An Alternative Maximum Likelihood Estimator of Long-Memory Processes Using Compactly Supported Wavelets

Mark J. Jensen
Department of Economics, University of Missouri
Columbia, MO 65211
econmj@showme.missouri.edu

September 17, 1997

Abstract

In this paper we apply compactly supported wavelets to the ARFIMA(p,d,q) long-memory process to develop an alternative maximum likelihood estimate of the differencing parameter, d, that is invariant to the unknown mean and model specification, and to the level of contamination. We show that this class of time series have wavelet transforms who's covariance matrix is sparse when the wavelet is compactly supported. It is shown that the sparse covariance matrix can be approximated to a high level of precision by a matrix equal to the covariance matrix except with the off-diagonal elements set to zero. This diagonal matrix is shown to reduce the order of calculating the likelihood function to an order smaller than those associated with the exact MLE method. We test the robustness of the wavelet MLE of the fractional differencing parameter to a variety of compactly supported wavelets, series length, and contamination by generating ARFIMA(p,d,q) processes for different values of p, d and q, and calculating the wavelet MLE estimate using only the main diagonal elements of its covariance matrix. In our simulations we find the wavelet MLE to be superior to the approximate frequency MLE when estimating contaminated ARFIMA(0,d,0), and uncontaminated ARFIMA(1,d,0) and ARFIMA(0,d,1) processes except when the MA parameter is close to one. We also find the wavelet MLE to be robust to model specification and as such is an attractive alternative semiparameter estimator to the Geweke-Hudak estimator.

Keywords: ARFIMA, Fractional Integration, Long-Memory, Maximum Likelihood, Wavelets

JEL Classification: C15; C22

1 Introduction

Wavelet analysis is a new development in the area of applied mathematics. They were ↓rst introduced in seismology [Morlet (1983)] to provide a time dimension to seismic analysis that Fourier analysis lacked. The wavelet's generality and strong results quickly made them useful in other disciplines, ranging from signal [Kronland-Martinet, Morlet and Grossman (1987)] to numerical analysis [Beylkin, Coifman, and Rokhlin (1991)].

By design the wavelet's usefulness is its ability to localize a process in time-scale space. At high scales, the wavelet has a small time support and is thus, better able to focus on short lived phenomena like singularity points. At low scales, the wavelet's time support is large, making it suited for identifying long periodic processes. By moving from high to low scales the wavelet zooms in on the process's behavior at a point in time, identifying either singularities or degrees of smoothness. Furthermore, the wavelet's ability to localize a series in the time-scale space directly leads to the computational e¡ciency of the wavelet representation of a $N \times N$ matrix operator by allowing the N largest wavelet coe¡cients to adequately represent the matrix [Devore, et. al. (1992a) and Devore, et. al. (1992b)].

In this paper we apply the class of wavelets with compact support to ARFIMA processes to produce a semiparametric Maximum Likelihood Estimator (MLE) of the fractional di↑erencing parameter. Unlike the semiparametric Geweke and Porter-Hudak (1983) estimator (GPH), which utilizes the periodogram of the series to approximate the power spectrum, we draw on the stationarity and self-similarity of the wavelet transform's second-order moments and the sparsity of the wavelet representation of the ARFIMA's covariance matrix to estimate the fractional di↑erencing parameter.

The wavelet MLE enjoys the advantage of having both the strengths of a MLE and a semiparametric estimator, while not su↑ering their known drawbacks. Boes, Davis and Gupta (1989) have shown that semiparametric methods perform more satisfactorily than the MLE methods when the model is misspeci↓ed. On the other hand, Cheung (1993) has found that under correct model speci↓cation the various MLE methods are superior to the semiparametric estimators. Like the GPH estimator and Tieslau et. al. (1996) nonparametric minimum distance estimator (MDE), but unlike Sowell's (1992) exact MLE or Fox and Taqqu's (1986) approximate MLE, the wavelet MLE's performance does not depend on correctly specifying the ARFIMA model. Nor does the wavelet MLE suffer the loss in efficiency nor the increase in bias the MDE exhibits under short-memory dynamics. Hence, the wavelet MLE enjoys the superiority of being an efficient maximum likelihood estimator and provides a estimator of the differencing parameter that can be calculated separately from the short-run parameters.[1]

Another of the wavelet MLE's many strengths is that it significantly reduces the order of calculating the likelihood function. Of the many empirical studies testing for long-memory, few have opted to use the more efficient, but computationally difficult exact MLE. Instead opting for the easy to use GPH approach. The reasoning, long-memory causes the covariance matrix of a ARFIMA process to be extremely dense and hence, difficult to invert. In order to calculate Sowell's (1992) exact MLE, the likelihood function must be maximized numerically. At each iteration of the optimization procedure the likelihood function is calculated, along with the inverse of the covariance matrix, a number of times. Such calculations are a mounting task even for today's fastest workstations, and a justifiable reason for not using Sowell's exact maximum likelihood on large financial time series data.[2]

We show that the sparsity of the wavelet transform's covariance matrix lowers the complexity of computing the wavelet transform's likelihood function below that of Sowell's MLE estimators to an order similar to the approximate frequency domain MLE of Fox and Taqqu (1986). More specifically, we find that the rapid decay in the elements of the wavelet representation of the ARFIMA covariance matrix enables us to avoid the taxing calculation of inverting the covariance matrix.

This sparcity also allows the wavelet MLE to be robust to white noise that may have contaminated the true time series. By its construction the wavelet transform captures the important features of the long-memory process in the few non-zero elements of its wavelet tranform's covariance matrix. Since white noise will be white noise in any orthogonal basis,

[1] **Although we do not attempt to estimate the ARMA parameters with the wavelet MLE, it is not beyond the scope of the wavelet MLE. Estimating the short-run parameters along with the fractional differencing parameter in a correctly identified model is a topic of future research.**

[2] **Both Sowell (1992) and Deriche and Tewfik (1993) use the Levinson algorithm to reduce the order of calculating a long-memory process's likelihood function by decomposing the covariance matrix. But even this leaves a daunting task for large data sets.**

in the wavelet basis the elements of the covariance matrix associate with the structure of long-memory will always stick out in the presence or non-presence of white noise.

An additional argument for using the wavelet MLE is its invariance to the mean of the ARFIMA process. When the fractional di↑erencing parameter is empirically estimated the mean of the ARFIMA process is unknown. In the past such situations required a tradeo↑ between the more e¡cient but computationally intensive feasible exact MLE (Sowell's MLE except with the data detrended by the sample mean, rather than by the true mean) and the less e¡cient approximate frequency domain MLE [Cheung and Diebold (1994)]. We eliminate this tradeo↑ by showing that the wavelet MLE, like the approximate frequency domain MLE, is invariant to the mean of the ARFIMA process and is at least as e¡cient as the feasible exact MLE.

The fact that the wavelet MLE is less intensive to calculate, robust to the process's unknown mean and possible contamination, and does not require the user to identify the model makes it an appealing alternative to the currently available fractional di↑erencing parameter estimators. The plan of the paper is as follows.

In Section 2, we present some basic wavelet theory as found in multiresolution analysis. Multiresolution analysis is an area of computer design concerned with graphing continuous functions or images in the discrete space of computer pixels. Section 3 de↓nes the class of ARFIMA processes and their long-memory behavior. Section 4 contains the second-order properties and the likelihood function of the wavelet coe¡cients from a ARFIMA series. In Section 5, an extensive array of simulations that ascertains the robustness of the wavelet estimate to sample size, zero padding, di↑erencing parameter, short-run dynamics, wavelet type, and contaminated data are presented. Section 6 summarizes our results.

2 Wavelet Theory

A wavelet is de↓ned as

$$\begin{aligned}\psi_{m,n}(t) &= a_0^{m/2}\psi\left(a_0^m(t - nb_0a_0^{-m})\right)\\ &= a_0^{m/2}\psi\left(a_0^m t - nb_0\right) \qquad (1)\end{aligned}$$

where $a_0 > 0$, $b_0 > 0$, and m and n are elements of $\boldsymbol{Z} = \{0, \pm 1, \pm 2, \ldots\}$ representing the dilation and translation of $\psi(t)$. As is commonly done we set $a_0 = 2$ and $b_0 = 1$, i.e., $\psi_{m,n}(t) = 2^{m/2}\psi(2^m t - n)$.[3]

The function $\psi(t) = \psi_{0,0}(t)$ is the 'mother' wavelet, which satisfies the admissibility condition, $\int \psi(t)dt = 0$, i.e., ψ oscillates and decreases rapidly to zero as $t \to \pm\infty$. The regularity of the wavelet can also be increased to include $\psi(t)$ with higher order vanishing moments, i.e., $\int x^r \psi(t)dt = 0$ where $r = 0, 1, 2, \ldots, M-1$, and/or $\psi(t) \in C^k$.

These necessary conditions insure that $\psi_{m,n}(t)$ is smooth and well localized in frequency and time space. If the Fourier transform, $\hat{\psi}(\omega)$, has sufficiently fast decay as $\omega \to \infty$, $\psi(t)$ can be interpreted as a high bandpass filter over the frequencies $|\omega| \in [C_1, C_2]$ where $0 < C_1 < C_2$. $\psi_{m,n}$ is then a high bandpass filter over $|\omega| \in [(C_1 + n)/2^m, (C_2 + n)/2^m]$.

By increasing and decreasing m and n the wavelet covers different frequencies and moments in time. At high frequencies (large m) the time translation, $n2^{-m}$, is small, enabling the wavelet to zoom in on jumps, cusps and singularity points. At low frequencies (small m) the translations are large, allowing $\psi_{m,n}$ to zoom out on the smoothness and periodicity of the series.

2.1 Multiresolution Analysis

Using multiresolution analysis, Mallat (1989) showed that a linear combination of the dilated and translated 'mother' wavelet form an orthogonal basis of the set of squared integrable functions, $L^2(\Re)$.[4] Multiresolution analysis decomposes $L^2(\Re)$ into a chain of closed subspaces

$$\cdots \subset V_{m-1} \subset V_m \subset V_{m+1} \subset \cdots$$

where V_m is the set of $L^2(\Re)$ functions with resolution 2^m. V_m can be thought of as a set of $L^2(\Re)$ continuous time series which have been discretely and uniformly sampled 2^m times.

[3] In the harmonic analysis literature $a_0 = 1/2$. Using 2 or its inverse is only a matter of preference and tradition.

