Wavelet Power: Wavelet Energy Ratio Unit Root Tests

Mirza Trokić[a]

[a] *Department of Economics, McGill University, Montreal, QC, Canada*

Abstract

This paper uses wavelet theory to propose a frequency domain nonparametric and tuning parameter free family of unit root tests indexed by the fractional parameter d. The proposed test exploits the wavelet power spectrum of the observed series and its fractional partial sum to construct a test of the unit root based on the ratio of the resulting scaling energies. The construction takes its inspiration from the variance ratio (VR) unit root test of Nielsen (2009) and Fan and Gençay (2010) (FG). The result is a statistic whose power properties virtually mimic that of the VR statistics but which drastically reduces the severe size distortions suffered by both the VR and FG test in the presence of serially correlated MA(1) errors when the MA parameter is close to negative unity. Moreover, the test is visibly more robust to size distortions arising from lowering d than its VR counterpart and unlike the FG test, requires no estimation for construction.

Keywords: Wavelets, Wavelet energy ratio, Time series, Fractional integration, Fractional Brownian motion, Variance ratio statistic, Unit root, Hypothesis test, Size distortion, Statistical power

1. Introduction

Testing for the presence of a unit root is an important empirical problem and has a long-established history in the econometric literature. Early theoretical developments include the seminal works of Dickey and Fuller (1979), Phillips (1987), and Phillips and Perron (1988). In fact, the literature has not suffered a shortage of contributions since. In this regard, some other prominent examples include Chan and Wei (1987), Sims et al. (1990), and Park and Fuller (1995). Unfortunately, whereas many contributions focus on constructing unit root tests for models of wider generality, fewer address the issues which seem to underlie the lot of them: low statistical power, poor size properties, and tuning parameter (eg. lag length, bandwidth, kernel choice, etc.) selection. The latter is a particularly salient problem in the presence of serial correlation which gives rise to test statistics which require tuning parameter specification (thereby rendering the test's performance highly dependent on values specified) although the latter are not reflected in the limiting distributions of the test statistics.

Email address: mirza.trokic@mail.mcgill.ca (Mirza Trokić)

September 9, 2012

Recently, there have been significant efforts to improve on the shortcomings of classical unit root tests. In this regard, the Elliott et al. (1996) contribution addresses both size and power issues through point optimal tests, power envelopes, and generalized least squares (GLS) detrending of augmented Dickey-Fuller (ADF) tests. Moreover, Ng and Perron (2001) and Perron and Qu (2007) address low power properties through optimized truncation lag selection. On the other hand, issues concerning tuning parameter selection prompted the development of tuning parameter free unit root tests as in Park and Choi (1988), Park (1990), Breitung (2002), and Nielsen (2009), the latter two having the additional advantage of being non-parametric.

It is interesting to note however that, apart from a handful of exceptions, unit root tests in the literature, and in fact all those mentioned above, are constructed directly in the time domain. This of course is not surprising considering that the pioneering unit root test of Dickey and Fuller (1979) (DF) was constructed similarly. On the other hand, some 13 years before the DF unit root test, Granger (1966) had observed that the vast majority of economic series are characterized by power spectra the most noticeable characteristic of which is the "overpowering importance of the low frequency components" which are amplified by the presence of trends in mean. Yet, intriguingly, the literature has produced little in the way of frequency domain unit root tests. Nevertheless, there are two important exceptions: Choi and Phillips (1993) and Fan and Gençay (2010).

It is important to remark that, although both Choi and Phillips (1993) and Fan and Gençay (2010) construct unit root tests directly in the frequency domain, they do so using two succeeding technologies. While the former relies on Fourier spectral analysis to demonstrate finite sample superiority over various time-domain counterparts, the latter relies on wavelet theory to do the same. This distinction is an important one as Fourier transforms are localized only in frequency whereas wavelet transforms are localized in both frequency and space. This makes Fourier analysis an excellent tool for studying stationary time series and wavelet analysis ideally adapted for the study of non-stationary series, thereby rendering wavelet transforms a *de facto* natural platform for the construction of unit root tests in the frequency domain. The present paper therefore continues this tradition and contributes unit root tests constructed using wavelet theory.

Motivated by developments in Nielsen (2009) and Fan and Gençay (2010), this article constructs a general family of non-parametric tuning parameter free wavelet based tests for the unit root hypothesis. These tests are indexed by the fractional parameter d and tend in distribution to the distribution of the variance ratio statistic of Nielsen (2009). Consequently, they have good asymptotic power properties, and as pointed out by Müller (2008), can consistently discriminate between a null and alternative hypothesis of stationarity. However, where the proposed tests truly shine is in their ability to improve upon various shortcomings of existing unit root tests.

The wavelet energy ratio (WER) unit root tests in this paper are a direct improvement over the wavelet unit root tests in Fan and Gençay (2010). The latter, although asymptotically nuisance parameter free, are only so after a transformation involving the Newey and West (1987) estimator with a Bartlett kernel correction of the long term variance of the errors. Even still, these tests are not tuning parameter free as they depend on a suitably chosen kernel bandwidth parameter q. The WER tests require no such estimation steps and are tuning parameter free by design. On the other hand, like many unit root tests in the literature, the tests of Nielsen (2009) and Fan and Gençay (2010) suffer from severe size distortions, particularly when the errors follow a moving average process with a highly negative MA parameter. Whereas Nielsen (2009) achieves substantial size distortion reductions through the sieve bootstrap algorithm of Chang and Park (2003), Fan and Gençay (2010) do not even consider size distortions arising from MA errors. This is rather strange since simulations in this paper clearly show that their test have unpalatable size distortions. In contrast, the WER tests suffers from size distortions which are impressively smaller than those of Nielsen (2009). Moreover, the WER test is more robust to size distortions arising from specifying very small d values than its variance ratio test counterpart. In terms of local asymptotic power however, the WER tests virtually mimic the variance ratio tests, although the power of both is visibly smaller than that of Fan and Gençay (2010).

The remainder of the article proceeds as follows. Section 2 presents an overview of the essential aspects of wavelet theory necessary for the development of the article. Section 3 reviews fractionally integrated process and the variance ratio statistic of Nielsen (2009). Section 4 presents the wavelet energy ratio tests, while Section 5 presents simulation evidence for the power and size distortion performance of the proposed tests. All proofs are contained in the Appendix.

2. Wavelet Power Spectrum

What distinguishes wavelet techniques from more classical spectral tools such as Fourier methods is that the latter can only extract frequency information form an input signal. The former on the other hand offers both frequency and temporal information. It is precisely this feature which makes wavelets an ideal tool for *multiresolution analysis* (MRA) - the ability to analyze a signal at different frequencies with varying resolutions. Essentially, moving along the time domain, MRA allows one to *zoom* to a desired level of detail. In fact, at high (low) frequencies, MRA by design yields good (poor) time resolutions and poor (good) frequency resolutions. Since economic (financial) time series often exhibit multiscale features, wavelet techniques are a powerful mechanism for decomposing a series into constituent processes associated with different time scales. In particular, since non-stationary series exhibit summary statistics which change with time, consequently amplifying their lower frequency components relative to stationary series, exploiting this distinction yields a platform for distinguishing a series as $I(1)$ or $I(0)$. It was precisely this feature which was exploited in Fan and Gençay (2010) and which will be exploited in the sections to follow.

2.1. Wavelets

A wavelet, as the diminutive form suggests, is a small wave. Unlike its augmentative counterpart, a wavelet oscillates in a strict subset of the infinite time domain. Formally, a wavelet is a real valued function $\psi(\cdot)$ satisfying two basic properties:

$$\int_{-\infty}^{\infty} \psi(t)\, dt = 0 \qquad\qquad \int_{-\infty}^{\infty} \psi^2(t)\, dt = 1$$

The first of the two properties says that the wavelet integrates to zero and consequently, any oscillations above zero must be counterbalanced by oscillations below zero. The second property says that the wavelet has unit energy. This implies that any nonzero activity of the wavelet is limited to a strict subset of the time domain, the length of which is vanishingly small relative to the entire time domain $(-\infty, \infty)$. Finally, if $y(t)$ is a signal (a time series in this case), the continuous wavelet transform (CWT) (or filter) of $y(t)$ is given by:

$$W(a, b) = \int_{-\infty}^{\infty} y(t)\, \psi^*_{a,b}(t)\, dt$$

where $\psi_{a,b} = \frac{1}{\sqrt{a}} \psi_{a,b}\left(\frac{t-b}{a}\right)$, and $*$ denotes the complex conjugate. See Percival and Walden (2006) for a detailed exposition.

2.2. Discrete Wavelet Transform

The CWT is not appropriate when dealing with empirical data however as time series are rarely given as continuous functions. Fortunately, the fundamental properties of the CWT also have discrete analogues. In this case, denote by $h = (h_0, \ldots, h_{L-1})$ a discrete wavelet (or high pass) filter of length $L - 1$. Similarly, let $g = (g_0, \ldots, g_{L-1})$ denote a discrete scaling (or low pass) filter of equal length. In this case, integration to zero and unit energy are satisfied by:

$$\sum_{l=0}^{L-1} h_l = 0 \qquad\qquad \sum_{l=0}^{L-1} h_l^2 = 1$$

The discrete wavelet transform (DWT) is an orthonormal transform. Thus, in addition to the two properties above, the wavelet filter satisfies the following orthogonality condition:

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0 \qquad \forall n \in \mathbf{Z}^+$$

In other words, the high pass filter is orthogonal to its even shifts. In fact, a similar set of conditions hold for the scaling filter g as well, namely:

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

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

for all positive integers n. Thus, the low pass filter is orthogonal to its even shifts, is orthogonal to even shifts in h, and has unit every. Furthermore, the relationship between the coefficients of the scaling and wavelet filters are determined through the *quadrature mirror relationship* which establishes that:

$$h_l = (-1)^l g_{L-1-l} \qquad g_l = (-1)^{l+1} h_{L-1-l} \qquad l = 0, \ldots, L-1$$

Filtering an observed series $\{y_t\}_{t=1}^{T}$ with the high pass filter h and the low pass filter g, yields a DWT of the original data. In turn, this transform yields two series - the first capturing the high frequency behaviour of y_t, the second capturing its low frequency behaviour. In fact, the entire process can be represented neatly in vector notation.

