An Approximate Wavelet MLE of Short and Long Memory Parameters

Mark J. Jensen
Department of Economics,University of Missouri
Columbia, Missouri, 65211
jensen@haar.econ.missouri.edu

Abstract. By design a wavelet's strength rests in its ability to localize a process simultaneously in time-scale space. The wavelet's ability to localize a time series in time-scale space directly leads to the computational efficiency of the wavelet representation of a $N \times N$ matrix operator by allowing the N largest elements of the wavelet represented operator to represent the matrix operator [Devore, et al. (1992a) and (1992b)]. This property allows many dense matrices to have sparse representation when transformed by wavelets.

In this paper we generalize the long-memory parameter estimator of McCoy and Walden (1996) to estimate simultaneously the short and long-memory parameters. Using the sparse wavelet representation of a matrix operator, we are able to approximate an ARFIMA models likelihood function with the series's wavelet coefficients and their variances. Maximization of this approximate likelihood function over the short and long-memory parameter space results in the approximate wavelet maximum likelihood estimates of the ARFIMA model.

By simultaneously maximizing the likelihood function over both the short and long-memory parameters and using only the wavelet coefficient's variances, the approximate wavelet MLE provides a fast alternative to the frequency-domain MLE. Furthermore, the simulation studies found herein reveal the approximate wavelet MLE to be robust over the invertible parameter region of the ARFIMA model's moving average parameter, whereas the frequency-domain MLE dramatically deteriorates as the moving average parameter approaches the boundaries of invertibility.

Keywords: Long-Memory, Fractional Integration, ARFIMA, Wavelets
JEL Classification: C22

1 Introduction

While long-memory and wavelets have only recently been used by econometricians, both have played an important role in other fields since the early 1900s.[1] Scientists in diverse fields have observed time series where observations that are far apart (in time or space) were correlated too strongly to be modeled as independent data or classical autoregressive, moving average models (ARMA). This concept of long-memory has grown rapidly and can be found in a broad scattering of fields such as agronomy, astronomy, chemistry, engineering, environmental sciences, geosciences, hydrology, mathematics, physics and statistics. Even

[1] For examples in economics and finance where wavelets have been used see Ramsey and Zhang (1997), Ramsey and Lampart (1998a,1998b), and Jensen (1999a,1999b).

in its infancy among economists, long-memory has been applied to a number of economic and financial time series. For example, real gross national product [Sowell (1992b), Diebold and Rudebusch (1991)], interest rates [Backus and Zin (1993)], consumer and wholesale price indices [Baillie et al. (1996), Hassler and Wolters (1995)], stock market returns [Ding, et al. (1993)], stock market prices [Lo (1991)], option prices [Baillie and Bollerslev (1994)] and exchange rates [Cheung (1993)] have all had long-memory ideas applied to them.

The empirical presence of long-memory is found in the persistence of the autocorrelations, $\gamma(s)$, which behave as $\gamma(s) \sim |s|^{2d-1}$, as $s \to \infty$, where $|d| < 1/2$ is the long-memory parameter. This slow decay by the autocorrelations is not consistent with either the stationary, short-memory, ARMA models ($d = 0$), nor the non-stationary, unit root models ($d = 1$). Instead, long-memory falls nicely in between these two knife-edge approaches. The drawback is the dense covariance matrix it creates, i.e., a large matrix with few zero elements. This dense matrix makes calculation of the exact likelihood function impossible for large data sets since inversion of the long-memory's covariance matrix is an exhaustive task, requiring on the order of cubed numerical operations.

Sowell (1992a) has derived an exact maximum likelihood estimator (MLE) for a general class of long-memory models possessing both short and long-memory parameters called the autoregressive, fractionally integrated, moving average model (ARFIMA). Although Sowell's method is theoretically appealing from a statistical inference standpoint, it is computationally exhaustive for large data sets since it requires inverting the ARFIMA model's dense covariance matrix at each step of the maximization routine. As a result, most economists use the approximate frequency-domain Whittle (1951) maximum likelihood estimator of Fox and Taqqu (1986). In a simulation study, Cheung and Diebold (1994) have shown that in large samples the approximate frequency-domain MLE performs just as well as the exact MLE.

By design the wavelet's strength rests in its ability to localize a process simultaneously in time-scale space. The wavelet's ability to localize a time series in time-scale space directly leads to the computational efficiency of the wavelet representation of a $N \times N$ matrix operator by allowing the N largest elements of the wavelet represented operator to represent

the matrix operator [Devore, et al. (1992a) and (1992b)]. This property allows many dense matrices to have sparse representation when transformed by wavelets.

In this paper we generalize the long-memory parameter estimator of McCoy and Walden (1996) to estimate simultaneously the ARFIMA model's short and long-memory parameters. Using the sparse wavelet representation of a matrix operator, we are able to approximate the likelihood function of an ARFIMA model with the series's wavelet coefficients and variances. Maximization of this approximate likelihood function over the short and long-memory parameter space results in the approximate wavelet maximum likelihood estimator of the ARFIMA model.

By simultaneously maximizing the likelihood function over both the short and long-memory parameters, and using only the wavelet coefficient's variances, the approximate wavelet MLE provides an equally fast alternative to the frequency-domain MLE. Furthermore, the simulation studies found herein reveal the approximate wavelet MLE to be robust over the invertible region of the ARFIMA model's moving average parameter, whereas the frequency-domain MLE dramatically deteriorates as the moving average parameter approaches the boundaries of invertibility.

The next section introduces the fractionally integrated, autoregressive, moving average model and its spectrum. In section 3, wavelet analysis and its transformation are presented through the wavelet's interpretation as a bandpass filter. Using the material of these two sections, section 4 calculates the wavelet coefficient's variances and covariances for an ARFIMA model and shows the uncorrelated nature of wavelet coefficients. In section 5, the approximate likelihood function is derived. The robustness and statistical inference of the approximate wavelet maximum likelihood estimators of the ARFIMA model's short and long-memory parameters are determined empirically in the Monte Carlo simulations of section 6. The final section summarizes our findings.

2 Fractionally Integrated ARMA Processes

Let $x(t)$ be the fractionally integrated, autoregressive, moving average, ARFIMA(p,d,q), process defined by

$$\Phi(L)(1-L)^d(x(t)-\mu) = \Theta(L)\epsilon(t) \tag{1}$$

where

$$\begin{aligned}\Phi(L) &= 1-\phi_1 L-\phi_2 L^2-\cdots-\phi_p L^p \\ \Theta(L) &= 1+\theta_1 L+\theta_2 L^2+\cdots+\theta_q L^q\end{aligned}$$

are polynomials of degrees p and q respectively, and whose roots lie outside the unit circle. The innovations $\epsilon(t) \sim i.i.d.\mathcal{N}(0,\sigma^2)$, the long-memory parameter, $|d| < 0.5$, μ is the unknown mean, and $(1-L)^d$ is the fractional differencing operator defined by the binomial expansion

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

As defined $x(t)$ is an invertible, causal, stationary process that generalizes the class of ARIMA models to non-integer values of d. If the differencing operator $(1-L)^d$ were applied to the time series, $x(t)$, the resulting series would be a ARMA(p,q) model. Hence, like an ARIMA model, if $0.5 \le d$, $x(t)$ can be differenced to obtain an ARFIMA$(p,d-1,q)$ model. Thus, the invertibility and stationarity condition, $|d| < 0.5$, can always be satisfied through differencing.