[4] In this section we limit ourselves to signals in $L^2(\Re)$. However, wavelets span many function spaces. For example, Sobolev, Lipschitz, Besov and Hölder spaces are all spanned by wavelets [Meyer (1990) and Mallat and Hwang (1991)].

V_m's composition permits us to go up and down the chain of subspaces by rescaling V_m's functions, i.e., $x(t) \in V_m$ if and only if $x(2^{k-m}t) \in V_k$.

Define $\phi(t)$ to be the scaling function such that $\int \phi(t)\ dt = 1$ is in C^k with every derivative up to order k rapidly decreasing, and $\{\phi(t-n) : n \in \boldsymbol{Z}\}$ is a orthonormal basis of V_0. By the definition of V_m, it follows that

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

Daubechies (1988), provides a sufficient set of non-zero values for the coefficient sequence $\{h_k\}_{k=0}^{2M-1}$ which enables ϕ to be an orthogonal basis of V_0 with supp$\{\phi\} = [0, 2\text{M}-1]$. From the rescaling property of V_0 it follows that $\{\phi_{m,n}(t) = 2^{m/2}\phi(2^m t - n) : n \in \boldsymbol{Z}\}$ is a orthonormal basis of V_m.

Projections of $x(t) \in L^2(\Re)$ onto V_m, i.e., a 2^m resolution approximation of $x(t)$, can be represented as

$$\text{Proj}_{V_m}\, x(t) = \sum_n < x, \phi_{m,n} > \phi_{m,n}(t) \tag{3}$$

where the scaling coefficients, $\{< x, \phi_{m,n} >= \int_{-\infty}^{\infty} x(t)\phi_{m,n}(t)dt\ :\ n \in \boldsymbol{Z}\}$, completely characterizes this projection.

From the definition of V_m it is easy to show that $L^2(\Re) = \overline{\bigcup_{m \in \boldsymbol{Z}} V_m}$. At first glance it would seem that a linear combination over m and n of the scaling functions, $\{\psi_{m,n}\}$, would form a orthonormal basis of $L^2(\Re)$. This thought, however, is incorrect. Although $V_m \subset V_{m+1}$, $\{\phi_{m,n}\}$ is not contained in $\{\phi_{m+1,n}\}$, and furthermore, $\{\phi_{m,n}\}$ contains elements that are not orthogonal to the basis elements of V_{m+1}. Fortunately, other multiresolution analysis properties show $\{\psi_{m,n}\}$ to be an orthonormal basis of $L^2(\Re)$.

Let W_m be the orthogonal complement of V_m in V_{m+1}, i.e., $V_{m+1} = V_m \oplus W_m$. W_m can be thought of as the level of information lost when every other observation of a discrete series is discarded. Define the 'mother' wavelet as

$$\psi(t) = \sqrt{2} \sum_{k=0}^{2M-1} g_k \phi(2t-k) \tag{4}$$

where $g_k = (-1)^k h_{2M-1-k}$. From the Daubechies sufficient set of non-zero coefficients, ψ has the same regularity properties as ϕ, $\{\psi(t-n) : n \in \boldsymbol{Z}\}$ is a orthonormal basis of W_0, and

$\{\psi_{m,n} : n \in \boldsymbol{Z}\}$ an orthogonal basis of W_m. Since $L^2(\Re) = \overline{\bigcup_{m\in \boldsymbol{Z}} V_m}$ and $\{0\} = \bigcap_{m\in \boldsymbol{Z}} V_m$, the Hilbert space direct sum of W_m equals $L^2(\Re)$, i.e., $L^2(\Re) = \oplus_{m\in \boldsymbol{Z}} W_m$. Thus, the orthogonality of W_m establishes $\{\psi_{m,n} : m, n \in \boldsymbol{Z}\}$ as an orthogonal basis of $L^2(\Re)$ and any $x(t) \in L^2(\Re)$ can be represented as

$$x(t) = \sum_m \sum_n < x, \psi_{m,n} > \psi_{m,n}(t) \tag{5}$$

where the equality of Eq. (5) holds in the L^2 sense.

Daubechies (1988) constructed scaling and wavelet functions from the sufficient non-zero sequences, $\{h_k\}_{k=0}^{2M-1}$, and $\{g_k\}_{k=0}^{2M-1}$, to have supp$\{\psi\} = [-(\mathrm{M}-1), \mathrm{M}]$,

$$\int x^r \psi(t) dt, \qquad r = 0, 1, \ldots, M-1.$$

and $\psi \in C^k$, where $k \leq 2M - 2$.

This class of wavelets is called Daubechies wavelets of order M. The Daubechies wavelet has many desirable properties, but for our purpose their most useful property is possessing the smallest support for a given number of vanishing moments, M. Even though higher ordered Daubechies wavelets have a larger number of non-zero coefficients, $\{g_k\}_{k=0}^{2M-1}$, and hence a larger support, for a given M, the Daubechies wavelet has the fewest coefficients of any class of wavelets.

Observed time series are not continuous functions, but are finite sequences dependent on how often $x(t)$ is sampled. Hence, we can remove the coarser and finer resolution spaces of $L^2(\Re)$ from Eq. (5) for a observed series. For example, suppose that $x(t)$ has only been sampled quarterly at $t = 0, 1, \ldots, 2^{max} - 1$. If the coarsest resolution desired is 2^{min}, where $0 \leq min < max$, $x(t)$ can be projected onto V_{min}.[5] The wavelet representation of $x(t)$ would equal

$$x(t) = \sum_n < x, \phi_{min,n} > \phi_{min,n}(t) + \sum_{m \geq min} \sum_n < x, \psi_{m,n} > \psi_{m,n}(t). \tag{6}$$

Since $x(t)$ has been sampled quarterly 2^{max} times, the monthly, weekly, and daily values of $x(t)$ are unknown, and because $< x, \psi_{m,n} >$ measures the degree of information lost when $x(t)$ is sampled less often (or in other words measures the degree of information gained in the

[5] For example if the coarsest resolution were yearly then $min = max - 2$.

time series when the series is sampled more often), the wavelet coefficients, $\{<x,\psi_{m,n}>:$ $m \geq max\}$, are trivially equal to zero. Hence, the 2^{min} resolution representation of $x(t)$ found in Eq. (6) reduces to

$$x(t) \;=\; \sum_{n} <x,\phi_{min,n}> \phi_{min,n}(t) + \sum_{m=min}^{max-1} \sum_{n} <x,\psi_{m,n}> \psi_{m,n}(t).$$

$\psi_{m,n}$'s support can be thought of as $[n2^{-m}, (n+1)2^{-m}]$.[6] By normalizing the time interval of $x(t)$ to the unit interval, when $m = 0$ the translation value $n = 0$ causes the wavelet to cover the entire unit interval time domain. If $m = max-1$, $n = 0, 1, 2, \ldots, 2^{max-1} - 1$ is required. Hence, for a given scale, m, the translation parameter can only take on the values $n = 0, 1, 2, \ldots, 2^{m} - 1$. The wavelet representation of a times series $x(t)$ with 2^{max} observations is

$$x(t) \;=\; <x,\phi> \phi(t) + \sum_{m=0}^{max-1} \sum_{n=0}^{2^m-1} <x,\psi_{m,n}> \psi_{m,n}(t)$$

with the wavelet coefficients

$$\left\{ <x,\psi_{m,n(m)}>:\, m \in \{0,1,\ldots,max-1\},\; n(m) \in \{0,1,2,\ldots,2^m-1\} \right\}.$$

For future reference let $\mathcal{M} = \{0, 1, 2, \ldots, max-1\}$ and $\mathcal{N}(m) = \{0, 1, 2, \ldots, 2^m - 1\}$.

2.2 Example

One of the simplest and well know wavelets is the Haar function

$$\psi(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

This wavelet is the Daubechies wavelet of order 1. The subspaces, V_m, associated with this wavelet are $L^2(\Re)$ signals that are piecewise constant functions with length 2^{-m}. The corresponding scaling function, $\phi(t)$, is the indicator function $\chi_{[0,1)}$

3 Long-Memory Processes

Let $x(t)$ be the ARFIMA(p,d,q) process defined by

$$\Phi(L)(1-L)^d(x(t)-\mu) = \Theta(L)\epsilon(t) \tag{8}$$

[6] This is actually the support of the Daubechies wavelet with one vanishing moments, i.e., $M = 1$.

where

$$\Phi(L) = 1 + \phi_1 L + \phi_2 L^2 + \cdots + \phi_p L^p$$
$$\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$$

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

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

It is well known that at large lags $x(t)$'s autocovariance equals

$$\gamma(s) = E[x(t)\, x(t+s)]$$
$$\sim C(d, \Phi, \Theta)|s|^{2d-1} \qquad \text{as } s \to \infty \tag{9}$$

where

$$C(d, \Phi, \Theta) = \frac{\sigma_\epsilon^2}{\pi} \frac{|\Theta(1)|^2}{|\Phi(1)|^2} \Gamma(1-2d) \sin \pi d$$

and is hence, not summable when $d > 0$. Correspondingly, $x(t)$'s spectrum equals

$$S(\omega) \sim 1/|\omega|^{2d} \qquad \text{as } \omega \to 0. \tag{10}$$

and is unbounded at the origin for $d > 0$ [See Granger and Joyeux (1980), Hosking (1981), Brockwell and Davis (1993) and Baillie (1996)]. These asymptotic properties satisfy the many different definitions of long-memory that exist. The slow hyperbolic decay exhibited by $\gamma(s)$ satisfies the long-memory definition of Resnick (1987), and the unbounded spectrum at the origin satisfies the definition of McLeod and Hipel (1978).

Mandelbrot and Van Ness' (1968) statistical self-similiarty property can also be seen in Eq. (9) and (10). For any constant a, $S(\omega) = |a|^{2d} S(a\omega)$ and $\gamma(s) = a^{-2(d-1/2)}\gamma(as)$, i.e., the statistical properties of $x(t)$ remain the same regardless of how rapidly or slowly $x(t)$ is sampled.

4 Wavelet Analysis and Estimation of ARFIMA(p,d,q)

From our discussion on multiresolution analysis, wavelets display a form of self-similarity where at di↑erent values of m, $\psi_{m,n}$ has the same properties as ψ but over smaller or larger time intervals. In this section we show how the wavelet's self-similarity causes a ARFIMA process's wavelet coe¡cients to be stationary and self-similar in time space and stationary in scale space. We ↓nd the wavelet coe¡cient's second-order moments to be less correlated at and between scales when the wavelet is compactly supported.[7] The autocovariance function of the wavelet coe¡cients exhibits exponential decay like that of a ARMA process but between both time and scale. This damping behavior leads to a sparse covariance matrix that aids in reducing the order of computing the MLE of d.

