Using Wavelets to Obtain a Consistent Ordinary Least Squares Estimator of the Long Memory Parameter

Mark J. Jensen
Department of Economics
University of Missouri
Columbia, MO 65211
e-mail: jensen@haar.econ.missouri.edu
Phone: (573) 882-9925

October 31, 1997

Abstract

We develop an ordinary least squares estimator of the long memory parameter from a fractionally integrated process that is an alternative to the Geweke Porter-Hudak estimator. Using the wavelet transform from a fractionally integrated process, we establish a log-linear relationship between the wavelet coefficients' variance and the scaling parameter equal to the long memory parameter. This log-linear relationship yields a consistent ordinary least squares estimator of the long memory parameter when the wavelet coefficients' population variance is replaced by their sample variance. We derive the small sample bias and variance of the ordinary least squares estimator and test it against the Geweke Porter-Hudak estimator and the McCoy Walden maximum likelihood wavelet estimator by conducting a number of Monte Carlo experiments. Based upon the criterion of choosing the estimator which minimizes the mean squared error, the wavelet OLS approach was superior to the Geweke Porter-Hudak estimator, but inferior to the McCoy Walden wavelet estimator for the processes simulated. However, given the simplicity of programming and running the wavelet OLS estimator and its statistical inference of the long memory parameter we feel the general practitioner will be attracted to wavelet OLS estimator.

Keywords: Fractionally Integrated Processes, Long-Memory, Wavelets

1 Introduction

Wavelet analysis is a relatively new development in the area of applied mathematics that has recently received the attention of statisticians [Donoho and Johnstone (1994), (1995a), (1995b), Donoho, et.al. (1995), Percival (1995) and McCoy and Walden (1996)]. The mathematical theory of wavelets has existed for over half a century, but only recently has its many different strains been pulled together and given the name wavelets.[1] For our purpose, wavelets were first used in time series analysis by seismologists to provide a time dimension to nonstationary seismic signals that Fourier analysis lacked [Morlet (1983)]. The generality and strong results of the wavelet quickly made it useful in other scientific areas, enriching each discipline with its unique combination of mathematics and applications.

By design the wavelet's strength rests in its ability to simultaneously localize a process in time and scale. At high scales, the wavelet has a small centralized time support enabling it to focus in on short lived time phenomena like a singularity point. At low scales, the wavelet has a large time support allowing it to identify long periodic behavior. By moving from low to high scales, the wavelet zooms in on a process's behavior at a point in time, identifying singularities, jumps and cusps. Alternatively, the wavelet can zoom out to reveal the long, smooth features of a series [Mallat and Zhong (1992), Mallat and Hwang (1992) and Wang (1995)].

Scientists in diverse fields have observed time series where observations that are far apart (in time or space) were correlated too strongly to be modeled as independent data or classical autoregressive, moving average models (ARMA). This concept of long memory has grown rapidly and can be found in a broad scattering of fields such as agronomy, astronomy, chemistry, engineering, environmental sciences, geosciences, hydrology, mathematics, physics and statistics. Even in its infancy among economists, long memory has been applied to a number of economic and financial time series. For example, real gross national product [Sowell (1992b), Diebold and Rudebusch (1991)], interest rates [Backus and Zin (1993)], consumer and wholesale price indices [Baillie et al. (1996), Hassler and Wolters (1995)], stock market returns [Ding, et al. (1993)], stock market prices [Lo (1991)], option

[1] See Meyer (1993) for a historical perspective of the wavelet.

prices [Baillie and Bollerslev (1994)] and exchange rates [Cheung (1993)] have all had long memory ideas applied to them.

The empirical presence of long memory is found in the persistence of the autocorrelations. This slow decay by the autocorrelations is not consistent with either the stationary, short-memory, ARMA models, nor the non-stationary, unit root models. Instead, long memory falls nicely in between these two knife-edge approaches. The drawback is the dense covariance matrix it creates, i.e., a large matrix with few zero elements. This dense matrix makes calculation of the exact maximum likelihood function (MLE) impossible for large data sets since inversion of the long memory's covariance matrix is an exhaustive task, requiring on the order of cubed numerical operations.

Using the logarithmic decay a long memory process's autocovariance function, we show that a log-linear relationship exists between the variance of the wavelet coefficient from the long memory process and its scale equal to the long memory parameter. This log-linear relationship lends itself nicely to the estimation of the long memory parameter of a fractional integrated process known as the fractional differencing parameter. We show that the wavelet OLS estimator yields a consistent estimator of the fractional differencing parameter.

In a heuristic manner, McCoy and Walden (1996) have shown the existence of this log-linear relationship between the wavelet coefficients' variance and its scale, but they show it graphically with a plot of $\log_2$ of the sample variance of the wavelet coefficients from a long memory process against the $\log_2$ of the frequency and compare it to the $\log_2$ of the process's power spectrum. McCoy and Walden use this log-linear relationship to calculate the maximum likelihood estimator of the fractional differencing parameter (MW estimator). By using only the wavelet coefficients' variance and ignoring their correlation, McCoy and Walden implicitly assume that the wavelet coefficients' covariance between scale and time are insignificantly different from zero, i.e., the wavelet coefficients are independent over time and scale. Hence, the MW estimator amounts to an approximate maximum likelihood estimator, the precision of which is dependent on how rapidly the wavelet coefficients' autocovariance function decays as the difference in scale and time increases.

The estimator of the fractional differencing parameter most often used is the Geweke, Porter-Hudak (1983) [GPH] estimator. The GPH utilizes a nonparametric approach which regresses the log values of the periodogram on the log Fourier frequencies to estimate the fractional differencing parameter. However, due to the inconsistency of the periodogram as an estimator of the spectrum [Priestley (1992) p. 425], and the normalized periodogram being neither asymptotically independent nor identically distributed [Hurvich and Beltrao (1993), and Robinson (1995)], the GPH estimator has no satisfactory asymptotic properties.

