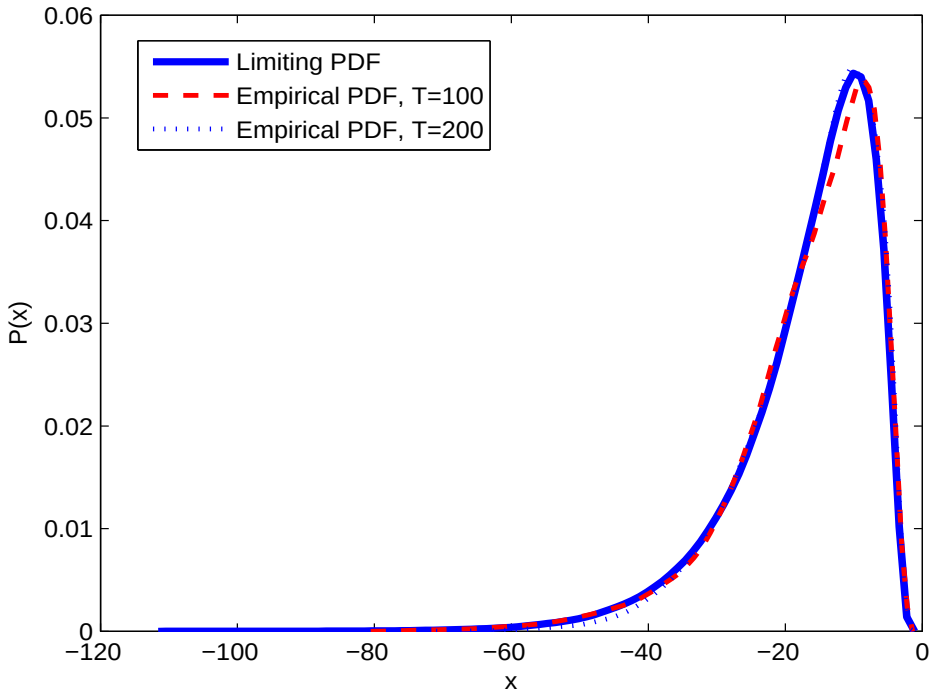


Figure 12: Limiting and Empirical Distributions of $\widehat{D}_{T,1}^{Ld}$

The limiting distribution of $-\frac{1}{\int_0^1 [V_\mu(r)]^2 dr}$ for 1 million replications. The empirical distribution of $\widehat{D}_{T,1}^{Ld}$ is with $T = 100$ and 200 observations and for 5,000 replications. The simulated data for the null distribution is generated from $y_t = \mu + \alpha t + y_t^s$, where $y_t^s = y_{t-1}^s + u_t$, $u_t \sim iidN(0, \sigma^2)$ and $y_0 \sim N(0,1)$.

References

Allan, D. W. (1966). Statistics of atomic frequency standards. *Proceedings of the IEEE*, **31**, 221–230.

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

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.

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.

Dacorogna, M. M., Gençay, R., Müller, U. A., Olsen, R. B., and Pictet, O. V. (2001). *An Introduction to High Frequency Finance*. Academic Press, San Diego.

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

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.

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.

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

Haar, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, **69**, 331–371. In German.

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

Howe, D. A. and Percival, D. B. (1995). Wavelet variance, Allan variance, and leakage. *IEEE Transactions on Intrumentation and Measurement*, **44**, 94–97.

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.

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.

Oppenheim, A. V. and Schafer, R. W. (1989). *Discrete-Time Signal Processing*. Prentice–Hall, Englewood Cliffs, New Jersey.

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. (1983). *The Statistics of Long Memory Processes*. Ph.D. thesis, Department of Statistics, University of Washington.

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

Percival, D. B. and Guttorp, P. (1994). Long-memory processes, the Allan variance and wavelets. In E. Foufoula-Georgiou and P. Kumar, editors, *Wavelets in Geophysics*, volume 4 of *Wavelet Analysis and its Applications*, pages 325–344. Academic Press.

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

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.

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.*