It has been shown [Brockwell and Davis (1993), p. 525] that the power spectrum of $x(t)$ equals

$$S_x(\omega) = \frac{\sigma^2}{2\pi}\frac{|\Theta(e^{-i\omega})|^2}{|\Phi(e^{-i\omega})|^2}|1-e^{-i\omega}|^{-2d} \tag{2}$$

$$\sim \frac{\sigma^2}{2\pi}\left(\frac{\Theta(1)}{\Phi(1)}\right)^2\omega^{-2d} \qquad \text{as } \omega \to 0 \tag{3}$$

where "$\sim$" means that the ratio of left- and right-hand sides tends to 1.

For $0 < d < 0.5$, the unbounded spectrum at the origin satisfies the long-memory definition of McLeod and Hipel (1978a). The infinite value of the spectrum as ω approaches

zero exhibits the strong positive dependence between distance observations. From Eq. (3), this strong dependence is the result of the differencing parameter d, hence, the name long-memory parameter. $x(t)$ is often referred to as an intermediate memory process when $-0.5 < d < 0$. For negative values of d, the spectrum declines to zero as ω approaches the origin.

3 Wavelets

A wavelet is defined as any $L^2(\Re)$ function, $\psi(t)$, that satisfies the admissibility condition

$$\int_{-\infty}^{\infty} \frac{d\omega}{\omega} |\hat{\psi}(\omega)|^2 < \infty$$

where $\hat{\psi}$ denotes the Fourier transform of ψ. Since ψ is a square integrable function, the admissibility condition is satisfied only if $\hat{\psi}(0) = 0$, i.e., $\int \psi(t)dt = 0$, and $\hat{\psi}$ has sufficiently fast decay as $|\omega| \to \infty$. In general, more restrictive conditions than the admissibility condition will hold for the wavelets that are used in practice.

A useful pedagogical example of a wavelet that satisfies these conditions and that facilitate the results of this paper is the ideal high-bandpass wavelet with frequency response

$$\hat{\psi}(\omega) = \begin{cases} 1 & \text{if } |\omega| \in (\pi, 2\pi], \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Let

$$\psi_{m,n}(t) \quad = \quad 2^{-m/2}\psi(2^{-m}t - n) \tag{5}$$

be the dilation and translation of ψ, where $m \in \boldsymbol{Z} = \{0, \pm 1, \pm 2, \ldots\}$ is the dilation parameter[2] and $n \in \boldsymbol{Z}$ is the translation parameter. It is easily shown that $\hat{\psi}_{m,n}(\omega) = 2^{m/2}e^{-i2^m\omega n}\hat{\psi}(2^m\omega)$ and the dilated ideal high-bandpass wavelet of Eq. (4) will have the frequency response

$$\hat{\psi}_{m,0}(\omega) \quad = \quad \begin{cases} 2^{m/2} & \text{if } |\omega| \in (2^{-m}\pi, 2^{-m+1}\pi], \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Because the translated wavelet's Fourier transform also includes the complex exponential term, $e^{-i2^m\omega n}$, $\hat{\psi}_{m,n}$ does not equal $\hat{\psi}_{m,0}$, but their supports are the same.

[2] m is also referred to as the scaling parameter which will soon become apparent.

Daubechies (1988) has shown the collection of dilations and translations, $\{\psi_{m,n}\}_{m,n\in \boldsymbol{Z}}$, to be a complete orthonormal basis of L^2, such that any $x(t) \in L^2(\Re)$ can be represented as

$$x(t) = \sum_{m\in \boldsymbol{Z}} \sum_{n\in \boldsymbol{Z}} w_{m,n}\psi_{m,n}(t) \tag{7}$$

where

$$w_{m,n} \quad = \quad \int x(t)\psi_{m,n}(t)\ dt \tag{8}$$

is the wavelet coefficient.

Drawing on the ideal high-bandpass wavelet of Eq. (4), the wavelet representation of $x(t)$ in Eq. (7) essentially splits the frequency space into the dyadic blocks, $(\pm 2^{-m}\pi, \pm 2^{-m+1}\pi]$. For each value of the dilation parameter, m, the dyadic frequency block moves down an octave (frequencies twice as small) from those with scale $m-1$, and is twice as small, but is log constant in size. Small (large) values of m are associated with dyadic blocks of large (small) support. This characteristic shows how wavelets are well suited to analyze piecewise smooth functions that may have sudden jumps, or function that are smooth and then suddenly rough. Since the wavelet is localized in time and frequency, wavelet coefficients over rough sections or over jumps in the function will be large relative to the coefficients over smooth sections.

3.1 Filter Banks

In theory the wavelet coefficient, $w_{m,n}$, equals the convolution of $x(t)$ with $\psi_{m,n}$. Since empirically $x(t)$ is only known on a discrete set of points, $w_{m,n}$ is calculated by a two-channel filter bank.[3] Thus, one never needs to explicitly calculate ψ, instead all the calculations are performed in terms of filter bank coefficients, h_k.[4]

A two-channel filter bank representation of the wavelet transform consists of a low-pass filter

$$\phi(t) \quad = \quad \sqrt{2} \sum_{k=0}^{2M-1} h_k\, \phi(2t-k) \tag{9}$$

[3] See Strang and Nguyen (1996) for a good introduction to wavelets and filter banks.

[4] The wavelet ψ can easily be calculated numerically from the values of filter bank coefficients.

where $\{h_k\}_{k=0}^{2M-1}$ are non-zero filter coefficients and a high-bandpass filter

$$\psi(t) \quad = \quad \sum_{k=0}^{2M-1} g_k \, \phi(2t-k) \tag{10}$$

where $g_k = (-1)^k h_{2M-1-k}$.

Daubechies (1988) has provided a sufficient set of non-zero values for $\{h_k\}_{k=0}^{2M-1}$ where ψ is compactly supported with the smallest possible support, $2M-1$, for a wavelet possessing M vanishing moments and whose regularity (number of derivatives and support size) increases linearly with M. Wavelets constructed with these filter coefficients are called Daubechies wavelets of order M. Larger values of M corresponds to smoother wavelets which contribute to better frequency location. However, this localization comes at the cost of a larger time support.

The low-pass filter coefficients, $\{h_k\}$, are a moving average filter that smoothes the high frequency traits (jumps, cusps, singularities) of a series. On the other hand, the high-bandpass filter coefficients, $\{g_k\}$, act as a differencing operator that captures the details filtered out by the low-pass filter.

The function ϕ is referred to as the scaling function since it does just that. Like the ideal high-bandpass wavelet, we can think of an ideal low-bandpass scaling function with the frequency response

$$\hat{\phi}(\omega) = \begin{cases} 1 & \text{if } \omega \in [-\pi, \pi], \\ 0 & \text{otherwise.} \end{cases}$$

Note that unlike ψ, ϕ is not required to have any vanishing moments. We do normalize ϕ so that $\int \phi(t)\, dt = 1$, i.e., $\sum_k h_k^2 = 1$.

Defining the dilations and translations of ϕ as

$$\phi_{m,n} \quad = \quad 2^{-m/2} \phi(2^{-m}t - n)$$

where $m, n \in \boldsymbol{Z}$, the filter bank definition of $\psi_{m,n}$ can be written as

$$2^{-m/2} \psi(2^{-m}t - n) \quad = \quad 2^{-m/2} \sum_{k=0}^{2M-1} g_k \, \phi(2^{-(m-1)}t - 2n - k).$$

Using this high-bandpass filter definition of $\psi_{m,n}$ it follows that

$$w_{m,n} \quad = \quad 2^{-m/2} \sum_{k=0}^{2M-1} g_k \int x(t) \, \phi(2^{-(m-1)}t - 2n - k) \, dt$$

$$= \quad 2^{-1/2} \sum_{k=0}^{2M-1} g_k \, s_{m-1,2n+k} \tag{11}$$

where $s_{m,n} = 2^{-m/2} \int x(t) \, \phi(2^{-m} - n) \, dt$ is the scaling coefficient. Thus, computing $w_{m,n}$ requires knowledge of $\{s_{m-1,n}\}_{n \in \boldsymbol{Z}}$.

