The Philippine Statistician, 2001
Vol. 50, Nos. 1-4, pp. 25-40.

Wavelet-based Inference for Long-memory Processes

Alex C. Gonzaga[1] and Daniel C. Bonzo[2]

ABSTRACT

A long-memory process may be characterized by its corresponding wavelet variance, an analogue of the spectrum, which decomposes the variance of a process with respect to a variable called scale. In this paper, we derive the variance of the logarithm of the maximal-overlap estimator – a relatively efficient estimator of the wavelet variance. We use this to obtain a weighted-least-square estimator and a test for the long-memory parameter. We show that this weighted-least-square estimator is more statistically efficient than the one based on the wavelet-transform estimator of the wavelet variance. Finally, we apply these estimators and tests to determine the long-memory parameter of the Nile river data, a well-known long-memory process.

KEYWORDS: Long-memory process, ARFIMA(p,d,q) process, wavelets, wavelet variance, maximal-overlap estimator

1. INTRODUCTION

In various areas of human endeavor, it is not uncommon to encounter phenomena that are subject to long-range dependence (LRD) or long memory. For instance, the minimum water level of the Nile river is characterized by its slowly decaying autocorrelations. Many hydrological, geophysical, climatological and economic phenomena have likewise exhibited LRD. See, e.g. Beran (1994) or Granger (1966). Studies on telecommunications traffic (e.g. Abry and Veitch, 1997), self-similar processes and fractals (Abry, Veitch and Flandrin, 1997), and unstable processes (Chan and Terrin, 1995) have also involved an analysis of long memory behavior.

Several approaches have been introduced for detection, estimation and testing for long-memory. These include the R/S statistic, variogram, periodogram-based least square estimator, maximum likelihood estimators, and M-estimators (Beran, 1994). Recently, Jensen (1995), introduced wavelet-based ordinary least-square estimator of the long-memory parameter. On the other hand, Beran (1992) proposed a test for long-memory processes based on the spectral density of the process. In this paper, we present a weighted-least-square estimator of the long-memory parameter based on the maximal-overlap estimator and wavelet-transform estimator, which are known to be unbiased and consistent estimators of the wavelet variance.

The organization of this paper is as follows. We present an introduction of wavelets, wavelet variance and long-memory process in Sections 2, 3 and 4, respectively. Our main results and applications are given in Sections 5 and 6. We give some concluding remarks in Section 7.

2. Wavelets

A *wavelet* is defined by

$$\psi_{a,b}(t) = |a|^{-1/2}\, \psi(a^{-1}(t-b))$$

[1]College of Arts and Sciences, University of the Philippines Manila, Padre Faura St., Manila
[2] School of Statistics, University of the Philippines Diliman, Quezon City

where $a, b \in R$ $(a \neq 0)$. The function $\psi(t) \in L^2(R)$ is often referred to as the *mother wavelet* and must satisfy the admissibility condition given by $\int_R |\Psi(w)|^2 |w|^{-1} dw < \infty$, where $\Psi(w)$ is the Fourier transform of $\psi(t)$. This admissibility condition is required so that wavelet transforms become invertible. If $\psi(t)$ has sufficient decay, then this condition is equivalent to

$$\Psi(0) = \int_R \psi(t)\, dt = 0.$$

This means that the positive and negative areas 'under' the curve of $\psi(t)$ must cancel out. Moreover, since the Fourier transform is zero at the origin and the spectrum decays at high frequencies, the wavelet has a bandpass behavior. It is often referred to as a bandpass filter function.

Example 1. (*Haar Wavelet*) Historically, the Haar wavelet is the earliest wavelet. It represents a piecewise constant function given by

$$\psi(t) = \begin{cases} 1 & 0 \le t \le 1/2 \\ -1 & 1/2 \le t \le 1 \\ 0 & \text{otherwise} \end{cases}$$

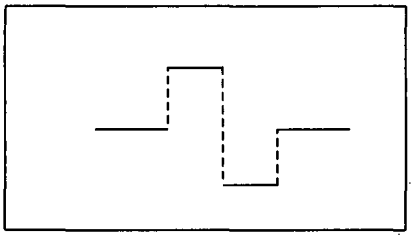

Figure 1. Haar Wavelet.