Because the frequency domain approximate MLE is invariant to the unknown mean [Priestley (1992, p. 417)], and since the sample mean is an imprecise estimator when long-memory is present [Beran (1994, p. 7)], Cheung and Diebold (1994) found the approximate MLE to be an e¡cient and attractive alternative to the exact MLE when μ is unknown. Like the frequency domain approximate MLE, the wavelet MLE of d will be una↑ected by the unknown μ since the wavelet coe¡cients autocovariance function is invariant to μ (See Lemma 1).

4.1 Covariance Structure of ARFIMA's Wavelet Coefficients

Let $\gamma_{<x,\psi>}(m, j; n, k) = E[< x, \psi_{m,n} >< x, \psi_{j,k} >]$ represent the wavelet transform's autocovariance function. Using only the admissibility condition, $\int \psi(t)\, dt$, and $x(t)$'s asymptotic autocovariance function, we derive the following result.

Theorem 1 *As* $|2^{-j}k - 2^{-m}n| \to \infty$, *the normalized wavelet coefficients,* $2^{m(d+1/2)} < x(t), \psi_{m,n} >$, *associated with a ARFIMA(p,d,q) process with unknown mean* μ *and* $|d| < 0.5$ *are*

i) self similar for any scale m *and stationary time sequences, i.e., for any scale* m *the autocovariance function,* $\gamma_{<x,\psi>}(m, m; n, k)$, *is a unique function of the time interval*

[7] Tewfik and Kim (1992) and Flandrin (1991) have derived similar properties for fractional Brownian motion.

$k-n$,

ii) stationary scale sequences, i.e., for any time interval associated with n and k, the autocovariance function, $\gamma_{<x,\psi>}(m,j;n,k)$, is a unique function of $m-j$.

Proof: See Appendix A.

Like the original time series, $< x, \psi_{m,n} >$ is stationary and self-similar for fixed m. This property is important since the tool chosen for analysis should preserve the important statistical properties of the original signal. But by themselves the results of Theorem 1 provide us with little if anything in making the calculation of d's MLE any easier or better. By choosing $\psi(t)$ to have M vanishing moments and a compact support, we find that $< x, \psi_{m.n} >$ has short-memory or weak dependence over time and scale. This short-memory causes the wavelet transform's covariance matrix to be sparse. Both of these properties are the results of the following theorem.

Theorem 2 *If $\psi(t)$ has $M \geq 1$ vanishing moments with support $[-K_1, K_2]$ where $K_1 \geq 0$ and $K_2 > 0$ and $x(t)$ is a ARFIMA(p,d,q) process with unknown mean μ and $|d| < 0.5$ then for fixed scales $m \geq j$ the autocovariance, $\gamma_{<x,\psi>}(m,j;n,k)$, decays as $O(|2^{-j}k - 2^{-m}n|^{2(d-M)-1})$ for all k and n such that $|2^{-j}k - 2^{-m}n| > max(2^{-j}K_1 + 2^{-m}K_2, 2^{-m}K_1 + 2^{-j}K_2)$.*

Proof: See Appendix B

Theorem 2 can also be proven by taking the Taylor-series expansion of $|t + 2^{m+j}k - 2n|^{2d-1}$ around the center of S, the support of $\int \psi(s)\psi(2^{m-1}s - t)ds$, and eliminating the first $2M$ terms, to obtain

$$|\gamma_{<x,\psi>}(m,n;j,k)| < C|S|^{M+1} \sup_{t \in S}(|\frac{\partial^{2M} x}{\partial t^{2M}}|)$$

where C is dependent on ψ. Clearly, the magnitude of $\gamma_{<x,\psi>}(m,j;n,k)$ is small if either the interval S or the $2M$th order derivative of x is small.

From Theorem 2 if $|d| < 0.5$ any positive integer M insures that $\gamma_{<x,\psi>}(m,j;n,k)$ will decay exponentially at and between scales. Only when $d > 0.5$ does the number of vanishing moments of ψ become critical to the decay of $\gamma_{<x,\psi>}$.

Although a larger M increases the rate of decay, it comes at the cost of decreasing the set $\{(m,j;n,k) : |2^{-j}k - 2^{-m}n| > max(2^{-j}K_1 + 2^{-m}K_2, 2^{-m}K_1 + 2^{-j}K_2)\}$. As stated in Section 2, if ψ has M vanishing moments the smallest its support width can be is $2M-1$, i.e., $|K_1 - K_2| = 2M - 1$. Fortunately, Daubechies (1988) and Tew↓k and Kim (1992) have found that the "e↑ective" support of the Daubechies class of wavelet is smaller than its theoretical support, and that its "e↑ective" support grows at a slower rate than $2M-1$.

4.2 Sparsity of the Wavelet Coefficient's Covariance Matrix

Let

$$\begin{aligned} < X, \psi >' \;&=\; [< x(t), \psi_{0,0} >, < x(t), \psi_{1,0} >, < x(t), \psi_{1,1} >, < x(t), \psi_{2,0} >, \ldots \\ &\quad \ldots, < x(t), \psi_{max-1, 2^{max-1}-2} >, < x(t), \psi_{max-1, 2^{max-1}-1} >] \end{aligned}$$

and $\mathbf{\Sigma}_{<x,\psi>} = E[< X, \psi >< X, \psi >']$ be the $2^{max} - 1 \times 2^{max} - 1$ covariance matrix containing the elements $\gamma_{<x,\psi>}(m,j;n,k)$, where $m, j \in \mathcal{M}$, $n \in \mathcal{N}(m)$ and $k \in \mathcal{N}(j)$.

Unlike the Toeplitz form of the time domain's covariance matrix, the covariance matrix of the wavelet transform is an assembly of Toeplitz matrices at di↑erent scales. Fig. 1 shows how $\mathbf{\Sigma}_{<x,\psi>}$ for $max = 4$ is comprised of the autocovariance function $\gamma_{<x,\psi>}(m,j;n,k)$, where G_m is a zero-o↑set $2^m \times 2^m$ Toeplitz matrix of the form

$$G_m = [\gamma_{<x,\psi>}(m,m;|i-j|)]$$

with $m \in \mathcal{M}$, $i, j \in \mathcal{N}(m)$, and G_j^m is the $2^m \times 2^j$ matrix

$$\begin{bmatrix} \gamma_{<x,\psi>}(m,j;0,0) & \gamma_{<x,\psi>}(m,j;0,1) & \cdots & \gamma_{<x,\psi>}(m,j;0,2^j-1) \\ \gamma_{<x,\psi>}(m,j;1,0) & \gamma_{<x,\psi>}(m,j;1,1) & \cdots & \gamma_{<x,\psi>}(m,j;1,2^j-1) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{<x,\psi>}(m,j;2^m-1,0) & \gamma_{<x,\psi>}(m,j;2^m-1,1) & \cdots & \gamma_{<x,\psi>}(m,j;2^m-1,2^j-1) \end{bmatrix}$$

where $m, j \in \mathcal{M}$ and $\gamma_{<x,\psi>}(m,j;n,k)$ is the $n+1, k+1$ element of G_j^m. In addition $G_j^m = G_m^{j'}$.

By Theorem 2, $\mathbf{\Sigma}_{<x,\psi>}$ is a sparse matrix whose elements decay exponentially as one moves away from the the diagonal elements of G_m and G_j^m. This decay creates ↓nger-like bands that eminate from the ↓rst row and column of $\mathbf{\Sigma}_{<x,\psi>}$. Both the ↓nger-like bands

