An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets

Mark J. Jensen

Department of Economics, 118 Professional Building, University of Missouri, Columbia, MO 65211, USA

Received 1 July 1997; accepted 1 February 1999

Abstract

In this paper we apply compactly supported wavelets to the ARFIMA($p,d,q$) long-memory process to develop an alternative maximum likelihood estimator of the differencing parameter, $d$, that is invariant to unknown means, model specification, and contamination. We show that this class of time series have wavelet transforms whose 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 equal 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 levels by generating ARFIMA($p,d,q$) processes for different values of $p$, $d$, and $q$, and calculating the wavelet MLE 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, Porter–Hudak estimator. © 2000 Elsevier Science B.V. All rights reserved.
1. Introduction

In this paper we apply wavelets with compact support to the autoregressive, fractionally integrated, moving average (ARFIMA) class of long-memory processes to produce a semiparametric maximum likelihood estimator (MLE) of the fractional differencing parameter. Jensen (1999) has already used wavelet analysis on the smaller long-memory class of fractional white noise processes to establish a log-linear relationship between the wavelet coefficient’s variance and the scaling parameter. This log-linear relationship equals the fractional differencing parameter and by replacing the wavelet coefficients variance by its sample analog, Jensen (1999) found the ordinary least squares estimator of this relationship to be a consistent semiparametric estimator of the differencing parameter. The wavelet based estimator developed in this paper draws on the stationarity and self-similarity of the wavelet transform’s second-order moments and the sparsity of the ARFIMA’s covariance matrix wavelet representation to construct a MLE of the fractional differencing parameter.

The wavelet MLE enjoys the advantage of having both the strengths of a MLE and a semiparametric estimator, but does not suffer their known drawbacks. Boes et al. (1989) have shown that semiparametric methods perform more satisfactorily than the MLE methods when the model is misspecified. On the other hand, Cheung (1993) has found that under correct model specification the various MLE methods are superior to the semiparametric estimators. Like the Geweke and Porter-Hudak (1983) log-periodogram estimator (GPH) 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, while at the same time providing an estimate of the differencing parameter that can be calculated separately from the short-memory parameters.¹

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. Jensen (1998) uses the spectral representation of the wavelet transform to simultaneously estimate a correctly identified ARFIMA model’s short and long-memory parameters.
Another of the wavelet MLE’s many strengths is its significant reduction in 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, most choose 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. Such calculations are a mounting task even for today’s fastest workstations and a justifiable reason for not using Sowell’s exact MLE.²

We show how 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 estimator to an order similar to the approximate frequency domain MLE of Fox and Taqqu (1986). More specifically, we show 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.

The wavelet MLE is also a robust estimator to white noise that may have contaminated the time series. By its construction the wavelet transform captures the long-memory characteristics of the ARFIMA model in the rate of decay of the wavelet transform’s autocovariance function. Since wavelets are an orthogonal basis, white noise is transformed into white noise whose autocovariance function’s rate of decay exhibits no structure. Hence, the rate of decay of the autocovariance function from a long-memory process’s wavelet transform will continue to identify the presence of long-memory even if the data is contaminated.

The plan of the paper is as follows. In Section 2 we present the necessary wavelet theory needed to derive the wavelet MLE.³ Section 3 defines the class of ARFIMA processes and their long-memory behavior. Section 4 contains the second-order properties and the likelihood function of the wavelet coefficients from a ARFIMA series. In Section 5, an extensive array of simulations that ascertains the robustness of the wavelet estimator to sample size, zero padding, differencing parameter, short-run dynamics, wavelet type, and contaminated data are presented. Section 6 summarizes our results.

2. Wavelets

Let \( \psi(t) \) be a wavelet satisfying the admissibility condition, \( \int \psi(t) \, dt = 0 \), i.e., \( \psi \) is an oscillating function that decreases rapidly to zero as \( t \to \pm \infty \). The

³ 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.

³ For a thorough introduction to wavelets see Hernández and Weiss (1996).
regularity of $\psi$ can be increased to include only those functions with a greater number of vanishing moments, i.e., $\int t^r \psi(t) \, dt = 0$ where $r = 0, 1, \ldots, M - 1$, and/or $\psi(t) \in C^r$. These necessary conditions insure that $\psi(t)$ is well localized in frequency and time space.

Define the dilated and translated wavelet as

$$\psi(t)_{m, n} = 2^{m/2} \psi(2^m t - n),$$

where $m$ and $n$ are elements of $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$. Then for various values of $m$ and $n$, $\psi_{m, n}$ covers different frequencies and periods of time. At high frequencies (large $m$) the time translation, $2^{-m} n$, 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.

Mallat (1989) shows that the set of dilated and translated wavelets form an orthonormal basis of the set of squared integrable functions, $L^2(\mathbb{R})$. However, in this work we are interested in long-memory stochastic processes.

Let $x(t, \omega)$ be a mean square convergent random process that is jointly measurable on the probability space $(\Omega, \mathcal{F}, P)$, where $t \in \mathbb{R}$ and $\omega \in \Omega$. Since $\psi$ is compactly supported and $x(t)$ is mean squared convergent, $x(t)$’s wavelet transform

$$\{ \langle x, \psi_{m, n} \rangle = \int x(t) \psi_{m, n}(t) \, dt : m, n \in \mathbb{Z}, \omega \in \Omega \}$$

exist with probability one and is a random process with finite second moments (Cambanis and Houdré, 1995). The scaling coefficient, $\langle x, \phi_{m, n} \rangle$, is defined in an analogous manner.

Daubechies (1988) has constructed a class of wavelet functions of the form

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

where $\phi$ is a function such that $\{\phi(t - n) : n \in \mathbb{Z}\}$ forms an orthonormal basis of piecewise constant functions of length one. Daubechies discovered a sufficient non-zero sequence, $\{g_k\}_{k=0}^{2^{M-1}}$, in which $\text{supp}\{\psi\} = [-(M-1), M]$,

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

---

4 We limit ourselves to signals in $L^2(\mathbb{R})$. However, wavelets span many function spaces. For example, Sobolev, Lipschitz, Besov and Hölder spaces are all spanned by wavelets (Meyer, 1990, 1992; Mallat and Hwang, 1991).

5 We simplify our notation by suppressing $\omega$ when writing $x$. 
and $\psi \in C^k$, for $k \leq 2M - 2$. 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, and hence a larger support, the Daubechies wavelet has the fewest coefficients of any class of wavelets for a given $M$.

The support of $\psi_{m,n}$ can be thought of as $[n2^{-m}, (n + 1)2^{-m}]$. By normalizing the time interval of $x(t)$ to the unit interval, in other words $t \in [0,1]$, the translation parameter $n = 0$ causes the wavelet to cover the entire unit interval when $m = 0$. If we observe realizations of $x(t)$ at $t = 0, \ldots, 2^{\text{max}} - 1$, then for $m = \text{max} - 1$, $n = 0, 1, 2, \ldots, 2^{\text{max}} - 1$ is needed. Hence, for a given scale, $m$, the translation parameter need only take on the values $n = 0, 1, 2, \ldots, 2^{\text{max}} - 1$. For future reference let $\mathcal{M} = \{0, 1, 2, \ldots, \text{max} - 1\}$ and $\mathcal{N}(m) = \{0, 1, 2, \ldots, 2^{m} - 1\}$.

