Unit Root Tests with Wavelets

Yanqin Fan & Ramo Gençay

http://www.sfu.ca/∼rgencay

Summary

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 unit root tests, and finally,
- We propose higher order wavelet filters to capture low-frequency stochastic trend parsimoniously and gain power against near unit root alternatives.

1 Granger (1966) – Spectral Shape

- 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.

- The 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.

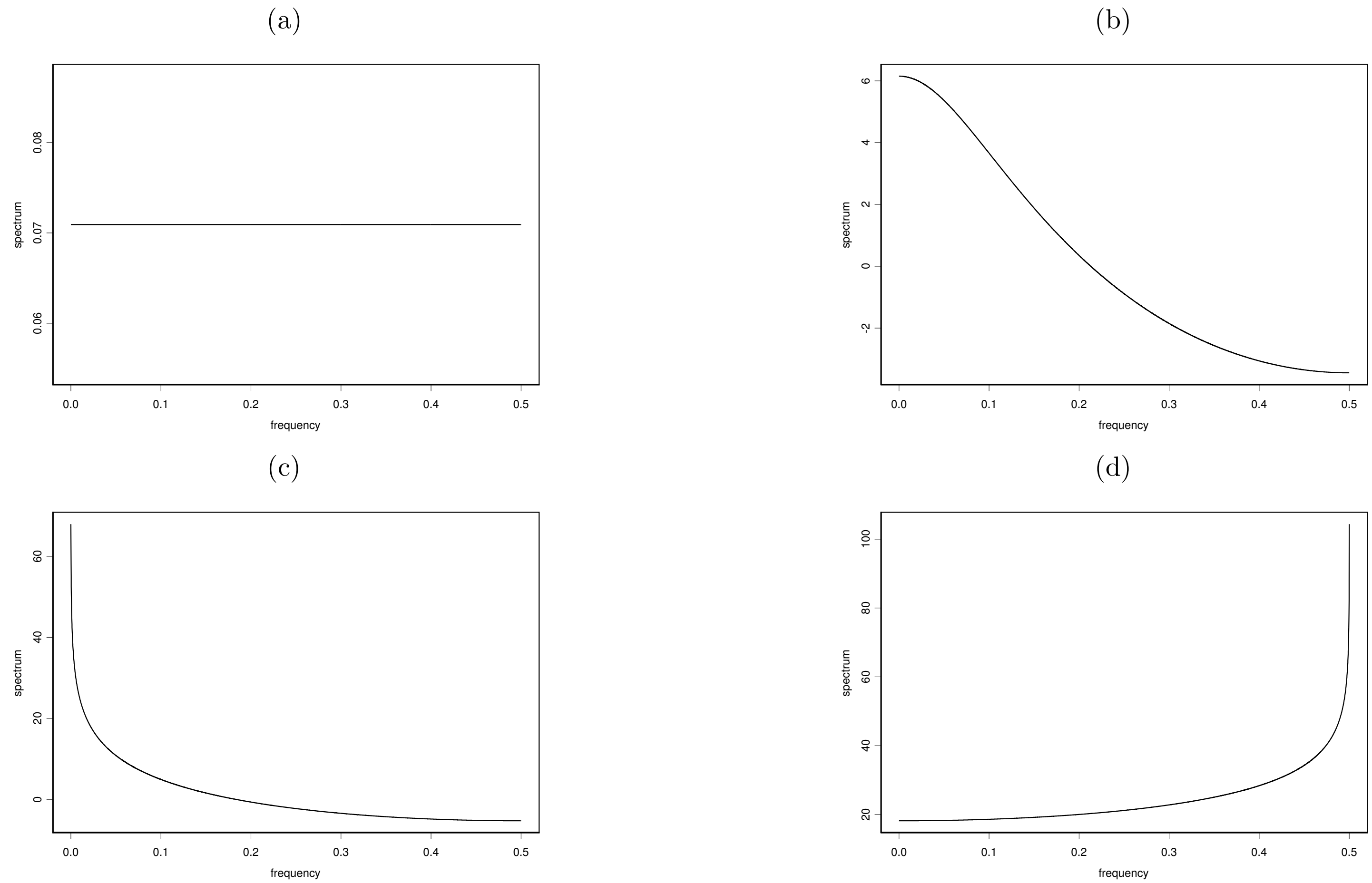


Figure 1: Spectrum of white noise and AR(1) processes. (a) White noise (b) AR(1) with $\phi = 0.5$ (c) AR(1) with $\phi \to 1$ (d) AR(1) with $\phi \to -1$.

- Since Nelson and Plosser (1982) argued that this persistence was captured by modeling the series as having a unit autoregressive root, designing tests for unit root has attracted the attention of many researchers.
- The well-known Dickey and Fuller (1979) unit root tests have limited power to separate a unit root process from near unit root alternatives in small samples.

- Phillips (1986) and Phillips (1987) pioneered the use of the functional central limit theorem to establish the asymptotic distribution of statistics constructed from unit root processes.
- To construct unit root tests with serially correlated errors, one approach is due to Phillips (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.

- 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.
- Cai and Shintani (2006) provide alternative VN tests based on combinations of consistent and inconsistent long run variance estimators. Phillips and Xiao (1998) and Stock (1999) provide a helpful review of the main tests and an extensive list of references.

- In this paper, we develop a general wavelet spectral approach to testing unit roots inspired by Granger (1966). The method of wavelets decomposes a stochastic process into its components, each of which is associated with a particular frequency band.