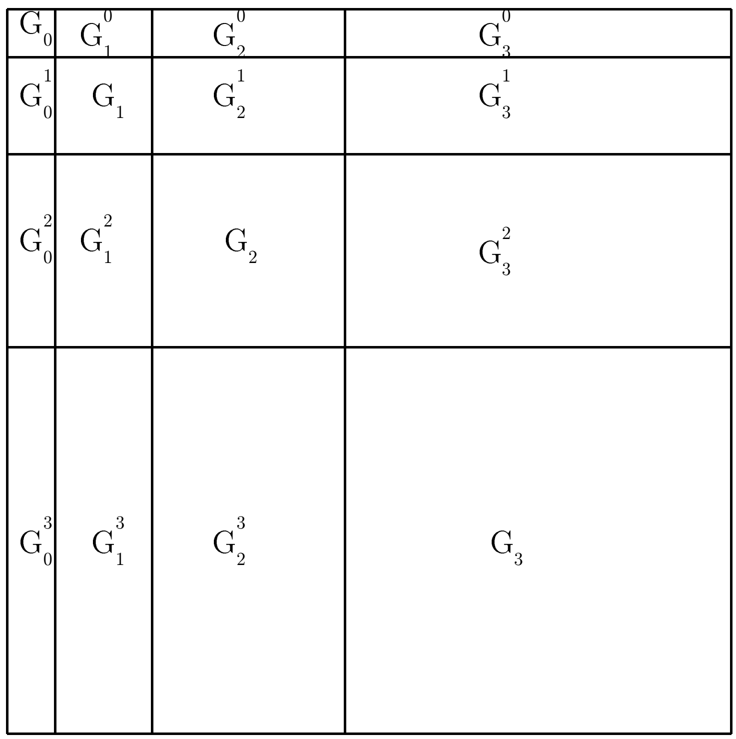

Figure 1: Wavelet Covariance Matrix $\boldsymbol{\Sigma}_{<x,\psi>}$

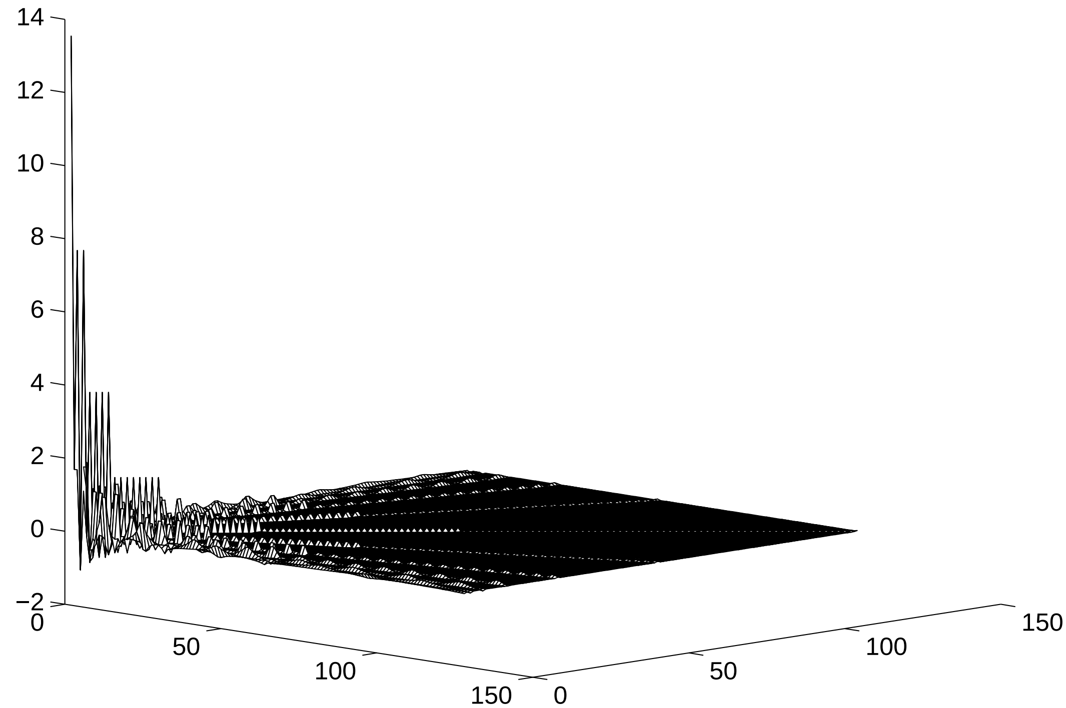

Figure 2: $\mathbf{\Sigma}_{<x,\psi>}$ of ARFIMA(1,0.35,1), $\phi_1 = \theta_1 = 0.8$, using the Haar wavelet

and decay of $\mathbf{\Sigma}_{<x,\psi>}$ are evident in Fig. 2 where the wavelet transform's autocovariance matrix for an ARFIMA(1,0.35,1) process with $\phi_1 = \theta_1 = 0.8$ is plotted.

Define $\mathbf{\Sigma}^B_{<x,\psi>}$ to be the matrix equal to $\mathbf{\Sigma}_{<x,\psi>}$ except with the elements,

$$\left\{\gamma_{<x,\psi>}(m,j;n,k) : |2^{-j}k - 2^{-m}n| \geq B > max(2^{-j}K_1 + 2^{-m}K_2, 2^{-m}K_1 + 2^{-j}K_2)\right\}$$

set equal to zero. $\mathbf{\Sigma}^1_{<x,\psi>}$'s only nonzero elements would be the main diagonal elements of $\mathbf{\Sigma}_{<x,\psi>}$, i.e., the main diagonal of G_m for $m = 0, 1, \ldots, max - 1$, and the finger-like diagonal elements of $\mathbf{\Sigma}_{<x,\psi>}$ associated with the main diagonal elements of G^m_j. The nonzero elements of $\mathbf{\Sigma}^B_{<x,\psi>}$ form a band of width B around the diagonals of $\mathbf{\Sigma}_{<x,\psi>}$.

4.3 Maximum Likelihood Estimate of d

The calculation time required to compute the covariance matrix and its inverse is substantially reduced when $\mathbf{\Sigma}^B_{<x,\psi>}$ is used to approximate $\mathbf{\Sigma}_{<x,\psi>}$. Hence, the sparsity of $\mathbf{\Sigma}_{<x,\psi>}$ provides a numerical benefit to calculating the likelihood function. The order of calculating

$(\mathbf{\Sigma}_{<x,\psi>})^{-1}$ drops from $\mathcal{O}(N^2)$ to nearly $\mathcal{O}(N)$ where $N = 2^{max} - 1$.[8] It follows that the order of calculating the likelihood function reduces to $\mathcal{O}(N)$.

Approximate the likelihood function

$$L_N(d \,|\, < X, \psi >) = (2\pi)^{-\left(\frac{2^{max}-1}{2}\right)} |\mathbf{\Sigma}_{<x,\psi>}(d)|^{-1/2} \exp\left[-\frac{1}{2} < X, \psi >' \mathbf{\Sigma}^{-1}_{<x,\psi>}(d) < X, \psi >\right] \tag{11}$$

with the function L_N^B equal to L_N except with $\mathbf{\Sigma}_{<x,\psi>}$ replaced by $\mathbf{\Sigma}^B_{<x,\psi>}$. Because ψ is compactly supported, the elements of $\mathbf{\Sigma}_{<x,\psi>}$ will be finite and hence, both L_N and L_N^B are well behaved and bounded from above.

The level of accuracy achieved with L_N^B is

$$\left\| L_N^B - L_N \right\| \leq \frac{C}{B^{2(M-d)}} \log_2 N \tag{12}$$

where C is a constant dependent on $\gamma(s)$ and ψ.[9] The level of accuracy associated with L_N^B is dependent on the value of M. A ψ possessing a larger M has faster decay in $\gamma_{<x,\psi>}$, reducing the approximation error of L_N^B.

Define the wavelet MLE of d as

$$\hat{d} = \arg \max_{|d|<0.5} L_N(d)$$

and the approximate wavelet MLE as

$$\tilde{d}_B = \arg \max_{|d|<0.5} L_N^B(d).$$

4.4 Banded MLE of d

In this section we draw on the results from the previous section, but rather than using $\mathbf{\Sigma}^B_{<x,\psi>}$ to approximate $\mathbf{\Sigma}_{<x,\psi>}$, we only use the main diagonal elements of $\mathbf{\Sigma}_{<x,\psi>}$ to

[8] The improvement from $\mathcal{O}(N \log N)$ to $\mathcal{O}(N)$ is based on the decay in the wavelet coefficient's second-order moments as the difference in their dilation parameters increase and the factorization of the Fast Wavelet Transform [Strang (1993)]. The covariance matrix then has closer to $\mathcal{O}(N \log(\log N))$ non-zero elements, which is nearly $\mathcal{O}(N)$. In a extensive study Meyer (1989) provides the conditions in which the wavelet representation of a matrix has $\mathcal{O}(N)$ non-zero elements.

[9] Beylkin, Coifman, and Rokhlin (1991) provide a similar result for the wavelet representation of linear Calderón-Zygmund operators. The result comes from approximating the sum of a decreasing sequence $\sum_{v=B}^{\infty} 1/v^m$, where $m > 1$, by $\int_B^{\infty} dv/v^m$, which equals $1/B^{m-1}$.

estimate d. To distinguish this estimator from $\tilde{d}_B$ we de↓ne this estimator as the banded MLE of d.

Let the observed time series be

$$\tilde{x}(t) = x(t) + \eta(t) \tag{13}$$

for $t = 0, 1, \ldots, 2^{max} - 1$ with the corresponding wavelet tranform vector,

$$\begin{aligned} < \tilde{X}, \psi >' \quad &= \quad [< \tilde{x}(t), \psi_{0,0} >, < \tilde{x}(t), \psi_{1,0} >, < \tilde{x}(t), \psi_{1,1} >, < \tilde{x}(t), \psi_{2,0} >, \ldots \\ &\quad \ldots, < \tilde{x}(t), \psi_{max-1,2^{max-1}-2} >, < \tilde{x}(t), \psi_{max-1,2^{max-1}-1} >\Big], \end{aligned}$$

where $x(t)$ is a ARFIMA(p,d,q) process with $|d| < 0.5$, 2^{max} observations and $\eta(t) \sim WN(0, \sigma_\eta^2)$ and is independent of $x(t)$. $\tilde{x}(t)$ can be thought of as a ARFIMA process that has been contaminated by some type of measurement or aggregation error. Although estimating d without contamination is a simpler problem, most economic time series su↑er from either aggregation or measurement error and hence, require a method that addresses such problems. Since wavelets measure the average 'uctuation of a signal at a given scale, wavelet transforms are less sensitive to the presence of noise. It seems only logical to make use of this attribute when constructing an estimator of d.

$< \tilde{X}, \psi >$'s covariance matrix is approximated by

$$\mathbf{\Sigma}_1 = \text{diag}(\tilde{\sigma}_{0,0}, \tilde{\sigma}_{1,0}, \tilde{\sigma}_{1,1}, \ldots, \tilde{\sigma}_{\max-1,2^{\max-1}-2}, \tilde{\sigma}_{\max-1,2^{\max-1}-1})$$

where

$$\begin{aligned} \tilde{\sigma}_{m,n} \quad &= \quad E\left[< x, \psi_{m,n} >^2\right] + E\left[< \eta, \psi_{m,n} >^2\right] \\ &= \quad C(d, \Phi, \Theta) 2^{-m(2d-1)-m} V_\psi(d) + \sigma_\eta^2 \\ &= \quad \sigma^2 2^{-m2d} + \sigma_\eta^2, \end{aligned} \tag{14}$$

$\sigma^2 = C(d, \Phi, \Theta) V_\psi(d)$, $V_\psi(d) = \int |t|^{2d-1} \Lambda(1,t)\, dt$, and $\Lambda(1,t) = \int ds\, \psi(s)\psi(s-t)$ for $m \in \mathcal{M}$ and $n \in \mathcal{N}(m)$. If follows that the approximate likelihood function is

$$L(\boldsymbol{\theta}) \quad = \quad \prod_{m \in \mathcal{M}} \prod_{n \in \mathcal{N}(m)} \frac{1}{\sqrt{2\pi \tilde{\sigma}_{m,n}}} \text{exp}\left[-\frac{< \tilde{x}, \psi_{m,n} >^2}{2\tilde{\sigma}_{m,n}}\right]$$

where $\hat{\theta} = (d, \sigma^2, \sigma_\eta^2)$, and the approximate log-likelihood is

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{m \in \mathcal{M}} \sum_{\mathcal{N}(m)} \left[\frac{< x, \psi_{m,n} >^2}{\tilde{\sigma}_{m,n}} + \ln 2\pi \tilde{\sigma}_{m,n} \right]$$

or as the concentrated approximate log-likelihood function

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{m \in \mathcal{M}} \# \mathcal{N}(m) \left[\frac{\hat{\tilde{\sigma}}_{m,n}}{\tilde{\sigma}_{m,n}} + \ln 2\pi \tilde{\sigma}_{m,n} \right] \tag{15}$$

where $\hat{\tilde{\sigma}}_{m,n} = \frac{1}{\#\mathcal{N}(m)} \sum_{n \in \mathcal{N}(m)} < x, \psi_{m,n} >^2$ and $\#\mathcal{N}(m)$ denotes the number of elements in the set.

