Unit Root and Cointegration 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.

- Consider the stylized trend-cycle decomposition of a time series y_t:

$$y_t = \mu + \alpha t + \epsilon_t, \quad \epsilon_t = \rho\epsilon_{t-1} + u_t, \quad u_t \sim WN(0, \sigma^2)$$

- If $|\rho| < 1$ then y_t is $I(0)$ about the deterministic trend. If $|\rho| = 1$, then ϵ_t follows a stichastic trend and y_t is $I(1)$ with drift.

- Autoregressive unit root tests are based on testing the null hypothesis that $\rho = 1$ (difference stationary) against the alternative hypothesis that $\rho < 1$ (trend stationary).

- They are called unit root tests because under the null hypothesis the autoregressive polynomial of ϵ_t has a root equal to unity.

Figure 1: Simulated Trend Stationary, I(0) and Difference Stationary, I(1) Processes.

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.

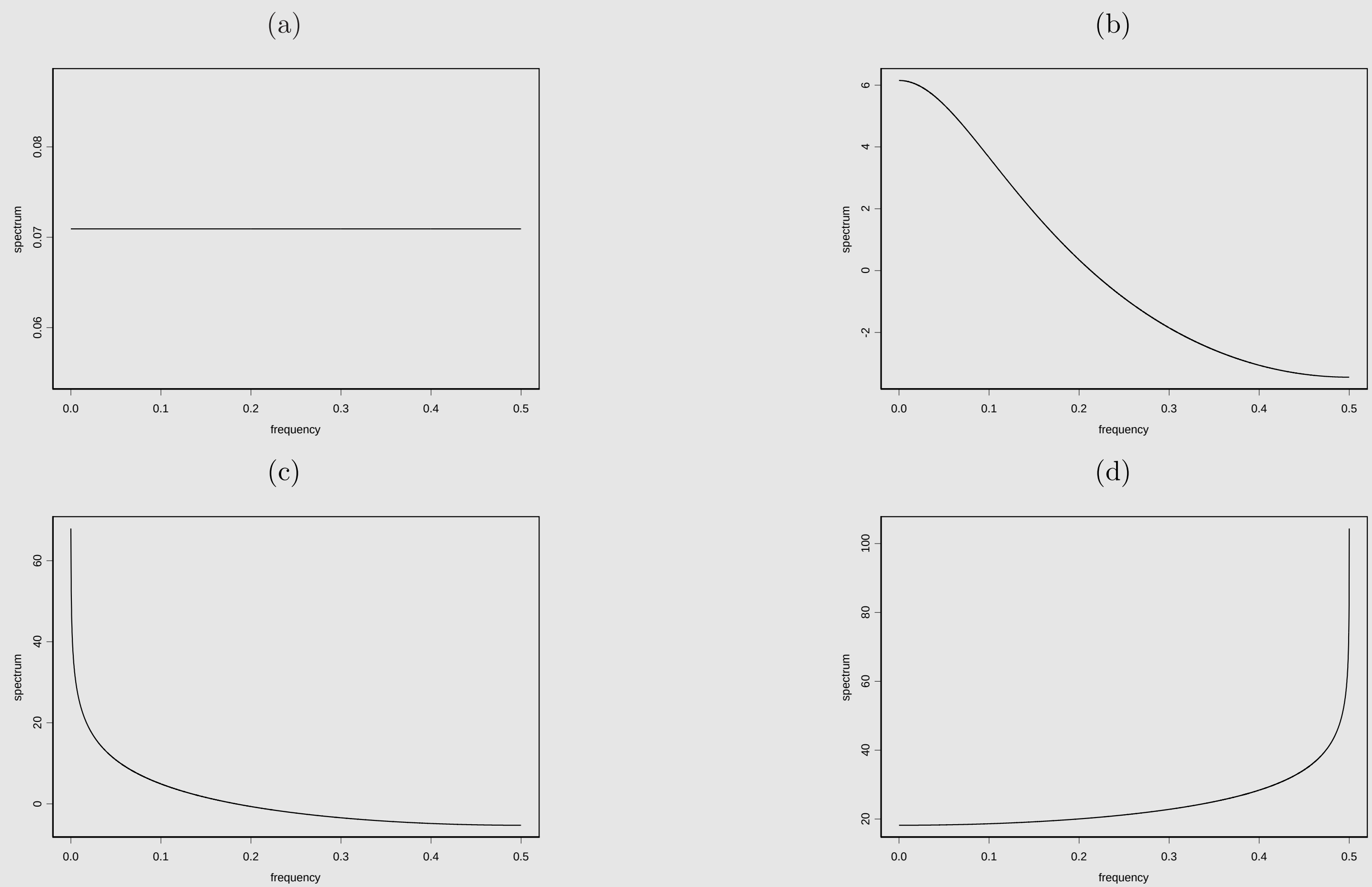


Figure 2: Spectrum of white noise and AR(1) processes. (a) White noise (b) AR(1) with $\phi = 0.5$ (c) AR(1) with $\phi \to 1$ (d) AR(1) with $\phi \to -1$.

- Unit root tests have limited power to separate a unit root process from near unit root alternatives in small samples.

- Discrete wavelet transformation (DWT) is an energy preserving transformation and able to disbalance energy (variance) across stationary and nonstationary components of a series.

- It is possible to isolate the most persistent component of a series in a small number of wavelet coefficients at a low-frequency scale.

- Our approach relies on the decomposition of variance across low and high-frequency scales and can be motivated with the 'integrate' and 'difference' operators.

- In 'integration' or sum, the high frequency components are filtered out and what remains are low frequencies, which is a feature of the unit root 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.

- Wavelet-based test statistics which utilize the energy in a low-frequency scale may therefore increase the power of a unit root test. In particular, we construct test statistics from the ratio of the energy from a low-frequency scale to the sum of the energy of high-frequency scales.

- We postulate that energy of a unit root process can be summarized parsimoniously by its low frequency component, establish the limiting distributions of the test statistics under this premise and demonstrate that size and power properties of these test statistics are superior to those currently available.

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

- 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.

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.

- 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,

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

- 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.