- The wavelet power spectrum measures the contribution of the variance at a particular frequency band relative to the overall variance of the process. If a particular band contributes substantially more to the overall variance relative to another frequency band, it is considered an important driver of this process.

- By decomposing the variance of the underlying process into the variance of its low frequency components and that of its high frequency components via the discrete wavelet transformation (DWT), we design wavelet-based unit root tests.

- Since DWT is an energy preserving transformation and able to disbalance energy across high and low frequency components of a series, it is possible to isolate the most persistent component of a series in a small number of coefficients referred to as the scaling coefficients.

- Our tests utilize the scaling coefficients of the unit scale. In particular, we construct test statistics from the ratio of the energy from the unit scale to the total energy (variance) of the time series. We establish asymptotic properties of our tests, including their asymptotic null distributions, consistency, and local power properties.

- Our tests are easy to implement, as their asymptotic null distributions are nuisance parameter free and the corresponding critical values can be tabulated.

- The Monte Carlo simulations are conducted to compare the empirical size and power of our tests to the Elliott *et al.* (1996) (ERS) and Ng and Perron (2001) (MPP) tests. Our tests have good size and comparable power against near unit root alternatives in finite samples.

- The DWT is an orthonormal transformation which may be relaxed through an oversampling approach termed as the maximum overlap DWT (MODWT), Percival and Mofjeld (1997).

- 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).

- 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.

2 Wavelets

- A wavelet is a small wave which grows and decays in a limited time period. The contrasting notion is a big wave such as the sine function which keeps oscillating indefinitely.

- 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,

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

- 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.

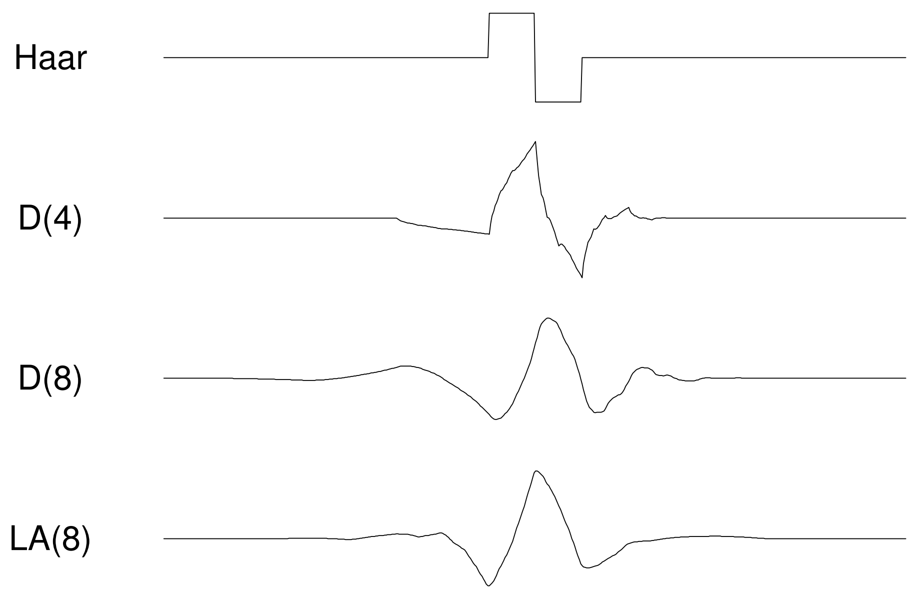


Figure 2: Daubechies wavelet filters of lengths $L \in \{2, 4, 8\}$ for level $j = 6$. From top to bottom, the first three rows are extremal phase Daubechies compactly supported wavelets (the Haar wavelet is equivalent to the D(2)), while the last row is a least asymmetric Daubechies compactly supported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

- 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. 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 basic properties of the scaling filter are

$$\sum_{l=0}^{L-1} g_l = \sqrt{2} \tag{6}$$

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

$$\sum_{l=0}^{L-1} g_l g_{l+2n} = \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0, \tag{8}$$

for all nonzero integers n, and

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

for all integers n. Equation (6) states that scaling coefficients are local weighted averages.

- 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.

- 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.