Assume that $T = 2^M$ and consider a series $\mathbf{y} = \{y_t\}_{t=1}^{T}$.[1] Then, denote by $\mathbf{W} = [\mathbf{W}_1, \ldots, \mathbf{W}_M, \mathbf{V}_M]^\top$ the matrix of DWT coefficients. Here, $\mathbf{W}_j$ is a vector of wavelet coefficients of length $T/2^j$ and is associated with changes on a scale of length $\lambda_j = 2^{j-1}$. Moreover, $\mathbf{V}_M$ is a vector of scaling coefficients of length $T/2^j$ and is associated with averages on a scale of length $\lambda_M = 2^{M-1}$. Then, $\mathbf{W}$ can be obtained by $\mathbf{W} = \mathcal{W}\mathbf{y}$ where $\mathcal{W}$ is an $T \times T$ orthonormal matrix which generates the DWT coefficients.

2.3. Pyramid Algorithm

In practice, DWT coefficients are derived through the pyramid algorithm of Mallat (1989). Formally, if $\mathbf{W}_j = \left(W_{1,1} \ldots W_{T/2^j,j}\right)^\top$ and $\mathbf{V}_j = \left(V_{1,1} \ldots V_{T/2^j,j}\right)^\top$, the j^{th} iteration of the algorithm then convolves an input signal with filters h and g respectively to derive the j^{th} level DWT matrix $[\mathbf{W}_1, \ldots \mathbf{W}_j, \mathbf{V}_j]^\top$. Explicitly, the convolution process takes the following form:

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

$$W_{t,j} = \sum_{l=0}^{L-1} h_l V_{2t-l \bmod T, j-1} \qquad V_{t,j} = \sum_{l=0}^{L-1} g_l V_{2t-l \bmod T, j-1} \qquad j = 2, \ldots, M$$

where $t = 1, \ldots, T/2^j$. In other words, the first iteration of the algorithm convolves the data series y_t with both the high pass and the low pass filters respectively. Each subsequent iteration however takes the scaling coefficients from the preceding step, namely $V_{t,j-1}$ as it's input signal, and convolves them with h and g respectively. The entire algorithm continues until the M^{th} iteration although it can be stopped at any earlier point.

2.4. Energy Decomposition

The orthonormality of the DWT generating matrix $\mathcal{W}$ has some important implications; the first is that $\mathcal{W} \times \mathcal{W} = I_T$, where I_T is an identity matrix of dimension T. A much more

[1] Requiring series to have dyadic length is certainly restrictive. Luckily, methods such as the *discrete wavelet packet transform* (DWPT) and the *lifting scheme* are designed to overcome this shortcoming.

important implication is that $\|\mathbf{y}\|^2 = \|\mathbf{W}\|^2$. To see this, recall that $\mathbf{y} = \mathcal{W}^\top \mathbf{W}$ and that $\|\mathbf{y}\|^2 = \mathbf{y}^\top \mathbf{y}$. In other words, the DWT is an energy (variance) preserving transformation. Furthermore, coupled with this preservation of energy is the decomposition of energy on a scale by scale basis. This is another consequence of the orthonormality of the DWT and is formalized as follows:

$$\|\mathbf{y}\|^2 = \sum_{j=1}^{M} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_M\|^2 \tag{1}$$

where $\|\mathbf{W}_j\|^2 = \sum_{t=t}^{T/2^j} W_{t,j}^2$ and $\|\mathbf{V}_M\|^2 = \sum_{t=t}^{T/2^M} V_{t,M}^2$. Thus, $\|\mathbf{W}_j\|^2$ represents the amount of energy of y_t accounted for at scale λ_j. Moreover, $\|\mathbf{W}_j\|^2/T$ is the contribution to the sample variance of y_t associated with scale λ_j. This decomposition is often referred to as the *wavelet power spectrum*, and is arguably the most insightful of the properties of the DWT for the exposition in Section 4.

3. Variance Ratio Tests

This section formalizes the variance ratio statistic of Nielsen (2009). Since the latter is constructed as ratio of a series and it's fractional partial sum, it is useful to briefly overview fractional processes first.

3.1. Fractional Processes

Recall that a general fractional process x_t of order d is defined as

$$(1-L)^d x_t = u_t, \quad u_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad t = 1, 2, \dots$$

where $d > -1/2$, ϵ_t are zero-mean, finite variance, IID random variables, and $(1-L)^d$ is defined by the Maclaurin series:

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)} L^j$$

In empirical work however, one does not observe the values of a series for $t \leq 0$. Thus, the theory to follow assumes that only terms with a positive time index are of interest. It can be shown (see Appendix 2 of Wang et al. (2002)) that when t is restricted to positive integers, x_t above reduces to:

$$x_t = (1-L)_+^{-d} u_t \equiv \Delta_+^{-d} u_t = \sum_{k=0}^{t-1} c_k^{(d)} u_{t-k}$$

where

$$c_0^{(0)} = 1, \quad c_k^{(0)} = 0, \quad k \geq 1, \quad c_k^{(d)} = \frac{\Gamma(d+k)}{\Gamma(d)\Gamma(k+1)}, \quad k \geq 0$$

As usual let $W(t)$ denote a standard Wiener process, denote by $B(t)$ the standard Brownian motion, and recall that a type II[2] fractional Brownian motion $B_d(t)$ for $d > -1/2$ is defined as:

$$B_d(t) = \int_0^t (t-s)^{d-1} dW(s), \quad B_d(0) = 0, \quad 0 \leq t \leq 1 \tag{2}$$

3.2. Fractional Variance Ratio Tests

Consider an univariate time series y_t and assume its DGP is the familiar autoregressive $AR(1)$ model $y_t = \phi y_{t-1} + u_t, \quad y_0 = 0$. The classical unit root hypothesis for y_t then posits that:

$$H_0: \ \phi = 1, \quad H_1: \ |\phi| < 1 \tag{3}$$

The Nielsen (2009) variance ratio (VR) statistic for testing H_0 vs. H_1 is a non-parametric test formed by generating an ancillary fractionally differenced series $\tilde{y}_t = \Delta^{-d} y_t$ and constructing the following scaled ratio of sample variances:

$$\rho(d) = T^{2d} \frac{\sum_{t=1}^T y_t^2}{\sum_{t=1}^T \tilde{y}_t^2}$$

The statistic $\rho(d)$ above is a generalization of the classical VR test statistics, is indexed by d and has the very desirable property of neither requiring the estimation of the long-run variance of y_t nor the short term serial correlation parameters in case the u_t exhibits autocorrelation (formalizations will be introduced in Section 4). Furthermore, if $\Rightarrow$ denotes weak convergence in $D[0,1]$, the asymptotic distribution of $\rho(d)$ is summarized below:

$$\rho(d) \Rightarrow \frac{\int_0^1 B^2(s)ds}{\frac{1}{\Gamma^2(d+1)} \int_0^1 B_{d+1}^2(s)ds}$$

Note that like the statistic itself, the asymptotic distribution is indexed by d and therefore this indexing parameter is not considered to be a tuning parameter. More importantly, as argued in Müller (2008), statistics akin the one considered above consistently discriminate between a null hypothesis of a unit root and an alternative hypothesis of stationarity. For further a detailed discussion on the performance of these statistics see Nielsen (2009).

[2]For a good reference on the differences between type I and type II fractional Brownian motions, see Davidson and Hashimzade (2009).

4. Wavelet Energy Ratio Tests

This section introduces a powerful new (spectral) test of the classical unit root hypothesis. Like Fan and Gençay (2010), the new approach exploits the wavelet power spectrum to construct a family of powerful, non-parametric unit root tests. Where these tests diverge however is in the mechanism by which power against H_1 is gained. Whereas the old test achieves power by relativizing the energy of the scaling coefficients to that of total energy, the new test derives its inspiration from the VR unit root test of Nielsen (2009) in that it gains power by exploiting the relative energy of the scaling coefficients of the original series to that of its fractionally differenced transform. The result is a new family of tuning parameter free, non-parametric unit root tests indexed by the fractional parameter d, which shall be referred to as *wavelet energy ratio* (WER) unit root tests.

4.1. Model Outline

Consider again the univariate $AR(1)$ model:

$$y_t = \phi y_{t-1} + u_t \tag{4}$$

$$u_t = \Psi(L)\epsilon_t, \quad \Psi(z) \equiv \sum_{j=0}^{\infty} \psi_j z^j \tag{5}$$

equations (4) and (5) describe a very general model capable of generating stationary and non-stationary series both with and without serial correlation. Because of this generality, in order to make the theory manageable, the following assumptions will be imposed.

Assumptions.

(a) *$\{\epsilon_t, \mathcal{F}_t\}$ is a MDS with respect to some filtration $\mathcal{F}_t$.*

(b) *$E\{\epsilon_t^2\} = \sigma^2 < \infty$ and $E\{|\epsilon_0|^{2/(2d+1)}\} < \infty$ for $d > -1/2$.*

(c) *$\sum_{j=0}^{\infty} |\psi_j| < \infty$, $\sum_{j=0}^{\infty} j|\psi_j| < \infty$, and $b_\psi = \sum_{j=0}^{\infty} \psi_j \neq 0$.*

Assumptions (a) through (c) establish regularity conditions for the error terms. Apart from convenience, some are required to invoke (fractional) functional central limit theorems when establishing limiting distributions. Particularly important here is the second condition in Assumption (b). This condition is necessitated by the FCLT for fractional processes and has been a standard requirement in the literature on limiting results for fractionally integrated processes since Gourieroux and Akonom (1988).

Finally, consider the fractionally differenced series:

$$\tilde{y}_t = (1 - L)_+^{-d} \equiv \Delta_+^{-d} y_t \tag{6}$$

and recall that under Assumptions (a) through (c) the following (fractional) FCLT results hold as $T \to \infty$:

$$T^{-1/2} y_{\lfloor Tr \rfloor} \Rightarrow b_\psi \sigma W(r)$$
$$T^{-(d+1/2)} \tilde{y}_{\lfloor Tr \rfloor} \Rightarrow \frac{b_\psi \sigma}{\Gamma(d+1)} B_{d+1}(r)$$

where $\lfloor \cdot \rfloor$ denotes the floor function, $0 \leq r \leq 1$, $(b_\psi \sigma)^2$ is the long-term variance of y_t, and $d > 1/2$. For a detailed exposition on this and several other limiting results for general fractionally integrated processes, see Wang et al. (2002).