3. Long-memory processes

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

$$\Phi(L)(1 - L)^d(x(t) - \mu) = \Theta(L)c(t),$$

where

$$\Phi(L) = 1 + \phi_1L + \phi_2L^2 + \cdots + \phi_pL^p,$$

$$\Theta(L) = 1 + \theta_1L + \theta_2L^2 + \cdots + \theta_qL^q$$

are polynomials of degrees $p$ and $q$ respectively whose roots lie outside the unit circle, $\varepsilon \sim \text{i.i.d.} \mathcal{N}(0, \sigma^2)$, $|d| < 0.5$, and $\mu$ is the unknown mean. $L$ is the lag operator, i.e., $L^jx(t) = x(t - j)$, where $j \in \mathbb{Z}$, 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 $x(t)$ is mean square convergent and for large lags has the autocovariance

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

where $\gamma(s)$ is the autocovariance function of $x(t)$, $C(d, \Phi, \Theta)$ is a constant, and $|s|^{2d - 1}$ is the decay rate of the autocovariance. This is actually the support of the Daubechies wavelet with one vanishing moments, i.e., $M = 1$. 
where
\[ C(d, \Phi, \Theta) = 2^{1+d} \frac{\sigma_x^2 |\Theta(1)|^2}{\pi |\Phi(1)|^2} \Gamma(1 - 2d) \sin \pi d \]
and is hence not summable when \( d > 0 \). Correspondingly, \( x(t) \)'s spectrum equals
\[ S(\omega) \sim \frac{1}{|\omega|^{2d}} \quad \text{as} \quad \omega \to 0, \quad (4) \]
and is unbounded at the origin for \( d > 0 \) (see Granger and Joyeux, 1980; Hosking, 1981; Brockwell and Davis, 1993; 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-similarity property can also be seen in Eqs. (3) and (4). 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 frequently \( x(t) \) is sampled.

4. Wavelet analysis and estimation of ARFIMA\((p, d, q)\)

As a result of their construction, wavelets display a form of self-similarity where at different values of \( m \), \( \psi_{m,n} \) has the same properties as \( \psi \) but over smaller or larger time intervals. In this section we show how the wavelet’s self-similarity causes an ARFIMA process’s wavelet coefficients to be stationary and self-similar in time space and stationary in scale space. We find the wavelet coefficient’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 coefficients exhibits exponential decay like that of a ARMA process but between both time and scale. This dampening behavior leads to a sparse covariance matrix that aids in reducing the order of computing the MLE of \( d \).

Because the approximate frequency domain MLE is invariant to an unknown mean (Priestly, 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 efficient and attractive alternative to the exact MLE when \( \mu \) is unknown. Like the approximate frequency domain MLE, the wavelet MLE of \( d \) is unaffected by the unknown \( \mu \) since the wavelet coefficients autocovariance function is invariant to \( \mu \) (see Lemma 1 in Appendix A).

\(^7\)Tewfik and Kim (1992) and Flandrin (1991) have derived similar properties for the continuous fractional Brownian motion process.
4.1. Covariance structure of ARFIMA’s wavelet coefficients

Let $\gamma_{\langle x, \psi \rangle}(m, j; n, k) = E[\langle x, \psi_{m, n} \rangle \langle x, \psi_{j, k} \rangle]$ represent the wavelet transform’s autocovariance function. Using only the admissibility condition, $\int \psi(t) dt = 0$, and $x(t)$’s asymptotic autocovariance function, we derive the following result.

Theorem 1. For any $n, k \in \mathbb{Z}$, as $|m - j| \to \infty$, or for any $m, j \in \mathbb{Z}$, as $|n - k| \to \infty$, the normalized wavelet coefficients, $2^{m(d+1/2)} \langle x(t), \psi_{m, n} \rangle$, associated with an ARFIMA($p, d, q$) process with unknown mean $\mu$ and $|d| < 0.5$ are

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

Proof. See Appendix B.

Like the original time series, $\langle x, \psi_{m, n} \rangle$ 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 that will make the calculation of $d$’s MLE any easier or better than existing maximum likelihood estimators. By choosing $\psi(t)$ to have $M$ vanishing moments and a compact support, we find that $\langle x, \psi_{m, n} \rangle$ has weak dependence over time and scale. This weak dependence 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 $\mu$ and $|d| < 0.5$, then for scales $m \geq j$ and $m - j \to \infty$, or $|n - k| \to \infty$, the autocovariance, $\gamma_{\langle x, \psi \rangle}(m, j; n, k)$, 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 C.

This theorem generalizes Jensen’s (1999) Theorem 2 to the ARFIMA class of long-memory processes and establishes the wavelet coefficients autocovariance function’s rate of decay in the time-scale domain. If $|d| < 0.5$ any positive integer $M$ insures that $\gamma_{\langle x, \psi \rangle}(m, j; n, k)$ will decay exponentially at and between scales. Only when $d \geq 0.5$ does the number of vanishing moments of $\psi$ become critical to the decay of $\gamma_{\langle x, \psi \rangle}$. 
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 $\psi$ 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 Tewfik and Kim (1992) have found the ‘effective’ support of the Daubechies class of wavelet to be smaller than its theoretical support, and that its ‘effective’ support grows at a slower rate than $2M - 1$.

4.2. Sparsity of the wavelet coefficient’s covariance matrix

Let

$$\langle X, \psi \rangle = [\langle x(t), \psi_{0, 0} \rangle, \langle x(t), \psi_{1, 0} \rangle, \langle x(t), \psi_{1, 1} \rangle, \langle x(t), \psi_{2, 0} \rangle, \ldots, \langle x(t), \psi_{\text{max}-1, 2^{\text{max}-1}-2} \rangle, \langle x(t), \psi_{\text{max}-1, 2^{\text{max}-1}-1} \rangle]$$

and $\Sigma_{\langle x, \psi \rangle} = E[\langle X, \psi \rangle \langle X, \psi \rangle^\prime]$ be the $2^{\text{max}} - 1 \times 2^{\text{max}} - 1$ covariance matrix containing the elements $\gamma_{\langle x, \psi \rangle}(m, j; n, k)$, 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 wavelet transform of a covariance matrix is an assembly of Toeplitz matrices at different scales. Fig. 1 shows how $\Sigma_{\langle x, \psi \rangle}$ for $\text{max} = 4$ is comprised of the autocovariance function, $\gamma_{\langle x, \psi \rangle}(m, j; n, k)$, where $G_m$ is a zero-offset $2^n \times 2^n$ Toeplitz matrix of the form

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

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

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

where $m, j \in \mathcal{M}$ and $\gamma_{\langle x, \psi \rangle}(m, j; n, k)$ is the $n + 1, k + 1$ element of $G^m_j$. In addition, $G^m_j = (G^m_j)^\prime$.