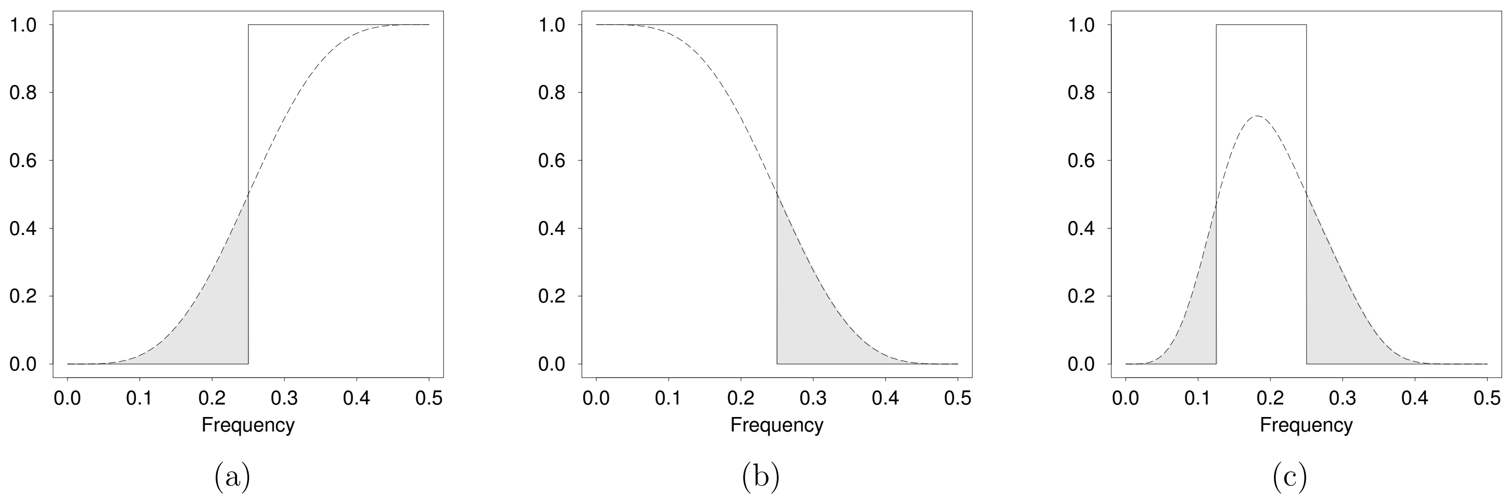


Figure 3: Squared gain functions for ideal filters (solid line) and their wavelet approximations (dotted line). The shaded regions represent *leakage*, meaning frequencies outside the nominal pass-band persist in the filtered output. (a) An ideal high-pass filter (solid line) over the frequency interval $f \in [1/4, 1/2]$ 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.

- The Haar wavelet filter is an excellent benchmark to illustrate $\{h_{j,l}\}$ and $\{g_{j,l}\}$.

- 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

$$h_0 = 1/\sqrt{2} \;\; \text{and} \;\; h_1 = -1/\sqrt{2}$$

through the inverse quadrature mirror relationship.

- Although the Haar wavelet filter is easy to visualize and implement, it is a poor approximation to an ideal band-pass filter.

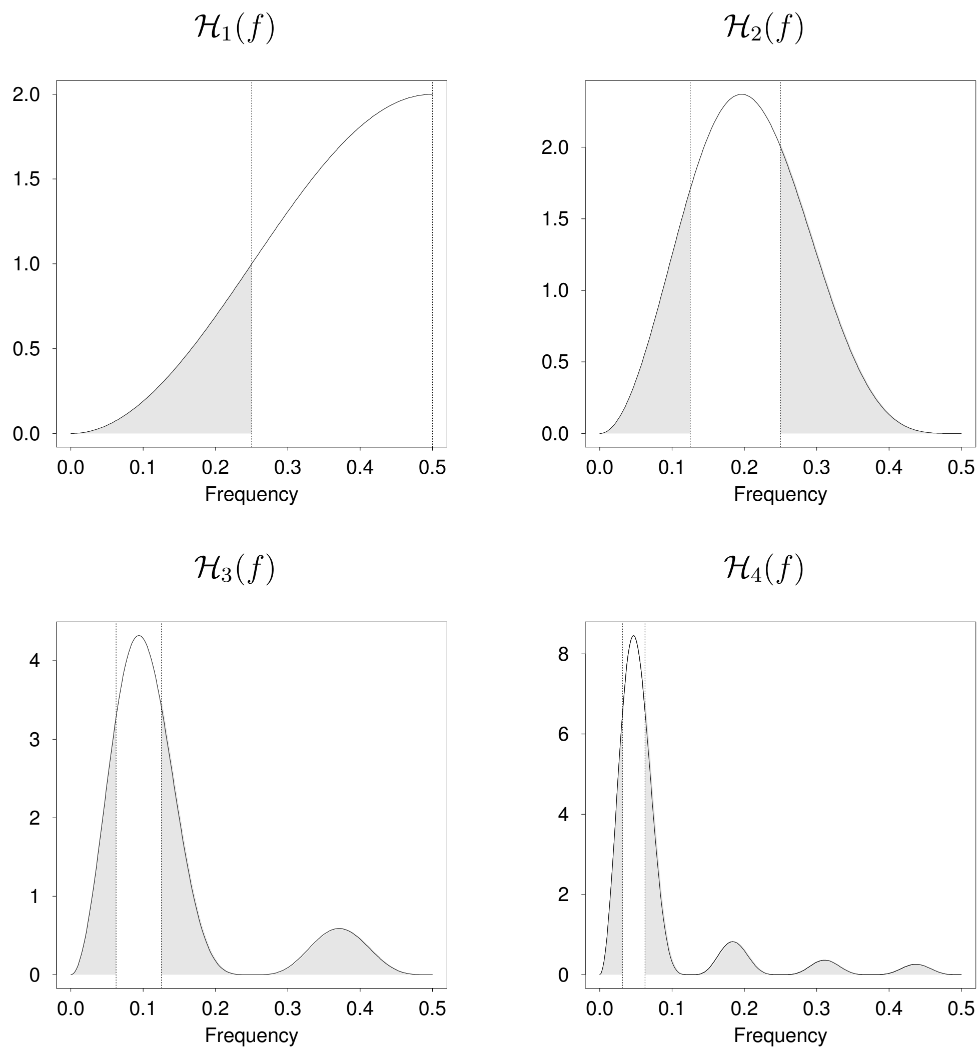


Figure 4: Frequency-domain representations of the Haar wavelet filter. Each plot shows the squared gain function corresponding to the wavelet coefficient vectors. An ideal band-pass filter would only exhibit positive values on the frequencies between the dotted lines. Frequencies with positive weight $\mathcal{H}(f) > 0$ 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*.