4.2. New Spectral Unit Root Tests

To motivate the derivation of the new spectral unit root test, recall that the wavelet power spectrum isolates contributions to the sample variance of $\{y\}_{t=1}^{T}$ associated with scales λ_j. This implies that the proportion of total energy generated on a particular scaling level J is $\|\mathbf{V}_J\|^2/\|\mathbf{y}\|^2$. To see just why this is so insightful, let $T = 2^{10} = 1024$ and consider in Figure 1 the plots of the level 6 DWT energy decomposition of four series: a standard Gaussian white noise process z_t, an $AR(1)$ unit root process $y_t = y_{t-1} + z_t$, the fractionally differenced series $\tilde{z}_t = \Delta_+^{-d} z_t$, and the fractionally differenced series $\tilde{y} = \Delta_+^{-d} y_t$.

Consider now the scaling energy ratio $\|\mathbf{V}_6\|^2/\|\widetilde{\mathbf{V}}_6\|^2$. If the model in equations (4) and (5) contains a unit root, then $\phi = 1$, both $\|\mathbf{V}_6\|^2$ and $\|\widetilde{\mathbf{V}}_6\|^2$ are close to one, but $\|\widetilde{\mathbf{V}}_6\|^2 < \|\mathbf{V}_6\|^2$. On the other extreme, if the model in question is white noise, then $\|\mathbf{V}_6\|^2 \approx \|\widetilde{\mathbf{V}}_6\|^2 \approx 0$. In the example above, this implies that $\|\mathbf{V}_6\|^2/\|\widetilde{\mathbf{V}}_6\|^2\Big|_{H_0} > 1$ whereas $\|\mathbf{V}_6\|^2/\|\widetilde{\mathbf{V}}_6\|^2\Big|_{H_1} \approx 1$. In other words, the test has power. The following subsection formalizes this intuition in the case of a Haar wavelet filter. The one immediately following it does the same for a Daubechies (1992) compactly supported wavelet filter.

4.3. WER: Haar Wavelet Filter

Recall that the wavelet and scaling coefficients of the unit scale Haar DWT transform of a series $\{y_t\}_{t=1}^{T}$ (where T is of dyadic length), are defined respectively as:

$$W_{t,1} = \frac{1}{\sqrt{2}} \left(y_{2t} - y_{2t-1} \right), \qquad t = 1, 2, \ldots, T/2 \tag{7}$$

$$V_{t,1} = \frac{1}{\sqrt{2}} \left(y_{2t} + y_{2t-1} \right), \qquad t = 1, 2, \ldots, T/2 \tag{8}$$

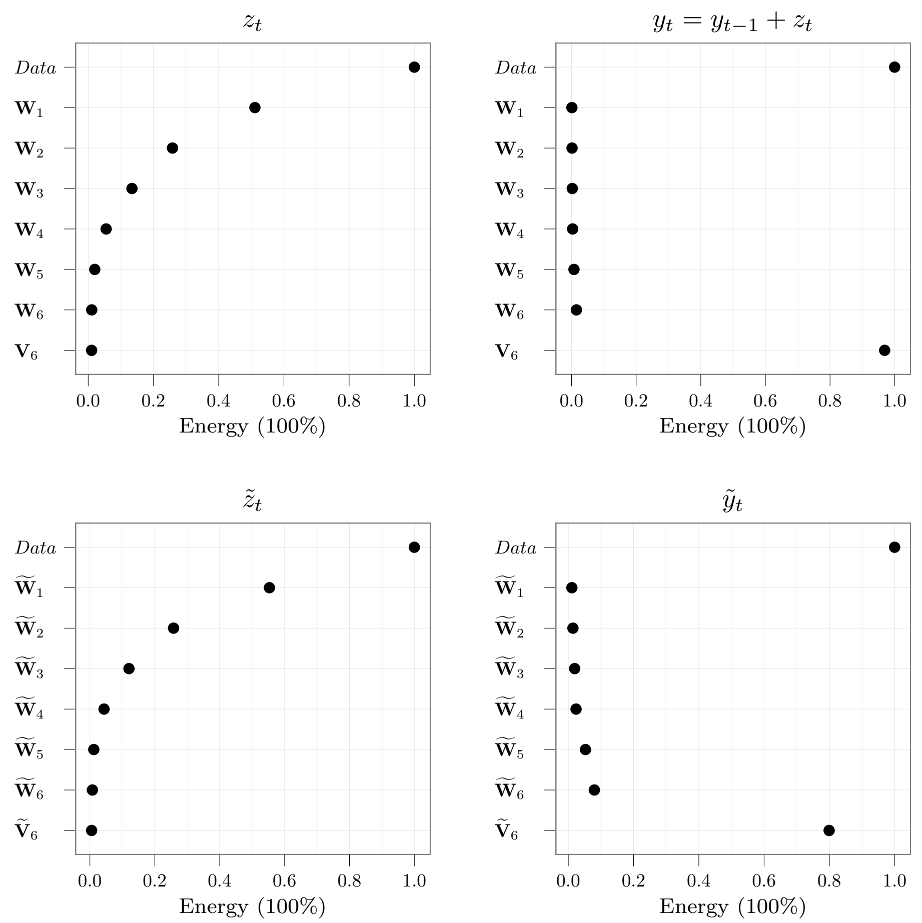

Figure 1: Haar wavelet filter Level 6 DWT energy decomposition of z_t, $y_t = y_{t-1} + z_t$, $\tilde{z}_t = \Delta_+^{-d} z_t$, and $\tilde{y} = \Delta_+^{-d} y_t$. Following the conclusions in Nielsen (2009), the fractional parameter in the exposition is $d = 0.10$ although any $d > -1/2$ is valid. Moreover, to allow for easy comparison with Fan and Gençay (2010), $J = 6$, although the sample size allows for any $J \leq 10$. Finally, note that each point represents the proportion of total energy accounted for by wavelet and scaling parameters on scales λ_1 through λ_6, where *Data* represents the total energy of the input signal.

Similarly, define the Haar DWT coefficients for the fractionally differenced series $\{\tilde{y}_t\}_{t=1}^{T}$ as follows:

$$\widetilde{W}_{t,1} = \frac{1}{\sqrt{2}} \left(\tilde{y}_{2t} - \tilde{y}_{2t-1} \right), \qquad t = 1, 2, \ldots, T/2 \tag{9}$$

$$\widetilde{V}_{t,1} = \frac{1}{\sqrt{2}} \left(\tilde{y}_{2t} + \tilde{y}_{2t-1} \right), \qquad t = 1, 2, \ldots, T/2 \tag{10}$$

Above, the wavelet coefficients $W_{t,1}$ and $\widetilde{W}_{t,1}$ extract the behaviour of y_t and $\tilde{y}_t$ respectively, in the high frequency band $[1/2, 1]$, whereas the scaling coefficients $V_{t,1}$ and $\widetilde{V}_{t,1}$ do the same for the low-frequency band $[0, 1/2]$. Section 4.2 heuristically argued that a ratio of energies of scaling coefficients of the original series and it's fractional transform has power to distinguish whether a series possesses a unit root. Nevertheless, using a naive ratio of the proposed energies is not appropriate unless it is suitably scaled. Lemma 1 formalizes this observation and establishes that the appropriate scaling factor is in fact T^{-2d}. This of course should not be too surprising considering that this paper in part derives its inspiration from the variance ratio statistic of Nielsen (2009).

Lemma 1. *Under the null hypothesis in equation* (3) *and Assumptions (a) through (c), for any $d > -1/2$ the following holds:*

1. $V_{t,1}^2$ *is* $O_p\left(T^2\right)$
2. $\widetilde{V}_{t,1}^2$ *is* $O_p\left(T^{2(1+d)}\right)$.

It follows from Lemma 1 that the presence of a unit root can therefore be tested using the statistic defined below:

$$\hat{\rho}_V^H(d) = T^{2d}\frac{\|\mathbf{V}_1\|^2}{\|\widetilde{\mathbf{V}}_1\|^2} = T^{2d}\frac{\sum_{t=1}^{T/2} V_{t,1}^2}{\sum_{t=1}^{T/2} \widetilde{V}_{t,1}^2} \tag{11}$$

The superscript H in $\hat{\rho}_V^H(d)$ stands as a reminder that the wavelets used in the construction above are those based on the Haar DWT filter and to distinguish it from other constructions to be considered later on.

The primary result of this section is the limiting distribution of the WER statistic which, as is to be expected is a ratio of functionals of the standard and fractional Brownian motions, respectively. Theorem 1 formalizes this result.

Theorem 1. *Provided Assumptions (a) through (c) hold, under H_0, $\hat{\rho}_V^H(d)$ is characterized by the following limiting distribution result.*

$$\hat{\rho}_V^H(d) \Rightarrow \rho_V^H(d) = \frac{\int_0^1 W^2(r)dr}{\frac{1}{\Gamma^2(d+1)}\int_0^1 B_{d+1}^2(r)dr} \tag{12}$$

Theorem 1 makes it clear that $\rho_V^H(d)$ generates a family of limiting distributions of $\hat{\rho}_V^d(d)$ indexed by the fractional parameter d. In other words, the family of tests described by $\hat{\rho}_V^H(d)$ are tuning parameter free. Here it is of importance to note that this is an inherent property of the nature of the statistic. Contrast this with Fan and Gençay (2010) who propose the statistic $\widehat{FG}_1 = T\frac{\hat{\lambda}_v^2}{\hat{\gamma}_0}\left[\hat{S}_{T,1} - 1\right]$ where $\hat{S}_{T,1} = \frac{\|\mathbf{V}_1\|^2}{\|\mathbf{W}_1\|^2 + \|\mathbf{V}_1\|^2}$, $\hat{\lambda}_v^2 = 4\hat{\omega}^2$, $\hat{\omega}^2$ is a consistent estimate of the long-run variance, ω^2, of $\{u\}_{t=1}^T$, and $\hat{\gamma}_0^2$ is a consistent estimate of $\gamma_0 = E\left\{u_{2t}^2\right\}$. Since neither ω^2 nor γ_0^2 are reflected in the limiting distribution of $\widehat{FG}_1$, the latter cannot be tuning parameter free.