The *continuous wavelet transform* of $x(t) \in L^2$ at the time-scale location (b,a) is defined by the inner product

$$< x, \psi_{a,b} > = |a|^{-1/2} \int x(t)\, \psi(a^{-1}[t-b])\, dt.$$

By introducing an appropriate constant $c>0$ (in frequency unit selected by the choice of $\psi(t)$), we have the following mapping from scale a to frequency w

$$f(a) = c/a = w.$$

One method to determine this constant c is to take the inverse wavelet transform (IWT) of a function with a single but unknown frequency and to match this value with scale axis.

The wavelet transforms, $< x, \psi_{a,b} >$, satisfy the property

$$\int |< x, \psi_{a,b} >|^2\, db = \int |x(t)|^2\, dt.$$

Hence, they completely characterize $x(t)$ in the L^2 sense. Moreover, $x(t)$ may be reconstructed by the inverse transform given by

$$x(t) = C_\psi^{-1} \int\int a^{-2} <x, \psi_{a,b}> \psi_{a,b}\, da\, db$$

where $C_\psi^{-1} = 2\pi \int |\Psi(\xi)|^2 |\xi|^{-1} d\xi < \infty$. The admissibility condition $\int \psi(t)\, dt=0$ is implied by $C_\psi^{-1} < \infty$ if $\psi(t)$ has sufficient decay.

The *discrete wavelet transform* (DWT) of $x(t) \in L^2(R)$ is the doubly indexed sequence $\{d_{j,k}: j,k \in Z\}$, such that

$$d_{j,k} = 2^{j/2} \int_R x(t)\ \psi(2^j(t-k/2^j))\, dt.$$

Note that $d_{j,k}$ is just the value of the continuous wavelet transform of $x(t)$ at the time-scale location $(k/2^j, 1/2^j)$ or at the time-frequency location $(k/2^j, c2^j)$, where $c>0$ is a constant that depends on the choice of $\psi(t)$. If the time interval is normalized to the unit interval, the support of the wavelet becomes $[(n-1)2^{-(m-1)}, n2^{-(m-1)}]$ so that the wavelet covers the entire time series. Hence, for a scaling parameter, m, the translation parameter has values $n = 1,2,3,\ldots,2^{m-1}$. Thus, for a time series of length $N = 2^r$, the discrete wavelet transform (wavelet coefficients) are

$$\{ d_{m,n}: m \in \{1,2,\ldots,r\}, n(m) \in \{1,2,\ldots,2^{m-1}\} \}.$$

The discrete wavelet transform (DWT) has a corresponding fast algorithm for signal decomposition and reconstruction, which is efficient for both computation and implementation on computers and processors. This algorithm is faster than the so-called Fast Fourier Transform (FFT) used in computing the discrete Fourier transform of long time series. Moreover, the information contained in the DWT is sufficient to determine the signal uniquely.

3. Wavelet Variance

The *wavelet variance* $v^2_Y(2^j)$ of a stochastic process $Y_t (t=0, \pm 1,\ldots)$ decomposes $var(Y_t)$ with respect to scale $\lambda = 2^j$, that is,

$$var(Y_t) = \sum_{j=0}^{\infty} v_Y^2(2^j).$$

This is similar to the property of the spectrum of Y_t, R_Y, that satisfies

$$var(Y_t) = \int_{-1/2}^{1/2} R_Y(w)dw.$$

Let

$$W_{t,\lambda} = \sum_{i=0}^{L_\lambda - 1} h_{i,\lambda} Y_{t-i}$$

represent the output obtained from filtering Y_t using the wavelet filter $h_{i,\lambda}$ of scale λ, where $L_\lambda = (2\lambda-1)(L-1)+1$, and L is the length of the wavelet filter h_i (Daubechies, 1992). The wavelet variance for the process Y_t at scale λ is defined by

$$v_Y^2(\lambda) = \frac{E(W_{t,\lambda}^2)}{2\lambda}.$$

The wavelet filter $h_{l,\lambda}$ for scale λ can be regarded as an approximation to a bandpass filter with passband given by $1/4\lambda < |w| \leq 1/2\lambda$ (Percival, 1995). Hence, the wavelet variance can be approximated by

$$v_Y^2(\lambda) \approx 2 \int_{1/4\lambda}^{1/2\lambda} R_Y(w)dw.$$

This approximation improves as the length of the wavelet filter increases.