2.1 Daubechies Wavelets

- 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.[1] In general, there are no explicit time-domain formulae for this class of wavelet filters.

- Daubechies first choose an *extremal phase* factorization,[2] 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).[3]

---

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

[2] 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).

[3] 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).

- The D(4) wavelets have a simple expression in the time domain via

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

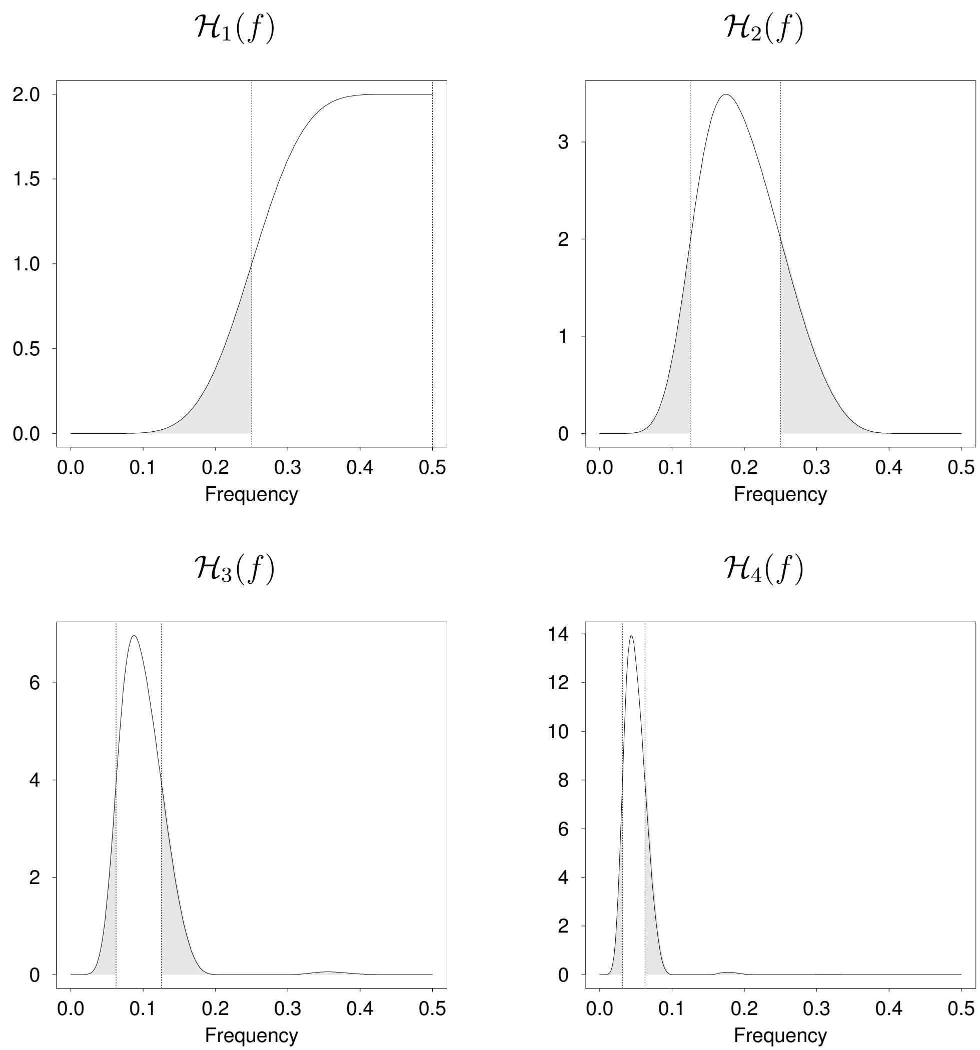


Figure 5: Frequency-domain representations of the LA(8) wavelet filter. Each plot shows the squared gain function corresponding to the wavelet coefficient vectors. Frequencies with positive weight $\mathcal{H}(f) > 0$ outside of the dotted lines (shaded regions) correspond to the leakage associated with 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.

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

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

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).[4]

- The vector of wavelet coefficients may be organized into $M + 1$ vectors,

$$\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M, \mathbf{v}_M]^T, \tag{10}$$

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 $2^M = 2\lambda_M$.