By Theorem 2, $\Sigma_{\langle x, \psi \rangle}$ is a sparse matrix whose elements decay exponentially as one moves away from the diagonal elements of $G_m$ and $G^m_j$. This decay creates finger-like bands that emanate from the first row and column of $\Sigma_{\langle x, \psi \rangle}$. Both the finger-like bands and decay of $\Sigma_{\langle x, \psi \rangle}$ 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 $\Sigma^B_{x,\psi}$ to be the matrix equal to $\Sigma_{x,\psi}$ except with the elements,

$$\{ \gamma_{x,\psi}(m,j; n,k) : (2^{-j}k - 2^{-m}n) \geq B$$
\vspace{0.1cm}
$$> \max(2^{-j}K_1 + 2^{-m}K_2, 2^{-m}K_1 + 2^{-j}K_2) \}$$

set equal to zero. $\Sigma^B_{x,\psi}$'s only nonzero elements are the main diagonal elements of $\Sigma_{x,\psi}$, i.e., the main diagonal of $G_m$ for $m = 0, 1, \ldots, \text{max} - 1$, and the finger-like diagonal elements of $\Sigma_{x,\psi}$ associated with the main diagonal elements of $G^m_m$. Thus, the nonzero elements of $\Sigma^B_{x,\psi}$ form a band of width $B$ around the diagonals of $\Sigma_{x,\psi}$.

**4.3. Maximum likelihood estimator of $d$**

Approximate the likelihood function

$$L_N(d|\langle X,\psi \rangle) = (2\pi)^{-\frac{(m-1)}{2}} |\Sigma_{x,\psi}(d)|^{-1/2}$$

\vspace{0.1cm}

$$\times \exp \left[ - \frac{1}{2} \langle X,\psi \rangle' \Sigma^{-1}_{x,\psi}(d) \langle X,\psi \rangle \right]$$

\vspace{0.1cm}

(5)
with the function $L_N^B$ equal to $L_N$ except with $\Sigma_{\langle x, \psi \rangle}$ replaced by $\Sigma_{\langle x, \psi \rangle}^B$, and define the wavelet MLE of $d$ as

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

and the approximate wavelet MLE as

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

Because $\psi$ is compactly supported, the elements of $\Sigma_{\langle x, \psi \rangle}$ will be finite and hence, both $L_N$ and $L_N^B$ are well behaved and bounded from above. Thus, $\hat{d}$ and $\tilde{d}_B$ will exist.

The level of accuracy achieved with $L_N^B$ is

$$||L_N^B - L_N|| \leq \frac{C}{B^{1/(M-\alpha)}} \log_2 N,$$

(6)

where $C$ is a constant dependent on $\gamma(s)$ and $\psi$. In Eq. (6) we see that the level of accuracy associated with $L_N^B$ depends on the value of $M$. Thus, $\psi$s with a larger

---

8 Beylkin et al. (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 $\sum_{m=1}^{\infty} 1/m^m$, where $m > 1$, by $\int_0^1 \psi \, e^{-t^m}$, which equals $1/B^{m-1}$. 
number of vanishing moments, $M$, have a faster rate of decay in $\gamma_{(x, \psi)}$, which reduces the approximation error of $L^B_N$.

The calculation time involved in computing the inverse of the covariance matrix and hence, the likelihood function, is substantially reduced when $\Sigma^B_{(x, \psi)}$ is used to approximate $\Sigma_{(x, \psi)}$. Since $\Sigma^B_{(x, \psi)}$ is the approximate wavelet representation of the covariance matrix in an orthonormal basis, multiplying it by another wavelet represented matrix, $S_\psi$, results in the product having the same submatrix composition illustrated in Fig. 1. In other words, the wavelet representation of the product, $S_\Sigma$, consists of the sum of the products of the $S_\psi$ and $\Sigma^B_{(x, \psi)}$ individual submatrices, $G^j_m$, $m = 0, 1, \ldots, \max - 1$. Assume that $S_\psi$ has the same banded characteristic as $\Sigma^B_{(x, \psi)}$, then each individual submatrix of $S_\psi \Sigma^B_{(x, \psi)}$ will possess this banded characteristic (Beylkin et al., 1990).

Computations for each of the $\max = \log_2 N$ submatrices, $G^j_m$ and $G_m$, associated with $S_\psi \Sigma^B_{(x, \psi)}$, require $N \log N$ calculations. Thus, the total cost of multiplying two matrices together in the wavelet domain is $O(N \log^2 N)$. Fortunately, as a result of Theorem 2, both $S_\psi$ and $\Sigma^B_{(x, \psi)}$ submatrices, $G^j_m$, can be set equal to zero when the scales $j$ and $m$ are well separated. This reduces the computing cost asymptotically to $O(N)$.

We can now draw on the results of Schulz (1933) for computing the inverse of a matrix and apply it to $\Sigma^B_{(x, \psi)}$. The Schulz algorithm is based on the sequence of matrices $X_k$ defined by

$$X_{k+1} = 2X_k - X_k A X_k$$

quadratically converging to the inverse of $A$, where $X_0 = xA'$, and $x$ is the largest eigenvalue of $A$. The Schulz method’s rapid convergence to the inverse is attractive, but is often not used because of the $O(N^3)$ cost associated with matrix multiplication. Fortunately, in the wavelet domain when $A$ is replaced by the sparse, banded matrix $\Sigma^B_{(x, \psi)}$, the Schulz algorithm becomes a manageable $O(N)$ operation. Thus, inverting $\Sigma^B_{(x, \psi)}$ with the Schulz algorithm is a $O(N)$ computation as is the wavelet domain calculation of the approximate likelihood function, $L^B_N$.

4.4. Banded MLE of $d$

In this section we draw on the results from the previous section, but rather than using $\Sigma^B_{(x, \psi)}$ to approximate $\Sigma_{(x, \psi)}$, we only use the main diagonal elements of $\Sigma_{(x, \psi)}$ to estimate $d$. To distinguish this estimator from $d_B$, we name it the banded MLE of $d$.

Let the observed time series be

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

(7)
for \( t = 0, 1, \ldots, 2^{max} - 1 \) with the corresponding wavelet transform vector,

\[
\langle \tilde{X}, \psi \rangle^t = \left[ \langle \tilde{x}(t), \psi_{0,0} \rangle, \langle \tilde{x}(t), \psi_{1,0} \rangle, \langle \tilde{x}(t), \psi_{1,1} \rangle, \langle \tilde{x}(t), \psi_{2,0} \rangle, \ldots \\
\ldots, \langle \tilde{x}(t), \psi_{max-1, 2^{max-1} - 2} \rangle, \langle \tilde{x}(t), \psi_{max-1, 2^{max-1} - 1} \rangle \right],
\]

where \( x(t) \) is an ARFIMA\((p,d,q)\) process with \(|d| < 0.5, 2^{max} \) observations and \( \eta(t) \sim W.N.(0, \sigma^2) \) and is independent of \( x(t) \).

\( \tilde{x}(t) \) can be thought of as an ARFIMA process that has been contaminated by some type of measurement or data error.\(^9\) Eq. (7) can also be viewed in the context of the logarithmic transformation of Breidt et al. (1998) long-memory stochastic volatility model, \( y(t) = \sigma(t)/\zeta(t) \), where \( \sigma(t) = \sigma \exp\{x(t)/2\} \), and \( \zeta(t) \sim \text{i.i.d.}(0, 1) \) and is independent of \( x(t) \).