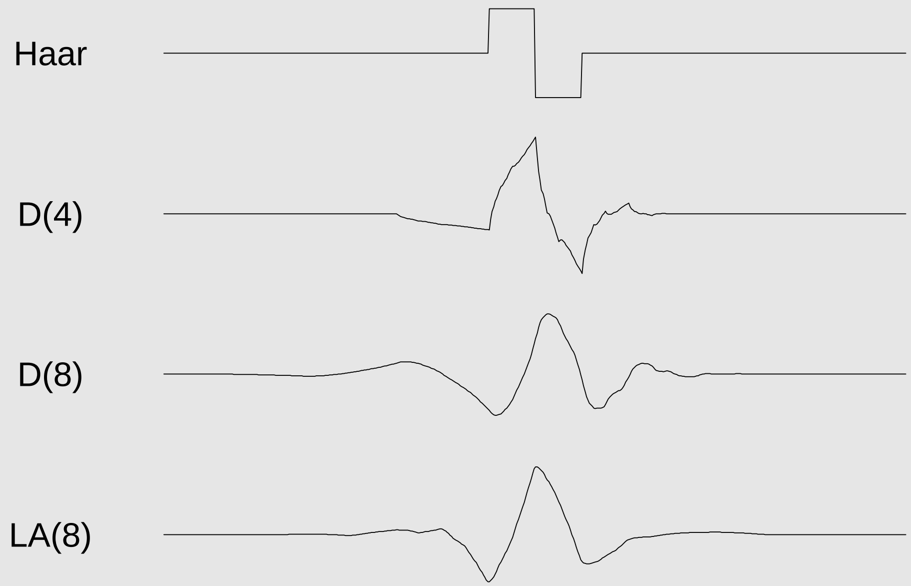


Figure 3: Daubechies wavelet filters of lengths $L \in \{2, 4, 8\}$ for level $j = 6$. From top to bottom, the first three rows are extremal phase Daubechies compactly supported wavelets (the Haar wavelet is equivalent to the D(2)), while the last row is a least asymmetric Daubechies compactly supported wavelet.

- 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. 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.

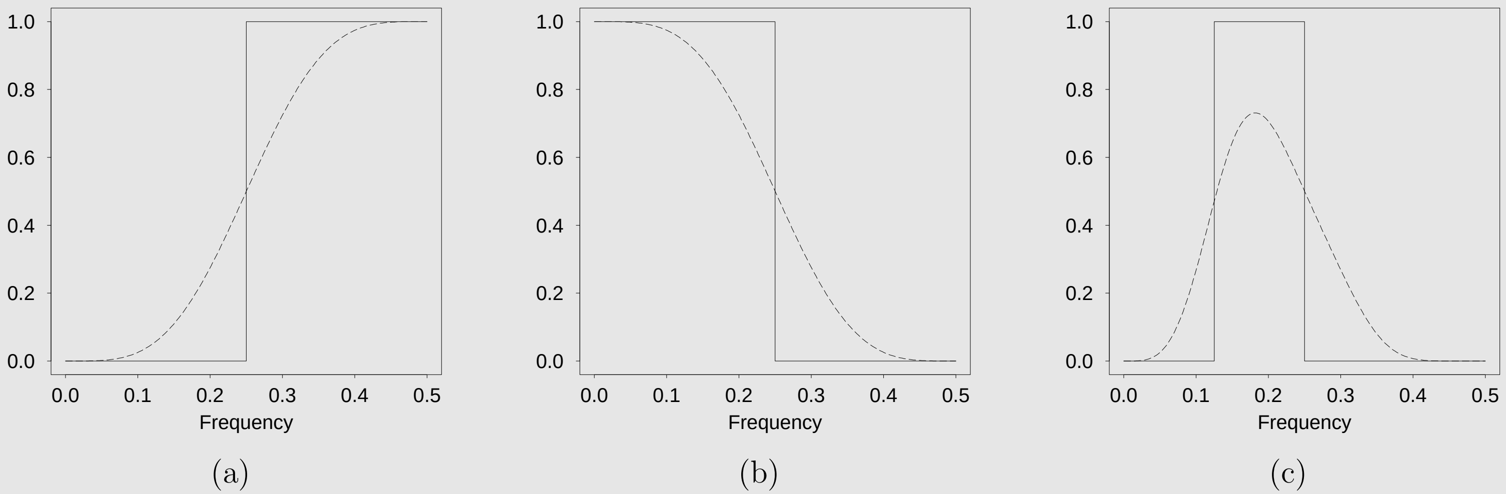


Figure 4: Squared gain functions for ideal filters (solid line) and their wavelet approximations (dotted line). The shaded regions represent *leakage*, meaning frequencies outside the nominal pass-band persist in the filtered output. (a) An ideal high-pass filter (solid line) over the frequency interval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

2.1 Haar Wavelet

- 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

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

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.