$\mathcal{L}(\boldsymbol{\theta})$ is bounded from above and has a maximum when

$$\boldsymbol{\theta} \in \boldsymbol{\Theta} = \left\{ (d, \sigma^2, \sigma_\eta^2) : |d| < 0.5, \sigma^2 \geq 0, \sigma_\eta^2 \geq 0 \right\}.$$

Occasionally, $\mathcal{L}(\boldsymbol{\theta})$ will have more than one maximum. If this occurs the additional maximums are from trivial boundary cases where $\sigma^2 = 0$ and/or $\sigma_\eta^2 = 0$. Thus, the behavior of Eq. (15) insures that a numerical algorithm will converge to the banded MLE if initialized with values from $\boldsymbol{\Theta}$. Wornell and Oppenheimer (1992) have provided an EM-algorithm that maximizes $\mathcal{L}(\boldsymbol{\theta})$, in which $< X, \psi >$ is the complete data and $< \tilde{X}, \psi >$ the censored data.[10]

5 Simulation

Since Theorem 1 and 2 rely on $|2^{-j} - 2^{-m}n| \to \infty$ and the short-run dynamics parameters of $\Phi(L)$ and $\Theta(L)$ are not parameterized in the wavelet MLE of d, we conduct Monte Carlo experiments and estimate the banded MLE of $\boldsymbol{\theta}$ from artificially generated ARFIMA processes for different wavelets, series length, and contamination levels.

In each experiment the wavelet coefficients, $< x, \psi_{m,n} >$, were computed with the compact Haar wavelet (Daubechies with $M = 1$) and the Daubechies wavelet with $M = 10$. From a calculation point of view we prefer the Haar wavelet since its coefficients can be calculated exactly at each scale and increased regularity is not needed in analysing long-memory process with wavelets. With finite data it is not always possible to exactly calculate

[10] McCoy and Walden (1996) have developed a estimator similar to the banded MLE, but include scaling coefficients in the likelihood function, and do not take into consideration contaminated data.

		BANDED				GPH		APPROX. MLE	
		HAAR		DAUB. M=10					
d	N	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS
0.05	2^7	0.0064	-0.0044	0.0060	0.0054	0.0799	0.0007	0.0029	-0.0033
0.15	2^7	0.0063	0.0058	0.0070	0.0253	0.0789	0.0075	0.0032	0.0137
0.25	2^7	0.0081	0.0163	0.0105	0.0473	0.0708	-0.0067	0.0044	0.0386
0.35	2^7	0.0217	0.0642	0.0243	0.1024	0.0779	0.0095	0.0063	0.0582
0.45	2^7	0.0229	0.1047	0.0285	0.1389	0.0741	0.0246	0.0086	0.0742
0.05	2^8	0.0031	0.0009	0.0032	0.0096	0.0486	-0.0006	0.0014	0.0026
0.15	2^8	0.0028	0.0089	0.0038	0.0295	0.0404	0.0062	0.0019	0.0250
0.25	2^8	0.0034	0.0213	0.0054	0.0497	0.0472	0.0014	0.0035	0.0458
0.35	2^8	0.0077	0.0467	0.0110	0.0830	0.0419	0.0010	0.0056	0.0647
0.45	2^8	0.0126	0.0885	0.0179	0.1168	0.0468	0.0042	0.0081	0.0822
0.05	2^9	0.0012	0.0022	0.0014	0.0099	0.0270	-0.0050	0.0007	0.0087
0.15	2^9	0.0015	0.0137	0.0023	0.0316	0.0297	0.0091	0.0016	0.0300
0.25	2^9	0.0021	0.0229	0.0040	0.0511	0.0307	0.0045	0.0032	0.0501
0.35	2^9	0.0039	0.0438	0.0079	0.0794	0.0279	0.0115	0.0051	0.0668
0.45	2^9	0.0082	0.0779	0.0129	0.1042	0.0271	-0.0021	0.0078	0.0846
0.05	2^{10}	0.0007	0.0045	0.0008	0.0102	0.0177	0.0069	0.0004	0.0088
0.15	2^{10}	0.0008	0.0130	0.0016	0.0309	0.0190	0.0030	0.0013	0.0314
0.25	2^{10}	0.0011	0.0228	0.0031	0.0501	0.0188	0.0070	0.0029	0.0511
0.35	2^{10}	0.0025	0.0410	0.0065	0.0767	0.0199	0.0032	0.0050	0.0682
0.45	2^{10}	0.0063	0.0713	0.0102	0.0944	0.0178	0.0067	0.0076	0.0854

Table 1: ARFIMA(0,d,0), no contamination

all of the wavelet coe¡cients. The more regular (large M) a wavelet is, the larger its support. Hence, at coarser scales the wavelet straddles the data, resulting in boundary a↑ects.

In the ↓rst experiment we generated 1000 ARFIMA(0,d,0), fractionally integrated white noise processes with and without contamination for $d = 0.05, 0.15, 0.25, 0.35, 0.45$ and $N = 2^7, 2^8, 2^9, 2^{10}$ observations. Since we found no ill-e↑ects from zero-padding (when N was at least half way to the next highest integer power of 2) or truncating (when N was less than half way to the next highest power of 2) for this series or any of the others generated, we only report the results from series with integer powers of 2. In each simulation the starting values for the EM algorithm were $\theta = (\log_2(0.1)/2, 0.1, 0.1)$ and the convergence criterion was 0.001.

		BANDED				GPH		APPROX. MLE	
		HAAR		DAUB. M=10					
d	N	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS
0.05	2^7	0.0066	-0.0144	0.0061	-0.0093	0.0799	-0.0074	0.0030	0.0104
0.15	2^7	0.0068	-0.0220	0.0064	-0.0066	0.0788	-0.0101	0.0058	0.0528
0.25	2^7	0.0082	-0.0351	0.0070	-0.0119	0.0713	-0.0315	0.0160	0.1141
0.35	2^7	0.0187	-0.0328	0.0174	-0.0031	0.0782	-0.0244	0.0612	0.2407
0.45	2^7	0.0465	-0.0398	0.0392	-0.0090	0.0758	-0.0452	0.0141	0.1041
0.05	2^8	0.0032	-0.0090	0.0031	-0.0015	0.0487	-0.0085	0.0016	0.0161
0.15	2^8	0.0031	-0.0188	0.0029	-0.0016	0.0404	-0.0101	0.0053	0.0632
0.25	2^8	0.0037	-0.0267	0.0029	-0.0049	0.0476	-0.0191	0.0158	0.1197
0.35	2^8	0.0080	-0.0389	0.0053	-0.0137	0.0424	-0.0248	0.0606	0.2491
0.45	2^8	0.0217	-0.0330	0.0175	-0.0164	0.0485	-0.0455	0.0154	0.1006
0.05	2^9	0.0013	-0.0075	0.0013	-0.0011	0.0272	-0.0125	0.0012	0.0220
0.15	2^9	0.0015	-0.0134	0.0014	0.0012	0.0296	-0.0050	0.0053	0.0677
0.25	2^9	0.0021	-0.0239	0.0014	-0.0020	0.3090	-0.0119	0.0158	0.1230
0.35	2^9	0.0037	-0.0376	0.0022	-0.0132	0.0278	-0.0077	0.0599	0.2432
0.45	2^9	0.0118	-0.0343	0.0091	-0.0205	0.0285	-0.0373	0.0110	0.1017
0.05	2^{10}	0.0007	-0.0023	0.0007	-0.0007	0.0177	-0.0012	0.0008	0.0220
0.15	2^{10}	0.0009	-0.0138	0.0007	0.0010	0.0190	-0.0097	0.0050	0.0688
0.25	2^{10}	0.0011	-0.0227	0.0006	-0.0018	0.0188	-0.0057	0.0155	0.1233
0.35	2^{10}	0.0023	-0.0378	0.0010	-0.0143	0.0201	-0.0116	0.0599	0.2441
0.45	2^{10}	0.0069	-0.0269	0.0050	-0.0170	0.0181	-0.0188	0.0106	0.1011

Table 2: ARFIMA(0,d,0), contamination

Tables 1 and 2 report the ↓ndings of the banded MLE's mean squared error (MSE) and its level of bias with and without contamination using the Haar, and Daubechies-10 wavelet. For comparison, the GPH and approximate frequency domain MLE's MSE and bias are also reported.

With or without contamination, there was no considerable di↑erence between the MSE of the Haar and Daubechies-10 wavelet based estimator. However, there was a sizeable improvement in the absolute value of the estimator's bias when the Daubechies-10 wavelet was used with contaminated data. For the non-contaminated data the bias of the Daubechies-10 wavelet estimator was noticeably larger than with the Haar wavelet.

For ↓xed N the bias of both wavelet based estimators increased as d moved away from the origin.[11] This behavior was also reported by Cheung and Diebold (1994) for the feasible exact MLE. But the most notable feature of Table 1 and 2 is how much larger the MSE of the GPH is than either the banded or approximate MLE. Although the MSE for the banded MLE grows as d increases, for every sample size the GPH estimate's MSE is at least twice as large, and in some cases twelve times as large, as the banded MLE estimate's MSE.[12]

In comparison with the approximate MLE, the banded MLE was superior in regards to its absolute bias with and without contamination. Out of the twenty Monte Carlo experiments without (with) contamination, the approximate MLE's absolute bias was smaller than the banded MLE in only four (one) cases (three of which occurred when $N = 2^7$).

Determining the "better" estimator from the MSE is not as easy, given that out of the twenty cases with uncontaminated data the approximate MLE produced the smallest MSE in twelve of them. However, this edge was reduced to six cases when the series were contaminated. Except for the extreme cases, $d = 0.05, 0.45$, the relative efficiency, $MSE_{banded}/MSE_{approx.}$, for the contaminated data were almost all substantially less than 0.6. And given the smaller relative absolute bias, $|Bias|_{banded}/|Bias|_{approx}$, at these extreme values of d, the banded MLE appears to be the "better" estimator under contamination. Where the approximate MLE's large bias was offset by its small MSE when the series were uncontaminated, the increase in its bias for the contaminated series was not countered by a smaller MSE, instead it was compounded by a larger MSE. In contrast, for the few cases where the absolute bias of the banded MLE increased under contamination the MSE was either the same or smaller.