Suppose that $(Y_1, Y_2, \ldots, Y_N)$ is a portion of the realization of the process Y_t. The *maximal-overlap estimator* of the wavelet variance (Percival, 1995) is defined by

$$\hat{v}_Y^2(\lambda) = \frac{1}{2\lambda N_{W_\lambda}} \sum_{t=L_\lambda}^{N} W_{t,\lambda}^2$$

where $N_{W_\lambda} = N - L_\lambda + 1$. The *wavelet-transform estimator* (Percival, 1995) is given by

$$\hat{v}_Y^{*2}(\lambda) = \frac{1}{2\lambda N_{v_\lambda}} \sum_{t=\lceil L_\lambda/2\lambda \rceil}^{\lfloor N/2\lambda \rfloor} V_{t,\lambda}^2$$

where

$$V_{t,\lambda} = W_{2t,\lambda} \text{ and } N_{V_\lambda} = \left\lfloor \frac{N}{2\lambda} \right\rfloor - \left\lceil \frac{L_\lambda}{2\lambda} \right\rceil + 1.$$

Percival (1995) has shown that the maximal-overlap estimator is more statistically efficient than the wavelet-transform estimator. In fact, for long-memory processes the asymptotic relative efficiency of the wavelet-transform estimator with respect to the maximal-overlap estimator is close to 0.5 for small values of L. Nevertheless, the former is more computationally efficient since it could be obtained from the discrete wavelet transform of the process. Moreover, both of these are unbiased and consistent estimators of the wavelet variance.

The following theorem will be used later to derive a weighted-least-square estimator of the long-memory parameter based on the maximal-overlap estimator.

Theorem 3.1(Percival, 1995) Let R_w be the spectrum of $W_{t,\lambda}$. If R_w is finitely integrable and strictly positive almost everywhere, then the maximal-overlap estimator $\hat{v}^2(\lambda)$ is asymptotically normally distributed with mean $v^2(\lambda)$ and large sample variance $A_{W_\lambda}/(2\lambda^2 N_{W_\lambda})$, where $N_{w_\lambda} = N - L_\lambda + 1$, $A_{w_\lambda} = \int_{-1/2}^{1/2} R_W^2(f)df$, *and* $L_\lambda = (2\lambda - 1)(L - 1) + 1$.

4. Long-memory Process

An ARMA process $\{X_t\}$ is usually referred to as a *short memory process* since the autocorrelation between X_t and X_{t+k} decreases rapidly at an exponential rate to zero as $k \to \infty$, that is, $\rho(k) \sim Cr^{-k}$, $k = 1,2,\ldots$, where $C>0$ and $0<r<1$. Brockwell (1987) defines a *long-memory process* as a stationary process for which $\rho(k) \sim Ck^{2d-1}$ as $k \to \infty$, where $C>0$ and $d < 0.5$. In this case, the autocorrelations decay to zero slowly at a hyperbolic rate. For our purpose, if $d < 0$ and $\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty$, we call $\{X_t\}$ an *intermediate-memory process*. It is a *long memory* process when $0 < d < 0.5$ and $\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty$.

The definition of long-range dependence is an asymptotic one. It tells us that the autocorrelations decrease slowly at a hyperbolic rate as the lag goes to infinity and not the size of each autocorrelation. Hence, a time series with arbitrarily small autocorrelations that tend to zero very slowly may be a long-memory process. Thus, to detect long-range dependence all autocorrelations must be considered simultaneously, instead of taking them separately. This requires a lengthy time series for detection of long-range dependence to be reliable. Nevertheless, unlike short-memory processes, long-range dependence allows for more reliable and precise prediction of remote future values of the series.