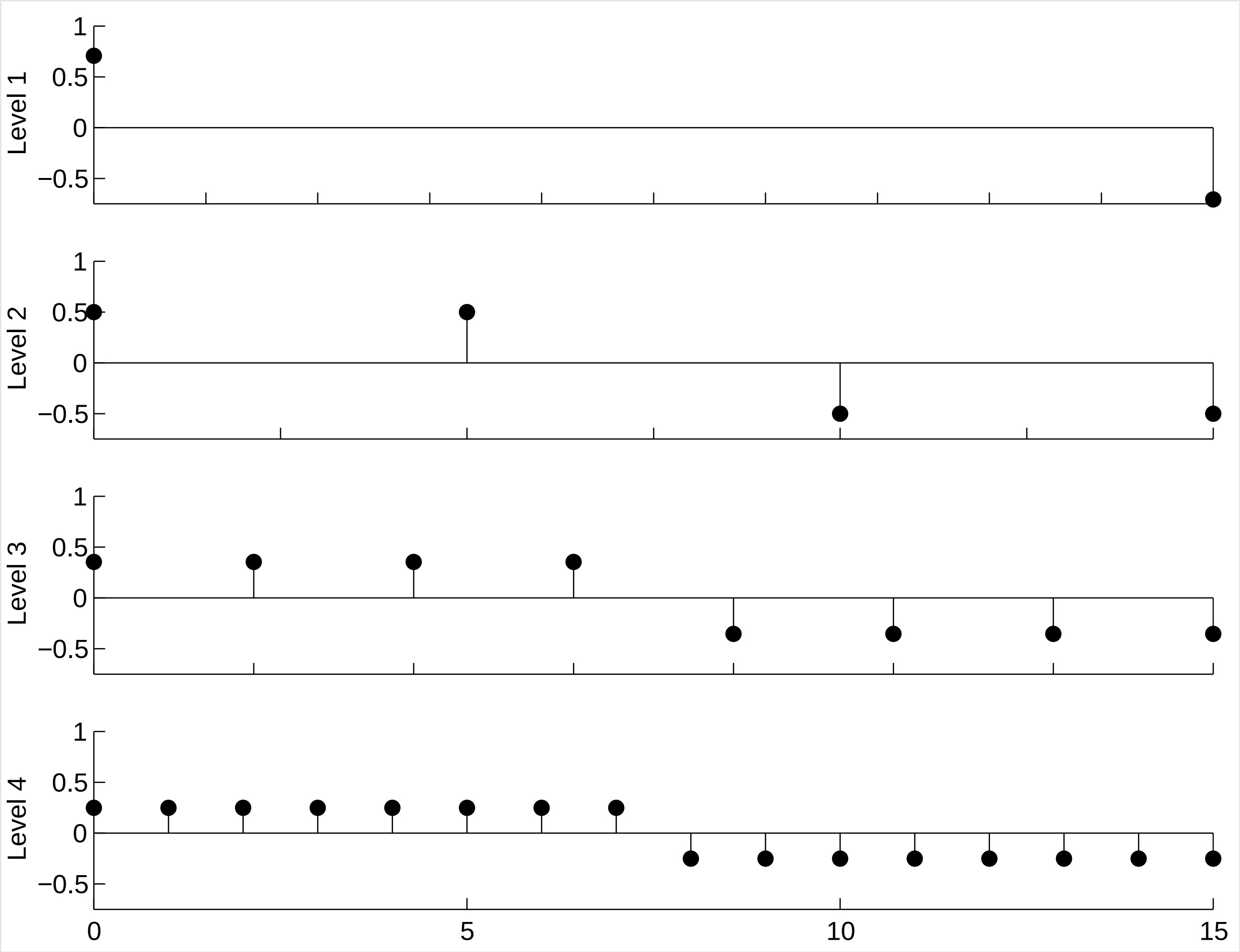


Figure 5: Haar wavelet filter coefficients for the first four scales. The nonzero coefficients indicated by vertical lines attached to the solid circles—from top to bottom—are given by $h_{1,l} = (1, -1)/\sqrt{2}$, $h_{2,l} = (1, 1, -1, -1)/2$, $h_{3,l} = (1/\sqrt{8} \cdot \mathbf{1}_4, -1/\sqrt{8} \cdot \mathbf{1}_4)$, and $h_{4,l} = (1/16 \cdot \mathbf{1}_8, -1/16 \cdot \mathbf{1}_8)$, where $\mathbf{1}_N$ is a length N vector of ones.

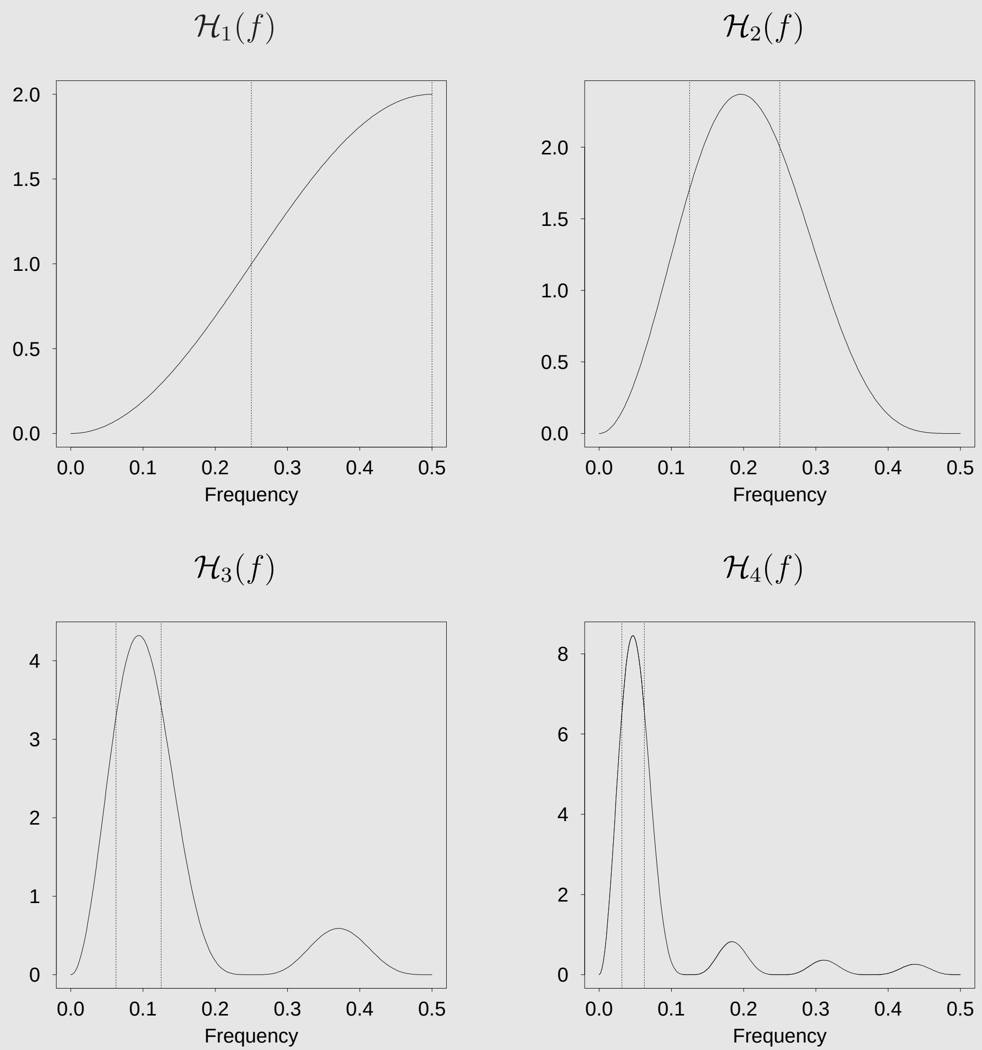


Figure 6: Frequency-domain representations of the Haar wavelet filter. Each plot shows the squared gain function corresponding to the wavelet coefficient vectors in Figure 5. An ideal band-pass filter would only exhibit positive values on the frequencies between the dotted lines. Frequencies with positive 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*.