Besides the GPH, the other estimators of the fractional differencing parameter that exist calculate either the exact or approximate maximum likelihood estimator of the fractional differencing parameter. Although the statistical properties of the MLE are well known, their calculation is computationally intensive, suffering from the burden of inverting a dense covariance matrix at each iteration of the numerical optimization algorithm [Deriche and Tewfik (1993), Li and McLeod (1986)], or are approximations of the likelihood function in frequency space [McCoy and Walden (1996), Fox and Taqqu (1986)]. An additional problem associated with the maximum likelihood approaches is their sensitivity to misidentified short memory parameters [Schmidt and Tschernig (1995)]. Unlike the MLEs, the GPH and wavelet OLS estimator do not require the inversion of the covariance matrix, nor the parameterization of the short memory parameters. Hence, they are easier to implement and take fewer cycles to compute.

In Section 2 we provide a brief theoretical background of the wavelet.[2] In Section 3, we define the particular long memory process we are interested in. We then establish in Section 4 the log-linear relationship between the variance of the wavelet coefficient and its scale, and provide some of the asymptotic properties of the wavelet OLS estimator of the long memory parameter. Lastly, in Section 5 we conduct a Monte Carlo simulation to determine the robustness of the wavelet OLS estimator to different values of the long memory parameter and signal length, and to compare these results with the GPH and MW estimators.

[2]For those interested in a basic introduction to wavelets see Strang (1993) or Strichartz (1993). For a broader view of wavelet theory see Daubechies (1992, 1988), Mallat (1989) or Meyer (1993).

2 Wavelet Theory

A wavelet is defined as any function, ψ, whose collection of dilations, j, and translations, k,

$$\psi_{j,k}(t) \quad = \quad 2^{j/2}\psi\left(2^j t - k\right) \tag{1}$$

where $j, k \in \mathbf{Z} = \{0, \pm 1, \pm 2, \ldots\}$, form an orthonormal basis of $L^2(\Re)$. Any continuous function qualifies if it is well localized around zero (decreases rapidly to zero as $t \to \pm\infty$) and oscillates ($\int \psi(t)dt = 0$). These conditions can be strengthened to include more vanishing moments and/or higher orders of continuous derivatives, i.e., $\int t^r \psi(t)dt = 0$ where $r = 0, 1, 2, \ldots, M-1$, and/or $\psi(t) \in C^r$, to enable $\{\psi\}_{m,n}$ to span other function spaces.

The translated and dilated $\psi_{j,k}$ is a well localized function in time around k that can be interpreted as a ideal highpass filter with energy concentrated in the intervals $[-\pi/2^j, -\pi/2^{j+1}] \cup [\pi/2^{j+1}, \pi/2^j]$. The dilated $\psi_{j,0}$ preserves the shape and oscillation of ψ, but $\psi_{j,0}$'s support in time and frequency space is larger ($j < 0$) or smaller ($j > 0$) than ψ. For this reason j is referred to as the scaling parameter.

Let $x(t)$ be a $L^2(\Re)$ real valued function and suppose that we have observations of $x(t)$ at $t = 0, 1, \ldots, 2^p - 1$, where $p \in \mathbf{Z}^+$. Define the inner product, $< ., . >$, by

$$< x, g > \quad = \quad \int x(t)g(t)\, dt$$

where $g \in L^2(\Re)$. The wavelet coefficient of $x(t)$ is a function of the scale parameter, j, and translations parameter, k, equal to

$$w_{j,k} \quad = \quad < x, \psi_{j,k} > = 2^{j/2} \int x(t)\psi\left(2^j t - k\right) dt. \tag{2}$$

The wavelet coefficient, $w_{j,k}$, represents how much information is lost (gained) if the series $x(t)$ is sampled less (more) often. For example, suppose that every two observations of the observed values of $x(t)$ are averaged together, i.e., $y(t/2) = (x(t) + x(t+1))/2$, for $t = 0, 2, 4, \ldots, 2^p$. The wavelet coefficients, $w_{p,k}$, where $k = 0, 1, \ldots, 2^{p-1}$, is the amount that would need to be added to $y(t)$ in order to obtain the original series $x(t)$, i.e., $w_{p,t} = (-x(t) + x(t+1))/2$, for $t = 0, 2, 4, \ldots, 2^p$.. Hence, $\psi_{j,k}$ has the interpretation of being a highpass filter and $w_{j,k}$ is the representation of $x(t)$ at different levels of resolution and

periods of time. This example also illustrates that in the continuous case $w_{j,k}$ involve integrals of the type found in (2), whereas in discrete time the wavelet transform requires matrix multiplication.

Because of the rapid decay in $\psi_{j,k}$, for each j, $\{\psi_{j,k} : k \in \mathbf{Z}\}$ covers the entire real line by shifting by an amount equal to $\psi_{j,0}$'s support. Hence, for a finite number of observations, j need only take on those integer values which keeps $\psi_{j,k}$'s time support equal to or smaller than the support of observed data. Since low values of j require fewer translations for $\psi_{j,k}$ to cover the entire support of the observed data, whereas high values of j require more translations, for a finite series the number of translation parameters will be a function of j.

Since $x(t)$ is a finite series it will have a minimum and maximum scale. The support of $\psi_{j,k}$ can be thought of as $[k2^{-j}, (k+1)2^{-j}]$. If we normalize the time interval of $x(t)$ to the unit interval, i.e., if $T = 2^p$ then $t = 0/T, 1/T, \ldots, (T-1)/T$, then $j \in \mathcal{J} = \{0, 1, \ldots, p-1\}$, and $k \in \mathcal{K}(j) = \{0, 1, \ldots, 2^j - 1\}$. A scaling parameter less than zero causes the wavelet's support to exceed the unit interval and a scaling parameter greater than $p-1$ causes the support of $\psi_{j,k}$ to land in between $x(t-1)$ and $x(t)$.

3 Fractionally Integrated Series

Let $x(t)$ be the fractionally integrated process, $I(d)$, defined by

$$(1-L)^d x(t) = \epsilon(t) \tag{3}$$

where $\epsilon(t) \sim i.i.d.\mathcal{N}(0, \sigma_\epsilon^2)$, $|d| < 1/2$, and $(1-L)^d$ is the fractional differencing operator defined by the binomial expansion

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

where L denotes the lag operator and Γ the gamma operator. $I(d)$ is a generalization of an integrated process, where instead of differencing a series as $(1-L)x(t) = x(t) - x(t-1)$ to obtain stationarity, the series is dth differenced.