Long-memory processes are often modeled by means of the *autoregressive fractionally integrated moving average* (ARFIMA) process. (For our purpose, we say that a stochastic process is *stationary* if it is covariance stationary.) An *ARFIMA(p,d,q) process* $\{X_t\}$ is a stationary process such that

$$\Phi(B)\,(1-B)^d X_t = \Theta(B)\varepsilon_t \qquad (1)$$

where ε_t is white noise , B is the backshift operator such that $BX_t = X_{t-1}$, $\Phi(B)=1+\phi_1B+\ldots+\phi_pB^p$ is the autoregressive operator, $\Theta(B) = 1+\theta_1B+\ldots+\theta_qB^q$ is the moving average operator, and $(1-B)^d$ is the fractional difference operator. If $d\in(0,0.5)$, $\{X_t\}$ is long-memory process (nonsummable autocorrelations). If $d\in(-0.5,0)$, $\{X_t\}$ is an intermediate-memory process (summable autocorrelations). If $d = 0$, equation (1) defines the usual ARMA(p,q), which is a short-memory process. If d is an integer $(1-B)^d$ becomes the usual differencing operator in Box-Jenkins models.

Clearly, $\{X_t\}$ is white noise process if $d = p = q = 0$. The upper bound $d<0.5$ is needed, because for $d \geq 0.5$, the process is not stationary. However, the case d>0.5 can be reduced to the case $-0.5<d<0.5$ by taking appropriate integer differencing. For instance, if equation (1) holds with $d = 1.4$, then the differenced process $(1-B)^d W_t$ is the stationary solution of equation (1) with $d = 0.4$ and $W_t = (1-B)X_t$. The parameter d determines the long-term behavior, whereas p, q, and the corresponding parameters $\phi(B)$ and $\psi(B)$ allow for more flexible modeling of short-range behavior.

A special case of ARFIMA(p,d,q) is the *fractionally integrated I(d)* process or ARFIMA(0,d,0). Note that an ARFIMA(p,d,q) is obtained by passing a fractional $I(d)$ process through an ARMA(p,q) filter, that is,

$$X_t = \Phi(B)^{-1}\Psi(B)X_t^*$$

where X_t^* is a fractional $I(d)$. Hence, the long-term behavior of an ARFIMA(p,d,q) process may be characterized by its corresponding fractional $I(d)$ process.

The spectral density of the ARFIMA(p,d,q) process X_t is given by

$$R(w) = |1-e^{iw}|^{-2d} R_{ARMA}(w),$$

where $R_{ARMA}(w)$ is the spectral density of ARMA(p,q) process given by

$$R_{ARMA}(w) = \frac{\sigma_\varepsilon^2 \left|\Psi(e^{iw})\right|^2}{2\pi\left|\Phi(e^{iw})\right|^2}.$$

The behavior of the spectral density of X_t at the origin is given by

$$R(w) \sim R_{ARMA}(0)|w|^{-2d}$$

Long-range dependence occurs for $0 < d < ½$.

5. Main Results

In this section, we derive the variance of the logarithm of the maximal-overlap estimator of the wavelet variance. We also present a weighted least square estimator and a test for the long-memory parameter d.

In the following lemma, we derive the variance of the logarithm of the maximal-overlap estimator of the wavelet variance.

Lemma 5.1. Let $\hat{v}^2(\lambda)$ be the maximal-overlap estimator of the wavelet variance. If the spectrum of $W_{t,\lambda}$, R_w, is finitely integrable and strictly positive almost everywhere, and $\log_2(\hat{v}^2(\lambda)$ is uniformly integrable, then

i) $\log_2 \hat{v}^2(\lambda) \xrightarrow{d} N(\log_2 v^2(\lambda), A_{w_\lambda} \log_2^2 e / 2\lambda^2 N_{w_\lambda} v^4(\lambda))$,

ii) $var\{\log_2(\hat{v}^2(\lambda)\} \approx \phi(p/2)/\ln 2$,

where $N_{w_\lambda} = N - L_\lambda + 1$, $A_{w_\lambda} = \int R_{w_\lambda}^2(f)df$, $L_\lambda = (2\lambda - 1)(L-1)+1$, and $\phi(.)$ is the trigamma function.