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

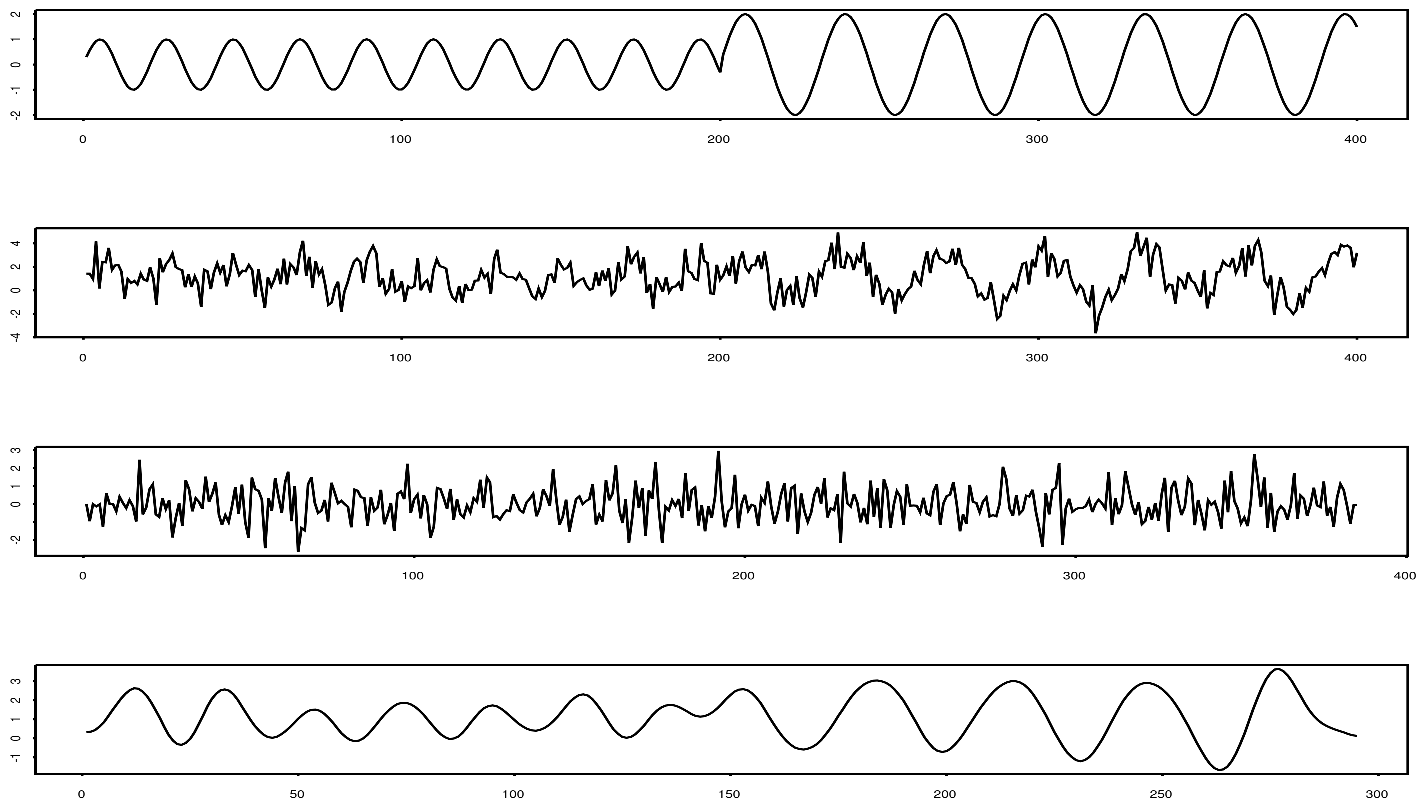


Figure 6:
$y_t = x_t + u_t$ where x_t is $sin(0.3t)$ for $t = 1, 2, \ldots, 200$, $2sin(0.2t)$ for $t = 201, 202, \ldots, 400$ and $u_t \sim N(0,\ 1.2)$. It is a third level MODWT decomposition with S16 filter.

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

$$\|\mathbf{y}\|^2 = \sum_{j=1}^{M} \|\mathbf{w}_j\|^2 + \|\mathbf{v}_M\|^2 , \tag{11}$$

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.

- The motivation behind a wavelet based unit root test can be illustrated through the energy (variance) decomposition of the process.

- For a white noise process,

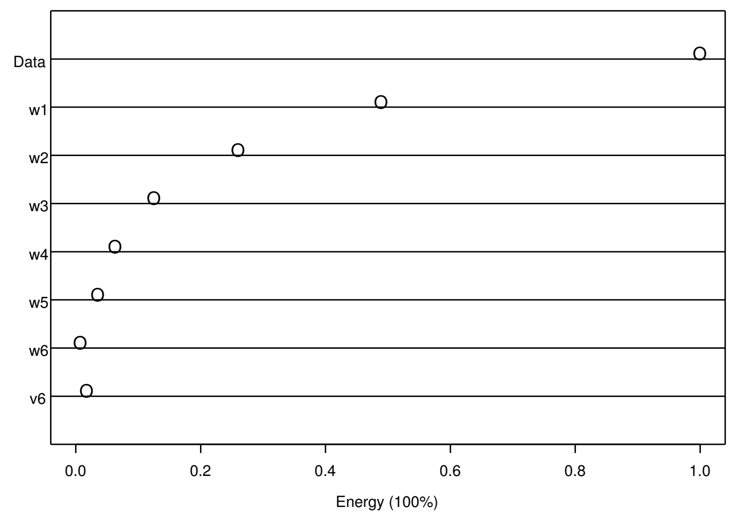
$$\left\|\mathbf{v}_J\right\|^2 / \left\|\mathbf{y}\right\|^2$$