Comparing the MSE of the banded MLE when $N = 2^7, 2^8, 2^9$ to the MSE of the exact, and feasible exact found in Cheung and Diebold (1994), we find the banded MLE to perform marginally better than the exact MLE and substantially better than the feasible exact MLE with respect to the MSE and level of bias when $0.05 \leq d \leq 0.25$. However, the same can not be said for $0.35 \leq d < 0.5$. For these parameter values the MSE and bias of the banded MLE are significantly larger than those for the exact MLE, but are only slightly larger than

[11] Though not reported in Table 1 this behavior was also found for the banded MLE when d moved in the negative direction away from zero.

[12] Sowell (1992) found similar results for the exact maximum likelihood estimate of d.

those found for the feasible exact MLE. Cheung and Diebold's (1994) simulation results also show that the exact MLE's level of bias decreases as N grows, whereas from Table 1 there is no distinguishable pattern for the banded MLE.

The empirical properties of the banded MLE for $0.35 \leq d < 0.5$ are very similar to those found by Tieslau, Schmidt and Baillie (1996) for their minimum distance estimator (MDE). Since the sample autocorrelation of a fractional integrated white noise process is not $\sqrt{N}$ consistent for $d \geq 0.25$, Tieslau, Schmidt and Baillie (1996) found their MDE to have zero asymptotic efficiency against the MLE when $0.25 \leq d < 0.5$. This might suggest that the sample autocovariance of the wavelet coefficients are $\sqrt{2^m}$ not consistent when $0.35 \leq d < 0.5$.

The results of the banded and approximate MLE for the ARFIMA(1,d,0) model $(1 + \phi L)(1 - L)^d x(t) = \epsilon(t)$ are presented in Table 4 and the results from the ARFIMA(0,d,1) model $(1-L)^d x(t) = (1+\theta L)\epsilon(t)$ are presented in Table 5. Like those from the ARFIMA(0,d,0) processes, the banded MLE's level of bias tended to increase under both models as d moved away from the origin. Unlike the findings of Sowell (1992) for the exact MLE, the banded MLE's level of bias did not increase for negative values of ϕ, but the bias did increase for values of θ near one. The MSE of the banded MLE for the ARFIMA(1,d,0) model approached the theoretical variance $\frac{6}{\pi^2 N}$ as N increased and regardless of ϕ being positive or negative its MSE stayed the same.

In comparison with the approximate MLE, the banded MLE's mean squared error and absolute bias were almost always smaller for the ARFIMA(1,d,0) processes and when the MA parameter, θ, of the ARFIMA(0,d,1) process equaled -0.8. Out of the 48 Monte Carlo experiments with ARFIMA(1,d,0) process in only two cases ($N = 2^{10}$ and $d = -0.25$ for $\phi = 0.8, -0.7$) was the appoximate MLE's mean square smaller than the banded MLE's with a relative efficiency close to 1.1. In only 8 experiments was the approximate MLE's absolute bias smaller than the banded MLEs. These cases occurred for large N showing that the difference between the banded and approximate MLE's MSE and bias diminished as N increased. But for small sample sizes the banded MLE clearly dominates the approximate MLE for ARFIMA(1,d,0) processes.

d	N	ϕ	BANDED MSE	BANDED BIAS	APPROX. MLE MSE	APPROX. MLE BIAS	ϕ	BANDED MSE	BANDED BIAS	APPROX. MLE MSE	APPROX. MLE BIAS
-0.25	2^7	0.8	0.0099	0.0401	0.1114	-0.1290	-0.7	0.0104	0.0376	0.1405	-0.2001
-0.15	2^7	0.8	0.0086	0.0193	0.1338	-0.1816	-0.7	0.0087	0.0169	0.1285	-0.1728
-0.05	2^7	0.8	0.0067	-0.0011	0.1400	-0.2012	-0.7	0.0070	-0.0019	0.1519	-0.2192
0.05	2^7	0.8	0.0064	-0.0148	0.1374	-0.2142	-0.7	0.0066	-0.0154	0.1440	-0.2183
0.15	2^7	0.8	0.0071	-0.0301	0.1645	-0.2480	-0.7	0.0071	-0.0296	0.1566	-0.2256
0.25	2^7	0.8	0.0079	-0.0382	0.1684	-0.2619	-0.7	0.0081	-0.0363	0.1534	-0.2174
-0.25	2^8	0.8	0.0052	0.0402	0.0311	-0.0682	-0.7	0.0052	0.0394	0.0299	-0.0679
-0.15	2^8	0.8	0.0036	0.0166	0.0320	-0.0760	-0.7	0.0037	0.0159	0.0315	-0.0742
-0.05	2^8	0.8	0.0032	0.0030	0.0365	-0.0864	-0.7	0.0032	0.0028	0.0372	-0.0871
0.05	2^8	0.8	0.0031	-0.0122	0.0221	-0.0762	-0.7	0.0030	-0.0118	0.0282	-0.0811
0.15	2^8	0.8	0.0034	-0.0231	0.0337	-0.0100	-0.7	0.0034	-0.0218	0.0329	-0.0941
0.25	2^8	0.8	0.0038	-0.0283	0.0354	-0.0103	-0.7	0.0037	-0.0280	0.0353	-0.0988
-0.25	2^9	0.8	0.0032	0.0384	0.0062	-0.0257	-0.7	0.0032	0.0372	0.0062	-0.0278
-0.15	2^9	0.8	0.0019	0.0180	0.0064	-0.0313	-0.7	0.0019	0.0178	0.0057	-0.0308
-0.05	2^9	0.8	0.0015	0.0038	0.0062	-0.0298	-0.7	0.0016	0.0037	0.0062	-0.0300
0.05	2^9	0.8	0.0015	-0.0103	0.0068	-0.0315	-0.7	0.0015	-0.0102	0.0068	-0.0313
0.15	2^9	0.8	0.0016	-0.0172	0.0070	-0.0322	-0.7	0.0016	-0.0169	0.0069	-0.0320
0.25	2^9	0.8	0.0019	-0.0237	0.0060	-0.0308	-0.7	0.0018	-0.0230	0.0061	-0.0308
-0.25	2^{10}	0.8	0.0022	0.0351	0.0020	-0.0066	-0.7	0.0022	0.0346	0.0019	-0.0072
-0.15	2^{10}	0.8	0.0011	0.0191	0.0021	-0.0100	-0.7	0.0011	0.0188	0.0022	-0.0114
-0.05	2^{10}	0.8	0.0007	0.0044	0.0021	-0.0106	-0.7	0.0007	0.0043	0.0020	-0.0107
0.05	2^{10}	0.8	0.0006	-0.0063	0.0022	-0.0183	-0.7	0.0006	-0.0063	0.0021	-0.0171
0.15	2^{10}	0.8	0.0009	-0.0158	0.0046	-0.0420	-0.7	0.0009	-0.0156	0.0044	-0.0408
0.25	2^{10}	0.8	0.0011	-0.0216	0.0084	-0.0515	-0.7	0.0011	-0.0214	0.0075	-0.0486

Table 3: Results for ARFIMA(1,d,0) process.

The most notable strength of the banded MLE over the approximate were the cases $|d| = 0.05$. Regardless of the sample size or short-memory parameter (excluding the ARFIMA(0,d,1) process with $\theta = 0.9$), the banded MLE's level of bias and MSE were sizeably smaller when $|d| = 0.05$ than those found with the approximate MLE.

Sowell (1992) found the exact MLE bias and MSE increased for ARFIMA(1,d,0) processes as ϕ approached negative one. In Table 3, the absolute bias and MSE of the banded MLE remains essentially constant for either positive or negative values of ϕ. Furthermore, the absolute bias and MSE of the banded and exact MLE are nearly identical when $\phi = 0.8$.

			BANDED		APPROX. MLE			BANDED		APPROX. MLE	
d	N	θ	MSE	BIAS	MSE	BIAS	θ	MSE	BIAS	MSE	BIAS
-0.25	2^7	0.9	0.0735	-0.1274	0.0298	-0.0705	-0.8	0.0099	0.0544	0.0298	-0.0705
-0.15	2^7	0.9	0.1191	-0.1464	0.0266	-0.0806	-0.8	0.0067	0.0267	0.0266	-0.0806
-0.05	2^7	0.9	0.0981	-0.0998	0.0260	-0.0898	-0.8	0.0058	0.0035	0.0260	-0.0898
0.05	2^7	0.9	0.0521	-0.0253	0.0247	-0.0885	-0.8	0.0058	-0.0191	0.0247	-0.0885
0.15	2^7	0.9	0.0290	0.0395	0.0262	-0.0882	-0.8	0.0071	-0.0379	0.0262	-0.0882
0.25	2^7	0.9	0.0328	0.0727	0.0248	-0.0828	-0.8	0.0081	-0.0489	0.0248	-0.0828
-0.25	2^8	0.9	0.0310	-0.1022	0.0094	-0.0412	-0.8	0.0058	0.0549	0.0094	-0.0412
-0.15	2^8	0.9	0.0452	-0.1080	0.0104	-0.0460	-0.8	0.0034	0.0270	0.0104	-0.0460
-0.05	2^8	0.9	0.0204	-0.0515	0.0099	-0.0459	-0.8	0.0025	0.0071	0.0099	-0.0459
0.05	2^8	0.9	0.0115	-0.0064	0.0103	-0.0485	-0.8	0.0029	-0.0139	0.0103	-0.0485
0.15	2^8	0.9	0.0117	0.0474	0.0097	-0.0466	-0.8	0.0033	-0.0293	0.0097	-0.0466
0.25	2^8	0.9	0.0143	0.0751	0.0096	-0.0400	-0.8	0.0042	-0.0389	0.0096	-0.0400
-0.25	2^9	0.9	0.0115	-0.0850	0.0040	-0.0216	-0.8	0.0045	0.0554	0.0040	-0.0216
-0.15	2^9	0.9	0.0130	-0.0821	0.0042	-0.0232	-0.8	0.0021	0.0287	0.0042	-0.0232
-0.05	2^9	0.9	0.0087	-0.0384	0.0041	-0.0224	-0.8	0.0014	0.0085	0.0041	-0.0224
0.05	2^9	0.9	0.0044	-0.0047	0.0040	-0.0231	-0.8	0.0013	-0.0117	0.0040	-0.0231
0.15	2^9	0.9	0.0063	0.0518	0.0042	-0.0226	-0.8	0.0018	-0.0243	0.0042	-0.0226
0.25	2^9	0.9	0.0081	0.0701	0.0043	-0.0231	-0.8	0.0025	-0.0369	0.0043	-0.0231
-0.25	2^{10}	0.9	0.0092	-0.0858	0.0017	-0.0101	-0.8	0.0035	0.0536	0.0017	-0.0101
-0.15	2^{10}	0.9	0.0079	-0.0753	0.0020	-0.0112	-0.8	0.0015	0.0301	0.0020	-0.0112
-0.05	2^{10}	0.9	0.0034	-0.0322	0.0018	-0.0117	-0.8	0.0006	0.0085	0.0018	-0.0117
0.05	2^{10}	0.9	0.0029	-0.0080	0.0018	-0.0113	-0.8	0.0006	-0.0091	0.0018	-0.0113
0.15	2^{10}	0.9	0.0039	0.0526	0.0019	-0.0131	-0.8	0.0011	-0.0238	0.0019	-0.0131
0.25	2^{10}	0.9	0.0058	0.0688	0.0019	-0.0124	-0.8	0.0018	-0.0344	0.0019	-0.0124

Table 4: Results for ARFIMA(0,d,1) processes.