2.2 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

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

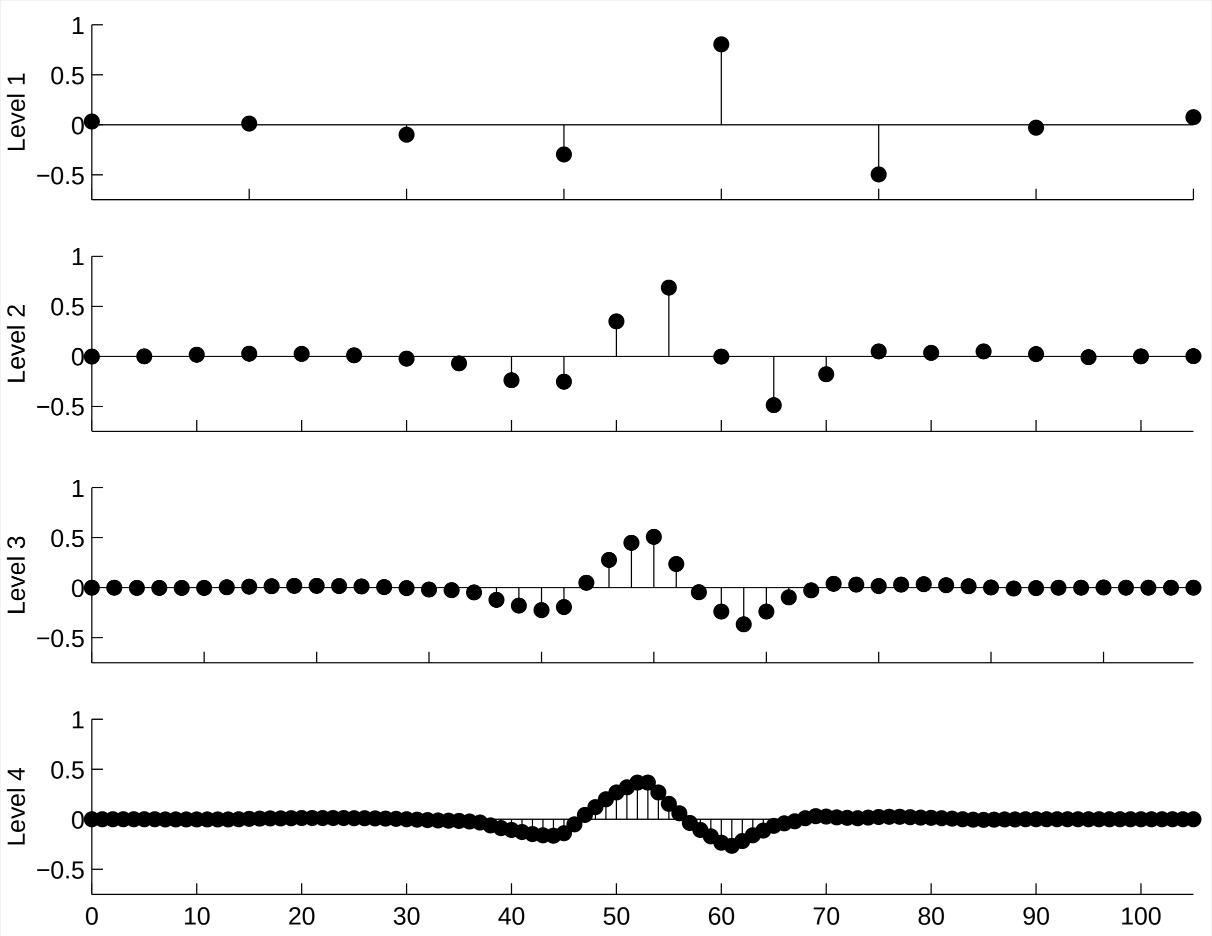


Figure 7: LA(8) wavelet filter coefficients for the first four scales $h_{1,l}, \ldots, h_{4,l}$. Starting with eight nonzero coefficients at the first scale, the LA(8) wavelet filter is smoother than the Haar and is nearly symmetric with a positive peak and a negative dip on either side. Unlike the Haar wavelet filter, the LA(8) coefficients do not have a convenient closed-form expression.

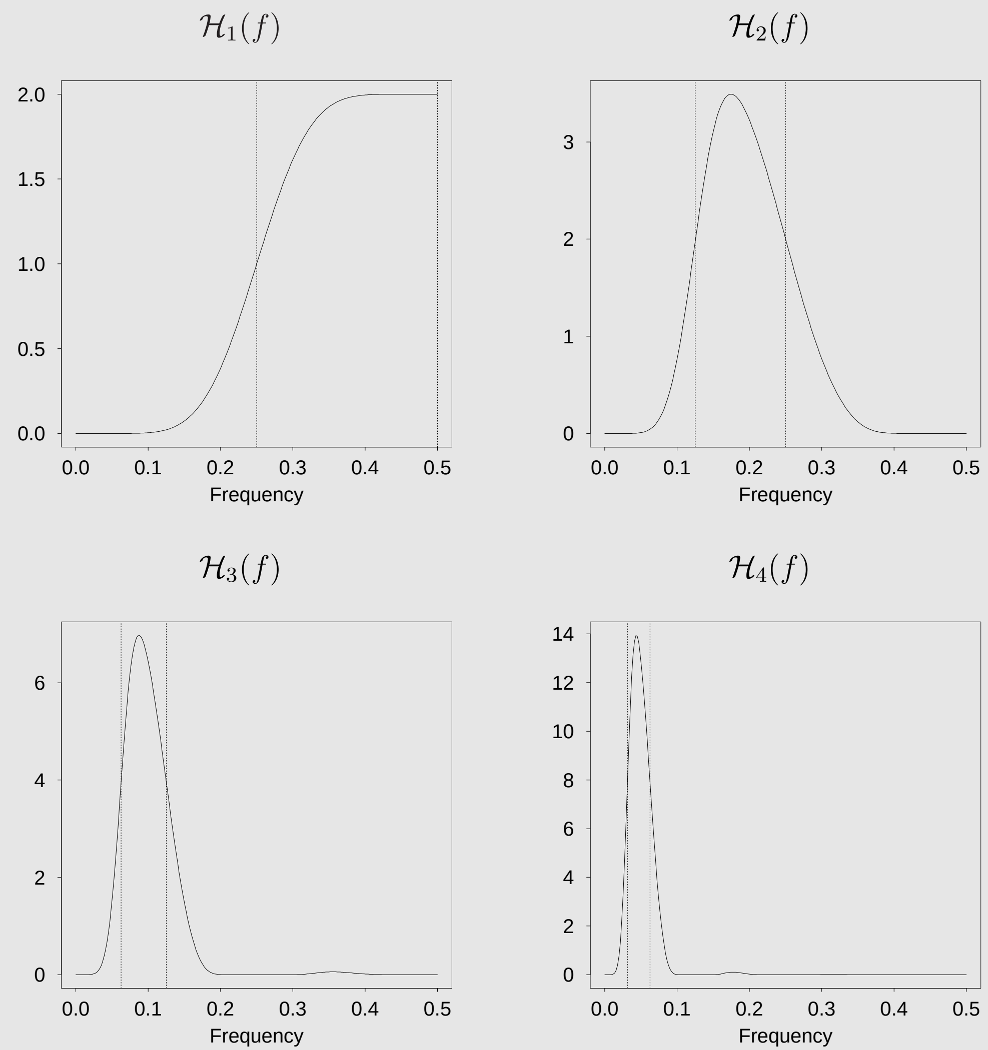


Figure 8: Frequency-domain representations of the LA(8) wavelet filter. Each plot shows the squared gain function corresponding to the wavelet coefficient vectors in Figure 7. Frequencies with positive weight $\mathcal{H}(f) > 0$ outside of the dotted lines (shaded regions) correspond to the leakage 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.