Calculation of the scaling coefficient, $s_{m,n}$, can be performed by writing $\phi_{m,n}$ in terms of the lowpass filter as

$$2^{-m/2} \phi(2^{-m} t - n) = 2^{-(m-1)/2} \sum_{k=0}^{2M-1} h_k \, \phi(2^{-(m-1)} t - 2n - k).$$

Convoluting $x(t)$ with the above equation we find that $s_{m,n}$ equals

$$\begin{aligned} s_{m,n} &= 2^{-(m-1)/2} \sum_{k=0}^{2M-1} h_k \int x(t) \, \phi(2^{-(m-1)} t - 2n - k) \, dt \\ &= \sum_{k=0}^{2M-1} h_k \, s_{m-1,2n-k}. \end{aligned} \tag{12}$$

Thus, both $s_{m,n}$ and $w_{m,n}$ are calculated recursively from the smallest to the largest scale with the simple multiplication and addition operators of a two-channel filter bank.[5]

As an example of the two-channel filter wavelet transform suppose we observe the time series $x(t)$ for $t = 1, 2, \ldots, 2^{\max}$. Let $s_{0,n}$ be the output from the low-pass filter at the lowest possible scale, i.e., let $s_{0,n} = x(n)$ for $n = 1, \ldots, 2^{\max}$.[6] The value of $w_{m,n}$ and $s_{m,n}$ for $m = 1, 2, \ldots, \max$ and $n = 1, 2, \ldots, 2^{\max - m}$ are recursively calculated from the $x(t)$s by applying over and over the filters of Eq. (11) and (12).

This recursive algorithm known as the Fast Wavelet Transform is illustrated in Fig. 1 where

$$\begin{aligned} \mathbf{x} &= (x(1), x(2), \ldots, x(2^{\max}))', \\ \mathbf{s}_m &= (s_{m,1}, s_{m,2}, \ldots, s_{m,2^{\max - m}})', \end{aligned}$$

[5] Calculation of the wavelet coefficients is even easier than this when the filter coefficients are strategically placed in a sparse matrix. The wavelet transform then becomes a fast $\mathcal{O}(T)$ calculation, which is more efficient than the Fast Fourier Transform's $\mathcal{O}(T \log_2 T)$ calculation.

[6] In general, a function $f \in L^2(\Re)$ cannot be sampled as is but must first be smoothed by a low-bandpass filter to eliminate those points where f is possibly undefined. By defining $x(n) = s_{0,n}$, we assume x has been passed through a low-bandpass filter causing the signals values between $x(t)$ and $x(t+1)$ to equal $x(t+1)$ for $t = 1, 2, \ldots, 2^{\max} - 1$.

and

$$\mathbf{w}_m \quad = \quad (w_{m,1}, w_{m,2}, \ldots, w_{m,2^{\max - m}})'.$$

The wavelet coefficients for any value of m, represents the information lost when $\mathbf{s}_{m-1}$ is filtered by $\{h_k\}$, i.e., $\mathbf{w}_m$ contains the details or information needed to obtain $\mathbf{s}_{m-1}$ from $\mathbf{s}_m$. The box $\boxed{\uparrow 2}$ in Fig. 1 represents the decimation of the output from the filter by 2, i.e., discarding the odd sampled observation. By their definition the ideal low and high-bandpass filter, ϕ and ψ, include this decimation. Because the filters coefficients $\{h_k\}$ and $\{g_k\}$ are both applied to $\mathbf{s}_m$, twice as many observations as the length of $\mathbf{s}_m$ are created. Only half of this output is needed to completely represent or recover $\mathbf{s}_m$.

Because of the orthogonality of the filter banks for the Daubechies wavelet, the arrows in Fig. 1 can be reversed to synthesize $\mathbf{x}$ from its wavelet transform. In this wavelet synthesis, adding the details of $\mathbf{w}_m$ to the smoothed series $\mathbf{s}_m$ provides us with the representation of $\mathbf{x}$ at the next degree of resolution $\mathbf{s}_{m-1}$ with the highest resolution being the series $\mathbf{x}$.

The decimation in Fig. 1 illustrates a problem with the Fast Wavelet Transform. Both the low and high-bandpass filters are comprised of multiplying and adding together $2M$ coefficients with $\mathbf{s}_m$. Except for $M = 1$, the Fast Wavelet Transform will thus suffer from boundary effects where the support of the filter straddles the end points. At each dilation there will be $2M - 2$ wavelets coefficients that suffers this edge effect.

Possible solutions to the boundary effect include; *i*) doing nothing at all, *ii*) reflecting the series at the end points, in other words, setting $x(t) = x(T - t)$ for $t > T$ and $x(t) = x(-t)$ for $t \leq 0$, or *iii*) periodizing $\mathbf{x}$.[7] The do nothing approach introduces large wavelet coefficients at small m near the end points and results in the undesirable situation of having redundant or "too many" wavelet coefficients, whereas the reflection approach typically does not lead to an orthonormal basis. Thus, we choose to compute the wavelet transform by periodizing $\mathbf{x}$ [Daubechies (1994)]. Unless $\mathbf{x}$ was already periodic, this construction introduces a discontinuity at $t = 1$ and $t = 2^{max}$, which will show up as slow decay in the wavelet coefficient with a small dilation parameter. However, since the estimator developed in this paper is based on the the frequency domain properties of the wavelet coefficients,

[7] See Cohen, et.al. (1993) for a discussion on these and other methods.

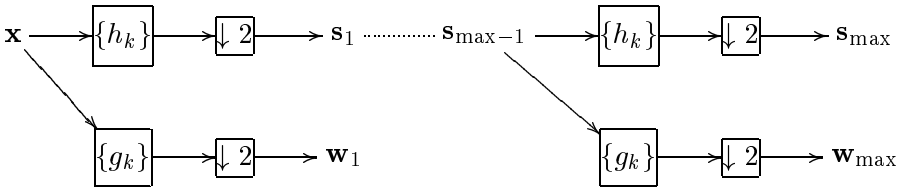

Figure 1: Schematic representation of the Fast Wavelet Transform.

these time dependent effects should not have to much of an affect on the properties of the estimator.

4 "Almost Diagonalized" Covariance Matrix

It is well known that the Fourier transform of a covariance matrix asymptotically obtains a diagonal matrix with the elements, $S(\omega_i)$, where $\omega_i = 2\pi i/T$, $i = 0, 1, \ldots, T-1$, on its diagonal [Grenander and Rosenblatt (1957, Chapter 7, pp. 226-240)]. This results shows that the complex exponentials, $e^{-i\omega t}$, are eigenfunctions to all Toeplitz matrices.

The wavelet transform of a covariance matrix results in a matrix that is "almost diagonal." The wavelet trades off exact diagonalization of the Toeplitz class of matrices in order to "almost diagonalize" the larger class of non-translations invariant matrix, e.g., non-Toeplitz matrices [Meyer and Coifman (1997)]. The Fourier transform is unable to diagonalize any other matrix operator but translation invariant ones. The wavelet transform gives up exact diagonals consisting of the spectrum evaluated at the Fourier frequencies, $2\pi i/T$, by providing an average of the spectrum over a particular octave. Thus, the wavelet is an "almost-eigenfunction" to both Toeplitz matrices and the larger class of non-translation invariant matrices.