Hence, when $\phi = -0.7$ the banded MLE's MSE and absolute bias was smaller than the exact MLE.

In Table 4, the MSE of the banded MLE is smaller than those found by Sowell (1992) for the exact MLE and the bias of the two estimators are almost the same when $N = 2^7$ and $\theta = -0.8$. Because of the banded MLE poor performance for ARFIMA(0,d,1) processes with $\theta = 0.9$, it is not surprising that the exact MLE's bias and MSE are smaller in this case.

Except for the ARFIMA(0,d,1) processes with large positive values of θ, the banded

MLE is generally as good if not better than the exact and approximate MLE. Furthermore, unlike Tieslau, Schmidt and Baillie (1996) MDE which is asymptotically biased and less efficient when estimating d in a ARFIMA(p,d,q), the banded MLE of d is robust to these short-run parameters. Where the exact and approximate MLE require the short-memory parameters to be identified correctly in order to keep their small sample bias down, the likelihood function maximized by the banded MLE is the same regardless of the order of the short-memory parameters. Hence, with the banded MLE the user is not troubled by model identification problems.[13]

5.0.1 Computation Time

If the issue of computation time is also considered the banded MLE estimate of the fractional differencing parameter is superior to the exact MLE. Even though Sowell (1992) recursively calculates the inverse of the covariance matrix with the $\mathcal{O}(N^2)$ Levinson algorithm, the banded MLE is only $\mathcal{O}(N)$. This difference in the number of computations is important when one recognizes that these calculations are performed at each iteration of the optimization procedure, and that tick-by-tick data with observations numbering 16,000 are being tested for long-memory.

In his comparison of the exact MLE's computation time, Sowell (1992) addressed the difference in the time required to calculate the exact MLE likelihood function versus the nonfractional ARMA model's likelihood function for the same N. Computational time comparisons performed in this manner are misleading. The comparison between two algorithms should consider how the calculation time of the two methods increases relative to one another as N increases. Two methods that take the same amount of time evaluating the likelihood function at small N does not guarantee that they will take the same amount of time at large N.

[13] See Boes et. al. (1989) and Schmidt and Tschernig (1995) for a discussion on the increase in the bias of the MLE when the model is misspecified.

6 Conclusion

In this paper we have synthesized wavelet analysis with ARFIMA processes to derive a maximum likelihood estimator of the differencing parameter. This estimator utilizes the second-order statistical properties of the ARFIMA's wavelet transform. We found that the wavelet transform's second-order moments for this class of signals were stationary and self-similar in time-space, stationary in scale-space, and invariant to the process's mean. It was also shown that with a compactly supported wavelet the transform's have weak dependence over both time and scale space. Rather than calculating the entire covariance matrix, this weak dependance allows us to approximate the matrix to high degrees of precision with a matrix who's off-diaganol elements are set to zero. We find that this approximating covariance matrix significantly reduced the order of calculating the likelihood function since inversion of the approximate matrix is of $\mathcal{O}(\mathcal{N})$ where before it was at least $\mathcal{O}(N^2)$.

We tested the banded wavelet MLE of d, which uses only the main diagonal covariance elements, by performing a number of Monte Carlo simulations. The robustness of the banded MLE was determined by testing it against different compactly supported wavelets with both high and low orders of regularity, a wide range of ARFIMA processes with and without contamination, and to different lengths of processes with and without zero-padding and truncation. The simulations revealed that the choice of the wavelet, its order of regularity, nor zero-padding had any substantial effect on the banded MLE and did not seem to be critical to the estimator.

The Monte Carlo experiments also revealed a smaller MSE for the banded MLE relative to the GPH estimator. When compared with the exact and approximate MLE, the banded MLE preformed better in regards to MSE and bias for contaminated ARFIMA(0,d,0) processes, and uncontamined ARFIMA(1,d,0) processes, and except for MA parameters near one ARFIMA(0,d,1) processes. In addition, whereas the approximate MLE's bias increased as the degree of misspecification of the short-memory parameters increases, the banded wavelet MLE is unaffected. Given that the MLE methods are superior to the GPH estimator when the model is correctly specified, these results make the banded MLE a strong candidate for replacing the GPH estimator as the desired semiparametric estimator of long-memory processes.

Considering that the exact MLE is a order N^2 calculation and the banded MLE is an order N calculation, and that the banded MLE is invariant to drift and model specification, we feel the banded MLE is superior to the exact and approximate MLE. Overall we believe the wavelet MLE is an attractive alternative estimator of the fractional differencing parameter.

A Lemmas

Lemma 1 enables us to generalize any ARFIMA model with an unknown mean μ to have mean zero.

Lemma 1 *Let $x(t)$ be a ARFIMA(p,d,q) process with unknown mean μ and $x'(t)$ be a ARFIMA(p,d,q) process with mean zero, then $< x'(t), \psi_{m,n} > = < x(t), \psi_{m,n} >$.*

Proof:

$$\begin{aligned} < x'(t), \psi_{m,n} > &= < x(t) - \mu, \psi_{m,n} > \\ &= \int (x(t) - \mu)\psi_{m,n} dt \\ &= \int x(t)\psi_{m,n} dt - \int \mu\psi_{m,n} dt \end{aligned}$$

By the admissibility condition $\int \mu\psi_{m,n} dt = 0$. Hence,

$$< x'(t), \psi_{m,n} > \;=\; < x(t), \psi_{m,n} > \text{ Q.E.D.}$$

In proving Theorem 2 we will need the following Lemma.

Lemma 2 *If $\psi(t)$ has $M \geq 1$ vanishing moments then $\Lambda(2^{m-j}, t)$ has $2M$ vanishing moments.*

Proof:

$$\begin{aligned} \int dt\, t^k \Lambda(2^{m-j}, t) &= \int dt\, t^k \int ds\, \psi(2^{m-j}s - t)\, \psi(s) \\ &= -\int dt \int ds\, (2^{m-j}s - t)^k \psi(t)\psi(s) \\ &= -\int dt \int ds \sum_n \left(\begin{array}{c} m \\ n \end{array} \right) (2^{m-j}s)^{k-n}(-t)^n \psi(t)\psi(s) \\ &= 0 \qquad \text{for } k < 2M. \text{ Q.E.D.} \end{aligned}$$

B Proof of Theorem 1

Without loss of generality let $x(t)$ be a ARFIMA(p,d,q) with mean zero, and $|d| < 0.5$. The autocovariance function of $< x, \psi_{m,n} >$ is

$$
\begin{aligned}
\gamma_{<x,\psi>}(m,j;n,k) &= E\left[< x(t), \psi_{m,n} >< x(s), \psi_{j,k} >\right] \\
&= E\left[2^{\frac{m+j}{2}} \int dt \int ds \, x(t)\, x(s)\, \psi(2^m t - n)\, \psi(2^j s - k)\right] \\
&= 2^{\frac{m+j}{2}} \int dt \int ds \, E[x(t)x(s)]\, \psi(2^m t - n)\psi(2^j s - k) \\
&= 2^{\frac{m+j}{2}} \int dt \int ds \, \gamma(|t - s|)\psi(2^m t - n)\, \psi(2^j s - k) \\
&= 2^{\frac{-(m+j)}{2}} \int dt \int ds \, \gamma(|2^{-m} t - 2^{-j} s + 2^{-m} n - 2^{-j} k|) \\
&\quad \psi(t)\psi(s).
\end{aligned}
$$

As $|2^{-j}k - 2^{-m}n| \to \infty$,

$$
\begin{aligned}
\gamma_{<x,\psi>}(m,j;n,k) &= C(d,\Phi,\Theta)\, 2^{\frac{-(m+j)}{2}} \int dt \int ds \, |2^{-m} t - 2^{-j} s + 2^{-m} n - 2^{-j} k|^{2d-1} \\
&\quad \psi(t)\psi(s) \\
&= C(d,\Phi,\Theta)\, 2^{\frac{-(m+j)}{2}}\, 2^{-m(2d-1)} \int dt \int ds \, |t + 2^{m-j} k - n|^{2d-1} \\
&\quad \psi(2^{m-j} s - t)\, \psi(s). \qquad (16)
\end{aligned}
$$

Define $\Lambda(2^{m-j}, t) = \int ds \, \psi(s)\, \psi(2^{m-j} s - t)$ and write Eq. (16) as

$$
\begin{aligned}
\gamma_{<x,\psi>}(m,j;n,k) &= C(d,\Phi,\Theta)\, 2^{\frac{-(m+j)}{2}}\, 2^{-m(2d-1)} \int dt \, |t + 2^{m-j} k - n|^{2d-1} \\
&\quad \Lambda(2^{m-j}, t) \\
&= C(d,\Phi,\Theta)\, 2^{\frac{-(m+j)}{2}}\, 2^{-m(2d-1)} \int dt \, |t|^{2d-1} \\
&\quad \Lambda(2^{m-j}, t - (2^{m-j} k - n)). \qquad (17)
\end{aligned}
$$

The wavelet coefficients, $< x, \psi_{m,n} >$, are stationary over time when $m = j$ since Eq. (17) is a unique function of the difference in translation parameters, $k - n$, and is stationary over scale since (17) is also a unique function of $m - j$ for any two translations parameters,

k and n. The property of self-similarity of the wavelet coe¡cients is also found in Eq. (17) since for any a, $a^{-2(d-1/2)}\gamma_{<x,\psi>}(m,m;ak,an) = \gamma_{<x,\psi>}(m,m;k,n)$. Q.E.D.