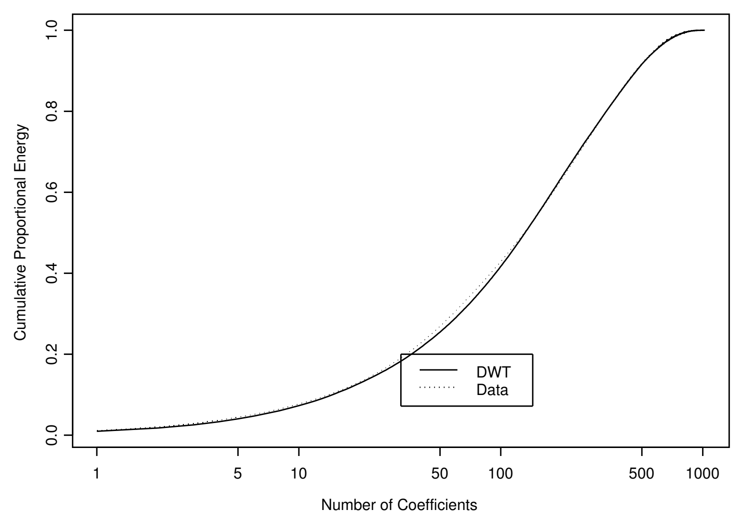
  is close to zero whereas

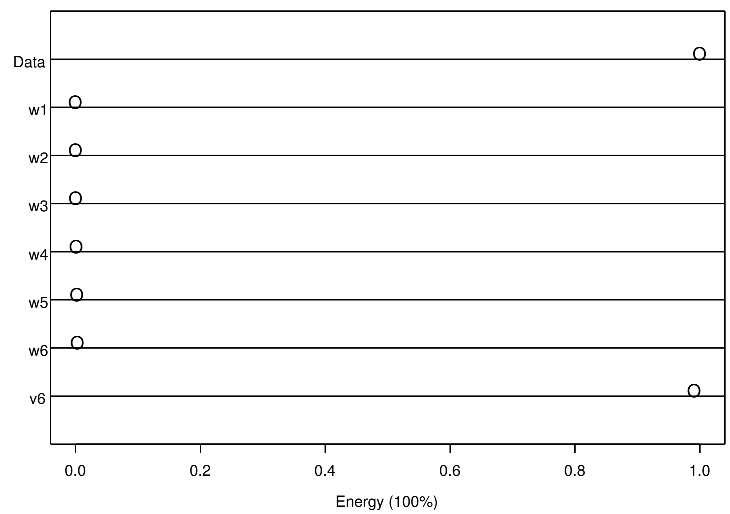
$$\left\|\mathbf{v}_J\right\|^2 / \left\|\mathbf{y}\right\|^2$$

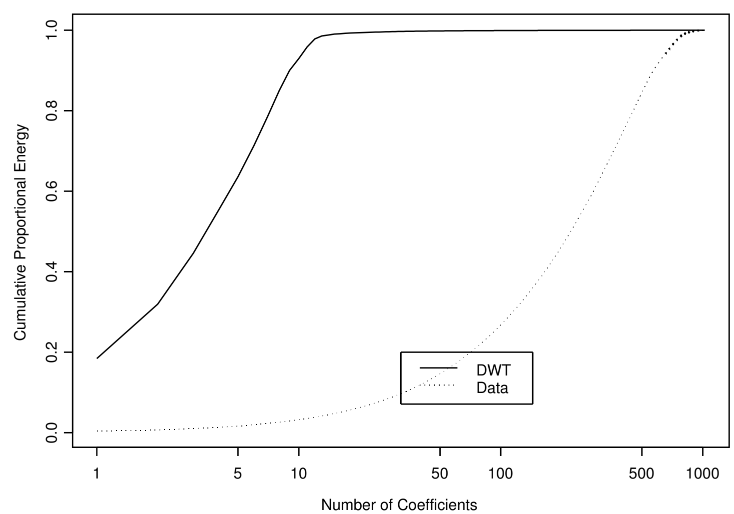
  is close to one for a unit root process.

- 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.


Data
w1
w2
w3
w4
w5
w6
v6
0.0
0.2
0.4
0.6
0.8
1.0
Energy (100%)


Cumulative Proportional Energy
0.0
0.2
0.4
0.6
0.8
1.0
DWT
Data
1
5
10
50
100
500
1000
Number of Coefficients


Data
w1
w2
w3
w4
w5
w6
v6
0.0
0.2
0.4
0.6
0.8
1.0
Energy (100%)


Cumulative Proportional Energy
0.0
0.2
0.4
0.6
0.8
1.0
DWT
Data
1
5
10
50
100
500
1000
Number of Coefficients

3 New Unit Root Tests — No Drift Case

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

$$y_t = \rho y_{t-1} + u_t, \tag{12}$$

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 $\gamma_j = E(u_t u_{t-j})$.

- In this section, we consider tests for

$$H_0 : \rho = 1 \quad \text{against} \quad H_1 : |\rho| < 1.$$

- Therefore, under the alternative hypothesis, $\{y_t\}$ is a zero-mean stationary process with the long run variance $\omega_y^2 = (1-\rho)^{-2}\,\omega^2$.

3.1 The first test — Haar 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

$$W_{t,1} = \frac{1}{\sqrt{2}}(y_{2t} - y_{2t-1}),\ t = 1, 2, \ldots, T/2, \tag{13}$$

$$V_{t,1} = \frac{1}{\sqrt{2}}(y_{2t} + y_{2t-1}),\ t = 1, 2, \ldots, T/2. \tag{14}$$

- 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 total energy of the scaling coefficients $\{V_{t,1}\}$ dominates that of the wavelet coefficients $\{W_{t,1}\}$, we propose the following test statistic:

$$\hat{S}_{T,1} = \frac{\sum_{t=1}^{T/2} V_{t,1}^2}{\sum_{t=1}^{T/2} V_{t,1}^2 + \sum_{t=1}^{T/2} W_{t,1}^2}. \tag{15}$$

- 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.

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

Note that:

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

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

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

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

The proof of Theorem 3.2 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.

- 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 $\hat{\gamma}_j = T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}$.

- 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

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

- 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:

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

  to approximate the null limiting distribution of FG_1.

3.2 General filter case: unit scale decomposition

- Let $\{h_l\}_{l=0}^{L-1}$ be an even length L sequence of Daubechies compactly supported wavelet filter coefficients.