\( \langle \tilde{X}, \psi \rangle \)'s covariance matrix is approximated by

\[
\Sigma_1 = \text{diag}(\tilde{\sigma}_{0,0}, \tilde{\sigma}_{1,0}, \tilde{\sigma}_{1,1}, \ldots, \tilde{\sigma}_{max-1, 2^{max-1} - 2}, \tilde{\sigma}_{max-1, 2^{max-1} - 1}),
\]

where

\[
\tilde{\sigma}_{m,n} = E[\langle x, \psi_{m,n} \rangle^2] + E[\langle \eta, \psi_{m,n} \rangle^2] \\
= C(d, \Phi, \Theta)2^{-m(2d-1)-mV(d)} + \sigma^2 \\
= \sigma^2 2^{-m2d} + \sigma^2,
\]

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

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

where \( \theta' = (d, \sigma^2, \sigma^2) \), and the approximate log-likelihood is

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

or as the concentrated approximate log-likelihood function

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

(9)

where \( \tilde{\sigma}_{m,n} = (1/\#\mathcal{N}(m))\sum_{n \in \mathcal{N}(m)} \langle \tilde{x}, \psi_{m,n} \rangle^2 \) and \( \#\mathcal{N}(m) \) denotes the number of elements in the set, \( \mathcal{N}(m) \).

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

\[
\theta \in \Theta = \{ (d, \sigma^2, \sigma^2) : |d| < 0.5, \sigma^2 \geq 0, \sigma^2 \geq 0 \}.
\]

---

\(^9\) Measurement error in the forward premium of foreign exchange rates has been proposed as one of the possible explanations for the failure of forward exchange rates to forecast future spot rates (Maynard and Phillips, 1998).
Occasionally, $\mathcal{L}(\theta)$ will have more than one maximum. If this occurs the additional maximums are for trivial boundary cases where $\sigma^2 = 0$ and/or $\sigma^2_g = 0$. Thus, the behavior of Eq. (9) insures that a numerical algorithm will converge to the banded MLE if initialized with values from $\Theta$. Wornell and Oppenheim (1992) have provided an EM-algorithm that maximizes $\mathcal{L}(\theta)$ in which $\langle X, \psi \rangle$ is the complete data and $\langle \hat{X}, \psi \rangle$ the censored data.$^{10}$

5. Simulation

Since Theorems 1 and 2 rely on $|j - m| \to \infty$ or $|k - n| \to \infty$, 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 $\theta$ from artificially generated ARFIMA processes for different wavelets, series length, and contamination levels.

We employ Hosking (1984) algorithm to simulate an ARFIMA process. This method generates an ARFIMA$(0,d,0)$ process, $y(t)$, and then recursively computes the ARFIMA$(p,d,q)$ process, $x(t)$, from the equation

$$x(t) = \sum_{i=1}^{p} \phi_j y(t - j) + y(t) - \sum_{j=1}^{q} \theta_j y(t - j).$$

To diminish the simulated processes dependence on initial conditions the first 50 realizations are discarded.

In each experiment the wavelet coefficients, $\langle x, \psi_{m,n} \rangle$, are 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 analyzing stationary long-memory process with wavelets. With finite data it is not always possible to exactly calculate all of the wavelet coefficients. The more regular (large $M$) a wavelet is, the larger its support. Hence, at coarser scales the wavelet straddles the data, resulting in boundary affects.

5.1. ARFIMA$(0,d,0)$ case

We generate 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. In each simulation the starting values for the

$^{10}$McCoy and Walden (1996) have developed an estimator similar to the banded MLE but include the scaling coefficients in the likelihood function and do not take into consideration contaminated data.
EM algorithm are $\theta^* = (\log_2(0.1)/2, 0.1, 0.1)$ and the convergence criterion is 0.001. Since we found no ill-effects from zero-padding (when $N$ was at least half way to the next highest integer power of 2) or truncation (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 those results from series with integer powers of 2 number of observations.

Tables 1 and 2 report the findings of the banded MLE’s mean squared error (MSE) and its level of bias with and without contamination. For comparison, the GPH and approximate frequency domain MLE’s MSE and bias are also reported.

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

For fixed $N$ the bias of both wavelet based estimators increases as $d$ moves away from the origin. This behavior was also reported by Cheung and Diebold (1994) for the feasible exact MLE. The most notable feature of Tables 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.

In comparison with the approximate MLE, the banded MLE is superior in regard to its absolute bias with and without contamination. Out of the twenty Monte Carlo experiments without (with) contamination, the approximate MLE’s absolute bias is smaller than the banded MLE in only four (one) cases, three of which occur 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 produces the smallest MSE in twelve of them. However, this edge reduces to six cases when the series is contaminated. Except for the extreme cases, $d = 0.05, 0.45$, the relative efficiency, $\text{MSE}_{\text{banded}}/\text{MSE}_{\text{approx}}$, for the contaminated data are almost all substantially less than 0.6. And given the smaller relative absolute bias, $|\text{Bias}|_{\text{banded}}/|\text{Bias}|_{\text{approx}}$, at these extreme values of $d$, the banded MLE appears to be the ‘better’ estimator under contamination. With uncontaminated data, we find the approximate MLE’s increase in bias to be offset by a smaller MSE,

---

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 $d$. 

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

<table>
<thead>
<tr>
<th>d</th>
<th>N</th>
<th>Banded Haar</th>
<th>Daub. M = 10</th>
<th>GPH</th>
<th>Approx. MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MSE</td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^7$</td>
<td>0.0064</td>
<td>-0.0044</td>
<td>0.0060</td>
<td>0.0054</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^7$</td>
<td>0.0063</td>
<td>0.0058</td>
<td>0.0070</td>
<td>0.0253</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^7$</td>
<td>0.0081</td>
<td>0.0163</td>
<td>0.0105</td>
<td>0.0473</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^7$</td>
<td>0.0217</td>
<td>0.0642</td>
<td>0.0243</td>
<td>0.1024</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^7$</td>
<td>0.0229</td>
<td>0.1047</td>
<td>0.0285</td>
<td>0.1389</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^8$</td>
<td>0.0031</td>
<td>0.0009</td>
<td>0.0032</td>
<td>0.0096</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^8$</td>
<td>0.0028</td>
<td>0.0089</td>
<td>0.0038</td>
<td>0.0295</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^8$</td>
<td>0.0034</td>
<td>0.0213</td>
<td>0.0054</td>
<td>0.0497</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^8$</td>
<td>0.0277</td>
<td>0.0467</td>
<td>0.0110</td>
<td>0.0830</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^8$</td>
<td>0.0126</td>
<td>0.0885</td>
<td>0.0179</td>
<td>0.1168</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^9$</td>
<td>0.0012</td>
<td>0.0022</td>
<td>0.0014</td>
<td>0.0099</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^9$</td>
<td>0.0015</td>
<td>0.0137</td>
<td>0.0023</td>
<td>0.0316</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^9$</td>
<td>0.0021</td>
<td>0.0229</td>
<td>0.0040</td>
<td>0.0511</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^9$</td>
<td>0.0039</td>
<td>0.0438</td>
<td>0.0079</td>
<td>0.0794</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^9$</td>
<td>0.0082</td>
<td>0.0779</td>
<td>0.0129</td>
<td>0.1042</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^{10}$</td>
<td>0.0007</td>
<td>0.0045</td>
<td>0.0008</td>
<td>0.0102</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^{10}$</td>
<td>0.0008</td>
<td>0.0130</td>
<td>0.0016</td>
<td>0.0309</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^{10}$</td>
<td>0.0011</td>
<td>0.0228</td>
<td>0.0031</td>
<td>0.0501</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^{10}$</td>
<td>0.0025</td>
<td>0.0410</td>
<td>0.0065</td>
<td>0.0767</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^{10}$</td>
<td>0.0063</td>
<td>0.0713</td>
<td>0.0102</td>
<td>0.0944</td>
</tr>
</tbody>
</table>
Table 2
ARFIMA(0, d, 0), contamination