2.3 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 $||\mathbf{w}||^2 = ||\mathbf{y}||^2$.

2.4 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,

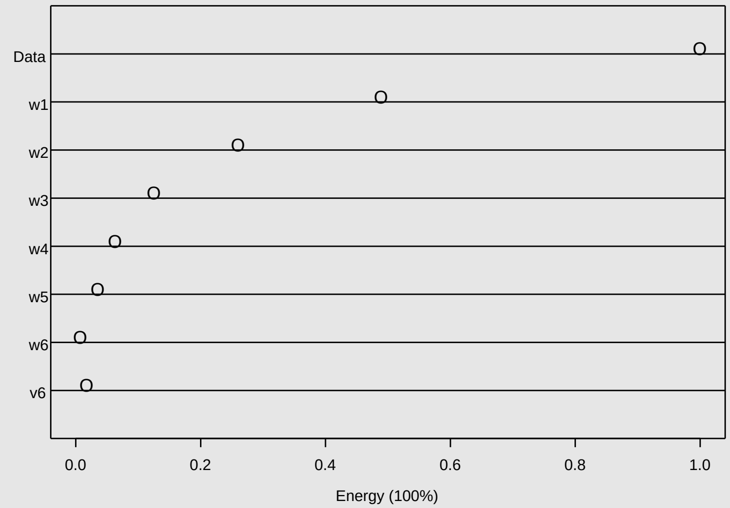
$$\|\mathbf{v}_J\|^2 / \|\mathbf{y}\|^2$$

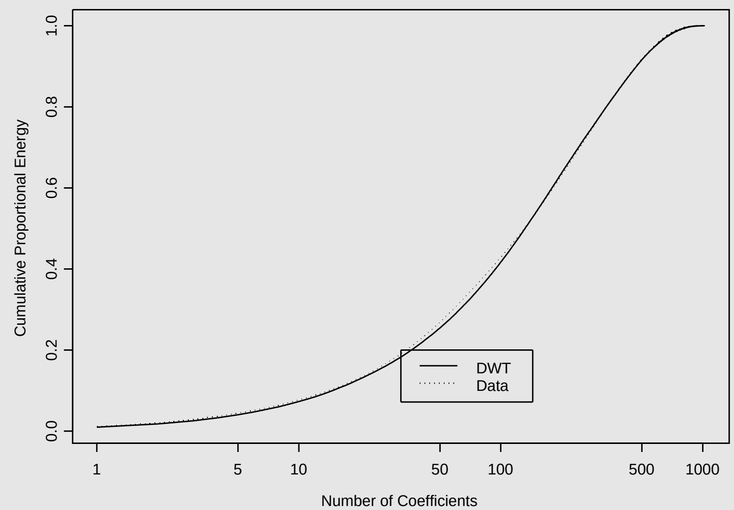
is close to zero whereas

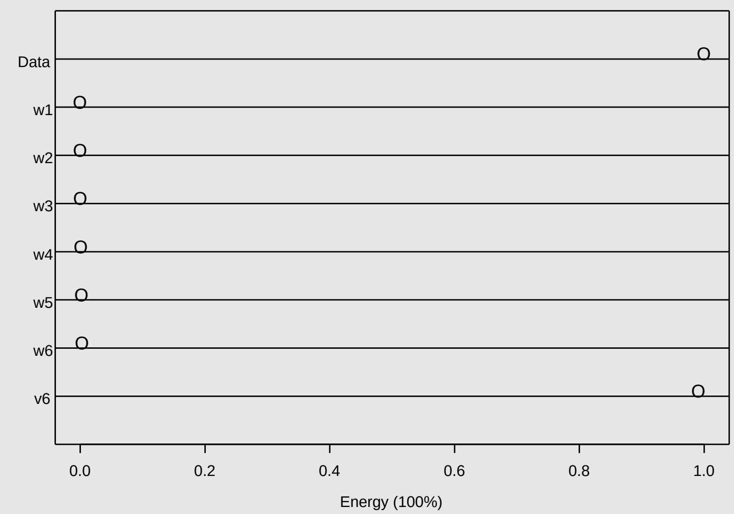
$$\|\mathbf{v}_J\|^2 / \|\mathbf{y}\|^2$$

is close to one for a unit root process.

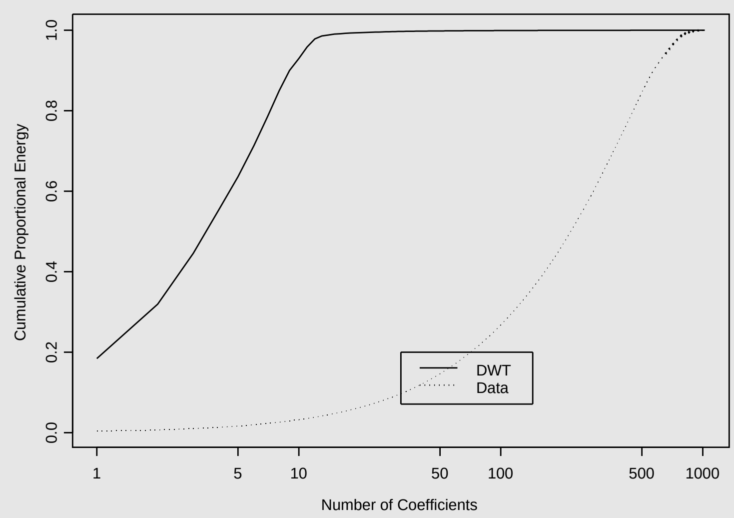
- Since a unit root process can be succinctly approximated by a few scaling coefficients and the energy of the scaling coefficients is almost equal to the total energy of the data, our statistical test for a unit root process is based on this principle of energy decomposition.