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)* $\left(\frac{T}{2}\right)\left(\hat{S}_{T,1}^{L} - 1\right) = -\frac{EW_{t,1}^2}{\lambda_v^2 \int_0^1 [W(r)]^2 dr} + o_p(1)$ *under the null hypothesis.*

- 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 L increases, the approximation of the Daubechies wavelet filter to the ideal high-pass filter improves[5], we expect tests based on $\hat{S}_{T,1}^{L}$ to gain power as L increases.

- On the other hand, as L increases, the number of BI wavelet and scaling coefficients will decrease which would have an adverse effect on the power of our tests. It might be possible to choose L based on some power criterion function, but this is beyond the scope of this paper.

- Define the test statistic:

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

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

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

3.3 Tests against trend stationarity

$$y_t = \alpha + \rho y_{t-1} + u_t, \tag{16}$$

where $\{u_t\}$ satisfies Assumption 1.

- Note that under H_0, model (16) 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 (16) 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.[6]

[6] 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:

$$y_t = \mu + \alpha t + y_t^s, \tag{17}$$

where $\{y_t^s\}$ is generated by model (12).

- 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 $\{\Delta y_t\}$.[7]

[7] Alternative expressions for the detrended series $\left\{\widetilde{y}_t - \overline{\widetilde{y}}\right\}$ can be found in Schmidt and Phillips (1992).

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

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

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

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

Theorem 3.4 *Under H_0, we have: (i) $T\left(\widehat{S}_{T,1}^{LM} - 1\right) \implies -\frac{E(W_{t,1}^{M})^2}{2\omega^2 \int_0^1 \left[W_\mu(r)\right]^2 dr}$; (ii) $T\left(\widehat{S}_{T,1}^{Ld} - 1\right) \implies -\frac{E(W_{t,1}^{d})^2}{2\omega^2 \int_0^1 \left[V_\mu(r)\right]^2 dr}$. We have: (i) $T\left(\widehat{S}_{T,1}^{LM} - 1\right) \implies -\frac{E(W_{t,1}^{M})^2}{2\omega^2 \int_0^1 \left[J_c^M(r)\right]^2 dr}$; (ii) $T\left(\widehat{S}_{T,1}^{Ld} - 1\right) \implies -\frac{E(W_{t,1}^{d})^2}{2\omega^2 \int_0^1 \left[J_c^d(r)\right]^2 dr}$, where $J_c^M(r) = \int_0^r \exp\{(r-u)c\}\, dW_\mu(u)$ and $J_c^d(r) = \int_0^r \exp\{(r-u)c\}\, dV_\mu(u)$.*

4 Maximum Overlap DWT

- MODWT has been demonstrated to have advantages over DWT in several situations including the estimation of wavelet variance.[8]

- It is interesting to note that when the Haar wavelet filter is used,

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

- This expression resembles that of the Sargan and Bhargava (1983) and Bhargava (1986) test.

- In fact, we can obtain the Sargan and Bhargava (1983) and Bhargava (1986) test from an extension of $\widehat{S}_{T,1}^{LM}$ by using MODWT instead of DWT.

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

- To see this, we recall that apart from a factor of $\sqrt{2}$, the unit scale MODWT wavelet and scaling coefficients of $\{y_t - \overline{y}\}$ are given by

$$\widetilde{W}_{t,1} = \sum_{l=0}^{L-1} h_l y_{t-l \bmod T}, \; \widetilde{V}_{t,1} = \sum_{l=0}^{L-1} g_l \left(y_{t-l \bmod T} - \overline{y}\right), \tag{18}$$

where $t = 1, \ldots, T$. It is easy to see that the DWT coefficients are obtained from the corresponding MODWT coefficients via downsampling by 2.

- At each scale, there are T MODWT wavelet coefficients and T MODWT scaling coefficients. Let

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

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

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

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

5 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:

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

- 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:

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

where $\left\{\widehat{W}_{t,1}\right\}$ is the unit scale wavelet coefficients of $\{\widehat{u}_t\}$.

Theorem 5.1 *Under the null hypothesis of no-cointegration,*

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

where $\eta' = (1, -a'_{21}A_{22}^{-1})$ *and* $\left\{W^z_{t,1}\right\}$ *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.