<table>
<thead>
<tr>
<th>d</th>
<th>N</th>
<th>Banded</th>
<th>Daub. M = 10</th>
<th>GPH</th>
<th>Approx. MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Haar</td>
<td>MSE</td>
<td>Bias</td>
<td>MSE</td>
</tr>
<tr>
<td>0.05</td>
<td>2^7</td>
<td>0.0066</td>
<td>0.0061</td>
<td>0.0093</td>
<td>0.0799</td>
</tr>
<tr>
<td>0.15</td>
<td>2^7</td>
<td>0.0068</td>
<td>0.0064</td>
<td>0.0066</td>
<td>0.0788</td>
</tr>
<tr>
<td>0.25</td>
<td>2^7</td>
<td>0.0082</td>
<td>0.0070</td>
<td>0.0119</td>
<td>0.0713</td>
</tr>
<tr>
<td>0.35</td>
<td>2^7</td>
<td>0.0187</td>
<td>0.0174</td>
<td>0.0031</td>
<td>0.0782</td>
</tr>
<tr>
<td>0.45</td>
<td>2^7</td>
<td>0.0465</td>
<td>0.0392</td>
<td>0.0090</td>
<td>0.0758</td>
</tr>
<tr>
<td>0.05</td>
<td>2^8</td>
<td>0.0032</td>
<td>0.0031</td>
<td>0.0015</td>
<td>0.0487</td>
</tr>
<tr>
<td>0.15</td>
<td>2^8</td>
<td>0.0031</td>
<td>0.0029</td>
<td>0.0016</td>
<td>0.0404</td>
</tr>
<tr>
<td>0.25</td>
<td>2^8</td>
<td>0.0037</td>
<td>0.0029</td>
<td>0.0049</td>
<td>0.0476</td>
</tr>
<tr>
<td>0.35</td>
<td>2^8</td>
<td>0.0080</td>
<td>0.0053</td>
<td>0.0137</td>
<td>0.0424</td>
</tr>
<tr>
<td>0.45</td>
<td>2^8</td>
<td>0.0217</td>
<td>0.0175</td>
<td>0.0164</td>
<td>0.0485</td>
</tr>
<tr>
<td>0.05</td>
<td>2^9</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0011</td>
<td>0.0272</td>
</tr>
<tr>
<td>0.15</td>
<td>2^9</td>
<td>0.0015</td>
<td>0.0014</td>
<td>0.0012</td>
<td>0.0296</td>
</tr>
<tr>
<td>0.25</td>
<td>2^9</td>
<td>0.0021</td>
<td>0.0014</td>
<td>0.0020</td>
<td>0.3090</td>
</tr>
<tr>
<td>0.35</td>
<td>2^9</td>
<td>0.0037</td>
<td>0.0022</td>
<td>0.0132</td>
<td>0.0278</td>
</tr>
<tr>
<td>0.45</td>
<td>2^9</td>
<td>0.0118</td>
<td>0.0091</td>
<td>0.0205</td>
<td>0.0285</td>
</tr>
<tr>
<td>0.05</td>
<td>2^{10}</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0177</td>
</tr>
<tr>
<td>0.15</td>
<td>2^{10}</td>
<td>0.0009</td>
<td>0.0007</td>
<td>0.0010</td>
<td>0.0190</td>
</tr>
<tr>
<td>0.25</td>
<td>2^{10}</td>
<td>0.0011</td>
<td>0.0006</td>
<td>0.0018</td>
<td>0.0188</td>
</tr>
<tr>
<td>0.35</td>
<td>2^{10}</td>
<td>0.0023</td>
<td>0.0010</td>
<td>0.0143</td>
<td>0.0201</td>
</tr>
<tr>
<td>0.45</td>
<td>2^{10}</td>
<td>0.0069</td>
<td>0.0050</td>
<td>0.0170</td>
<td>0.0181</td>
</tr>
</tbody>
</table>
while for contaminated data this is not the case. Instead in these cases the approximate MLE’s larger bias also comes with a larger MSE. In contrast, for the few cases of contaminated data where the absolute bias of the banded MLE increases its MSE either stays the same or becomes 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 cannot 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 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.

5.2. ARFIMA$(p, d, q)$ case

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 3 and the results for the ARFIMA$(0, d, 1)$ model $(1 - L)^d x(t) = (1 + \theta L)\epsilon(t)$ are presented in Table 4. Like those for the ARFIMA$(0, d, 0)$ processes, the banded MLE’s level of bias tends to increase under both models as $d$ moves away from the origin. Unlike the findings of Sowell (1992) for the exact MLE, the banded MLE’s level of bias does not increase for negative values of $\phi$, but the bias does increase for values of $\theta$ near one. The MSE of the banded MLE for the ARFIMA$(1, d, 0)$ model approaches the theoretical variance $6/(\pi^2 N)$ as $N$ increases and regardless of $\phi$ being positive or negative its MSE stays the same.

In comparison with the approximate MLE, the banded MLE’s mean squared error and absolute bias are almost always smaller for the ARFIMA$(1, d, 0)$ processes and when the MA parameter, $\theta$, of the ARFIMA$(0, d, 1)$ process equals $-0.8$. Out of the 48 Monte Carlo experiments with the ARFIMA$(1, d, 0)$ process, in only two cases ($N = 2^{10}$ and $d = -0.25$, for $\phi = 0.8$, $-0.7$) is the approximate MLE’s mean square error smaller than the banded MLE. In only 8 experiments is the approximate MLE’s absolute bias smaller than the banded MLEs. These cases occur for large $N$ showing that the difference between the banded and approximate MLE’s MSE and bias diminishes as $N$ increases. But for small sample sizes the banded MLE clearly dominates the approximate MLE for ARFIMA$(1, d, 0)$ processes.