4.1 Wavelet Coefficients Variance for ARFIMA Processes

From Eq. (8), the variance of $w_{m,n}$ can be written as

$$\begin{aligned} \mathrm{V}ar[w_{m,n}] &= E[w_{m,n}^2] && (13) \\ &= \int\int \psi_{m,n}(t)\,\gamma(|t-s|)\,\psi_{m,n}(s)\,dt\,ds && (14) \end{aligned}$$

where $\gamma(|s|) = E[x(t)x(t+|s|)]$ is the autocovariance function of $x(t)$. Because of the admissibility condition, the wavelet coefficients from a mean zero ARFIMA model will be the same as those from a nonzero mean process. Hence, without loss of generality we assume that $\mu = 0$. By a change of variable and application of Parseval's identity

$$\begin{aligned} \mathrm{V}ar[w_{m,n}] &= \int S_x(\omega)\, \hat{\psi}_{m,n}(\omega)\, \overline{\hat{\psi}_{m,n}(\omega)}\, d\omega \qquad (15) \\ &= 2^m \int S_x(\omega) \hat{\psi}(2^m\omega) \overline{\hat{\psi}(2^m\omega)}\, d\omega. \qquad (16) \end{aligned}$$

Using the definition of the ideal high-bandpass wavelet found in Eq. (6), the variance of the wavelet coefficient equals

$$\mathrm{V}ar[w_{m,n}] = 2^{m+1} \int_{2^{-m}\pi}^{2^{-m+1}\pi} S_x(\omega)\, d\omega. \qquad (17)$$

Substituting the ARFIMA process's spectrum found in Eq. (2) for $S_x(\omega)$, the variance of the wavelet coefficient from an ARFIMA process can be written as

$$\mathrm{V}ar[w_{m,n}] = \frac{2^{m+1}\sigma^2}{2\pi} \int_{2^{-m}\pi}^{2^{-m+1}\pi} \frac{|\Theta(e^{-i\omega})|^2}{|\Phi(e^{-i\omega})|^2} |1 - e^{-i\omega}|^{-2d}\, d\omega.$$

Note that $\mathrm{V}ar[w_{m,n}]$ is only a function of m. For a given dilation parameter, m, the wavelet coefficients of different translations, n, will have the same variance. In other words, the wavelet coefficients at a given dilation are homoskedastic, but are hetroskedastic over scale.

4.2 Uncorrelated Nature of Wavelet Coefficients with Different Scales

From Eq. (8) define the correlation between wavelet coefficients of different dilations, $m \neq m'$, and translations, $n, n' \in \boldsymbol{Z}$, as

$$\begin{aligned} \mathrm{C}ov(w_{m,n}, w_{m',n'}) &= E[w_{m,n} w_{m',n'}] \\ &= \int\int \psi_{m,n}(t)\, \gamma(|t-s|)\, \psi_{m',n'}(s)\, dt\, ds. \qquad (18) \end{aligned}$$

As in the previous section, by a change of variable and application of Parseval's identity, the wavelet coefficient's covariances frequency representation is

$$\mathrm{C}ov(w_{m,n}, w_{m',n'}) = 2^{(m+m')/2} \int S_x(\omega) e^{-i(2^m n - 2^{m'} n')\omega} \hat{\psi}(2^m\omega)\, \overline{\hat{\psi}(2^{m'}\omega)}\, d\omega. \qquad (19)$$

By the definition of the ideal high-bandpass wavelet found in Eq. (4),

$$\begin{aligned} Cov(w_{m,n}, w_{m',n'}) &= 2 \cdot 2^{(m+m')/2} \left[\int_{2^{-m}\pi}^{2^{-m+1}\pi} S_x(\omega)\ e^{-i(2^m n - 2^{m'} n')\omega}\ \cdot 0\ d\omega \right. \\ &\quad \left. + \int_{2^{-m'}\pi}^{2^{-m'+1}\pi} S_x(\omega)\ 0 \cdot e^{-i(2^m n - 2^{m'} n')\omega}\ d\omega \right] \\ &= 0 \end{aligned}$$

Thus, wavelet coefficients at different scales are uncorrelated.

4.3 Uncorrelated Nature of Wavelet Coefficients with Same Scales

For an ARFIMA process, Jensen (1999b) has shown that the autocovariance function of $w_{m,n}$ for a given value of m is a function of the absolute difference in the translation parameters, and for a fixed interval of the translation parameters, a function of the difference in the dilation parameters, i.e., the wavelet coefficients from an ARFIMA process are covariance stationary in time and scale. Jensen (1999b) also shows the wavelet coefficients autocovariance function decaying at the geometric rate, $\mathcal{O}(|2^{-j}k - 2^{-m}n|^{2(d-M)-1})$, where k and n are translation parameters and j and m are dilation parameters. Hence, the wavelet transform of an ARFIMA processes covariance matrix is a near diagonalization of the ARFIMA covariance matrix.

5 Approximate Maximum Likelihood Estimator

Let $\mathbf{w} = (\mathbf{w}_1', \mathbf{w}_2', \ldots, \mathbf{w}_{\max}')'$ be the $2^{\max} - 1 \times 1$ vector containing the wavelet coefficients. Since the wavelet coefficients are approximately uncorrelated, $\mathbf{w}$ will be distributed multivariate normal with vector mean zero and covariance matrix

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_1 \boldsymbol{I}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_2 \boldsymbol{I}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{\max} \end{bmatrix}$$

where $\sigma_m = Var[w_{m,n}]$ and $\boldsymbol{I}_m$ is a $2^{\max-m} \times 2^{\max-m}$ identity matrix.

It follows that the approximate log-likelihood function is

$$\mathcal{L}(d, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{w}) = -\frac{2^{\max} - 1}{2} \log 2\pi - \frac{1}{2} \sum_{m=1}^{\max} \left[2^{\max-m} \log(\sigma_m) + \frac{\mathbf{w}_m' \mathbf{w}_m}{\sigma_m} \right], \quad (20)$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$. Differentiating Eq. (20) with respect to σ^2 and setting the result equal to zero, the maximum likelihood estimator of σ^2 is found to equal

$$\hat{\sigma}^2 = \frac{1}{2^{\max} - 1} \left[\sum_{m=1}^{\max} \frac{\mathbf{w}'_m \mathbf{w}_m}{\sigma'_m} \right],$$

where $\sigma'_m = \sigma_m / \sigma^2$. Substituting $\hat{\sigma}^2$ into Eq. (20) we obtain the concentrated, approximate, log-likelihood function

$$\mathcal{L}(d, \boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{w}) \quad = \quad -\frac{2^{\max} - 1}{2} \log 2\pi - \frac{1}{2} \sum_{m=1}^{\max} \left[2^{\max - m} \log(\hat{\sigma}^2 \sigma'_m) + \frac{\mathbf{w}'_m \mathbf{w}_m}{\hat{\sigma}^2 \sigma'_m} \right]. \qquad (21)$$

The concentrated, approximate, log-likelihood function can be numerically maximized over the parameter space of d, $\boldsymbol{\phi}$, and $\boldsymbol{\theta}$ to obtain their approximate wavelet maximum likelihood estimators.