4.4. WER: Daubechies Compactly Supported Wavelet Filter

The Daubechies wavelet filters define DWTs indexed by the maximal number of vanishing moment conditions for a given support. Using the notation in this paper, such wavelet filters are often denoted as DL, where L denotes the number of coefficients (length of the wavelet filter) and $L_1 = L/2$ is the number of vanishing moment conditions. The latter represents the maximal order of the polynomial behaviour which can be extracted from an input signal. For example, the Haar wavelet filter is in fact the $D2$ filter and can only capture constant signal components. Clearly, if richer component behaviour is desired, a more general approach is necessary. To this end, recall that the boundary-independent (BI) unit scale wavelet and scaling coefficients of the Daubechies compactly supported wavelet filter are defined as:

$$W_{t,1} = \sum_{l=0}^{L-1} h_l y_{2t-l} \qquad V_{t,1} = \sum_{l=0}^{L-1} g_l y_{2t-l}, \qquad t = L_1, L_1+1, \ldots, T/2 \tag{13}$$

Similarly, define the above for the fractionally differenced series $\{\tilde{y}_t\}_{t=1}^{T}$ as follows:

$$\widetilde{W}_{t,1} = \sum_{l=0}^{L-1} h_l \tilde{y}_{2t-l} \qquad \widetilde{V}_{t,1} = \sum_{l=0}^{L-1} g_l \tilde{y}_{2t-l}, \qquad t = L_1, L_1+1, \ldots, T/2 \tag{14}$$

As before, first level wavelet coefficients extract behaviour of an input in the high frequency band, whereas first level scaling coefficients do so for the low-frequency band. Since the Haar DWT is a special case of the Daubechies DWT, the intuition from Sections 4.2 and 4.3 carry over. Adapting equation (11) to the general Daubechies framework, the following form of the WER statistic emerges:

$$\hat{\rho}_V^L(d) = T^{2d} \frac{\|\mathbf{V}_1\|^2}{\|\widetilde{\mathbf{V}}_1\|^2} = T^{2d} \frac{\sum_{t=L_1}^{T/2} V_{t,1}^2}{\sum_{t=L_1}^{T/2} \widetilde{V}_{t,1}^2} \tag{15}$$

The following theorem justifies the scaling factor T^{2d} in equation (15) and establishes the limiting distribution of the WER statistic.

Theorem 2. *Under the null hypothesis equation* (3)*, Assumptions (a) through (c), and any* $d > -1/2$

$$\frac{V_{t,1}^2}{\widetilde{V}_{t,1}^2} = O_p\left(T^{2d}\right)$$

and

$$\hat{\rho}_V^L(d) \Rightarrow \rho_V^L(d) = \frac{\int_0^1 W^2(r)dr}{\frac{1}{\Gamma^2(d+1)} \int_0^1 B_{d+1}^2(r)dr} \tag{16}$$

There a few things to note here. First, as in the case of the Haar wavelet filter, inherent in the construction of the $\hat{\rho}_V^L(d)$ is the fact that the statistic is tuning parameter free. This should again be contrasted with Fan and Gençay (2010) whose proposed statistic $\widehat{FG}_1^L$ is not tuning parameter free as it requires consistent estimates of the long-run variance of $\{u\}_{t=1}^T$ and $E\left\{u_{2t}^2\right\}$, although these parameters are not reflected in the asymptotic distribution. Second, Fan and Gençay (2010) argue in favour of power gains as L increases since the approximation of the Daubechies wavelet filter approaches the ideal high-pass filter as L grows. Since the $\hat{\rho}_V^L(d)$ statistics proposed in this paper depend only indirectly on the high-pass wavelet coefficients, increasing L should produce only slight power gains.

4.5. Detrended WER Tests

Thus far the discussion has been limited to a simple zero mean $AR(1)$ DGP with no trend. Clearly this can be very restrictive in practical work and a richer model is needed. This section therefore adapts the WER test to a general non-zero mean $AR(1)$ DGP with a linear trend. As is the norm with unit root tests in the time domain, the simplest way to deal with such models is to introduce some sort of detrending procedure. Here the focus will be on the simplest such procedure, namely ordinary least squares (OLS) detrending, although more advanced techniques such as GLS detrending are possible as well. Moreover, since the Haar wavelet filter is a special case of the Daubechies wavelet filter, the latter more general setup will be considered here.

The model under consideration here is of the following form:

$$y_t^{(i)} = \delta_t^{(i)}\gamma^{(i)} + q_t, \qquad i = 0, 1, 2 \tag{17}$$

$$q_t = \phi q_{t-1} + u_t \tag{18}$$

where $u_t = \Psi(L)\epsilon_t$, $\Psi(z) \equiv \sum_{j=0}^{\infty}\psi_j z^j$, $\delta_t^{(0)} = \gamma^{(0)} = 0$ when $i = 0$, $\delta_t^{(1)} = 1$ and $\gamma^{(1)} = \gamma_0$ when $i = 1$, and when $i = 2$, define $\delta_t^{(2)}$ as the 1×2 vector $[1, t]$ and $\gamma^{(2)}$ as the 2×1 vector $[\gamma_0, \gamma_1]^\top$. Thus, when $i = 0$ the above model reduces to that looked at in Section 4.1.

For $i = 1, 2$, let $\hat{\gamma}^{(i)}$ denote the OLS estimator of $\gamma^{(i)}$ from regression equation (17) or its fractional transform $\tilde{y}_t^{(i)} = \delta_t^{(i)}\gamma^{(i)} + \tilde{q}_t$. The residuals from said regressions can then be expressed as:

$$\hat{y}_t^{(i)} = y_t^{(i)} - \delta_t^{(i)}\hat{\gamma}^{(i)} \tag{19}$$

$$\hat{\tilde{y}}_t^{(i)} = \tilde{y}_t^{(i))} - \delta_t^{(i)}\hat{\gamma}^{(i)} \tag{20}$$

Equations (19) and (20) above say that $\hat{y}_t^{(i)}$ and $\hat{\tilde{y}}_t^{(i)}$ are in fact the detrended versions of $y_t^{(i)}$ and $\tilde{y}_t^{(i)}$ respectively. Thus, in accounting for detrending, introduce a modified WER statistic presented below:

$$\hat{\rho}_V^{L,(i)}(d) = T^{2d}\frac{\|\widehat{\mathbf{V}}_1^{(i)}\|^2}{\|\widehat{\widehat{\mathbf{V}}}_1^{(i)}\|^2} = T^{2d}\frac{\sum_{t=L_1}^{T/2}\widehat{V}_{t,1}^{(i)^2}}{\sum_{t=L_1}^{T/2}\widehat{\widehat{V}}_{t,1}^{(i)^2}} \tag{21}$$

where $i = 0, 1, 2$ and

$$\widehat{W}_{t,1}^{(i)} = \sum_{l=0}^{L-1} h_l \hat{y}_{2t-l}^{(i)} \qquad \widehat{V}_{t,1}^{(i)} = \sum_{l=0}^{L-1} g_l \hat{y}_{2t-l}^{(i)}$$

$$\widehat{\widehat{W}}_{t,1}^{(i)} = \sum_{l=0}^{L-1} h_l \hat{\hat{y}}_{2t-l}^{(i)} \qquad \widehat{\widehat{V}}_{t,1}^{(i)} = \sum_{l=0}^{L-1} g_l \hat{\hat{y}}_{2t-l}^{(i)}$$

The following theorem is the main result of this section and establishes the limiting distribution of $\hat{\rho}_V^{L,(i)}(d)$.

Theorem 3. *Let y_t be generated as in equations* (17) *and* (18)*. Suppose Assumption (a) through (c) hold and the null hypothesis in equation* (3) *is in effect. Then, for any $d > -1/2$ and $i = 0, 1, 2$*

$$\hat{\rho}_V^{L,(i)}(d) \Rightarrow \rho_V^{L,(i)}(d) = \frac{\int_0^1 B^{(i)}(t)^2 dt}{\frac{1}{\Gamma^2(d+1)}\int_0^1 B_{d+1}^{(i)}(t)^2 dt} \tag{22}$$

where

$$B^{(i)}(t) = B(t)^2 - \delta^{(i)}(t)\left(\int_0^1 \delta^{(i)}(s)^\top \delta^{(i)}(s)ds\right)^{-1}\left(\int_0^1 \delta^{(i)}(s)^\top B(s)ds\right)$$

$$B_{d+1}^{(i)}(t) = B_{d+1}(t)^2 - \delta^{(i)}(t)\left(\int_0^1 \delta^{(i)}(s)^\top \delta^{(i)}(s)ds\right)^{-1}\left(\int_0^1 \delta^{(i)}(s)^\top B_{d+1}(s)ds\right)$$

and

$$\delta^{(1)}(s) = 1 \quad \textit{and} \quad \delta^{(2)}(s) = [1, s]^\top$$

4.6. Asymptotic Local Power Analysis

Since the parameter which indexes the WER family of statistics is d, it is natural to ask whether there exists such a d which maximizes power for said family? Simulation analysis in Nielsen (2009) addresses the same question and suggests that the family-wise "power maximizing" choice of d for the variance ratio test should be $d = 0.1$. Although the maximizing value here is not a maximum in the theoretical sense as choices of $d < 0.1$ yield uniformly (in c) higher asymptotic local power, this choice is guided by the fact that choosing d too small can result in size distortions. Figure 2 paints a similar conclusion in the case of the WER statistic.

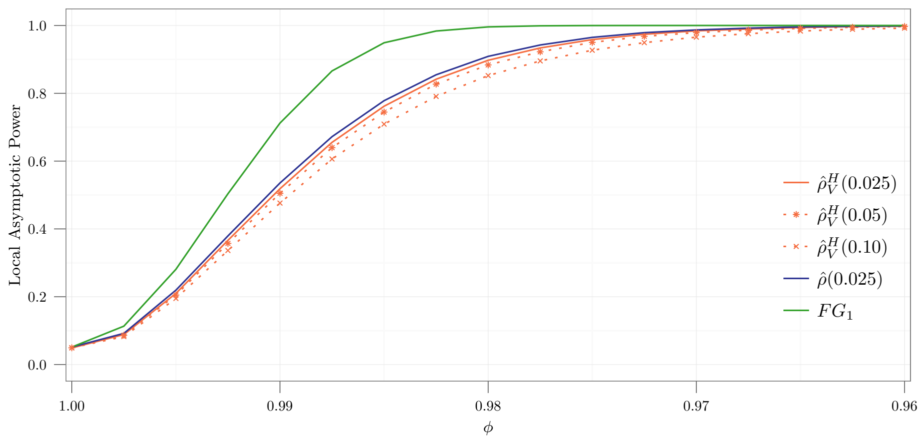

Figure 2: Local Asymptotic Power for $\hat{\rho}_V^H(d)$, $\hat{\rho}(d)$ and FG_1. Each power curve is derived for a sample of size $T = 1,000$ over $50,000$ Monte Carlo replications. The critical values were computed for the same sample size but over $100,000$ MC replications.