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

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

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

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

- 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 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 $L \geq 4$.

- 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.

4 Incorporating a Drift Term

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

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

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

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 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:

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

- 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 $T^*/T = o(1)$.

4.2 General filter case: unit scale decomposition

- 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

$$y_t = \alpha + \rho y_{t-1} + u_t, \tag{17}$$

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

- Note that under H_0, model (17) 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 (17) 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.[5]

[5]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:

$$y_t = \mu + \alpha t + y_t^s, \tag{18}$$

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 $\{\Delta y_t\}$.[6]

[6]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 $\left\{\widetilde{y}_t - \overline{\overline{y}}\right\}$.

- We will construct our tests based on

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

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

- The above theorem can be easily extended to (i) DWT of any finite scale J and (ii) a general trend in (18) using the detrending procedure discussed in Phillips and Xiao (1998).

5 Maximum Overlap DWT

- MODWT has been demonstrated to have advantages over DWT in several situations including the estimation of wavelet variance.[7]

- First, consider the *no intercept* case. 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:

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

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.

- Now consider the case with an *intercept*. By following the same arguments as in Section 4, we get the following lemma, the basis of our tests for H_0.

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

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

- 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 $\left\{\widetilde{W}_{t,1}^M\right\}$ and $\left\{\widetilde{V}_{t,1}^M\right\}$ 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

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

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

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

- One can easily show that under H_0, $T\widetilde{D}_{T,1}^{LM} \implies -\frac{E(\widetilde{W}_{t,1}^M)^2}{2\omega^2\int_0^1 [W_\mu(r)]^2 dr}$ and $T\widetilde{D}_{T,1}^{Ld} \implies -\frac{E(\widetilde{W}_{t,1}^d)^2}{2\omega^2\int_0^1 [V_\mu(r)]^2 dr}$.[8]

[8] 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 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:

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

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

Theorem 6.1 *Under the null hypothesis of no-cointegration,*

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

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.

6.1 No Drift Case

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

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

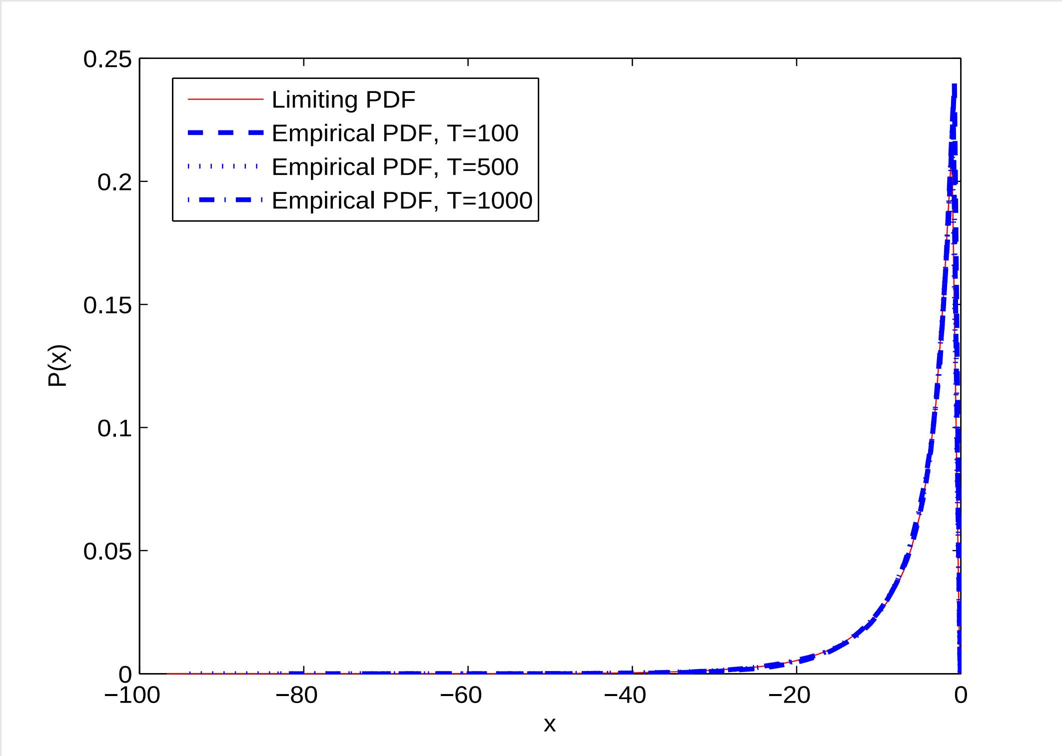


Figure 9: The limiting distribution of $-\frac{1}{\int_0^1 [W(r)]^2 dr}$ for 1 million replications. The empirical distribution of $FG_L(Haar)$ is with $T = 100$ & 500 and 1,000 observations and for 20,000 replications. The simulated data for the null distribution is generated from $y_t = y_{t-1} + u_t$ where $u_t \sim i.i.d.N(0, \sigma^2)$ where $y_0 \neq 0$.

T=100

ρ	1%	5%	10%	1%	5%	10%
		FG_M^L			ADF	
1.00	0.006	0.042	0.087	0.099	0.051	0.102
0.95	0.034	0.197	0.381	0.030	0.126	0.237
0.90	0.174	0.610	0.830	0.096	0.336	0.529
0.85	0.475	0.899	0.979	0.264	0.647	0.828
0.80	0.766	0.990	1.000	0.573	0.901	0.968

$T = 500$

ρ	1%	5%	10%	1%	5%	10%
		FG_M^L			ADF	
1.00	0.009	0.047	0.097	0.012	0.049	0.097
0.99	0.052	0.229	0.414	0.027	0.122	0.232
0.98	0.243	0.667	0.871	0.091	0.325	0.515
0.97	0.577	0.941	0.994	0.247	0.620	0.810

$T = 1,000$

ρ	1%	5%	10%	1%	5%	10%
		FG_M^L			ADF	
1.00	0.009	0.049	0.098	0.009	0.048	0.097
0.99	0.243	0.681	0.883	0.091	0.325	0.517
0.98	0.870	0.997	1.000	0.488	0.861	0.957

$T = 10,000$

ρ	1%	5%	10%	1%	5%	10%
		FG_M^L			ADF	
1.00	0.012	0.050	0.096	0.011	0.050	0.095
0.999	0.257	0.684	0.890	0.081	0.346	0.532
0.998	0.888	0.997	1.000	0.508	0.865	0.958