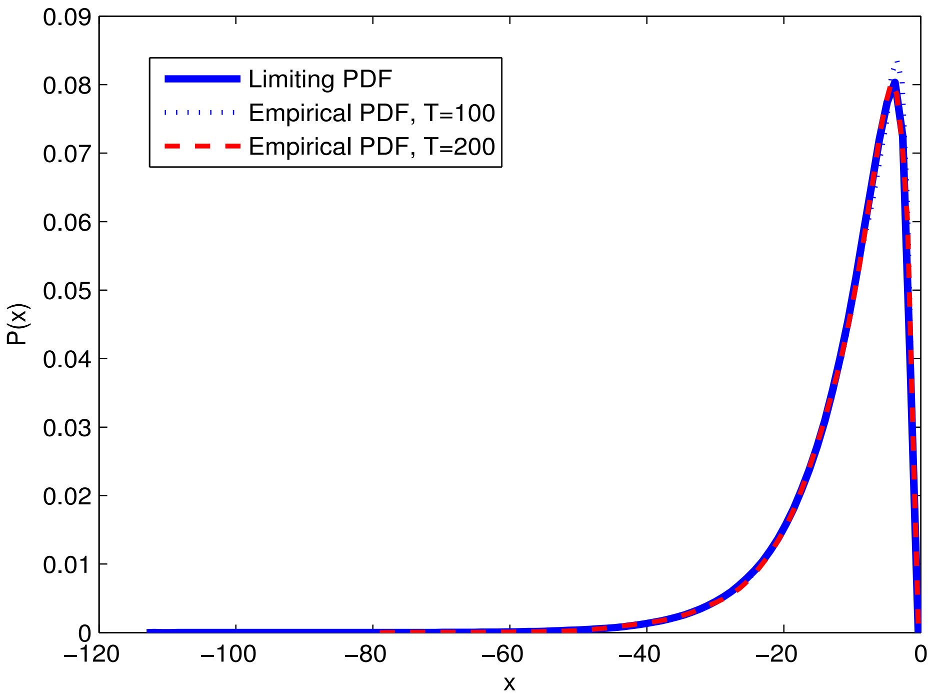


Figure 7: Limiting and Empirical Distributions of $\widehat{S}_{T,1}^{LM}$

The limiting distribution of $-\frac{1}{\int_0^1 [W_\mu(r)]^2 dr}$ for 1 million replications. The empirical distribution of $\widehat{S}_{T,1}^{LM}$ 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 $y_0 \sim N(0, 1)$.

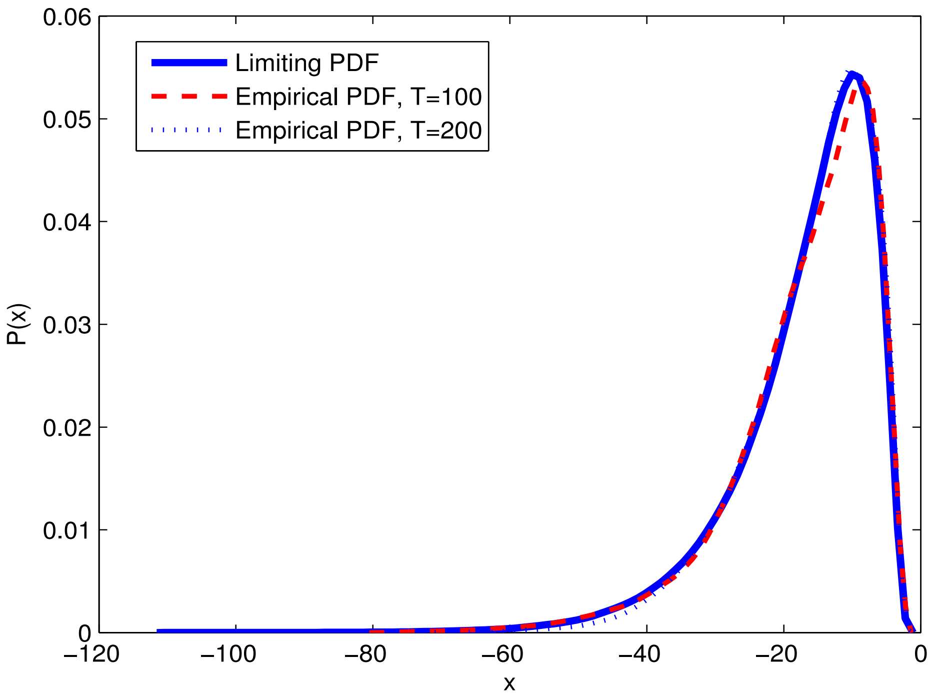


Figure 8: Limiting and Empirical Distributions of $\widehat{S}_{T,1}^{Ld}$

The limiting distribution of $-\frac{1}{\int_0^1 [W_\mu(r)]^2 dr}$ for 1 million replications. The empirical distribution of $\widehat{S}_{T,1}^{Ld}$ 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 $y_0 \sim N(0, 1)$.

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

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

Table 1: Critical Values of Wavelet Tests

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

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

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

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

Table 3: SIZE AND POWER OF THE $\widehat{S}_{T,1}^{Ld}$ - GLS DETRENDED SERIES WITH SERIALLY CORRELATED ERRORS

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

6 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.

- In our tests, the intuitive construction and simplicity are worth emphasizing. The simulation studies demonstrate the comparable power of our tests with reasonable empirical sizes.

- Several extensions of our tests are possible. First, our tests make use of the unit scale DWT only $(J = 1)$ and hence of the energy decomposition of $\{y_t\}$ into frequency bands $[0, 1/2]$ and $[1/2, 1]$. Heuristically, these tests are suitable for testing a unit root process against alternatives that have most energy concentrated in the frequency band $[1/2, 1]$.

- To distinguish between a unit root process and a 'strongly' dependent process that has substantial energy in frequencies close to zero, we need to further decompose the low frequency band $[0, 1/2]$. DWT of higher scales $(J > 1)$ provides a useful device. The choice of J thus depends on the energy concentration of the alternative process against which the unit root hypothesis is being tested.

- Second, we show in the paper that the Sargan and Bhargava test is a special case of wavelet based tests with MODWT using unit scale Haar wavelet filter.

- MODWT has proven to have advantages over DWT in various situations including wavelet variance estimation. It would be interesting to see if it also has advantages in the context of testing unit root.

- Thirdly, the unit root tests developed in this paper can be extended to residual-based tests for cointegration in the same way that other unit root tests have been extended, see e.g., Phillips and Ouliaris (1990) and Stock (1999).

References

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

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.

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.

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.

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

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.

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

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

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

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

Percival, D. B. (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. (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*.