Sowell (1992) found the exact MLE bias and MSE increased for ARFIMA$(1, d, 0)$ processes as $\phi$ 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 $\phi$. Furthermore, the absolute bias and MSE of the banded and exact MLE are nearly identical when $\phi = 0.8$. Hence, when
<table>
<thead>
<tr>
<th>$d$</th>
<th>$N$</th>
<th>$\phi$</th>
<th>Banded MSE</th>
<th>Banded Bias</th>
<th>Approx. MLE MSE</th>
<th>Approx. MLE Bias</th>
<th>$\phi$</th>
<th>Approx. MLE MSE</th>
<th>Approx. MLE Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>$2^7$</td>
<td>0.8</td>
<td>0.0099</td>
<td>0.0401</td>
<td>0.1114</td>
<td>-0.1290</td>
<td>0.7</td>
<td>0.0104</td>
<td>0.0376</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^7$</td>
<td>0.8</td>
<td>0.0086</td>
<td>0.0193</td>
<td>0.1338</td>
<td>-0.1816</td>
<td>0.7</td>
<td>0.0087</td>
<td>0.0169</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^7$</td>
<td>0.8</td>
<td>0.0067</td>
<td>-0.0011</td>
<td>0.1400</td>
<td>-0.2012</td>
<td>0.7</td>
<td>0.0070</td>
<td>-0.0019</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^7$</td>
<td>0.8</td>
<td>0.0064</td>
<td>-0.0148</td>
<td>0.1374</td>
<td>-0.2142</td>
<td>0.7</td>
<td>0.0066</td>
<td>-0.0154</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^7$</td>
<td>0.8</td>
<td>0.0071</td>
<td>-0.0301</td>
<td>0.1645</td>
<td>-0.2480</td>
<td>0.7</td>
<td>0.0071</td>
<td>-0.0296</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^7$</td>
<td>0.8</td>
<td>0.0079</td>
<td>-0.0382</td>
<td>0.1684</td>
<td>-0.2619</td>
<td>0.7</td>
<td>0.0081</td>
<td>-0.0363</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^8$</td>
<td>0.8</td>
<td>0.0052</td>
<td>0.0402</td>
<td>0.0311</td>
<td>-0.0682</td>
<td>0.7</td>
<td>0.0052</td>
<td>0.0394</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^8$</td>
<td>0.8</td>
<td>0.0036</td>
<td>0.0166</td>
<td>0.0320</td>
<td>-0.0760</td>
<td>0.7</td>
<td>0.0037</td>
<td>0.0159</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^8$</td>
<td>0.8</td>
<td>0.0032</td>
<td>0.0030</td>
<td>0.0365</td>
<td>-0.0864</td>
<td>0.7</td>
<td>0.0032</td>
<td>-0.0028</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^8$</td>
<td>0.8</td>
<td>0.0031</td>
<td>-0.0122</td>
<td>0.0221</td>
<td>-0.0762</td>
<td>0.7</td>
<td>0.0030</td>
<td>-0.0118</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^8$</td>
<td>0.8</td>
<td>0.0034</td>
<td>-0.0231</td>
<td>0.0337</td>
<td>-0.0100</td>
<td>0.7</td>
<td>0.0034</td>
<td>-0.0218</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^8$</td>
<td>0.8</td>
<td>0.0038</td>
<td>-0.0283</td>
<td>0.0354</td>
<td>-0.0103</td>
<td>0.7</td>
<td>0.0037</td>
<td>-0.0280</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^9$</td>
<td>0.8</td>
<td>0.0032</td>
<td>0.0384</td>
<td>0.0062</td>
<td>-0.0257</td>
<td>0.7</td>
<td>0.0032</td>
<td>0.0372</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^9$</td>
<td>0.8</td>
<td>0.0019</td>
<td>0.0180</td>
<td>0.0064</td>
<td>-0.0313</td>
<td>0.7</td>
<td>0.0019</td>
<td>0.0178</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^9$</td>
<td>0.8</td>
<td>0.0015</td>
<td>0.0038</td>
<td>0.0062</td>
<td>-0.0298</td>
<td>0.7</td>
<td>0.0016</td>
<td>-0.0037</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^9$</td>
<td>0.8</td>
<td>0.0015</td>
<td>-0.0103</td>
<td>0.0068</td>
<td>-0.0315</td>
<td>0.7</td>
<td>0.0015</td>
<td>-0.0102</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^9$</td>
<td>0.8</td>
<td>0.0016</td>
<td>-0.0172</td>
<td>0.0070</td>
<td>-0.0322</td>
<td>0.7</td>
<td>0.0016</td>
<td>-0.0169</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^9$</td>
<td>0.8</td>
<td>0.0019</td>
<td>-0.0237</td>
<td>0.0060</td>
<td>-0.0308</td>
<td>0.7</td>
<td>0.0018</td>
<td>-0.0230</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^{10}$</td>
<td>0.8</td>
<td>0.0022</td>
<td>0.0351</td>
<td>0.0020</td>
<td>-0.0066</td>
<td>0.7</td>
<td>0.0022</td>
<td>0.0346</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^{10}$</td>
<td>0.8</td>
<td>0.0011</td>
<td>0.0191</td>
<td>0.0021</td>
<td>-0.0100</td>
<td>0.7</td>
<td>0.0011</td>
<td>0.0188</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^{10}$</td>
<td>0.8</td>
<td>0.0007</td>
<td>0.0044</td>
<td>0.0021</td>
<td>-0.0106</td>
<td>0.7</td>
<td>0.0007</td>
<td>0.0043</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^{10}$</td>
<td>0.8</td>
<td>0.0006</td>
<td>-0.0063</td>
<td>0.0022</td>
<td>-0.0183</td>
<td>0.7</td>
<td>0.0006</td>
<td>-0.0063</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^{10}$</td>
<td>0.8</td>
<td>0.0009</td>
<td>-0.0158</td>
<td>0.0046</td>
<td>-0.0420</td>
<td>0.7</td>
<td>0.0009</td>
<td>-0.0156</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^{10}$</td>
<td>0.8</td>
<td>0.0011</td>
<td>-0.0216</td>
<td>0.0084</td>
<td>-0.0515</td>
<td>0.7</td>
<td>0.0011</td>
<td>-0.0214</td>
</tr>
<tr>
<td>$d$</td>
<td>$N$</td>
<td>$\phi$</td>
<td>Banded MSE</td>
<td>Banded Bias</td>
<td>Approx. MLE MSE</td>
<td>Approx. MLE Bias</td>
<td>$\phi$</td>
<td>Approx. MLE MSE</td>
<td>Approx. MLE Bias</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>--------</td>
<td>------------</td>
<td>-------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>--------</td>
<td>----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>-0.25</td>
<td>$2^7$</td>
<td>0.9</td>
<td>0.0735</td>
<td>-0.1274</td>
<td>0.0298</td>
<td>-0.0705</td>
<td>-0.8</td>
<td>0.0099</td>
<td>0.0544</td>
</tr>
<tr>
<td>-0.15</td>
<td>$2^7$</td>
<td>0.9</td>
<td>0.1191</td>
<td>-0.1464</td>
<td>0.0266</td>
<td>-0.0806</td>
<td>-0.8</td>
<td>0.0067</td>
<td>0.0267</td>
</tr>
<tr>
<td>-0.05</td>
<td>$2^7$</td>
<td>0.9</td>
<td>0.0981</td>
<td>-0.0998</td>
<td>0.0260</td>
<td>-0.0898</td>
<td>-0.8</td>
<td>0.0058</td>
<td>0.0035</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^7$</td>
<td>0.9</td>
<td>0.0521</td>
<td>-0.0253</td>
<td>0.0247</td>
<td>-0.0885</td>
<td>-0.8</td>
<td>0.0058</td>
<td>0.0191</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^7$</td>
<td>0.9</td>
<td>0.0290</td>
<td>0.0395</td>
<td>0.0262</td>
<td>-0.0882</td>
<td>-0.8</td>
<td>0.0071</td>
<td>-0.0379</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^7$</td>
<td>0.9</td>
<td>0.0328</td>
<td>0.0727</td>
<td>0.0248</td>
<td>-0.0828</td>
<td>-0.8</td>
<td>0.0081</td>
<td>-0.0489</td>
</tr>
<tr>
<td>-0.25</td>
<td>$2^8$</td>
<td>0.9</td>
<td>0.0310</td>
<td>-0.1022</td>
<td>0.0094</td>
<td>-0.0412</td>
<td>-0.8</td>
<td>0.0058</td>
<td>0.0549</td>
</tr>
<tr>
<td>-0.15</td>
<td>$2^8$</td>
<td>0.9</td>
<td>0.0452</td>
<td>-0.1080</td>
<td>0.0104</td>
<td>-0.0460</td>
<td>-0.8</td>
<td>0.0034</td>
<td>0.0270</td>
</tr>
<tr>
<td>-0.05</td>
<td>$2^8$</td>
<td>0.9</td>
<td>0.0204</td>
<td>-0.0515</td>
<td>0.0099</td>
<td>-0.0459</td>
<td>-0.8</td>
<td>0.0025</td>
<td>0.0071</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^8$</td>
<td>0.9</td>
<td>0.0115</td>
<td>-0.0064</td>
<td>0.0103</td>
<td>-0.0485</td>
<td>-0.8</td>
<td>0.0029</td>
<td>-0.0139</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^8$</td>
<td>0.9</td>
<td>0.0117</td>
<td>0.0474</td>
<td>0.0097</td>
<td>-0.0466</td>
<td>-0.8</td>
<td>0.0033</td>
<td>-0.0293</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^8$</td>
<td>0.9</td>
<td>0.0143</td>
<td>0.0751</td>
<td>0.0096</td>
<td>-0.0400</td>
<td>-0.8</td>
<td>0.0042</td>
<td>-0.0389</td>
</tr>
<tr>
<td>-0.25</td>
<td>$2^9$</td>
<td>0.9</td>
<td>0.0115</td>
<td>-0.0850</td>
<td>0.0040</td>
<td>-0.0216</td>
<td>-0.8</td>
<td>0.0045</td>
<td>0.0554</td>
</tr>
<tr>
<td>-0.15</td>
<td>$2^9$</td>
<td>0.9</td>
<td>0.0130</td>
<td>-0.0821</td>
<td>0.0042</td>
<td>-0.0232</td>
<td>-0.8</td>
<td>0.0021</td>
<td>0.0287</td>
</tr>
<tr>
<td>-0.05</td>
<td>$2^9$</td>
<td>0.9</td>
<td>0.0087</td>
<td>-0.0384</td>
<td>0.0041</td>
<td>-0.0224</td>
<td>-0.8</td>
<td>0.0014</td>
<td>-0.0085</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^9$</td>
<td>0.9</td>
<td>0.0044</td>
<td>-0.0047</td>
<td>0.0040</td>
<td>-0.0231</td>
<td>-0.8</td>
<td>0.0013</td>
<td>-0.0117</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^9$</td>
<td>0.9</td>
<td>0.0063</td>
<td>0.0518</td>
<td>0.0042</td>
<td>-0.0226</td>
<td>-0.8</td>
<td>0.0018</td>
<td>-0.0243</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^9$</td>
<td>0.9</td>
<td>0.0081</td>
<td>0.0701</td>
<td>0.0043</td>
<td>-0.0231</td>
<td>-0.8</td>
<td>0.0025</td>
<td>-0.0369</td>
</tr>
<tr>
<td>-0.25</td>
<td>$2^{10}$</td>
<td>0.9</td>
<td>0.0092</td>
<td>-0.0858</td>
<td>0.0017</td>
<td>-0.0101</td>
<td>-0.8</td>
<td>0.0035</td>
<td>0.0536</td>
</tr>
<tr>
<td>-0.15</td>
<td>$2^{10}$</td>
<td>0.9</td>
<td>0.0079</td>
<td>-0.0753</td>
<td>0.0020</td>
<td>-0.0112</td>
<td>-0.8</td>
<td>0.0015</td>
<td>0.0301</td>
</tr>
<tr>
<td>-0.05</td>
<td>$2^{10}$</td>
<td>0.9</td>
<td>0.0034</td>
<td>-0.0322</td>
<td>0.0018</td>
<td>-0.0117</td>
<td>-0.8</td>
<td>0.0006</td>
<td>0.0085</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^{10}$</td>
<td>0.9</td>
<td>0.0029</td>
<td>-0.0080</td>
<td>0.0018</td>
<td>-0.0113</td>
<td>-0.8</td>
<td>0.0006</td>
<td>-0.0091</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^{10}$</td>
<td>0.9</td>
<td>0.0039</td>
<td>0.0526</td>
<td>0.0019</td>
<td>-0.0131</td>
<td>-0.8</td>
<td>0.0011</td>
<td>-0.0238</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^{10}$</td>
<td>0.9</td>
<td>0.0058</td>
<td>0.0688</td>
<td>0.0019</td>
<td>-0.0124</td>
<td>-0.8</td>
<td>0.0018</td>
<td>-0.0344</td>
</tr>
</tbody>
</table>
\( \phi = -0.7 \) the banded MLE’s MSE and absolute bias is 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 \( \theta \), the banded MLE is generally as good if not better than the exact and approximate MLE. Furthermore, unlike Tieslau et al. (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 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.\(^{13}\)