For $|d| < 0.5$, $I(d)$ has an infinite moving average representation in which the moving average coefficients decay at the rate j^{d-1} and an infinite autoregressive representation in

which the coefficients decay at the rate j^{-d-1}. Hence, $x(t) \in L^2$ since the moving average (autoregressive) coefficients are square summable when $0 < d < 0.5$ $(-0.5 < d < 0)$.

It is well known [Granger and Joyeux (1980), Hosking (1981), Brockwell and Davis (1993) and Beran (1994)] that the $I(d)$ process's autocovariance function is

$$\begin{aligned} R_x(t,s) &= E[x(t)\,x(s)] \\ &= \frac{\sigma_\epsilon^2 \Gamma(1-2d)\Gamma(|t-s|+d)}{\Gamma(d)\Gamma(1-d)\Gamma(|t-s|+1-d)} \qquad (4) \\ &\approx |t-s|^{2d-1} \qquad \text{as } |t-s| \to \infty. \qquad (5) \end{aligned}$$

The slow hyperbolic decay of $R_x(t,s)$ satisfies the long-memory definition of Resnick (1987).

4 Wavelet OLS Estimator of d

Let $x(t)$ be a mean zero $I(d)$ process with $|d| < 1/2$. Using the autocovariance function of the $I(d)$ process found in Section 3, we arrive at the follow theorem.

Theorem 1 *As $j \to 0$, the wavelet coefficients, $w_{j,k}$, associated with a mean zero $I(d)$ process with $|d| < 1/2$ are distributed $\mathcal{N}(0, \sigma^2 2^{-2jd})$, where σ^2 is a finite constant.*

Proof: See Appendix A.

From Theorem 1, the wavelet coefficients from a $I(d)$ process have a variance that is a function of the scaling parameter, j, but is independent of the translation parameter, k. Hence, define $R(j)$ to be the wavelet coefficient's variance at scale j, i.e., $R(j) = \sigma^2 2^{-2jd}$. Taking the logarithmic transformation of $R(j)$, we obtain the relationship

$$\ln R(j) = \ln \sigma^2 - d \ln 2^{2j} \qquad (6)$$

where $\ln R(j)$ is linearly related to $\ln 2^{-2j}$ by the fractional differencing parameter, d. Hence, the unknown d of a fractionally integrated series can be estimated by the ordinary least squares estimator, $\hat{d}$.

To perform this OLS regression we require a estimate of the wavelet coefficient's population variance, $R(j)$. At scale j, define the sample variance of the wavelet coefficients as

$$\bar{R}(j) = \frac{1}{2^j} \sum_{k=0}^{2^j-1} w_{j,k}^2. \qquad (7)$$

As we will see, if a large number of wavelet coefficients are available for scale j, the wavelet coefficient's sample variance provides a consistent estimate of the true variance, $R(j)$.

To determine the statistical properties of $\hat{d}$, we expand $\ln \bar{R}(a)$ around $\ln R(a)$ in the following Taylor series

$$\ln \bar{R}(a) \quad = \quad \ln R(a) + \frac{\bar{R}(a) - R(a)}{R(a)} - \frac{1}{2}\left[\frac{(\bar{R}(a) - R(a))^2}{R(a)^2}\right]. \tag{8}$$

We require the following theorem showing $w_{j,k}$ to be asymptotically independent.

Theorem 2 *If $\psi(t)$ has $M \geq 1$ vanishing moments with support $[-K_1, K_2]$ where $K_1 \geq 0$ and $K_2 \geq 0$ and $x(t)$ is $I(d)$ with $|d| < 1/2$ then $w_{j,k}$ is asymptotically independent in both time and scale space since $w_{j,k}$'s correlation decays as $\mathcal{O}(|k_1 - k_2|^{2(d-M)-1})$ in time space and as $\mathcal{O}(2^{2j(d-M)-1})$ in scale space, for all k_1 and k_2 such that $|k_1 - k_2| > K_1 + K_2$.*

Proof: See Appendix B.

From Theorem 2, the correlation of the wavelet coefficients from a $I(d)$ process decay exponentially over time and scale space since $|d| < 1/2$ and $M \geq 1$. However, the larger M is, the wider the wavelet's support and fewer are the number of wavelet coefficients that satisfy the condition, $|k_1 - k_2| > K_1 + K_2$. Thus, by choosing a wavelet with a large M, the rate of decay in $w_{j,k}$'s autocovariances increases, but over a subset of $\mathcal{K}(j)$.

In theory the decay of the wavelet correlation should only occur when the difference in the translation parameters are outside the cone, $|k_1 - k_2| > K_1 + K_2$. However, simulations studies have shown the effective support of a wavelet to be much smaller than their theoretical support. Daubechies (1988), Tewfik and Kim (1992), Kaplan and Kuo (1993), and Flandrin (1991) have all found rapid decay in the wavelet coefficient's covariance for translations and dilations inside the cone, $|k_1 - k_2| > K_1 + K_2$.

By Theorems 1 and 2, $R(j)^{-1/2} w_{j,k} \sim \mathcal{N}(0,1)$ and is asymptotically independent as $j \to \infty$. It follows that $R(j)^{-1} \sum_k w_{j,k}^2 \sim \chi^2_{2^j}$, where 2^j is the number of degrees of freedom. Hence,

$$\text{var}\left[\bar{R}(j)\right] \quad = \quad \frac{1}{2^{2j}} \text{var}\left[\sum_{k \in \mathcal{K}(j)} w_{j,k}^2\right]$$

$$
\begin{aligned}
&= \frac{2}{2^{2j}}\left(\sigma^2 2^{-j(2d-1/2)}\right)^2 \\
&= 2\left(\sigma^2 2^{-2j(d+1/4)}\right)^2 \\
&\to 0 \qquad \text{as } j \to \infty
\end{aligned}
$$

when $d > -1/4$. By Markov's law of large numbers, $\bar{R}(j)$ will tend in probability to $R(j)$ as $j \to \infty$.