Table 1: Size and Power of the FG_M^L (M=1, Haar) - No Drift

6.2 Drift with i.i.d. Errors

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

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

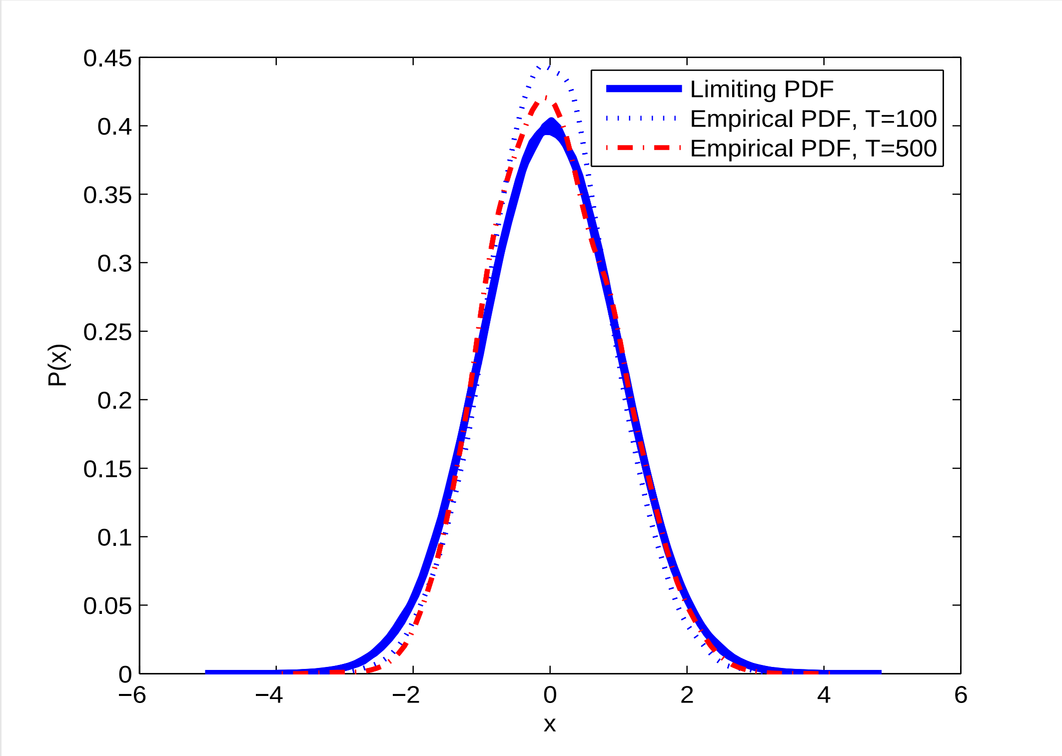


Figure 10: The limiting distribution is standard normal. The empirical distribution of $FG_L(drift)(Haar)$ is with $T = 100$ and $T = 500$ for 10,000 replications. The simulated data for the null distribution is generated from $y_t = \alpha + y_{t-1} + u_t$, $\alpha = 5$, $u_t \sim i.i.d.N(0, \sigma^2)$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

T=100

ρ	1%	5%	10%	1%	5%	10%
		$FG_M^L(drift)$			ADF	
1.00	0.005	0.041	0.072	0.011	0.052	0.108
0.99	0.205	0.255	0.295	0.005	0.051	0.104
0.95	0.416	0.497	0.542	0.015	0.084	0.164
0.90	0.681	0.747	0.787	0.049	0.194	0.334
0.85	0.825	0.870	0.894	0.128	0.409	0.606
0.80	0.897	0.922	0.935	0.302	0.671	0.837

$T = 500$

ρ	1%	5%	10%	1%	5%	10%
		$FG_M^L(drift)$			ADF	
1.00	0.005	0.042	0.073	0.012	0.054	0.103
0.99	0.234	0.302	0.349	0.016	0.085	0.160
0.98	0.506	0.587	0.632	0.050	0.187	0.332
0.97	0.690	0.767	0.797	0.119	0.395	0.575

$T = 1,000$

ρ	1%	5%	10%	1%	5%	10%
		$FG_M^L(drift)$			ADF	
1.00	0.005	0.043	0.074	0.008	0.051	0.094
0.99	0.422	0.506	0.557	0.046	0.184	0.310
0.98	0.756	0.815	0.845	0.268	0.636	0.812

Table 2: Size and Power of the $FG_M^L(drift)$ (M=1, Haar) - With Drift

The test statistic and its variance are calculated with discrete wavelet transformation. The data generating process is $y_t = \alpha + y_{t-1} + u_t$, $u_t \sim iidN(0, \sigma^2)$ where $y_0 \sim N(1,1)$. Under the alternative, $y_t = \alpha + \rho y_{t-1} + u_t$, $u_t \sim iidN(0, \sigma^2)$. For both models, $\alpha = 5$ and $\sigma^2 = \alpha^2$. The asymptotic null distribution 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. All simulations are with 2,000 replications. Under the alternative the first 2,000 observations are discarded.

6.3 Drift with AR(1) Errors

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

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

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 3: 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 $T^* = T^{0.95}$.

6.4 Drift with MA(1) Errors

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

$$y_t = \alpha + \rho y_{t-1} + u_t, \quad u_t = \epsilon_t + \theta \epsilon_{t-1}, \epsilon_t \sim iidN(0, \sigma^2)$$

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 4: 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 $T^* = T^{0.95}$.

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

The data generating process is

$$y_t = \mu + y_t^s, \text{ where } y_t^s = \rho y_{t-1}^s + u_t, u_t \sim iidN(0, \sigma^2) \text{ and } y_0 \sim N(0, \sigma^2)$$