Proof.

i) Let $\hat{v}^2(\lambda)$ be the maximal overlap estimator of the wavelet variance. Since R_w is finitely integrable and strictly positive almost everywhere, from Theorem 3.1,

$$\hat{v}^2(\lambda) \xrightarrow{d} N(v^2(\lambda), A_{w_\lambda} / 2\lambda^2 N_{w_\lambda}).$$

Now

$$f'(x) = \frac{\log_2 e}{x},$$

exists and nonzero for $x>0$. Since $v^2(\lambda)>0$, then by the Delta Method

$$\log_2 \hat{v}^2(\lambda) \xrightarrow{d} N(\log_2 v^2(\lambda), A_{w_\lambda} \log_2^2 e / 2\lambda^2 N_{w_\lambda} v^4(\lambda)).$$

ii) From Percival(1995),

$$\frac{s\hat{v}^2(\lambda)}{v^2(\lambda)}$$

is approximately equal in distribution to a chi-square random variable with s degrees of freedom for large N, where

$$s = \max\{1, 4\lambda^2 N_{w_\lambda} v^4(\lambda)/A_{w_\lambda}\}.$$

Since $log_2(\hat{v}^2(\lambda))$ is uniformly integrable, then for sufficiently large N

$$Var(log_2\hat{v}^2(\lambda)) = Var(log_2\hat{v}^2(\lambda) + log_2 s - log_2 v^2(\lambda))$$

$$= Var(log_2\frac{s\hat{v}^2(\lambda)}{v^2(\lambda)}) = Var(log_2 X),$$

where X is a chi-square random variable with s degrees of freedom. The distribution of X is a member of the exponential family. Its probability density function can be written in canonical form as

$$f(x) = \exp\{\eta \ln x - A(\eta)\},$$

where $A(\eta) = ln(\Gamma(\eta)2^{\eta})$, and $\eta = s/2$. Hence,

$$Var(lnx) = \frac{\partial^2 A(\eta)}{\partial\eta^2} = \frac{\Gamma(\eta)\Gamma''(\eta) - [\Gamma'(\eta)]^2}{[\Gamma(\eta)]^2} = \phi(\eta) = \phi(s/2).$$

where $\phi(.)$ is the trigamma function. Therefore,

$$var\{log_2(\hat{v}^2(\lambda))\} = \phi(s/2)/ln2$$ QED.

In certain practical problems that R_w is bandlimited and flat over its nominal passband, then $s \approx \max\{1, N_{w_\lambda}/2\lambda$, which is a function of λ (Percival, 1995).

In the following lemma, we show that the logarithms of the maximal-overlap estimators are uncorrelated. From Beran (1994) we use the fact that if the covariance between X_j and X_{j+k} is $\gamma(k)$, then the corresponding covariance for the mth Hermite polynomial is given by

$$\gamma_m(k) = cov(H_m(X_j), H_m(X_{j+k})) = m!\, \gamma^m(k). \quad (2)$$

Lemma 5.2. Under the assumptions of Theorem 3.1, let $\hat{v}^2(\lambda)$ be the maximal-overlap estimator of the wavelet variance at scale λ of an ARFIMA(p,d,q) process, then if $j \neq k$

$$cov(log_2\hat{v}^2(2^j), log_2\hat{v}^2(2^k)) \approx 0.$$

Proof.

Suppose that we have a time series of length $N = 2^k$ that can be considered as a realization of a portion $(Y_1, Y_2, \ldots, Y_N)$ of the ARFIMA(p,d,q) process Y_t with mean zero. Hence, $E(W_{t,\lambda}) = 0$. The maximal-overlap estimator of the wavelet variance is given by

$$\hat{v}^2(\lambda)=\frac{1}{2\lambda N_{W_\lambda}}\sum_{t=L_\lambda}^{N}W_{t,\lambda}^2$$

where $N_{w_\lambda}=N-L_\lambda+1$, $A_{w_\lambda}=\int R^2_{w_\lambda}(f)df$, $L_\lambda=(2\lambda-1)(L-1)+1$.(Percival, 1995). Hence,

$$|\operatorname{cov}(W_{t,\lambda_1},W_{s,\lambda_2})|=|E(W_{t,\lambda_1}W_{s,\lambda_2})|=|\sum_{m=0}^{L_{\lambda 1}-1}\sum_{n=0}^{L_{\lambda 2}-1}h_{m,\lambda_1}h_{n,\lambda_2}R(t-s+m-n)|,$$