Equation (8) can now be written as

$$\ln \bar{R}(j) = \ln R(j) + o_p(1).$$

Substituting $\sigma^2 2^{-2jd}$ for $R(j)$, we find

$$\ln \bar{R}(j) = \ln \sigma^2 - d \ln 2^{2j} + o_p(1). \tag{9}$$

In other words, as $j \to \infty$ the OLS estimate of the log-log relationship's slope provides a consistent estimate of the fractional differencing parameter, d.

4.1 First-order asymptotic properties of $\hat{d}$

Let

$$y_j = \ln 2^{-2j} - \frac{1}{p}\sum_{j=0}^{p-1} \ln 2^{-2j}.$$

The wavelet OLS estimate of the fractional differencing parameter is

$$\hat{d} = \left[\sum_{j=0}^{p-1} y_j^2\right]^{-1}\left[\sum_{j=0}^{p-1} y_j \ln \bar{R}(j)\right].$$

Expanding $\hat{d}$ in a Taylor series around $R(j)$ results in

$$
\begin{aligned}
\hat{d} = &\left[\sum_{j=0}^{p-1} y_j^2\right]^{-1}\left[\sum_{j=0}^{p-1} y_j \ln R(j)\right] + \left[\sum_{j=0}^{p-1} y_j^2\right]^{-1} \\
&\times \left[\sum_{j=0}^{p-1} y_j \frac{\bar{R}(j) - R(j)}{R(j)}\right] + \mathcal{O}_p\left(\frac{\operatorname{var}\bar{R}(j)}{R(j)^2}\right).
\end{aligned} \tag{10}
$$

Substituting $\sigma^2 2^{-2jd}$ for the first $R(j)$ in the RHS of (10), we find the bias of $\hat{d}$ to equal

$$\hat{d} - d = \left[\sum_{j=1}^{p-1} y_j^2\right]^{-1}\left[\sum_{j=0}^{p-1} y_j \frac{\bar{R}(j) - R(j)}{R(j)}\right] + \mathcal{O}_p\left(\frac{\operatorname{var}\bar{R}(j)}{R(j)^2}\right). \tag{11}$$

Because $\bar{R}(j)$ tends to $R(j)$ as $j \to \infty$ and $\sum_j y_j^2$ is bounded away from zero, Equation (11) shows $\hat{d}$ to be a consistent estimate of the fractional differencing parameter.

The variance of $\hat{d}$ can be found by calculating the variance of the first and second terms of Equation (11). Because $w_{j,k}$ is a asymptotically independent, normally distributed random variable with mean zero and variance $\sigma^2 2^{-2jd}$, and since $R(j) = \sigma^2 2^{-2jd}$, it follows that

$$\begin{aligned} \frac{\mathrm{var}[\bar{R}(j)]}{R(j)^2} &= \frac{2\left(\sigma^2 2^{-j(2d+1/2)}\right)^2}{\left(\sigma^2 2^{-2jd}\right)^2} \\ &= 2^{1-j} \end{aligned} \tag{12}$$

as $j \to \infty$, and in a similar, but more tedious, manner

$$\mathrm{var}\left(\left[\sum_{j=1}^{p-1} y_j^2\right]^{-1}\left[\sum_{j=0}^{p-1} y_j \frac{\bar{R}(j) - R(j)}{R(j)}\right]\right) = \theta 2^{-j} \tag{13}$$

where $\theta = \theta(1, 2^{-1}, \ldots, 2^{1-p})$ is a constant. Combining (13), (12) and (11), we arrive at

$$\hat{d} - d = \theta^{1/2} 2^{-j/2} Z + o_p(2^{-j/2}) \tag{14}$$

where Z is a random variable with unit variance.

5 Simulations

To determine the robustness of $\hat{d}$ to different values of d and T, and to compare its statistical properties to the GPH and MW estimators, we conducted a Monte Carlo experiment where 1000 artificial $I(d)$ processes were generated. Generating a series that exhibits long-memory has been a synthesis problem where many of the known methods required large amounts of computer memory and were computationally intensive [McLeod and Hipel (1978) and Hosking (1984)]. With this in mind we chose the Davies and Harte (1987) algorithm because of its computational and memory efficiencies.[3]

To insure that our simulations reported only the statistical properties of $\hat{d}$ and not how zero-padding or boundary effects adversely affects $\hat{d}$, the generated $I(d)$ processes had

[3] McCoy and Walden (1996) use the log-linear relationship between the wavelet coefficients' variance and scale as a means of efficiently generating a fractionally integrated process. We chose not to use McCoy and Walden's method so as to minimize the chance that the generated data would help the wavelet OLS estimator when estimating this relationship.

$T = 2^p$ observations, where $p = 7, 8, 9, 10$, and $\bar{R}(j)$ was calculated from the Daubechies wavelet with $M = 1$, for the scales $j = 2, 3, \ldots, p-1$. With finite data it is not always possible to precisely calculate all of the wavelet coefficients. The more regular (larger M) a wavelet is, the larger its support. Hence, at lower scales the wavelet straddles the data, resulting in boundary affects. Since the Daubechies wavelet with $M = 1$ has the smallest possible support, $K_1 = 0$, $K_2 = 1$, no boundary affects could occur.

6 Results

6.1 Mean-squared error

The simulation results for the wavelet OLS and GPH estimators are graphed in the box-plots of Fig. 1 (wavelet OLS) and Fig. 2 (GPH), and tabularized along with the MW estimator in Table 1 by their bias and mean squared error (MSE). Fig. 3 plots the MSE found in Table 1 for the wavelet OLS and GPH estimators against d.

For each value of d and T, the relative precision MSE(OLS)/MSE(GPH) was close to 0.2, suggesting that the small and large sample properties of the wavelet OLS estimator are superior to the GPH estimators. When d was held constant, both estimator's MSE declined as T increased. Whereas, for fixed T the wavelet OLS estimator's MSE was not sensitive to changes in the value of d, while for $T = 2^7, 2^8$ the GPH MSE was sensitive.