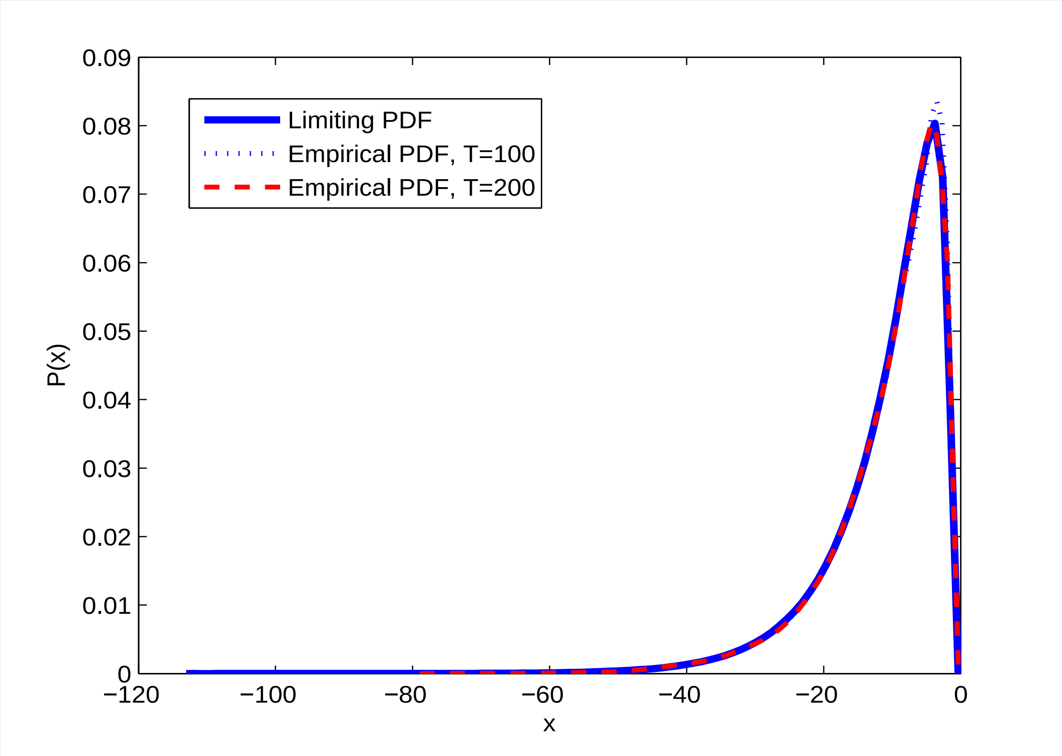


Figure 11: Limiting and Empirical Distributions of $\widehat{D}_{T,1}^{LM}$

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 $y_0 \sim N(0,1)$.

ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
		$\widehat{D}_{T,1}^{Ld}$			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 5: Size and Power of the $\widehat{D}_{T,1}^{LM}$ - 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 $y_t = \mu + y_t^s$, where $y_t^s = \rho y_{t-1}^s + u_t$, $u_t \sim iidN(0, \sigma^2)$ 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 $\widehat{D}_{T,1}^{LM}$ 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.

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

The data generating process is

$$y_t = \mu + \alpha t + y_t^s, \text{ where } y_t^s = \rho y_{t-1}^s + u_t, u_t \sim iidN(0, \sigma^2) \text{ and } y_0 \sim N(0, \sigma^2)$$

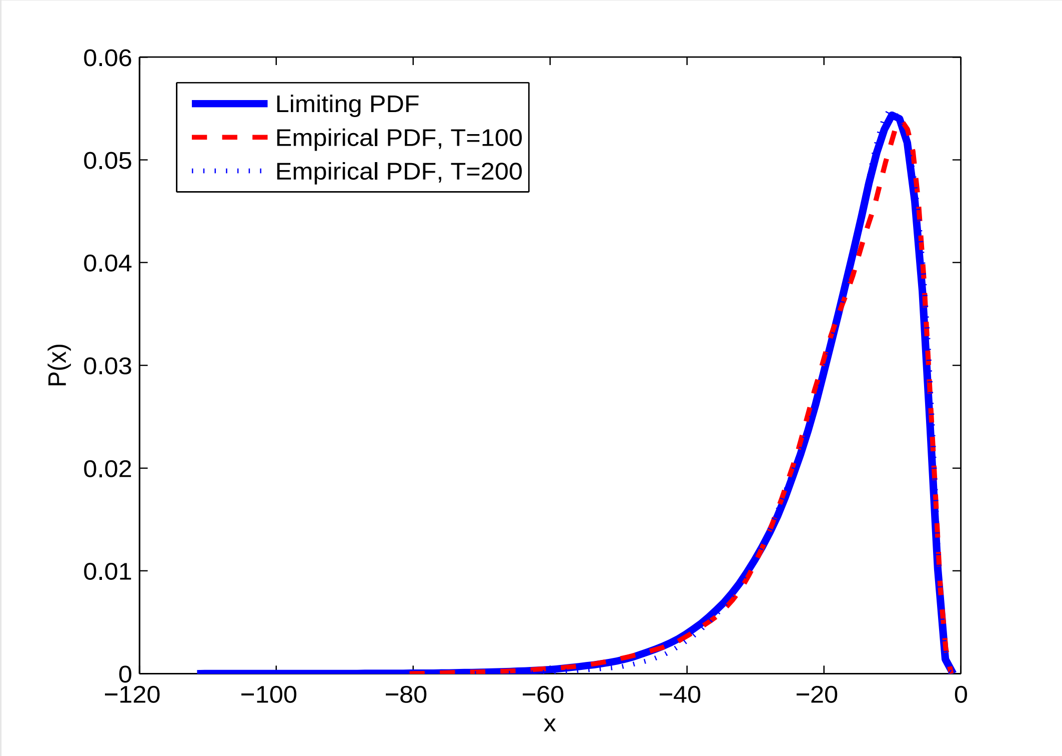


Figure 12: Limiting and Empirical Distributions of $\widehat{D}^{Ld}_{T,1}$

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

ρ	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
		$\widehat{D}_{T,1}^{Ld}$			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 6: SIZE AND POWER OF THE $\widehat{D}_{T,1}^{Ld}$ - 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 $y_t = \mu + \alpha t + y_t^s$, where $y_t^s = \rho y_{t-1}^s + u_t$, $u_t \sim iidN(0, \sigma^2)$ 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 $\widehat{D}_{T,1}^{Lt}$ 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.

	Wavelet	ERS	MPP	Observations	Dates	Frequency
SP-500 ($\widehat{D}_{T,1}^{Ld}$)	0.02	0.03	0.03	3,541	1983/01/03 - 1996/12/31	Daily
SP-500 ($\widehat{D}_{T,1}^{Ld}$)	0.15	0.24	0.26	2,292	1987/12/09 - 1996/12/31	Daily
DJIA ($\widehat{D}_{T,1}^{Ld}$)	0.25	0.43	0.55	27,567	1897/01/02 - 1996/12/31	Daily
USD-EUR ($\widehat{D}_{T,1}^{LM}$)	0.30	0.36	0.36	168	1993/01 - 2006/12	Monthly
USD-JPY ($\widehat{D}_{T,1}^{LM}$)	0.21	0.83	0.93	17,568	1996/01/01 - 1996/31/12	30-Minutes

Table 7: p-values of Unit Root Tests with Equity Index and FX 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.

7 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.

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

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.

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.

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.

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

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.

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 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.

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*.