To formalize matters, note first that the variance ratio test rejects for large values of $\hat{\rho}(d)$ and recall that Theorem 2 in Nielsen (2009) establishes the rejection region and consistency of this test. Lemma 2 derives the same in the case of the WER statistic.

Lemma 2. *Under the assumptions of Theorem 3, if α denotes the significance level, $\rho_V^{L,(i)}(d)$ then rejects H_0 in equation (3) whenever $\rho_V^{L,(i)}(d) > \xi_{i,\alpha}$, where $\xi_{i,\alpha}$ is the critical value obtained from*

$$P\left\{\rho_V^{L,(i)}(d) > \xi_{i,\alpha}\right\} = \alpha$$

Moreover, $\rho_V^{L,(i)}(d)$ has asymptotic size α and is consistent against the alternative H_1.

Turning now to the power, in order to avoid the computation of exact power functions, the asymptotic local power is dealt with using local-to-unity asymptotics. This implies that the DGP considered in equation (17) is modified to be of the form:

$$y_t = \phi_T y_{t-1} + q_t \quad \text{and} \quad \phi_T = 1 - \frac{c}{T} \tag{23}$$

for some $c \geq 0$. In other words, as $T \to \infty$, $\phi_T \to 1$ and the unit root DGP obtains. On the other hand, for any fixed T, the values of $c/T \in (0, 2)$ imply that y_t is stationary.

It is well known that under nearly integrated alternatives, the limiting distribution of the rescaled sample variance statistics used above follow an Ornstein-Uhlenbeck (O-U) process. In the context of this paper however, such processes need to be interpreted in the context of wavelet scaling coefficients and the fact that some series are fractionally integrated. This leads then to the standard and fractional O-U process below:

$$J_c^{(i)}(t) = B^{(i)}(t) - c\int_0^t e^{-c(t-r)}B^{(i)}(r)dr$$
$$J_{d+1,c}^{(i)}(t) = B_{d+1}^{(i)}(t) - c\int_0^t e^{-c(t-r)}B_{d+1}^{(i)}(r)dr$$

Given that the results in Theorem 1 and 2 indicate that the WER statistic has the same limiting distribution as the variance ratio statistic of Nielsen (2009), it shouldn't be too difficult to see that the limiting distribution of the WER statistic under nearly-integrated dynamics ought to follow a ratio of two O-U processes as is summarized in the following theorem.

Theorem 4. *Let y_t be generated by equation (23) and let $\tilde{y}_t$ be the fractionally differenced version of y_t. Define the detrended versions of y_t and $\tilde{y}_t$ by equations (19) and (20) respectively. Then, for any $d > -1/2$ and $i = 0, 1, 2$, as $T \to \infty$,*

$$\hat{\rho}_V^{L,(i)}(d) \Rightarrow \frac{\int_0^1 J_c^{(i)}(t)^2 dt}{\frac{1}{\Gamma^2(d+1)}\int_0^1 J_{d+1,c}^{(i)}(t)^2 dt}$$

where

$$J_c^{(i)}(t) = J_c^{(i)}(t)^2 - \delta^{(i)}(t)\left(\int_0^1 \delta^{(i)}(s)^\top \delta^{(i)}(s)ds\right)^{-1}\left(\int_0^1 \delta^{(i)}(s)^\top J_c^{(i)}(s)ds\right)$$
$$J_{d+1,c}^{(i)}(t) = J_{d+1,c}^{(i)}(t)^2 - \delta^{(i)}(t)\left(\int_0^1 \delta^{(i)}(s)^\top \delta^{(i)}(s)ds\right)^{-1}\left(\int_0^1 \delta^{(i)}(s)^\top J_{d+1,c}^{(i)}(s)ds\right)$$

and

$$\Pi_{1,r} = 1 \quad \textit{and} \quad \Pi_{2,r} = [1, r]^\top$$

5. Simulation Analysis

Whereas asymptotic results provide a nice overview of a test's properties, the ultimate performance benchmark of any test is it's finite sample behaviour. As mentioned in the introduction, where the WER test shines is in it's ability to significantly reduce size distortions of the original variance ratio test. Furthermore, the WER test suffers from smaller size distortions due to lower d values than its variance ratio counterpart. Thus, the simulations which follow focus on these performance benchmarks for various configurations.

The DGP of choice for the simulations considered in this section is given by

$$
\begin{aligned}
q_t &= \phi q_{t-1} + u_t \\
u_t &= \epsilon_t + \theta\epsilon_{t-1}, \qquad \epsilon_y \sim N(0,1) \\
y_t &= \delta_t^{(i)}\gamma^{(i)} + q_t, \qquad i = 0,1,2 \qquad \text{and} \qquad y_0 = 0
\end{aligned}
$$

where as before, $\delta_t^{(0)} = \gamma^{(0)} = 0$ when $i = 0$, $\delta_t^{(1)} = 1$ and $\gamma^{(1)} = \gamma_0$ when $i = 1$, and when $i = 2$, define $\delta_t^{(2)}$ as the 1×2 vector $[1, t]$ and $\gamma^{(2)}$ as the 2×1 vector $[\gamma_0, \gamma_1]^\top$. In other words, when $\phi = 1$, the data is generated from an AR(1) model with a unit root and serially correlated MA(1) errors. Choosing to study the MA model is only natural considering that it arises in many economic time series configurations in addition to generating a platform where many unit root tests are known to be severely size distorted, particularly when the MA parameter is close to negative unity, see Ng and Perron (2001) and Nielsen (2008) for further discussion. Throughout, the nominal significance level is $\alpha = 0.05$, sample sizes considered are $T = 100$ and $T = 250$, and all experiments, except for the generation of critical values, are performed over $50,000$ Monte Carlo replications. Each simulation compares the WER statistic to the corresponding variance ratio statistic of Nielsen (2009) and the wavelet unit root test of Fan and Gençay (2010). All simulations were performed in R, R Core Team (2012).

Table 1: Critical Values

Deterministics	$d = 0.025$	$d = 0.05$	$d = 0.010$	FG
$\delta_t^{(0)}\gamma^{(0)}$	1.13305	1.27977	1.61864	-17.62483
$\delta_t^{(1)}\gamma^{(1)}$	1.16587	1.35526	1.81603	$-26.30335^\star$
$\delta_t^{(2)}\gamma^{(2)}$	1.19071	1.41379	1.97824	$-174.09772^\star$

Values are derived for $T = 1,000$ over $100,000$ MC replications. Since $\rho_V^H(d)$ has the same limiting distribution as $\rho(d)$, the values for the latter statistic are those reported.

*These values are very different from those reported in Fan and Gençay (2010). Simulations in this paper indicate that the correct values are those reported above. The code used is directly replicated from Ramazan Gençay's website: `http://www.sfu.ca/~rgencay/wunit.html`.

Since finite sample size distortion and power depend on asymptotic critical values of tests, it is necessary to obtain the latter first. Table 1 summarizes these numbers for various configurations. Although critical values are tabulated in both Nielsen (2009) and Fan and Gençay (2010), the configurations considered in this paper are slightly different and so it is appropriate to list them for reference.

Consider now the small sample size distortions for the base model without a mean or linear trend. Figures 3 and 4 display the size distortion for samples of size $T = 100$ and $T = 250$ respectively. Notice first the very poor performance of the Fan and Gençay (2010) (FG) statistic. Negative MA parameters are particularly problematic for this statistic as the size distortion is quite violent for these configurations. In fact, the problem is even amplified as the sample size increases. Although size distortions are much better as one moves toward positive MA parameters, the problem is still very much persistent.

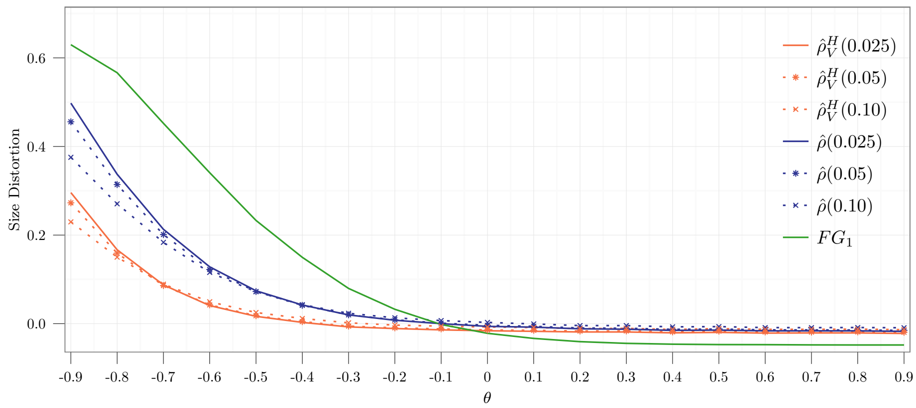

Figure 3: Size Distortion for T=100 and $\delta_t^{(0)}\gamma^{(0)}$

On the other hand, notice the very impressive performance of the WER tests. Although the limiting distribution of these statistics is the same as limiting distribution of the corresponding variance ratio statistic, (and in this sense the performance of these two statistics is indistinguishable asymptotically), in finite samples, the WER test has a significant edge over its counterpart. Another interesting feature of the WER statistic is its ability to damp the impact on size distortions arising from lowering d. As mentioned in Nielsen (2009), although lowering d can produce significant power gains, it should not be lowered too much as it then behaves as it is proportional to increasing sample size and thereby creates size distortions. Both Figures 3 and 4 demonstrate that the impact of doing this is diminished in the case of the WER statistic. This suggests that by choosing, say $d = 0.05$ instead of the recommended $d = 0.10$ in Nielsen (2009), in finite samples one can both increase power over the corresponding variance ratio test, in addition to have a significant reduction in size distortions. In other words, it seems that in the case of the WER statistic one can have a cake and eat it too.