The box-plots in Fig. 1 and 2 are informative in their ability to reveal the precision of the wavelet OLS estimator. In Fig. 1 the upper and lower quartiles of the wavelet OLS estimator only overlap with the quartiles from $d \pm 0.1$ when $T = 2^7$. This is in contrast with the GPH estimator where, except for the sample $T = 2^{10}$ (Fig. 2d), its upper and lower quartiles overlapped with those from $d \pm 0.2$.

6.2 Bias

From the finite-sample bias listed in Table 1 four results emerge. First, the bias of the wavelet OLS is alway negative, i.e., d tends to be underestimated by the wavelet OLS estimator, while the bias of the GPH estimator is most often positive. Secondly, for fixed T the bias of the wavelet OLS estimator decreases as d increases. Under these conditions, the

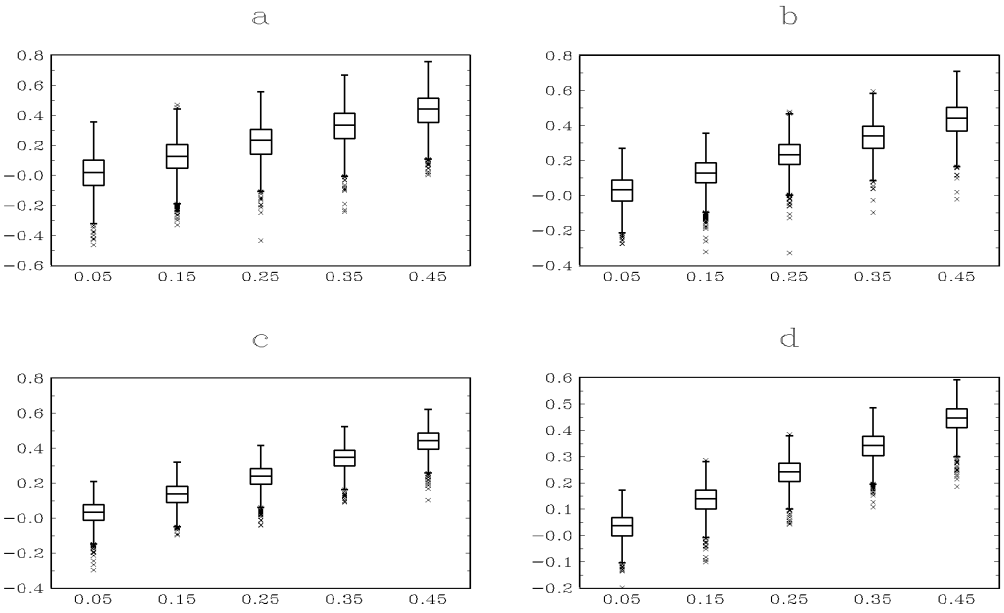

Figure 1: Box-plots of the wavelet OLS estimator from 1000 simulated $I(d)$ processes, where the x-axis is d and the y-axis $\hat{d}$, a) $T = 2^7$, b) $T = 2^8$, c) $T = 2^9$, d) $T = 2^{10}$

GPH estimator's bias did not exhibit any consistent pattern. Thirdly, holding d constant, the absolute value of the wavelet OLS estimator's bias diminished as T increased. Depending on the value of d, the absolute value of the GPH estimator's bias would either go up or down when T increased by a factor of 2, and then possibly reversing this trend for the next increase in T.

Lastly, the absolute value of the wavelet OLS estimator's bias is significantly larger than the bias of the GPH estimator. This larger bias is acceptable given that the wavelet OLS estimator's MSE is significantly smaller than the GPH estimators. Because the mean squared error is comprised of the estimator's level of bias and variance, the bias found in the wavelet OLS estimator is offset by its lower variance. Each of these points can be seen in the box-plots of Fig. 1 and Fig. 2.

		Wavelet OLS		GPH		MW	
T	d	BIAS	MSE	BIAS	MSE	BIAS	MSE
2^7	0.05	-0.0387	0.0164	0.0007	0.0799	0.0169	0.0037
	0.15	-0.0332	0.0172	0.0075	0.0789	0.0158	0.0054
	0.25	-0.0285	0.0174	-0.0067	0.0708	0.0190	0.0054
	0.35	-0.0225	0.0168	0.0095	0.0779	0.0236	0.0045
	0.45	-0.0207	0.0163	0.0246	0.0741	0.0060	0.0017
2^8	0.05	-0.0255	0.0087	-0.0006	0.0486	0.0106	0.0020
	0.15	-0.0273	0.0092	0.0062	0.0404	0.0142	0.0029
	0.25	-0.0215	0.0092	0.0014	0.0472	0.0232	0.0033
	0.35	-0.0186	0.0083	0.0010	0.0419	0.0276	0.0030
	0.45	-0.0179	0.0101	0.0042	0.0468	0.0141	0.0012
2^9	0.05	-0.0239	0.0054	-0.0050	0.0270	0.0065	0.0012
	0.15	-0.0172	0.0048	0.0091	0.0297	0.0167	0.0016
	0.25	-0.0166	0.0053	0.0045	0.0307	0.0247	0.0019
	0.35	-0.0118	0.0051	0.0115	0.0279	0.0314	0.0021
	0.45	-0.0126	0.0055	-0.0021	0.0271	0.0225	0.0010
2^{10}	0.05	-0.0181	0.0030	0.0069	0.0177	0.0050	0.0010
	0.15	-0.0147	0.0031	0.0030	0.0190	0.0157	0.0009
	0.25	-0.0110	0.0032	0.0070	0.0188	0.0244	0.0012
	0.35	-0.0109	0.0033	0.0032	0.0199	0.0321	0.0017
	0.45	-0.0088	0.0034	0.0067	0.0178	*	*

Table 1: Bias and mean-squared error (MSE) of the wavelet OLS, GPH and MW estimator of the fractional differencing parameter, d, from 1000 $I(d)$ processes with T observations. The approximate standard error for the bias is given by $\sqrt{6/(1000T\pi^2)}$ and for the MSEs $\sqrt{72/(1000T^2\pi^4)}$. *MW estimator failed to bracket a maximum in each of the simulations.