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 \( O(N^2) \) Levinson algorithm, the banded MLE is only \( O(N) \). This difference in the number of computations is important when one recognizes that these calculations are performed at each iteration of the optimization procedure.

6. Conclusion

In this paper we have synthesized wavelet analysis with the long-memory ARFIMA process 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 dependence allows us to approximate the matrix to high degrees of precision with a matrix whose off-diagonal elements are set to zero. We find that this approximating covariance matrix significantly reduced

\(^{13}\) 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.
the order of calculating the likelihood function since inversion of the approximate matrix is of $\mathcal{O}(N)$ 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 simulations revealed that neither the choice of the wavelet, its order of regularity, nor zero-padding had any substantial affect 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 performed better in regards to MSE and bias for contaminated ARFIMA$(0,d,0)$ processes and uncontaminated 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 was 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 an 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.

Acknowledgements

An earlier version of this paper was presented at the 1994 European Conference Series in Quantitative Economics and Econometrics EC2. This paper has benefited from the communication with two anonymous referees. I would like to thank Victor Wickerhauser for introducing me to wavelets and the University of Missouri Research Board for their financial support.

Appendix A. Lemmas

Lemma A.1 enables us to generalize any ARFIMA model with an unknown mean $\mu$ to have mean zero.

**Lemma A.1.** Let $x(t)$ be a ARFIMA$(p,d,q)$ process with unknown mean $\mu$ and $x'(t)$ be a ARFIMA$(p,d,q)$ process with mean zero, then $\langle x'(t), \psi_{m,n} \rangle = \langle x(t), \psi_{m,n} \rangle$. 
Proof. For almost every \( \omega \) in the probability space, \((\Omega, \mathcal{F}, \mathcal{P})\), we have

\[
\langle x'(t), \psi_{m,n} \rangle = \langle x(t) - \mu, \psi_{m,n} \rangle
\]

\[
= \int (x(t) - \mu) \psi_{m,n} \, dt
\]

\[
= \int x(t) \psi_{m,n} \, dt - \int \mu \psi_{m,n} \, dt.
\]

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

\[
\langle x'(t), \psi_{m,n} \rangle = \langle x(t), \psi_{m,n} \rangle.
\]

In proving Theorem 2 we will need the following lemma.