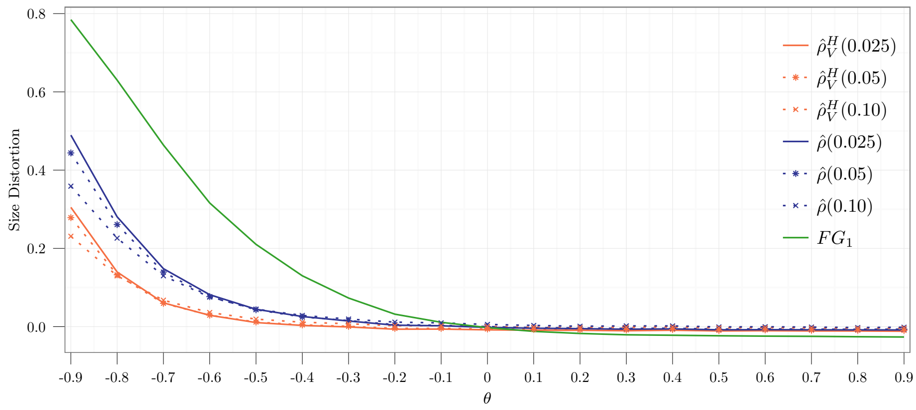

Figure 4: Size Distortion for T = 250 and $\delta_t^{(0)}\gamma^{(0)}$

A similar story holds when one includes a mean and/or a linear trend, although in the case of these configurations, size distortions are quite severe for all three statistics when the MA parameter approaches the negative unit root. For sake of brevity, only the model with both a mean and a linear trend is considered. Figures 5 and 6 plot the size distortion in the case of this configuration.

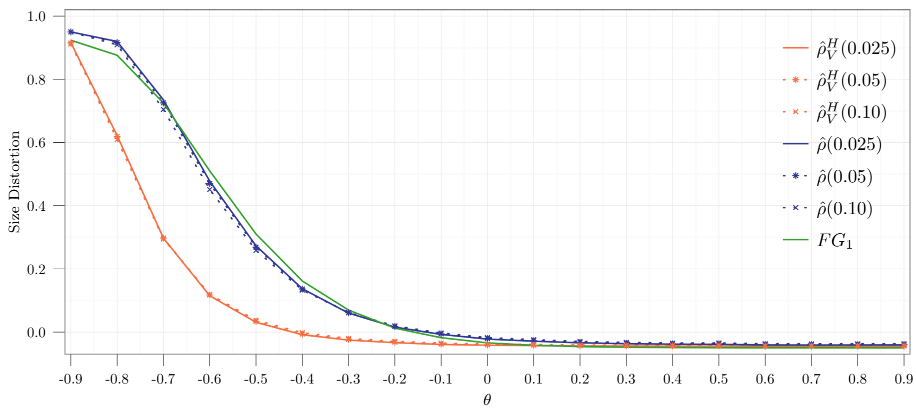

Figure 5: Size Distortion for T=100 and $\delta_t^{(2)}\gamma^{(2)}$

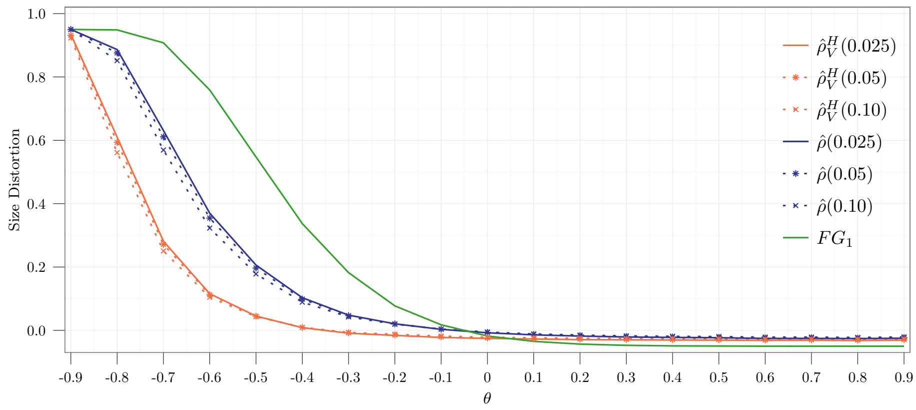

Figure 6: Size Distortion for T = 250 and $\delta_t^{(2)}\gamma^{(2)}$

The general conclusion remains the same throughout all the experiments presented above: the WER statistics are considerably less size distorted. Moreover, although size distortions instigated through very small choices of d tend to diminish as one considers models with a constant mean and a mean and trend, the WER statistic continues to be more robust to this effect than its variance ratio counterpart. An interesting observation is that in very small samples, the variance ratio statistic suffers from size distortions which are as bad as those of the FG statistic. As the sample size increases however, the variance ratio statistic becomes considerably less size distorted than the FG statistic. In either case, it is clear that the WER can offer non trivial size distortion reductions.

6. Conclusion

The wavelet energy ratio (WER) unit root test presented in this paper is a contribution to the rather sparse literature on unit root testing using wavelet theory. The WER draws its inspiration from the variance ratio unit root test of Nielsen (2009) by exploiting the wavelet power spectrum of the observed series and its fractional partial sum to construct a unit root test based on the ratio of the norms of the scaling energies obtained from a unit scale DWT. The result is a family of non-parametric, completely tuning parameter free, unit root tests indexed by the fractional parameter d and constructed entirely in the spectral domain. In this sense, this is a direct improvement over the existing unit root test of Fan and Gençay (2010) which, although nuisance parameter free (albeit after a transformation), is still not tuning parameter free. In fact, the latter test requires the estimation of the long-term variance of the error terms using the Newey and West (1987) estimator using a Bartlett kernel with a bandwidth tuning parameter. No such estimation procedure is necessary in the case of the WER statistic and insofar as this is concerned, the test is easier to implement.

Results established in this paper demonstrate that the WER statistics have the same limiting distributions as the limiting distributions of the corresponding variance ratio statistics of Nielsen (2009). Furthermore, these results also establish the limiting distribution of the WER statistics in the presence of a mean and linear trend after OLS detrending has been applied. In this sense, asymptotically, the WER tests exhibit the same performance gains and flaws as their variance ratio counterparts. In particular, using local-to-unity asymptotics, simulations demonstrate that in terms of power, the WER test virtually mimic the corresponding variance ratio unit root tests, although the power of both is visibly weaker than that exhibited by the Fan and Gençay (2010) test. However, where the WER tests truly shine is in their finite sample performance.

Simulation experiments in this paper clearly show that the WER tests exhibit non trivial size distortion reductions, particularly when the MA parameters are close to the negative unit root. In this regard, the WER tests are shown to cut size distortions of the traditional variance ratio unit root test anywhere from 25% - 50%. On the other hand, the Fan and Gençay (2010) tests are severely size distorted to the point where for certain configurations, size distortions can easily reach levels of 95% for extreme negative MA parameters. Just why this test behaves so poorly in this sense is not really clear and some type of investigative analysis might be useful.

It is also interesting to note that size distortions of the WER statistic are less impacted by lowering d than in the case of the corresponding variance ratio test. This problem was mentioned in Nielsen (2009) and simulations conducted here confirm this. Nevertheless, since the WER is less sensitive to these changes than the variance ratio test, one can afford to lower d beyond the suggested $d = 0.10$ mark with little sacrifice in terms of size distortion but with a noticeable gain in power. In this regard, it should be safe to choose $d = 0.05$.

It should also be noted that Nielsen (2009) suggested the use of the sieve bootstrap to reduce size distortions. No such procedure was considered here as the topic is being investigated by the author. However, it is safe to say that, given the success of the bootstrapping procedure in the case of the variance ratio unit root test, some sort of bootstrapping procedure in the case of the WER can very well reduce size distortions to zero. Finally, it is not difficult to see the potential of the WER test in tests for the cointegration rank in fractionally integrated systems as considered in Nielsen (2010). This too is being researched by the author.

References

J. Breitung. Nonparametric tests for unit roots and cointegration. *Journal of Econometrics*, 108(2): 343–363, 2002.

N.H. Chan and C.Z. Wei. Asymptotic inference for nearly nonstationary ar (1) processes. *The Annals of Statistics*, 15(3):1050–1063, 1987.

Y. Chang and J.Y. Park. A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis*, 24(4):379–400, 2003.

I. Choi and P.C.B. Phillips. Testing for a unit root by frequency domain regression. *Journal of Econometrics*, 59(3):263–286, 1993.

I. Daubechies. Ten lectures on wavelets, cbms-nsf conference series in applied mathematics. *SIAM Ed*, pages 122–122, 1992.

J. Davidson and N. Hashimzade. Type i and type ii fractional brownian motions: A reconsideration. *Computational Statistics & Data Analysis*, 53(6):2089–2106, 2009.

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

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

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

C.S. Gourieroux and J. Akonom. Functional limit theorem for fractional processes (a). *CEPREMAP Working Papers (Couverture Orange)*, 1988.

C.W.J. Granger. The typical spectral shape of an economic variable. *Econometrica: Journal of the Econometric Society*, pages 150–161, 1966.

J.D. Hamilton. *Time series analysis*, volume 2. Cambridge Univ Press, 1994.

S.G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):674–693, 1989.

U.K. Müller. The impossibility of consistent discrimination between i (0) and i (1) processes. *Econometric Theory*, 24(03):616–630, 2008.

W. Newey and K. West. A simple positive semidefinite heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, pages 55,150–161, 1987.

S. Ng and P. Perron. Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6):1519–1554, 2001.

M.Ø. Nielsen. A powerful tuning parameter free test of the autoregressive unit root hypothesis. *Working Papers*, 2008.

M.Ø. Nielsen. A powerful test of the autoregressive unit root hypothesis based on a tuning parameter free statistic. *Econometric Theory*, 25(06):1515–1544, 2009.

M.O. Nielsen. Nonparametric cointegration analysis of fractional systems with unknown integration orders. *Journal of Econometrics*, 155(2):170–187, 2010.

H.J. Park and W.A. Fuller. Alternative estimators and unit root tests for the autoregressive process. *Journal of Time Series Analysis*, 16(4):415–429, 1995.

J.Y. Park. Testing for unit roots and cointegration by variable addition. *Advances in Econometrics*, 8 (2):107–133, 1990.

J.Y. Park and B. Choi. A new approach to testing for a unit root. *Center for Analytic Economics, Cornell University, Working Paper*, 88:23, 1988.