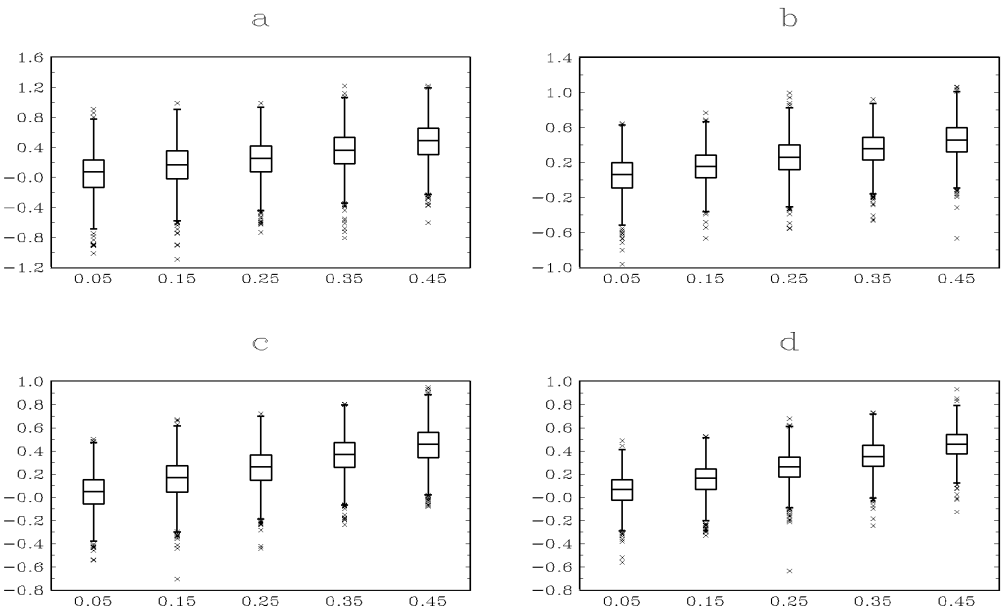

Figure 2: Box-plots of the GPH estimator from 1000 simulated $I(d)$ processes, where the x-axis is d and the y-axis $\hat{d}$, a) $T = 2^7$, b) $T = 2^8$, c) $T = 2^9$, d) $T = 2^{10}$

6.3 Comparison with MW estimator

Before comparing our simulation results with those found for the McCoy and Walden (1996) approximate wavelet MLE, we expected the MW estimator's mean squared error to be smaller than the wavelet OLS. In addition to the wavelet coefficients, the MW estimator includes the maximum scaling coefficient (a measure of the signal's average value) with its corresponding variance in the calculation of the likelihood function. In Table 2 the MW estimators MSE is three to four times smaller than the wavelet OLS's. How much of this improvement is dependent on the inclusion of the scaling coefficient is unknown. We, however, feel it is unlikely that including the scaling coefficient could alone be the reason for the MW estimator's smaller MSE.

Unlike the constant MSE of the wavelet OLS estimator, the MW estimator's MSE did increase for $d = 0.25, 0.35$ at each T. The MW estimator's bias also exhibited this same behavior, with the smallest levels of bias occurring when $d = 0.05, 0.45$.

Increasing the number of observations did not always lead to a decrease in the bias of

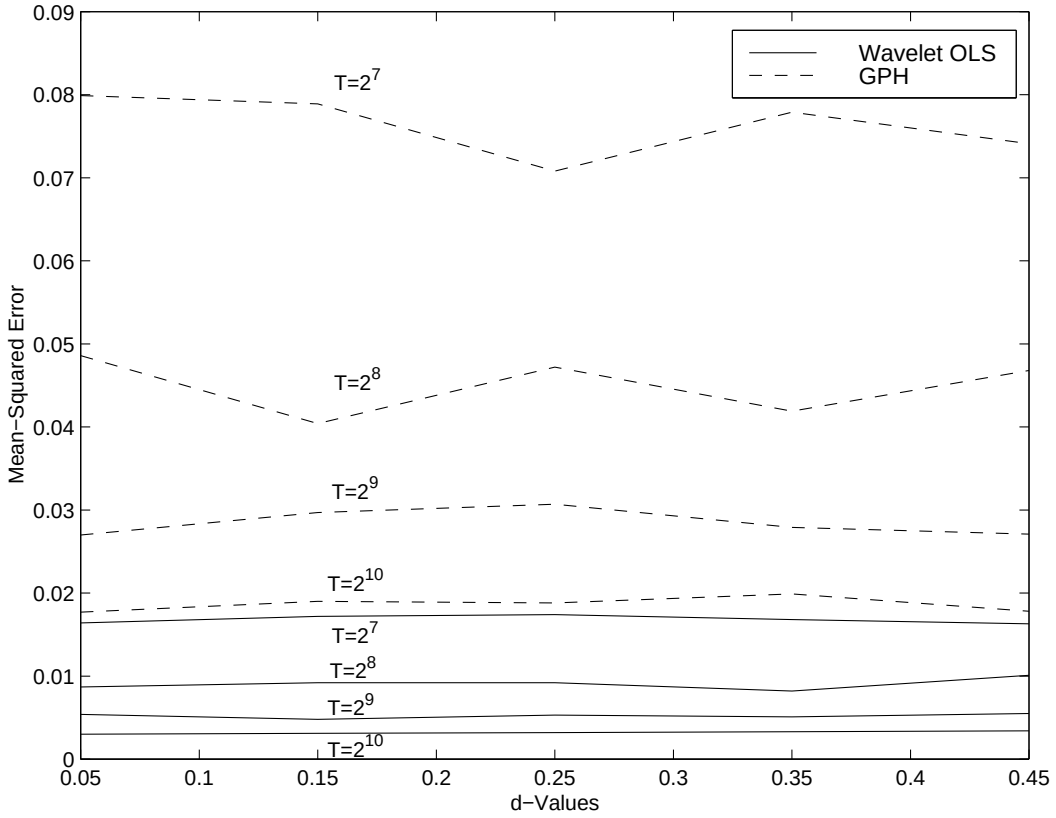

Figure 3: Mean-squared error as a function of d and T.

the MW estimator. Other than $d = 0.05$, the MW estimator showed a increase in its bias ($d = 0.25, 0.35, 0.45$) as T grew larger, or at least a decrease followed by an increase in its bias. The most apparent increase in the bias of the MW estimator was $d = 0.45$. In this case, as T grew the level of bias also grew, until we suspect the bias grew so large that at $T = 1024$ the value of d that maximizes the likelihood function was greater than 0.5. In order to alleviate this problem an alternative maximization algorithm to that provided by McCoy and Walden (1996) is needed.