6 Simulations

In order to gauge the statistical inference of the approximate wavelet maximum likelihood estimators of d, $\boldsymbol{\phi}$, and $\boldsymbol{\theta}$ and its robustness to different parameter values, we test it against Fox and Taqqu's (1986) [FT] approximate frequency-domain MLE in a number of Monte Carlo simulations. We employ Hosking's (1984) algorithm to generate 100 mean zero ARFIMA processes corresponding to $d = \pm 0.25, \pm 0.15, \pm 0.05$ and one short memory parameter of either an ARFIMA(1,d,0) model corresponding to $\phi = -0.7, -0.2, 0.3, 0.8$, or an ARFIMA(0,$d$,1) model with $\theta = -0.8, -0.3, 0.5, 0.9$. For each ARFIMA model series of length $T = 2^9, 2^{10}, 2^{11}$ are generated and estimated with the approximate wavelet MLE and Fox and Taqqu MLE. The Monte Carlo mean-squared error (MSE) and bias of the two approximate MLE are then calculated.[8]

One of the concerns with wavelet analysis is answering the question of which wavelet to use. We prefer the Daubechies (1988) class of wavelets given their maximum regularity for a given number of vanishing moments. However, the question remains as to how many vanishing moments a wavelet should possess when used to estimate the parameters of an ARFIMA model.

[8] Tables reporting the empirical bias and mean squares of each estimator can be downloaded in *pdf*-format from ????

To answer this question we compute the wavelet coefficients for the simulated processes using a Daubechies wavelet with low regularity (Haar, M = 1) and a Daubechies wavelet with high regularity (D20, M = 10). The empirical bias and mean square error of the more regular approximate D20 wavelet MLE are significantly smaller for both the short and long memory parameters than the less regular approximate Haar wavelet MLEs. This comes as no surprise given that increasing the number of vanishing moments reduces the size of the wavelet's transition band (the set of frequencies $\{\omega : \hat{\psi}(\omega) \in (0,1)\}$), moving it closer to the ideal bandpass wavelet.

In most of the simulations there is hardly any noticeable difference between the approximate wavelet and approximate frequency-domain MLE's empirical bias and mean squared error of the short and long-memory parameters. Only when the moving average parameter of the ARFIMA(0,d,1) model approaches the unit circle does a clear difference appear in the empirical properties of the two estimators. When the moving average parameter is near the boundary of invertibility, the approximate frequency-domain MLE deteriorates as an estimator of both the fractional differencing and the moving average parameter. On the other hand, the empirical properties of the approximate wavelet MLE do not change from those estimated from models with the moving average parameter well within its parameter space.

Fig. 2 and 3 contain the boxplots of the approximate D20 wavelet estimator (d_W in the figure) and the approximate frequency-domain MLE (d_FT) of the differencing parameter, d, for ARFIMA(0,d,1) models with $\theta = -0.8,\ 0.9$, respectively. In the experiments with series of lenght $T = 512$, the boxplots illustrate how the frequency-domain approach often produces differencing parameter estimates that are significant different from their true value. When $\theta = 0.9$, the Fox and Taqqu estimator of d produces such estimates more than 5% of the time. Often these estimates fall outside of d's parameter space and in the region of nonstationarity, i.e., $\hat{d} \geq 0.5$. In the cases where $\theta = -0.8$, the number of these occurences decreases as the series length increases. However, when $\theta = 0.9$ the approximate frequency-domain MLE of those models have a negative differencing parameters produces nonstationary estimates of d even when $T = 1024, 2048$ observations.

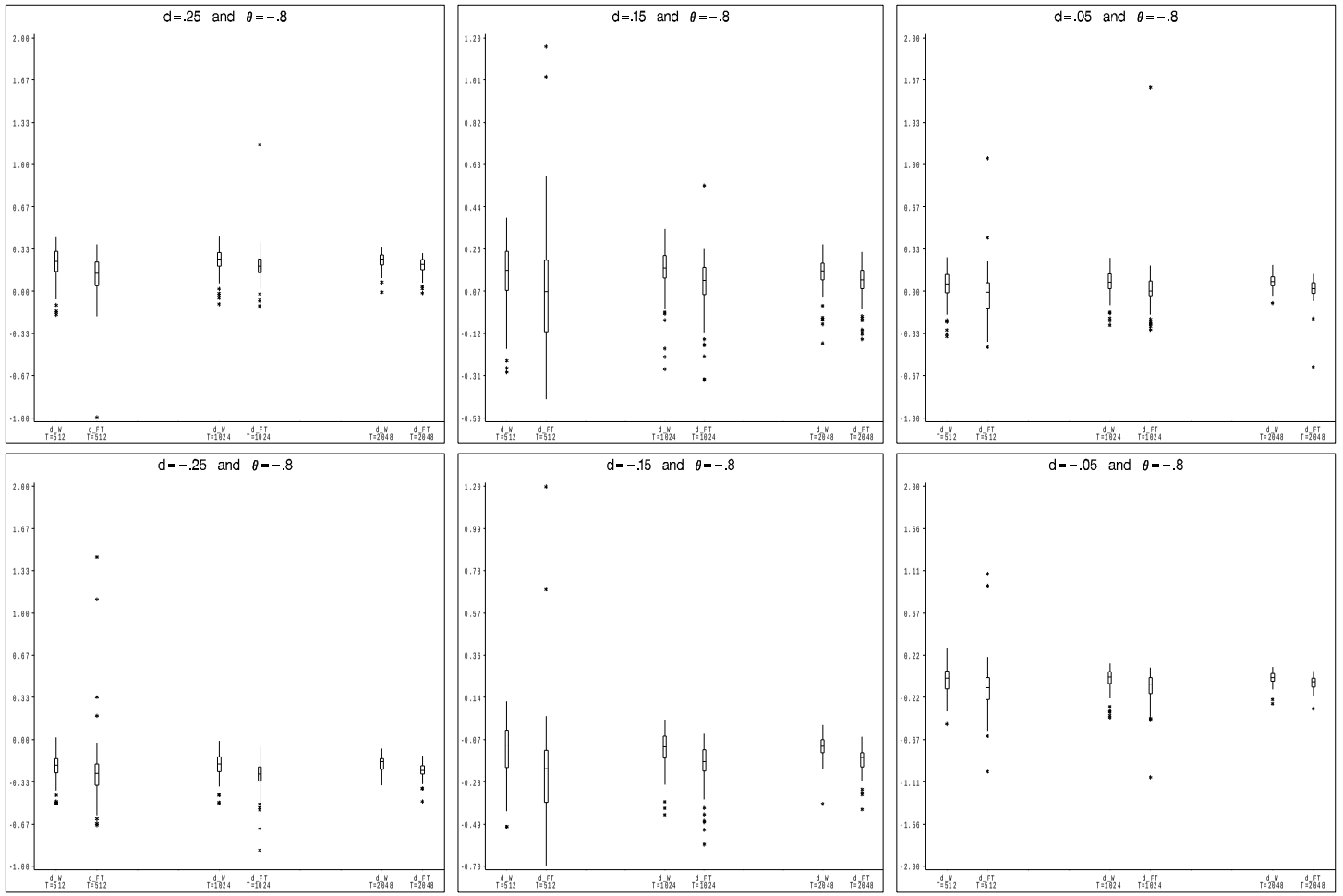

Figure 2: Boxplots of the D20 wavelet (d_W) and Fox and Taqqu estimator (d_FT) of d for the ARFIMA(0,d,1) model with $\theta = -0.8$ and $T = 512, 1024, 2048$ observations.