D.B. Percival and A.T. Walden. *Wavelet methods for time series analysis*, volume 4. Cambridge Univ Pr, 2006.

P. Perron and Z. Qu. A simple modification to improve the finite sample properties of ng and perron's unit root tests. *Economics Letters*, 94(1):12–19, 2007.

P.C.B. Phillips. Time series regression with a unit root. *Econometrica: Journal of the Econometric Society*, pages 277–301, 1987.

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

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2012. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

C.A. Sims, J.H. Stock, and M.W. Watson. Inference in linear time series models with some unit roots. *Econometrica: Journal of the Econometric Society*, pages 113–144, 1990.

Q. Wang, Y.X. Lin, and C. Gulati. Asymptotics for general nonstationary fractionally integrated processes without prehistoric influence. *Journal of Applied Mathematics and Decision Sciences*, 6(4): 255–269, 2002.

Appendix

Proof of Lemma 1. The first part of the lemma was proven in Fan and Gençay (2010). To prove the second statement, note that if $\{y\}_{t=1}^{T}$ is a simple unit root process, then $y_t = y_{t-1} + u_t$ and therefore $\tilde{y}_t = \tilde{y}_{t-1} + \tilde{u}_t$. This implies that under the null hypothesis H_0, $\widetilde{V}_{t,1}$ is given by:

$$\widetilde{V}_{t,1} = \frac{1}{\sqrt{2}}\left(2\tilde{y}_{2t-1} + \tilde{u}_{2t}\right)$$

Following Fan and Gençay (2010), argue that

$$\begin{aligned}\sum_{t=1}^{T/2} \widetilde{V}_{t,1}^2 &= \frac{1}{2}\left\{4\sum_{t=1}^{T/2} \tilde{y}_{2t-1}^2 + 4\sum_{t=1}^{T/2} \tilde{u}_{2t}\tilde{y}_{2t-1} + \sum_{t=1}^{T/2} \tilde{u}_{2t}^2\right\} \\ &\equiv 2\widetilde{A}_T + 2\widetilde{B}_T + \frac{1}{2}\widetilde{C}_T\end{aligned}$$

where

$$\widetilde{A}_T = \sum_{t=1}^{T/2} \tilde{x}_t^2 \qquad \widetilde{B}_T = \sum_{t=1}^{T/2} \tilde{u}_{2t}\tilde{x}_t \qquad \widetilde{C}_T = \sum_{t=1}^{T/2} \tilde{u}_t^2$$

and $\tilde{x}_t = \tilde{y}_{2t-1}$ for $t = 1, 2, \ldots, T/2$. Assume next that Assumptions (a) through (c) hold. Then, using Proposition 17.2 in Hamilton (1994), obtain the following expansion:

$$\tilde{x}_t = \tilde{x}_0 + \sum_{j=1}^{t} \tilde{v}_j = \tilde{x}_0 + \sum_{j=0}^{2t-1} \tilde{u}_j = \tilde{x}_0 \left\{\tilde{u}_0 + b_\psi \sum_{j=1}^{2t-1} \tilde{u}_j + \tilde{\eta}_{2t-1} - \tilde{\eta}_0\right\}$$

Next, let $T_1 = T/2$ and define the partial sum processes associated with $\tilde{v}_t$ as follows:

$$\widetilde{X}_{T_1}(r) = \frac{1}{T_1} \sum_{t=1}^{\lfloor T_1 r \rfloor} \tilde{v}_t \qquad 0 \leq r \leq 1$$

Now, let $\stackrel{\mathcal{L}}{=}$ denote equality in law and note that the above partial sum satisfies the following relation:

$$\widetilde{X}_{T_1}(r) \stackrel{\mathcal{L}}{=} b_\psi \frac{1}{T_1} \sum_{t=1}^{2\lfloor T_1 r \rfloor - 1} \tilde{\epsilon}_j = 2b_\psi \frac{1}{T} \sum_{t=1}^{\lfloor Tr \rfloor - 1} \tilde{\epsilon}_j$$

The functional central limit theorem (FCLT) for fractional process without prehistoric influence, see Wang et al. (2002), then ensures the following convergence result:

$$T^{1/2-d}\widetilde{X}_{T_1}(\cdot) \Rightarrow 2\frac{b_\psi \sigma}{\Gamma(d+1)} B_{d+1}(\cdot)$$

where $\Rightarrow$ denotes weak convergence. Note however that

$$\sum_{t=1}^{T_1} \tilde{x}_t^2 = \frac{T_1^2}{2} \int_0^1 T\widetilde{X}_{T_1}^2(r)dr$$

Invoking the continuous mapping theorem (CMT) to the construction above, a little algebra demonstrates that:

$$\begin{aligned} T^{-2d}\frac{1}{T_1^2}\sum_{t=1}^{T_1} x_t^2 &= \frac{1}{2}T^{-2(d+1)}\sum_{t=1}^{T_1} x_t^2 \\ &\Rightarrow 2\frac{b_\psi^2\sigma^2}{\Gamma^2(d+1)}\int_0^1 B_{d+1}^2(r)dr \end{aligned}$$

In other words, $\widetilde{A}_T$ is $O_p\left(T^{2(1+d)}\right)$.

Next, note that Corollary 1 in Wang et al. (2002) implies that:

$$T^{-2d}\widetilde{C}_T \Rightarrow 2\frac{b_\psi^2\sigma^2}{\Gamma^2(d)}\int_0^r B_d^2(s)ds$$

Thus, $\widetilde{C}_T$ is $O_p\left(T^{2d}\right)$. Finally, arguing along similar lines as Fan and Gençay (2010) and invoking Theorem 2 in Wang et al. (2002), it is not too difficult to show that $\widetilde{B_T}$ is $O_p\left(T^{2d}\right)$ as well. Conclude therefore that $\widetilde{V}_{t,1}^2$ is $O_p\left(T^{2(1+d)}\right)$. □

Proof of Theorem 1. Suppose Assumptions (a) through (c) hold. Then if H_0 is in effect, note that $\hat{\rho}_V^H(d)$ can be rewritten as follows:

$$\hat{\rho}_V^H(d) = \frac{T^{-2}\sum_{t=1}^{T/2} V_{t,1}^2}{T^{-2(d+1)}\sum_{t=1}^{T/2}\widetilde{V}_{t,1}^2}$$

The limiting form of the denominator follows immediately from Lemma 1. It was established there that $\sum_{t=1}^{T/2}\widetilde{V}_{t,1}^2 = 2\widetilde{A_T}+2\widetilde{B_T}+\frac{1}{2}\widetilde{C_T}$ and that $\widetilde{A_T}$ is $O_p\left(T^{2(d+1)}\right)$ whereas both $\widetilde{B_T}$ and $\widetilde{C_T}$ are $O_p\left(T^{2d}\right)$. The result follows by noting that $\widetilde{A_T} \Rightarrow \frac{1}{2}\frac{b_\psi^2\sigma^2}{\Gamma^2(d+1)}\int_0^1 B_{d+1}^2(r)dr$.

Similarly, the limiting form of the numerator flows from Lemma 1 in Fan and Gençay (2010). The latter established that $\sum_{t=1}^{T/2} V_{t,1}^2 = 2A_T + 2B_T + \frac{1}{2}C_T$ and that A_T is $O_p\left(T^2\right)$ whereas both B_T and C_T are $O_p(T)$. The result follows by recalling that

$T^{-2}A_T \Rightarrow 2b_\psi^2\sigma^2\int_0^1 W^2(r)dr$.

The desired result is evident after cancelling factors of 2 from the ratio. □

Proof of Theorem 2. Suppose Assumptions (a) through (c) hold and H_0 is in effect. Theorem 2 in Fan and Gençay (2010) shows that $\sum_{t=L_1}^{T/2} V_{t,1}^2$ is $O_p\left(T^2\right)$ and that $T^{-2}\sum_{t=L_1}^{T/2} V_{t,1}^2 \Rightarrow b_\psi^2\sigma^2\int_0^1 W^2(r)dr$. The rest of the proof focuses on establishing similar results for the fractional counterpart $\widetilde{V}_{t,1}$.

Consider first $\sum_{t=L_1}^{T/2} \widetilde{V}_{t,1}^2$, and note that:

$$\begin{aligned}\widetilde{V}_{t,1} &= \tilde{y}_{2t+1-L}\sum_{l=0}^{L-1} g_l + \sum_{l=0}^{L-2} g_l\left(\sum_{j=0}^{L-2-l}\tilde{u}_{2t-j-l}\right)\\ &= \sqrt{2}\tilde{y}_{2t+1-L} + \sum_{l=0}^{L-2} g_l\left(\sum_{j=0}^{L-2-l}\tilde{u}_{2t-j-l}\right)\end{aligned}$$

Then the following expansion holds:

$$\begin{aligned}T^{-2d}\frac{1}{T_1^2}\sum_{t=L_1}^{T_1}\widetilde{V}_{t,1}^2 &= T^{-2d}\frac{1}{T_1^2}\sum_{t=L_1}^{T_1}\left(\sqrt{2}\tilde{y}_{2t+1-L} + \sum_{l=0}^{L-2} g_l\left(\sum_{j=0}^{L-2-l}\tilde{u}_{2t-j-l}\right)\right)^2\\ &= 8T^{-2(1+d)}\sum_{t=L_1}^{T_1}\tilde{y}_{2t+1-L}^2\\ &\quad +8\sqrt{2}T^{-2(1+d)}\sum_{t=L_1}^{T_1}\tilde{y}_{2t+1-L}\left(\sum_{l=0}^{L-2} g_l\left(\sum_{j=0}^{L-2-l}\tilde{u}_{2t-j-l}\right)\right)\\ &\quad +4T^{-2(1+d)}\sum_{t=L_1}^{T_1}\left(\sum_{l=0}^{L-2} g_l\left(\sum_{j=0}^{L-2-l}\tilde{u}_{2t-j-l}\right)\right)^2\\ &= 8\widetilde{A}_T^{L_1} + 8\sqrt{2}\widetilde{B}_T^{L_1} + 4\widetilde{C}_T^{L_1}\end{aligned}$$

where

$$
\begin{aligned}
\widetilde{A}_T^{L_1} &= \sum_{t=L_1}^{T_1} \tilde{y}_{2t+1-L}^2 \\
\widetilde{B}_T^{L_1} &= \sum_{t=L_1}^{T_1} \tilde{y}_{2t+1-L} \left(\sum_{l=0}^{L-2} g_l \left(\sum_{j=0}^{L-2-l} \tilde{u}_{2t-j-l} \right) \right) \\
\widetilde{C}_T^{L_1} &= \sum_{t=L_1}^{T_1} \left(\sum_{l=0}^{L-2} g_l \left(\sum_{j=0}^{L-2-l} \tilde{u}_{2t-j-l} \right) \right)^2
\end{aligned}
$$