Because the wavelet OLS estimator's bias decreased with larger T, the absolute level of the wavelet OLS's bias was smaller than the MW estimators in 10 out of the 20 experiments. Six of these cases came when d and T were large. This suggests that for large processes with greater long memory dynamics (d closer to 0.5) the wavelet OLS estimator is an attractive alternative to McCoy and Walden's approximate wavelet maximum likelihood estimator.

7 Conclusion

In this paper we have shown that a log-linear relationship exists between the variance of the wavelet coefficient and the scaling parameter equal to the fractional differencing parameter of a fractionally integrated process. This log-linear relationship provides a simple least squares approach to estimating the differencing parameter. The wavelet OLS estimator of the fractional differencing parameter is shown to be consistent when the sample variance of the wavelet coefficient is used in the regression.

To obtain a consistent estimator of the fractional differencing parameter from a simple OLS regression is a substantial improvement over the popular GPH estimator. The wavelet coefficients' variance is a regularization of the spectrum [Percival (1995), McCoy and Walden (1996)]. Like the spectrum, which decomposes the variance of a series across different frequencies, the wavelet coefficients' variance decomposes the variance of the series across different scales. Those scales which contribute the most to the series' variance are associated with those wavelet coefficients with the largest variance. Hence, the wavelet coefficients' sample variance provides a more intuitive parametric estimate of its population variance than the nonparametric periodogram does of the power spectrum. More importantly, whereas the periodogram is a inconsistent estimator of the spectrum, the wavelet coefficients' sample variance is a consistent estimator of the population variance that enables the wavelet OLS estimator to be a consistent estimator of the fractional differencing parameter.

The Monte Carlo simulations bore this out and showed that the wavelet OLS estimator possesses a smaller mean square error than the GPH estimator for small and large sample sizes and for different values of d. Our simulations also showed the mean squared error of the wavelet OLS estimator to be slightly larger than McCoy and Walden's approximate wavelet MLE. However, the MW estimator's level of bias increased for $d = 0.25, 0.35, 0.45$ as T grew. This led to the MW estimator failure to find the parameter value of d that maximizes the likelihood function when $d = 0.45$ and $T = 1024$. This large bias suggests the maximum likelihood estimator may be greater than 0.5.

We conclude that given the ease of implementing the wavelet OLS estimator and its

non-numerical nature, many practitioners will be attracted to the wavelet OLS estimator. The most apparent attraction being the wavelet OLS estimator's substantial improvement over the often used GPH estimator. Furthermore, its small bias and reasonable mean square error with large data sets also makes it competitive with the MW estimator.

A Proof of Theorem 1

Let $x(t)$ be a mean zero $I(d)$ process with $|d| < 1/2$ and $\sigma_\epsilon^2 = 1$. The expected value of $w_{j,k}$ can easily be shown to equal zero, since

$$E[w_{j,k}] = 2^{j/2} \int E[x(t)]\psi(2^j t - k)dt.$$

The variance of the wavelet coefficients equals

$$\begin{aligned} \text{var}[w_{j,k}] &= E\left[w_{j,k}^2\right] \\ &= 2^j \int dt \int ds\, E[x(t)x(s)]\, \psi(2^j t - k)\psi(2^j s - k). \end{aligned}$$

Using the fractionally integrated processes' autocovariance function found in Equation (4) and by a change of variables

$$\text{var}[w_{j,k}] = K 2^{-j} \int dt \int ds\, \frac{\Gamma(2^{-j}|t-s| + d)}{\Gamma(2^{-j}|t-s| + 1 - d)} \psi(t)\psi(s) \tag{15}$$

Because $\Gamma(k+a)/\Gamma(k+b)$ is approximated well by k^{a-b} for large k, and for normalization purposes $j \in \mathcal{J} = \{0, 1, 2, \ldots, p-1\}$

$$\text{var}[w_{j,k}] = K 2^{-2jd} \int dt \int ds\, |t-s|^{2d-1} \psi(t)\, \psi(s)$$

as $j \to 0$. By another change of variables

$$\text{var}[w_{j,k}] = K' 2^{-2jd} \int dt\, |t|^{2d-1} \Lambda(1,t)$$

where $\Lambda(1,t) = \int \psi(s)\psi(s-t)\, ds$ is the wavelet transform of the 'mother' wavelet. Collecting terms we find

$$\text{var}[w_{j,k}] = \sigma^2 2^{-2jd}$$

where $\sigma^2 = K' \int dt\, |t|^{2d-1} \Lambda(1,t) < \infty$, since $\Lambda(1,t)$ is finite. Thus, $w_{j,k} \sim \mathcal{N}(0, \sigma^2 2^{-2jd})$ as $j \to 0$. Q.E.D.

B Proof of Theorem 2

Let $\psi(t)$ have $M \geq 1$ vanishing moments and $\alpha = 2^j(k_1 - k_2)$. Using the steps found in the proof of Theorem 1, the $\text{corr}(w_{j,k_1}, w_{j,k_2})$ can be written as

$$\begin{aligned} \text{corr}(w_{j,k_1}, w_{j,k_2}) &= \frac{\int dt \int ds\, |t-s+\alpha|^{2d-1}\psi(t)\psi(s)}{\int dt \int ds\, |t-s|^{2d-1}\psi(t)\psi(s)} \\ &= K' \int dt \int ds\, |t+\alpha|^{2d-1}\Lambda(1,t) \end{aligned} \tag{16}$$

where $\Lambda(1,t) = \int ds\, \psi(s-t)\psi(s)$ and K' is a finite constant. Let $|\alpha| > K_1 + K_2$, i.e. $\alpha \notin \text{supp}(\Lambda(1,t))$, so that $|t+\alpha|^{2d-1}$ is continuously differentiable. By the binomial theorem

$$\begin{aligned} |t+\alpha|^{2d-1} &= |\alpha|^{2d-1}\left|1+\frac{t}{\alpha}\right|^{2d-1} \\ &= |\alpha|^{2d-1}\left\{1+\sum_{i=1}^{\infty}\binom{2d-1}{i}\left(\frac{t}{\alpha}\right)^i\right\} \end{aligned} \tag{17}$$