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

**Proof.**

\[
\int dt \, t^k \Lambda(2^{m-j}, t) = \int dt \int ds \psi(2^{m-j}s - t) \psi(s)
\]

\[
= - \int dt \int ds (2^{m-j}s - t)^k \psi(t) \psi(s)
\]

\[
= - \int dt \int ds \sum_n \binom{k}{n} (2^{m-j}s)^{k-n} (-t)^n \psi(t) \psi(s)
\]

\[
= 0 \quad \text{for } k < 2M.
\]

**Appendix 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 \( \langle x, \psi_{m,n} \rangle \) is

\[
\gamma_{x,\psi}(m, j; n, k) = E[\langle x(t), \psi_{m,n} \rangle \langle x(s), \psi_{j,k} \rangle]
\]

\[
= E[2^{\frac{m-j}{2}} \int dt \int ds x(t)x(s)\psi(2^{m-t} - n)\psi(2^j s - k)]
\]

\[
= 2^{\frac{m-j}{2}} \int dt \int ds E[x(t)x(s)]\psi(2^{m-t} - n)\psi(2^j s - k)
\]

\[
= 2^{\frac{m-j}{2}} \int dt \int ds \gamma(t-s)\psi(2^{m-t} - n)\psi(2^j s - k)
\]

\[
= 2^{\frac{m-j}{2}} \int dt \int ds \gamma(2^{-m}t - 2^{-j}s + 2^{-m}n - 2^{-j}k)\psi(t)\psi(s).
\]
For any \( n, k \in \mathbb{Z} \), as \( |m - j| \to \infty \), or for any \( m, j \in \mathbb{Z} \), as \( |n - k| \to \infty \),

\[
\gamma_{(x, \psi)}(m, j; n, k)
= C(d, \Phi, \Theta) 2^{-\frac{m+n}{d}} \int \int ds \left| 2^{-m}t - 2^{-j}n + 2^{-m}n - 2^{-j}k \right|^{2d-1} \psi(t) \psi(s)

= C(d, \Phi, \Theta) 2^{-\frac{m+n}{d}} 2^{-m(2d-1)} \int \int ds \left| t + 2^{m-j}k - n \right|^{2d-1}
\times \psi(2^{m-j}s - t) \psi(s).
\]

(B.1)

Define \( A(2^{m-j}, t) = \int ds \psi(s) \psi(2^{m-j}s - t) \) and write Eq. (B.1) as

\[
\gamma_{(x, \psi)}(m, j; n, k) = C(d, \Phi, \Theta) 2^{-\frac{m+n}{d}} 2^{-m(2d-1)} \int dt \left| t + 2^{m-j}k - n \right|^{2d-1}
\times A(2^{m-j}, t)

= C(d, \Phi, \Theta) 2^{-\frac{m+n}{d}} 2^{-m(2d-1)} \int dt \left| t \right|^{2d-1}
\times A(2^{m-j}, t - (2^{m-j}k - n)).
\]

(B.2)

The wavelet coefficients, \( \langle x, \psi_{m,n} \rangle \), are stationary over time when \( m = j \) since Eq. (B.2) is a unique function of the difference in translation parameters, \( k - n \), and is stationary over scale since Eq. (B.2) is also a unique function of \( m - j \) for any two translations parameters, \( k \) and \( n \). The property of self-similarity of the wavelet coefficients is also found in Eq. (B.2) since for any \( a, a^{-2(d-1/2)} \gamma_{(x, \psi)}(m, m; ak, an) = \gamma_{(x, \psi)}(m, m; k, n) \).

Appendix C. Proof of Theorem 2

Let \( a = 2^{m-j}k - n \) and \( m \geq j \). Eq. (B.1) can be rewritten as

\[
\gamma_{(x, \psi)}(m, j; n, k)
= C(d, \Phi, \Theta) 2^{-\frac{m+n}{d}} 2^{-m(2d-1)}
\times \int dt \int ds \left| t + a \right|^{2d-1} \psi(2^{m-j}s - t) \psi(s)

= C(d, \Phi, \Theta) 2^{-\frac{m+n}{d}} 2^{-m(2d-1)} \int ds \left| t + a \right|^{2d-1} A(2^{m-j}, t),
\]

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

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

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

$$f(t + x) = |x|^{2d-1} \left\{ 1 + \sum_{i=1}^{\infty} \left( \begin{array}{c} 2d - 1 \\ i \end{array} \right) \left( \frac{t}{x} \right)^i \right\}.$$ (C.2)

Substituting Eq. (C.2) into Eq. (C.1), the autocovariance is

$$\gamma_{<x,y>}(m,j; n, k) = C(d, \Phi, \Theta)2^{-m+1}2^{m(2d-1)}|x|^{2d-1} \left\{ \int_S dt \Lambda(2^{m-j}, t) \right.$$  

$$+ \int_S dt \sum_{i=1}^{\infty} \left( 2d - 1 \right) \left( \frac{t}{x} \right)^i \Lambda(2^{m-j}, t), \right\}. \quad (C.3)$$

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

$$\gamma_{<x,y>}(m,j; n, k) = C_1|x|^{2d-1-2M} + R_{2M+1}, \quad (C.4)$$

where

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

$$R_{2M+1} = C(d, \Phi, \Theta)2^{-m(2d-1)}2^{m+1}x|^{2d-1} \left\{ \sum_{i=2M+1}^{\infty} \left( 2d - 1 \right) \right.$$  

$$\times \int_{-K_1}^{K_1} \int_{-K_1}^{K_1} \left( \frac{2^{j-k}S - t}{x} \right)^i \psi(t)\psi(s) \, dt \, ds \right\}.$$ 

Since $M \geq 1$ and $|d| < 0.5$

$$|R_{2M+1}| \leq C_2|x|^{2d-1} \sum_{i=1}^{\infty} \beta^{2M+i}, \quad (C.5)$$
where

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

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

where \( \Omega = \{(t, s) : -K_1 \leq t, s \leq K_2\} \). Since \( \beta < 1 \), Eq. (C.5) equals

\[ |R_{2M+1}| \leq C_3|\tau|^{2d-1-2M-1}, \tag{C.6} \]

where \( C_3 \) is a finite constant. From Eqs. (C.4) and (C.6)

\[ \gamma_{\langle x, \psi \rangle}(m; n, k) = \mathcal{O}(2^{-jK_1} 2^{-mK_2})^{-1} \tag{C.7} \]

for all \( k \) and \( n \) such that \( 2^{-jK_1} 2^{-mK_2} > \max(2^{-jK_1} 2^{-mK_2}, 2^{-mK_1} 2^{-jK_2}) \).

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

\[ \left| \gamma_{\langle x, \psi \rangle}(m, n; j, k) \right| < C|S|^{M+1} \sup_{t \in S} \left( \frac{\mathcal{O}(2^{2M}(|t + 2^{m-jk} - n|^{2d-1}))}{\mathcal{O}(2^{M})} \right), \]

where \( C \) is dependent on \( \psi \). Clearly, the magnitude of \( \gamma_{\langle x, \psi \rangle}(m, n; j, k) \) is small if either the interval \( S \) or the \( 2M \)th order derivative of \( |t + 2^{m-jk} - n|^{2d-1} \) is small.

References