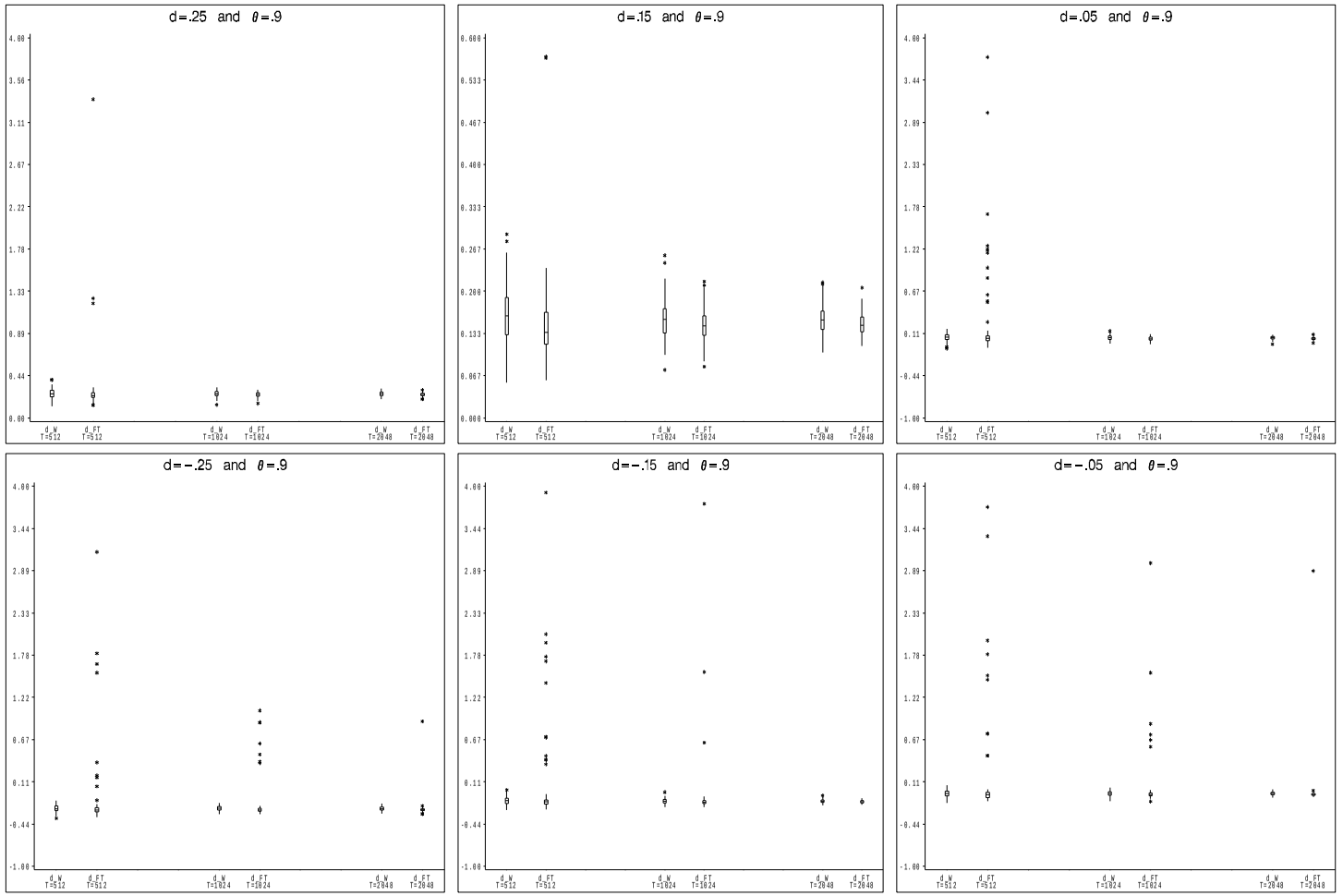

Figure 3: Boxplots of the D20 wavelet (d_W) and Fox and Taqqu estimator (d_FT) of d for the ARFIMA(0,d,1) model with $\theta = 0.9$ and $T = 512, 1024, 2048$ observations.

Fig. 4 and 5 contain the boxplots of the approximate D20 wavelet (t_W) and frequency-domain (t_FT) MLE of θ when $\theta = -0.8, 0.9$. In the boxplots of Fig. 4 the approximate wavelet MLE of θ is the estimator of choice. In each case where the ARFIMA(0,d,1) model has $\theta = -0.8$, the spread between the approximate wavelet estimator's upper and lower quartiles of θ is smaller than the frequency-domain approach. In those experiments where the differencing parameter of the ARFIMA(0,d,1) model are positive, the approximate frequency-domain MLE of θ regularly produces estimates that are outside the region of invertiblity. Inverting these moving average parameter estimates produces invertible ARFIMA models but at the cost of estimates that are indistiguishable from zero and a sizable distance away from the true moving average parameter value. These are the $*$ at the top of the Fig. 4 boxplots and at the bottom of Fig. 5 boxplots.

It is not as clear which of the two estimators is better when the simulated process follows a ARFIMA(0,d,1) model with $\theta = 0.9$. The boxplots of Fig. 5 show the spread between the approximate frequency-domain estimator's upper and lower quartile being sizeably smaller than the wavelet estimator. However, when ARFIMA(0,d,1) models differencing parameter is negative and $T = 512, 1024, 2048$, the Fox and Taqqu estimator's range (the difference between its max and min) is signficantly larger than the approximate wavelet estimators. This is also the case for postive values of the differencing parameter, but only when the series contains 512 observations.

In those experiments with $T = 512$ observations and negative values of d, the Fox and Taqqu estimate of the true moving average parameter of $\theta = 0.9$ is less than 0.1 over ten percent of the time, and three percent of the time when d is positive. When the number of observations increases to $T = 1024$, the number of these occurances drops for the negative values of d, but only to 5%. Given the strength of the wavelet estimator of d and the Fox and Taqqu MLEs proneness to produce unreasonably small estimates of θ, we feel the approximate wavelet maximum likelihood estimator is the superior estimator.