Substituting Equation (17) into Equation (16), the correlation can be written as

$$\begin{aligned} \text{corr}(w_{j,k_1}, w_{j,k_2}) &= K'|\alpha|^{2d-1}\left\{\int dt\, \Lambda(1,t)\right. \\ &\quad \left. + \int dt \sum_{i=1}^{\infty}\binom{2d-1}{i}\left(\frac{t}{\alpha}\right)^i \Lambda(1,t)\right\} \end{aligned} \tag{18}$$

Since $\psi(t)$ has M vanishing moments, $\Lambda(1,t)$ first $2M$ moments are zero (See Tewfik and Kim (1992) for the proof of this result). Hence,

$$\text{corr}(w_{a,k_1}, w_{a,k_2}) = C_1 2^{2j(d-M)+1}|k_1-k_2|^{2(d-M)-1} + R_{2M+1} \tag{19}$$

where

$$C_1 = K'\frac{(2d-1)!}{2M!\,(2(d-M)-1)!}\left(\int dt\, t^M \psi(t)\right)^2$$

and

$$R_{2M+1} = K'|\alpha|^{2d-1}\left\{\sum_{i=2M+1}^{\infty}\binom{2d-1}{i}\left(\frac{s-t}{\alpha}\right)^i \psi(t)\,\psi(s)\,dt\,ds\right\} \tag{20}$$

Since $M \geq 1$ and $|d| < 1/2$

$$|R_{2M+1}| \leq C_2|\alpha|^{2d-1}\sum_{i=1}^{\infty}\sup_{(t,s)\in\Omega}\left|\frac{s-t}{\alpha}\right|^{2M+i}$$

where

$$C_2 \quad = \quad K' \left| \begin{pmatrix} 2d-1 \\ 2M \end{pmatrix} \right| \left(\int |\psi(t)| dt \right)^2$$

and the set $\Omega = \{(t,s) : -K_1 \leq t, s \leq K_2\}$. Since

$$\sup_{(t,s)\in\Omega} \left| \frac{s-t}{\alpha} \right| \quad < \quad 1.$$

it then follows that

$$|R_{2M+1}| \quad \leq \quad C_3 2^{2j(d-M)} |k_1 - k_2|^{2(d-M)} \tag{21}$$

where C_3 is a finite constant. It follows from Equation (19) and Equation (21) that

$$\text{corr}(w_{j,k_1}, w_{j,k_2}) \quad = \quad \mathcal{O}\left(|k_1 - k_2|^{2(d-M)-1} \right)$$

and

$$\text{corr}(w_{j,k_1}, w_{j,k_2}) \quad = \quad \mathcal{O}\left(2^{2j(d-M)-1} \right) \tag{22}$$

for all k_1 and k_2 such that $|k_1 - k_2| > K_1 + K_2$. Q.E.D.

References

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

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

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

[4] Beran, J. (1994) *Statistics for Long-Memory Processes,* (Chapman and Hall: New York).

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

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

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

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

[9] Davies, R. and D. Harte (1987) "Tests for Hurst Effect," *Biometrika,* 74, 95-101.

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

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

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

[13] Donoho, D. L., and I.M. Johnstone, (1994) "Ideal Spatial Adaptation Via Wavelet Shrinkage," *Biometrika,* 81, 425-455.

[14] Donoho, D. L., and I.M. Johnstone, (1995a) "Minimax Estimation Via Wavelet Shrinkage," *Annals of Statistics,*, to be published.

[15] Donoho, D. L., and I.M. Johnstone, (1995b) "Adapting to Unknown Smoothness Via Wavelet Shrinkage," *Journal of the American Statistical Association,* to be published.

[16] Donoho, D. L., I.M. Johnstone, G. Kerkyacharian and D. Picard (1995) "Wavelet Shrinkage: Asymptopia?" *Journal of the Royal Statistical Society, B*, 57, 301-337.

[17] 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).

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

[19] 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.

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

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

[22] Hosking, J. R. (1981) "Fractional Differencing," *Biometrika*, 68, 165-176.

[23] Hurvich C. and K. Beltrao (1993) "Asymptotics for the Low-Frequency Ordinates of the Periodogram of a Long-Memory Time Series," *Journal of Time Series Analysis*, 14, 455-472.

[24] Kaplan, L. and C. Kuo, (1993) "Fractal Estimation from Noisy Data via Discrete Fractional Gaussian Noise and the Haar Basis," *IEEE Transactions on Signal Processing,* 41, 3554-3562.

[25] Li, W.K. and A.I. McLeod (1986) "Fractional Time Series Modelling," *Biometrika*, 73, 217-21.

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

[27] 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.

[28] 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.

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

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

[31] Mallat, S. and S. Zhong (1992) "Characterization of Signals from Multiscale Edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 14, 710-732.

[32] Meyer, Y. (1993) *Wavelets: Algorithms and Applications*, trans. Robert D. Ryan, (SIAM: Philadelphia).

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

[34] Percival, D.P. (1995) "On Estimation of the Wavelet Variance," *Biometrika,* 82, 619-631.

[35] Priestley, M.B. (1992) *Spectral Analysis and Time Series*, (Academic Press: San Diego).

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

[37] Robinson, P.M. (1995) "Log-Periodogram Regression of Time Series with Long Range Dependence," *Annals of Statistics,* 23, 1040-1072.

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

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

[40] Tewfik, A.H., and M. Kim (1992) "Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motion," *IEEE Transaction on Information Theory,* 38, 904-909.

[41] Wang, Y. (1995) "Jump and Sharp Cusp Detection by Wavelets," *Biometrika*, 82, 385-397.