A similar derivation to the one used in Lemma 1 demonstrates that $\widetilde{A}_T^{L_1}$ is $O_p\left(T^{2(1+d)}\right)$. Moreover, an argument analogous to the one made in Theorem 1 demonstrates that:

$$
T^{-2(1+d)} \widetilde{A}_T^{L_1} \Rightarrow \frac{1}{2} \frac{b_\psi^2 \sigma^2}{\Gamma^2(d+1)} \int_0^1 B_{d+1}^2(r) dr
$$

On the other hand, since $\widetilde{B}_T^{L_1}$ and $\widetilde{C}_T^{L_1}$ are weighted linear combination of $\widetilde{B}_T$ and $\widetilde{C}_T$, respectively, and since both $\widetilde{B}_T$ and $\widetilde{C}_T$ are $O_p\left(T^{2d}\right)$, this implies that both $T^{-2(1+d)} \widetilde{B}_T^{L_1}$ and $T^{-2(1+d)} \widetilde{C}_T^{L_1}$ are $o_p(1)$. Both results of the theorem follow immediately. □

Proof of Theorem 3. Begin with the non-fractional case and consider the $\mathcal{D}[0,1]$ approximation of the residuals regression of $y_t^{(i)}$ on $\delta_t^{(i)} \gamma^{(i)}$ for $t = 1 \ldots T$, by noting the following for $s \in [0,1]$:

$$
\begin{aligned}
\hat{y}_{\lfloor Ts \rfloor}^{(i)} &= y_{\lfloor Ts \rfloor}^{(i)} - \delta_{\lfloor Ts \rfloor}^{(i)} \hat{\gamma}^{(i)} \\
&= q_{\lfloor Ts \rfloor} - \delta_{\lfloor Ts \rfloor}^{(i)} \left(\hat{\gamma}^{(i)} - \gamma^{(i)} \right)
\end{aligned}
$$

Define $N^{(1)}(T) = 1$ and $N^{(1)}(T) = \left[\begin{smallmatrix} 1 & 0 \\ 0 & T^{-1} \end{smallmatrix}\right]$ and note the following:

$$
\begin{aligned}
&\left(b_\psi \sigma T^{1/2} \right)^{-1} \delta_{\lfloor Ts \rfloor}^{(i)} \left(\hat{\gamma}^{(i)} - \gamma^{(i)} \right) \\
&= N^{(i)}(T) \delta_{\lfloor Ts \rfloor}^{(i)} \left(N^{(i)}(T) \left(b_\psi^2 \sigma^2 \right)^{-1} T^{-1/2} \left(\hat{\gamma}^{(i)} - \gamma^{(i)} \right) \right) \\
&= N^{(i)}(T) \delta_{\lfloor Ts \rfloor}^{(i)} \left(T^{-1} \sum_{t=1}^{T} N^{(i)}(T) \delta_t^{(i)\top} N^{(i)}(T) \delta_t^{(i)} \right)^{-1} \left(T^{-1} \sum_{t=1}^{T} N^{(i)}(T) \delta_t^{(i)\top} \left(b_\psi^2 \sigma^2 \right)^{-1} T^{-1/2} q_{\lfloor Ts \rfloor} \right) \\
&= N^{(i)}(T) \delta_{\lfloor Ts \rfloor}^{(i)} \left(T^{-1} \sum_{t=1}^{T} \delta_{t/T}^{(i)\top} \delta_{t/T}^{(i)} \right)^{-1} \left(T^{-1} \sum_{t=1}^{T} \delta_{t/T}^{(i)\top} \left(b_\psi^2 \sigma^2 \right)^{-1} T^{-1/2} q_{\lfloor Ts \rfloor} \right) \\
&\Rightarrow \delta^{(i)}(s) \left(\int_0^1 \delta^{(i)}(s)^\top \delta^{(i)}(s) ds \right)^{-1} \left(\int_0^1 \delta^{(i)}(s)^\top B(s) ds \right)
\end{aligned}
$$

The above follows from Theorem 2, the continuous mapping theorem, and from the fact that as $T \to \infty$, $t/T \to r$ and therefore $N^{(i)}(T)\delta^{(i)}_{\lfloor Ts \rfloor} \to \delta^{(i)}(s)$. Thus, conclude that

$$\left(b_\psi \sigma T^{1/2}\right)^{-1} \hat{y}^{(i)}_{\lfloor Ts \rfloor} \Rightarrow B(s) - \delta^{(i)}(s) \left(\int_0^1 \delta^{(i)}(r)^\top \delta^{(i)}(r) dr\right)^{-1} \left(\int_0^1 \delta^{(i)}(r)^\top B(r) dr\right) \tag{24}$$

A similar argument can be used for the fractional case. Consider therefore the following:

$$\hat{\tilde{y}}^{(i)}_{\lfloor Ts \rfloor} = \tilde{y}_{\lfloor Ts \rfloor} - \delta^{(i)}_{\lfloor Ts \rfloor}\left(\hat{\gamma}^{(i)} - \gamma^{(i)}\right)$$

and note that

$$\begin{aligned}
&\left(\frac{b_\psi \sigma}{\Gamma(d+1)} T^{(1/2+d)}\right)^{-1} \delta^{(i)}_{\lfloor Ts \rfloor}\left(\hat{\gamma}^{(i)} - \gamma^{(i)}\right) \\
&= N^{(i)}(T)\delta^{(i)}_{\lfloor Ts \rfloor}\left(N^{(i)}(T)\left(\kappa^2(d)T^{2(d+1/2)}\right)^{-1/2}\left(\hat{\gamma}^{(i)} - \gamma^{(i)}\right)\right) \\
&= N^{(i)}(T)\delta^{(i)}_{\lfloor Ts \rfloor}\left(T^{-1}\sum_{t=1}^{T} N^{(i)}(T)\delta_t^{(i)\top} N^{(i)}(T)\delta_t^{(i)}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T} N^{(i)}(T)\delta_t^{(i)\top}\left(\kappa^2(d)T^{2(d+1/2)}\right)^{-1/2}\tilde{q}_{\lfloor Ts \rfloor}\right) \\
&= N^{(i)}(T)\delta^{(i)}_{\lfloor Ts \rfloor}\left(T^{-1}\sum_{t=1}^{T} \delta_{t/T}^{(i)\top}\delta_{t/T}^{(i)}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T} \delta_{t/T}^{(i)\top}\left(\kappa^2(d)T^{2(d+1/2)}\right)^{-1/2}\tilde{q}_{\lfloor Ts \rfloor}\right) \\
&\Rightarrow \delta^{(i)}(s)\left(\int_0^1 \delta^{(i)}(s)^\top \delta^{(i)}(s) ds\right)^{-1}\left(\int_0^1 \delta^{(i)}(s)^\top B_{d+1}(s) ds\right)
\end{aligned}$$

The above of course follows from a standard application of the fractional CLT on $\tilde{q}_{\lfloor Ts \rfloor}$ and the continuous mapping theorem. Thus, conclude that:

$$\left(\frac{b_\psi \sigma}{\Gamma(d+1)} T^{(1/2+d)}\right)^{-1} \hat{\tilde{y}}^{(i)}_{\lfloor Ts \rfloor} \Rightarrow B_{d+1}(s) - \delta^{(i)}(s)\left(\int_0^1 \delta^{(i)}(r)^\top \delta^{(i)}(r) dr\right)^{-1}\left(\int_0^1 \delta^{(i)}(r)^\top B_{d+1}(r) dr\right) \tag{25}$$

The results of the theorem follow immediately from the proof of Theorem 2 by adapting $A_T^{L_1}$ and $\widetilde{A}_T^{L_1}$ to equations (24) and (25) respectively, and by noting that the appropriate orders which derive the results of Theorem 2, continue to hold here as well. □

Proof of Lemma 2. Since the limiting distribution the $\rho_V^{L,(i)}(d)$ tends to the limiting distribution $\rho(d)$, the methods of proof in Theorem 3 can be analogously adapted to the proof of Theorem 2 in Nielsen (2009) to derive the stated result. □

Proof of Theorem 4. Note first that Theorem 3 of Nielsen (2009) shows that

$$\left(b_\psi \sigma T^{1/2}\right)^{-1} \hat{y}^{(i)}_{\lfloor Ts \rfloor} \Rightarrow J_c^{(i)}(t) \tag{26}$$

$$\left(\frac{b_\psi \sigma}{\Gamma(d+1)} T^{(1/2+d)}\right)^{-1} \hat{\hat{y}}^{(i)}_{\lfloor Ts \rfloor} \Rightarrow J_{d+1,c}^{(i)}(t) \tag{27}$$

First the limiting distribution of $A_T^{L_1}$ and $\widetilde{A}_T^{L_1}$ is established. This is essentially identical to the proof outlined in Theorem 2 whereby the (fractional) FCLT and the CMT is applied to equations (26) and (27) to yield:

$$T^{-2} A_T^{L_1} \Rightarrow \frac{1}{2} b_\psi^2 \sigma^2 \int_0^1 J_c^{(i)}(t)^2 dr$$

$$T^{-2(1+d)} \widetilde{A}_T^{L_1} \Rightarrow \frac{1}{2} \frac{b_\psi^2 \sigma^2}{\Gamma^2(d+1)} \int_0^1 J_{d+1,c}^{(i)}(t)^2 dr$$

The results of the theorem then follow by noting that the order of $A_T^{L_1}$ and $\widetilde{A}_T^{L_1}$ are $O_p\left(T^2\right)$ and $O_p\left(T^{2(1+d)}\right)$ respectively, whereas $B_T^{L_1}$ and $C_T^{L_1}$ are $O_p\left(T^2\right)$ and $\widetilde{B}_T^{L_1}$ and $\widetilde{C}_T^{L_1}$ are $O_p\left(T^{2d}\right)$. □