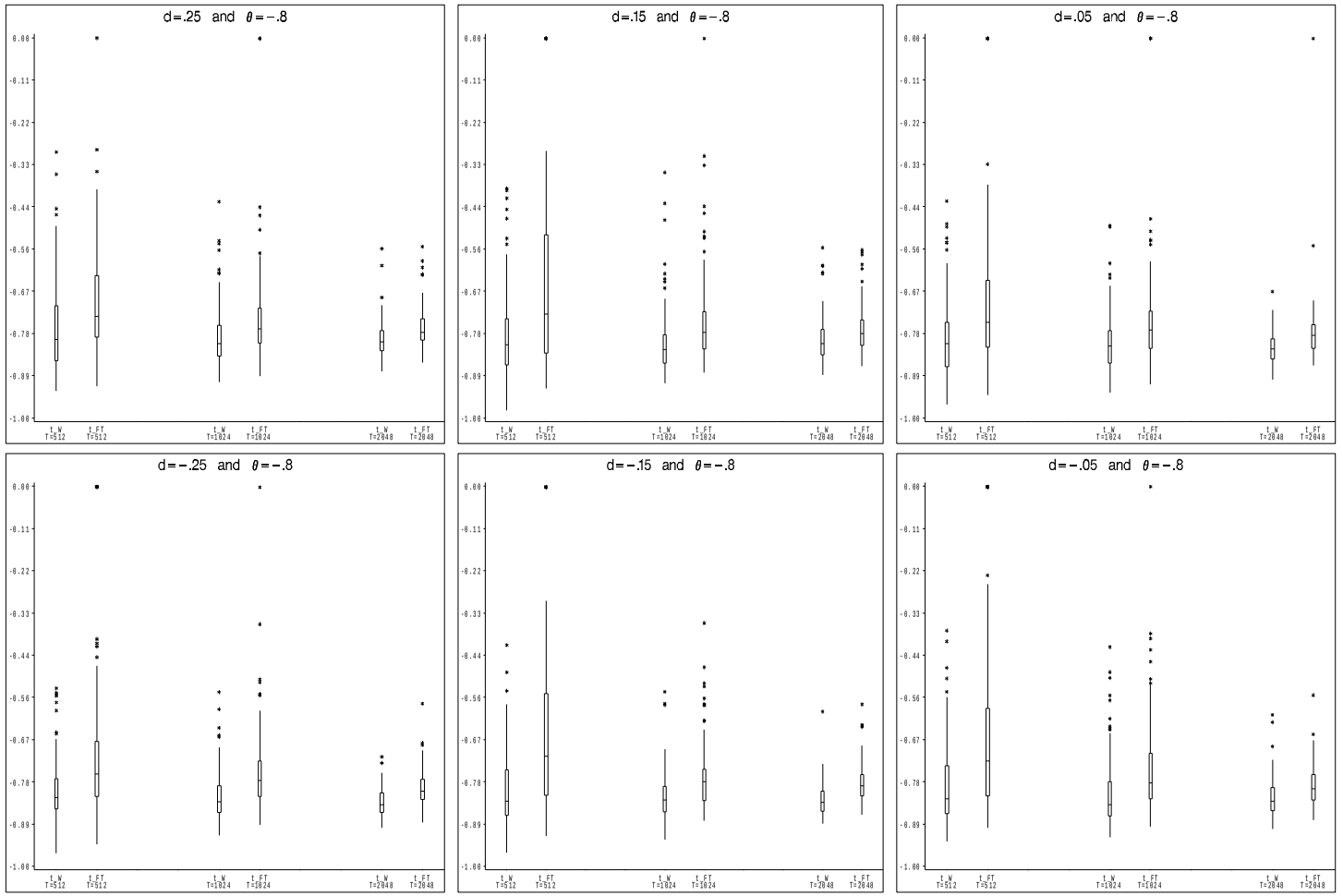

Figure 4: Boxplots of the D20 wavelet (t_W) and Fox and Taqqu estimator (t_FT) of θ for the ARFIMA(0,d,1) model with $\theta = -0.8$ and $T = 512, 1024, 2048$ observations.

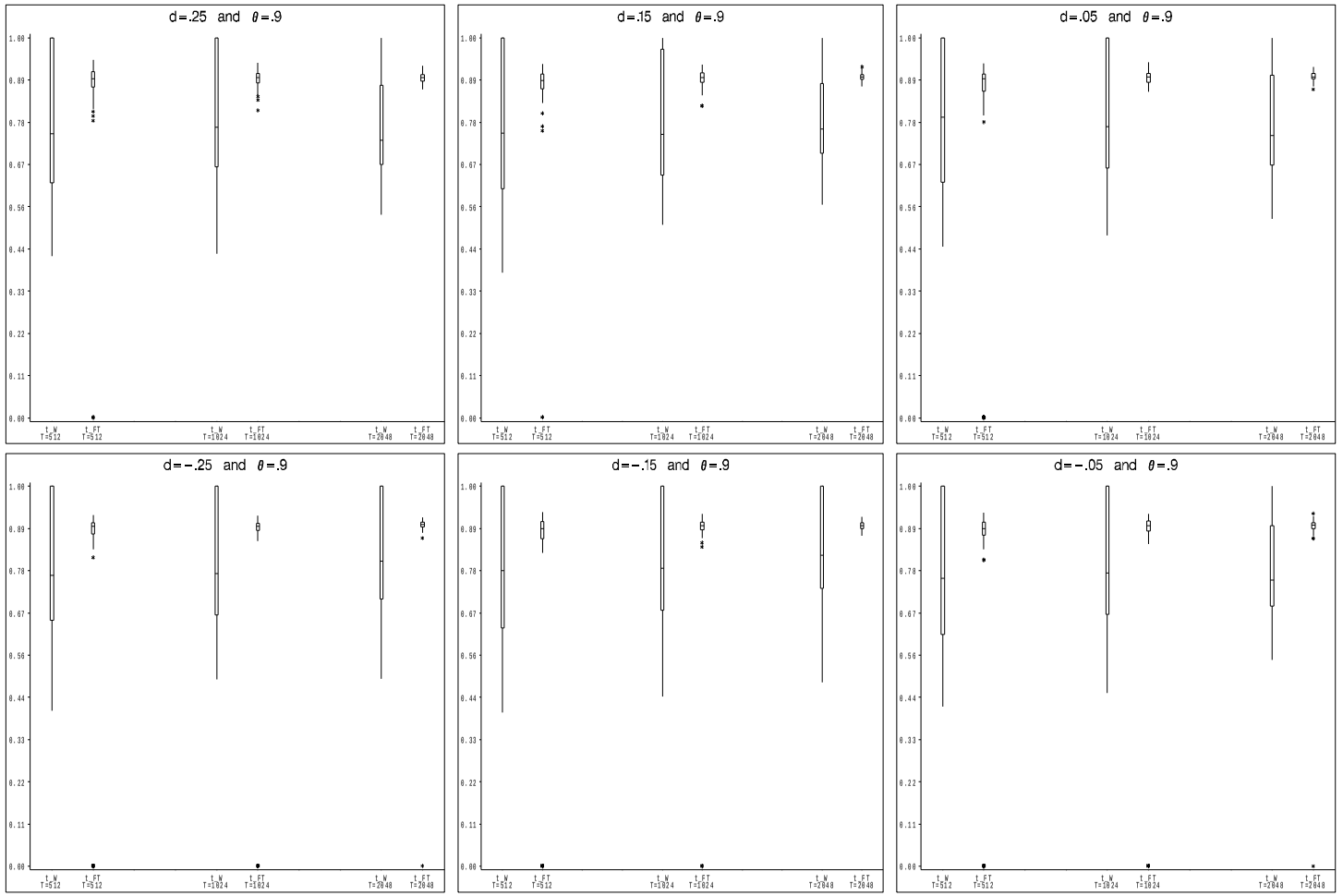

Figure 5: Boxplots of the D20 wavelet (t_W) and Fox and Taqqu estimator (t_FT) of θ for the ARFIMA(0,d,1) model with $\theta = 0.9$ and $T = 512, 1024, 2048$ observations.

To insure that our findings for the two approximate MLE are not dependent on Hosking's (1984) data generating mechanism, we preformed the simulations for the ARFIMA(0,d,1) models using the exact simulation algorithm of McLeod and Hipel (1978b). Calculating the autocovariance function with Sowell's (1992a) formula, we construct the ARFIMA models covariance matrix and use its Cholesky decomposition to premultiply a vector of standard normal innovations to obtain an exact ARFIMA series. We find no significant difference in the results of the two approximate MLE using the exact ARFIMA series. Once again the approximate frequency-domain MLE of the moving average parameter performs poorly as the parameter approaches the unit circle.

A possible explaination for the difference in the two approximate MLEs is found in their objective functions. A moving average model has a power spectrum that equals zero at frequency zero when $\theta = -0.8$, and at π when $\theta = 0.9$. Since the approximate frequency-domain approach minimizes the objective function

$$\frac{1}{T}\sum_{j=}^{T/2}\frac{I(\omega_j)}{g_x(\omega_j)}$$

where $I(\cdot)$ is the periodogram of $\mathbf{x}$, $g_x(\cdot) = 2\pi S_x(\cdot)/\sigma^2$, and $\omega_j = 2\pi j/T$, frequencies where S_x is zero cause the objective function to be undefined. In the case of the ARFIMA(0,d,1) model the objective function becomes unusually large.