C Proof of Theorem 2

Let $\alpha = 2^{m-j}k - n$ and $m \geq j$. Eq. (16) can be rewritten as

$$\begin{aligned}\gamma_{<x,\psi>}(m,j;n,k) &= C(d,\Phi,\Theta)\, 2^{\frac{-(m+j)}{2}} 2^{-m(2d-1)} \\ &\quad \int dt \int ds\, |t+\alpha|^{2d-1} \psi(2^{m-j}s - t)\psi(s) \\ &= C(d,\Phi,\Theta)\, 2^{\frac{-(m+j)}{2}} 2^{-m(2d-1)} \int_S dt\, |t+\alpha|^{2d-1} \Lambda(2^{m-j},t) \qquad (18)\end{aligned}$$

where S is the support of $\Lambda(2^{m-j},t)$. Let $f(t+\alpha) = |t+\alpha|^{2d-1}$ and to insure that $f(t+\alpha)$ is everywhere continuously di↑erentiable with respect to $t \in S$ let $|\alpha| > max(K_1 + 2^{j-m}K_2, 2^{j-m}K_1 + K_2)$, i.e., $-\alpha \notin S$. Write $f(t+\alpha)$ as

$$f(t+\alpha) = |\alpha|^{2d-1}\left|1+\frac{t}{\alpha}\right|^{2d-1}.$$

Since $|\alpha| > max(K_1 + 2^{j-m}K_2, 2^{j-m}K_1 + K_2)$ and $t \in S$, $|t/\alpha| < 1$. Hence, we write $|1+t/\alpha|^{2d-1}$ as the binomial power series and

$$f(t+\alpha) = |\alpha|^{2d-1}\left\{1+\sum_{i=1}^{\infty}\binom{2d-1}{i}\left(\frac{t}{\alpha}\right)^i\right\}. \qquad (19)$$

Substituting Eq. (19) into Eq. (18), the autocovariance is

$$\begin{aligned}\gamma_{<x,\psi>}(m,j;n,k) &= C(d,\Phi,\Theta)\, 2^{\frac{-(m+j)}{2}} 2^{-m(2d-1)} |\alpha|^{2d-1}\left\{\int_S dt\, \Lambda(2^{m-j},t)\right. \\ &\quad \left. + \int_S dt \sum_{i=1}^{\infty}\binom{2d-1}{i}\left(\frac{t}{\alpha}\right)^i \Lambda(2^{m-j},t)\right\}. \qquad (20)\end{aligned}$$

Since $\psi(t)$ has $M \geq 1$ vanishing moments, the ↓rst $2M$ moments of $\Lambda(2^{m-j},t)$ are equal to zero (see Lemma 2) and

$$\gamma_{<x,\psi>}(m,j;n,k) = C_1|\alpha|^{2d-1-2M} + R_{2M+1} \qquad (21)$$

where

$$\begin{aligned}C_1 &= (-1)^{M+1} C(d,\Phi,\Theta)\, 2^{-(M+1/2)(m+j)}\, 2^{-2m(d-1/2-M)} \frac{(2d-1)!}{(M!)^2(2d-1-2M)!} \\ &\quad \times \left(\int_{-K_1}^{K_2} t^M \psi(t) dt\right)^2\end{aligned}$$

AND

$$R_{2M+1} = C(d,\Phi,\Theta)\, 2^{-m(2d-1)} 2^{\frac{-(m+j)}{2}} |\alpha|^{2d-1} \left\{ \sum_{i=2M+1}^{\infty} \binom{2d-1}{i} \times \int_{-K_1}^{K_2} \int_{-K_1}^{K_2} \left(\frac{2^{j-k}s - t}{\alpha} \right)^{i} \psi(t)\psi(s) dt\, ds \right\}.$$

SINCE $M \geq 1$ AND $|d| < 0.5$

$$|R_{2M+1}| \leq C_2 |\alpha|^{2d-1} \sum_{i=1}^{\infty} \beta^{2M+i} \tag{22}$$

WHERE

$$C_2 = C(d,\Phi,\Theta)\, 2^{-m(2d-1)} 2^{\frac{-(m+j)}{2}} \left| \binom{2d-1}{2M} \right| \left(\int_{-K_1}^{K_2} |\psi(t)|\, dt \right)^2$$

AND

$$\beta = \sup_{(t,s)\in\Omega} \left| \frac{2^{m-j}s - t}{\alpha} \right| < 1$$

WHERE $\Omega = \{(t,s) : -K_1 \leq t,\ s \leq K_2\}$. SINCE $\beta < 1$, EQ. (22) EQUALS

$$|R_{2M+1}| \leq C_3 |\alpha|^{2d-1-2M-1} \tag{23}$$

WHERE C_3 IS A FINITE CONSTANT. FROM EQ. (21) AND EQ. (23)

$$\gamma_{<x,\psi>}(m,j;n,k) = \mathcal{O}\left(|2^{-j}k - 2^{-m}n|^{2(d-M)-1} \right) \tag{24}$$

FOR ALL k AND n SUCH THAT $|2^{-j}k - 2^{-m}n| > max(2^{-j}K_1 + 2^{-m}K_2, 2^{-m}K_1 + 2^{-j}K_2)$. Q.E.D.

References

[1] Baillie R.T. (1996) "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics,* 73, 5-59.

[2] Beylkin, G., R. Coifman, and V. Rokhlin (1991) "Fast Wavelet Transforms and Numerical Algorithms," *Communications on Pure and Applied Mathematics,* 44, 141-183.

[3] Beylkin, G. (1993) "Wavelets and Fast Numerical Algorithms," in *Proceedings of Symposia in Applied Mathematics*, eds. I. Daubechies, (American Mathematical Society: Rhode Island).

[4] Boes, D.C., R.A. Davis and S.N. Gupta (1989) "Parameter Estimation in Low Order Fractionally Di↑erenced ARMA Processes," *Stochastic Hydrology and Hydraulics,* 3, 97-110.

[5] Brockwell, P. and R. Davis (1993) *Time Series: Theory and Methods,* 2nd ed. (Springer-Verlag: New York).

[6] Cheung, Y.W. (1993) "Test For Fractional Integration: A Monte Carlo Investigation," *Journal of Time Series Analysis,* 14, 331-345.

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

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

[9] Daubechies, I. (1992) *Ten Lectures on Wavelets,* (SIAM: Philadelphia) 1992.

[10] Deriche, M. and A.H. Tew↓k (1993) "Maximum Likelihood Estimation of the Parameters of Discrete Fractionally Di↑erenced Gaussian Noise Process," *IEEE Transaction on Signal Processing,* 41, 2977-2989.

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

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

[13] Flandrin, P. (1991) "Fractional Brownian Motion and Wavelets," in *Wavelets, Fractals, and Fourier Transformations: New Developments and New Applications,* eds. M. Farge, J.C.R. Hunt, and J.C. Vassilicos, (Oxford Press: Oxford, UK).

[14] Flandrin, P. (1992) "Wavelet Analysis and Synthesis of Fractional Brownian Motion," *IEEE Transaction on Information Theory,* 38, 910-917.

[15] Geweke, J. and S. Porter-Hudak (1983) "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221-238.

[16] Granger, C. and R. Joyeux (1980) "An Introduction to Long-Memory Time Series Models and Fractional Di↑erencing," *Journal of Time Series Analysis*, 1, 15-29.

[17] Hosking, J. R. (1981) "Fractional Di↑erencing," *Biometrika*, 68, 165-176.

[18] Kronland-Martinet, R., J. Morlet, and A. Grossman (1987) "Analysis of Sound Patterns Through Wavelet Transforms," *International Journal of Pattern Recognition and Artificial Intelligence*, 1, 273-301.

[19] Mallat, S. (1989) "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 674-693.

[20] Mallat, S. and W. L. Hwang (1991) "Singularity Detection and Processing with Wavelets," *IEEE Transactions on Information Theory*, 38, 617-643.

[21] Mandelbrot, B. and J. Van Ness (1968) "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, 10, 422-437.

[22] McLeod, B. and K. Hipel (1978) "Preservation of the Rescaled Adjusted Range, I. A Reassessment of the Hurst Phenomenon," *Water Resources Research*, 14, 491-518.

[23] 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.

[24] Meyer, Y. (1989) "Wavelets and Operators," *Analysis at Urbana*, vol. 1 ed. E. Berkson, N.T. Peck and J. Uhl (Cambridge University Press: Cambridge), 256-365.

[25] Meyer, Y. (1990) *Ondelettes ét Operérateurs, I: Ondelettes, II: Opérateurs de Calderón-Zygmund, III: Opérateurs Multilinéaires*, (Paris: Hermann).

[26] Morlet, J. (1983) "Sampling Theory and Wave Propagation," in *Acoustic Signal/Image Processing and Recognition*, ed. C. H. Chen (Springer-Verlag: Berlin), 233-261.

[27] Resnick (1987) *Extreme Values, Regular Variation and Point Processes,* (Springer-Verlag: New York).

[28] Schmidt, C.M. and R. Tschernig (1995) "The Identi↓cation of Fractional ARIMA Models," Discussion Paper #8, Humboldt-Universität zu Berlin.

[29] Sowell, F. (1992) "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models," *Journal of Econometrics,* 53, 165-188.

[30] Strang, G. (1993) "Wavelet Transforms Versus Fourier Transforms," *Bulletin of the American Mathematical Society,* 23, 288-305.

[31] Tew↓k, A.H., and M. Kim (1992) "Correlation Structure of the Discrete Wavelet Coe¡cients of Fractional Brownian Motion," *IEEE Transaction on Information Theory,* 38, 904-909.

[32] Tieslau, M.A., P. Schmidt and R.T. Baillie (1996) "A Minimum Distance Estimator for Long-Memory Processes," *Journal of Econometrics,* 71, 249-264.

[33] Wornell, G. and A. Oppenheim (1992) "Estimation of Fractal Signals from Noisy Measurements Using Wavelets," *IEEE Transactions on Signal Processing*, 40, 611-623.