The objective function of the approximate wavelet MLE found in Eq. (21) is a function of the integration of the power spectrum over dyadic blocks of the frequencies space. Hence, it does not suffer from the power spectrum equallying zero. The value of the power spectrum at those frequencies where the spectrum is zero are integrated together with the value of the power spectrum at other frequencies. The approximate wavelet likelihood function is thus better behaved over the ARFIMA model's parameter space, resulting in better estimates.

7 Summary

In this paper we provide a fast alternative maximum likelihood estimator to Fox and Taqqu (1986) frequency-domain MLE of the ARFIMA models short and long-memory parameters that relies on the high-bandpass interpretation of the wavelet and its near diagonalization of

the covariance matrix. Called the approximate wavelet MLE, this estimator comes from the approximation of the wavelet transforms likelihood function where the off diagonal elements of the wavelet coefficients covariance matrix are set to zero.

From our Monte Carlo simulations we determined that wavelets possessing a greater number of vanishing moments provide better statistical inference about the unknown short and long memory parameter values as measured by their empirical bias and mean square error. In comparison with the Fox and Taqqu MLE, the approximate wavelet MLE is robust to moving average parameters close to the boundary of invertibility. For series from a ARFIMA(0,d,1) model a moving average parameter close to either ± 1 caused, the empirical properties of the approximate frequency-domain estimator to deteriorate, whereas the approximate wavelet MLE were unaffected. Thus, we find the approximate wavelet MLE to be the more desirable of the two approximate MLE.

Acknowledgment

The author would like to thank the financial support of the University of Missouri Research Board. The contents of the paper have been presented at the Midwest Econometric Group Meetings, the Saint Louis Federal Reserve Bank, the International Workshop: *Statistics in Wavelets*, The Royal Society Meeting: *Wavelets: The Key to Intermittent Information?* and the Society for Nonlinear Dynamics and Econometrics.

References

Backus, D.K. and S.E. Zin (1993): "Long-Memory Inflation Uncertainty: Evidence from the Structure of Interest Rates," *Journal of Money, Credit, and Banking,* 681-700.

Baillie, R.T. and T. Bollerslev (1994): "Cointegration, Fractional Cointegration and Exchange Rate Dynamics," *Journal of Finance,* 49, 737-745.

Baillie, R.T., C.F. Chung, and M.A. Tieslau (1996): "Analysing Inflation by the Fractionally Integrated ARFIMA-GARCH Model," *Journal of Applied Econometrics,* 11, 23-40.

Brockwell, P.J. and R.A. Davis (1991): *Time Series: Theory and Methods,* 2nd ed., New

York: Springer-Verlag.

Cheung, Y.W. (1993): "Long Memory in Foreign-exchange Rates," *Journal of Business and Economic Statistics,* 11, 93-101.

Cheung, Y.W. and F.X. Diebold (1994): "On Maximum Likelihood Estimation of the Differencing Parameter of Fractionally-Integrated Noise with Unknown Mean," *Journal of Econometrics,* 62, 301-316.

Cohen, A., I. Daubechies, and P. Vial (1993): "Wavelets on the Interval and Fast Wavelet Transforms," *Applied and Computational Harmonic Analysis,* 1, 54-81.

Daubechies, I. (1988): "Orthonormal Bases of Compactly Supported Wavelets," *Communications on Pure and Applied Mathematics*, 41, 909-996.

Daubechies, I, (1994): "Two Recent Results on Wavelets: Wavelet Bases for the Interval, and Biorthogonal Wavelets Diagonalizing the Derivative Operator," in *Recent Advances in Wavelet Analysis,*, L.L Schumaker and G. Webb (ed.) Academic Press, San Diego, 237-257.

Devore, R., B. Jawerth, and B. Lucier (1992a): "Image Compression Through Wavelet Transform Coding," *IEEE Journal on Information Theory,* 38, 719-747.

Devore, R., B. Jawerth, and V. Popov (1992b): "Compression of Wavelet decompositions," *American Journal of Mathematics,* 114, 737-785.

Diebold, F. and G. Rudebusch (1991): "Is Consumption Too Smooth? Long Memory and the Deaton Paradox," *Review of Economics and Statistics,* 73, 1-9.

Ding, Z., C.W.J. Granger and R.F. Engle (1993): "A Long Memory Property of Stock Returns and a New Model," *Journal of Empirical Finance,* 1, 83-106.

Fox, R. and M.S. Taqqu (1986): "Large Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series," *Annals of Statistics,* 14, 517-532.

Grenander, U., and Rosenblatt, M. (1957): *Statistical Analysis of Stationary Time Series,* New York: Wiley.

Hassler, U. and J. Wolters (1995): "Long Memory in Inflation Rates: International Evidence," *Journal of Business and Economic Statistics,* 13, 37-45.

Hosking, J.R.M. (1984): "Modeling Persistence in Hydrological Time Series Using Fractional Differencing," *Water Resources Research,* 20, 1898-1908.

Jensen, M.J. (1999a): "Using Wavelets to Obtain a Consistent Ordinary Least Squares Estimator of the Long Memory Parameter," *Journal of Forecasting*, 18, 17-32.

Jensen, M.J. (1999b): "An Alternative Maximum Likelihood Estimator of Long-Memory Processes Using Compactly Supported Wavelets," *Journal of Economic Dynamics and Control,* (forthcoming).

Lo, A.W. (1991): "Long Term Memory in Stock Market Prices," *Econometrica,* 59, 1279-1313.

McCoy, E.J. and A.T. Walden (1996): "Wavelet Analysis and Synthesis of Stationary Long-Memory Processes," *Journal of Computational and Graphical Statistics*, 5, 1-31.

McLeod, A.J. and K.W. Hipel (1978a): "Simulation Procedures for Box-Jenkins Models," *Water Resources Research,* 14, 969-975.

McLeod, A.J. and K.W. Hipel (1978b): "Preservation of the Rescaled Adjusted Range, I. A Reassessment of the Hurst Phenomenon," *Water Resources Research,* 14, 491-519.

Meyer, Y. and R. Coifman (1997): *Wavelets: Calderón-Zygmund and Multilinear Operators,* trans. D. Salinger, New York: Cambridge University Press.

Ramsey, J.B. & Z. Zhang (1997): "The Analysis of Foreign Exchange Data Using Waveform Dictionaries," *Journal of Empirical Finance,* 4, 341-372.

Ramsey, J.B. & C. Lampart (1998a): "The Decomposition of Economic Relationships by Time Scale Using Wavelets: Money and Income," *Macroeconomics Dynamics,* 2, 49-71.

Ramsey, J.B. & C. Lampart (1998b): "The Decomposition of Economic Relationships by Time Scale Using Wavelets: Expenditure and Income," *Studies in Nonlinear Dynamics and Econometrics,* 3, 23-42.

Sowell, F.B. (1992a): "Maximum Likelihood Estimation of Stationary Univariate Fraction-

ally Integrated Time Series Models," *Journal of Econometrics,* 53, 165-188.

Sowell, F.B. (1992b): "Modeling Long Run Behavior with the Fractional ARIMA Model," *Journal of Monetary Economics,* 29, 277-302.

Strang, G. and T. Nguyen (1996): *Wavelets and Filter Banks,* Wellesley:Wellesley-Cambridge Press.

Whittle, P. (1951): *Hypothesis Testing in Time Series Analysis,* New York: Hafner.