where $R(k)$ is the autocovariance function of Y_t at lag k. Thus,

$$\begin{aligned}|\operatorname{cov}(W_{t,\lambda_1},W_{s,\lambda_2})| &= \left|\sum_{m=0}^{L_{\lambda 1}-1}\sum_{n=0}^{L_{\lambda 2}-1}h_{m,\lambda_1}h_{n,\lambda_2}\int_{-1/2}^{1/2}e^{i2\pi(t-s+m-n)}dF_Y(v)\right| \\ &= \left|\int_{-1/2}^{1/2}(\sum_{m=0}^{L_{\lambda 1}-1}h_{m,\lambda_1}e^{i2\pi mv})(\sum_{n=0}^{L_{\lambda 2}-1}h_{n,\lambda_2}e^{i2\pi(-n)v})e^{i2\pi(t-s)}dF_Y(v)\right| \\ &= |\int_{-1/2}^{1/2}|H_{\lambda_1}(-v)||H_{\lambda_2}(v)||e^{i2\pi(t-s)}|dF_Y(v) \\ &\le \int_{-1/2}^{1/2}|H_{\lambda_1}(-v)||H_{\lambda_2}(v)|dF_Y(v),\end{aligned}$$

where H is the transfer function of h. From Percival (1995), the wavelet filter $h_{m,\lambda}$ can be regarded as an approximation to a bandpass filter with passband given by

$$\frac{1}{4\lambda}<|v|\le\frac{1}{2\lambda}.$$

Hence, for large $L_\lambda=(2\lambda-1)(L-1)+1$, the supports of $H_{\lambda_1}(.)$ and $H_{\lambda_2}(.)$ are $\left[\frac{-1}{2\lambda_1},\frac{-1}{4\lambda_1}\right)\cup\left(\frac{1}{4\lambda_1},\frac{1}{2\lambda_1}\right]$ and $\left[\frac{-1}{2\lambda_2},\frac{-1}{4\lambda_2}\right)\cup\left(\frac{1}{4\lambda_2},\frac{1}{2\lambda_2}\right]$, respectively. Thus,

$$|\operatorname{cov}(W_{t,\lambda_1},W_{s,\lambda_2})|\le\int_{-1/2}^{1/2}|H_{\lambda_1}(-v)||H_{\lambda_2}(v)|dF_Y(v)=0$$

if $\lambda_1\ne\lambda_2$. Now,

$$|\operatorname{cov}(W^2_{t,\lambda_1},W^2_{t,\lambda_2})|=|\operatorname{cov}(W^2_{t,\lambda_1}-1,W^2_{t,\lambda_2}-1)|.$$

Since the second Hermite polynomial is $H_2(x)=x^2-1$, from Equation(2) we have

$$|\operatorname{cov}(W^2_{t,\lambda_1},W^2_{t,\lambda_2})|=2|\operatorname{cov}(W_{t,\lambda_1},W_{t,\lambda_2})|^2=0.$$

Hence,

$$|cov\,(\hat{v}^2(\lambda_1), \hat{v}^2(\lambda_2)| \le \frac{1}{2\lambda_1 N_{W_{\lambda 1}}} \frac{1}{2\lambda_2 N_{W_{\lambda 2}}} \sum_{t=L_{\lambda 1}}^{N} \sum_{s=L_{\lambda 2}}^{N} |\operatorname{cov}(W_{t,\lambda_1}^2, W_{t,\lambda_2}^2)| = 0.$$

By Lemma 5.1, $E(log_2 \hat{v}^2(\lambda))^2 < \infty$. Since $log_2 X$ is a measurable function for $X>0$, from Beran(1994), $log_2(X_\lambda)$ can be written as

$$log_2 \hat{v}^2(\lambda) = G(X_\lambda) = \sum_{k=m}^{\infty} \frac{a_k}{k!} H_k(X_\lambda),$$

where H_k is the kth Hermite polynomial and $a_k = <G, H_k> = E[G(x)H_k(x)]$. Since $<H_k, H_k> = k!$, $<H_k, H_r> = 0$ for $k \neq r$, and from Equation(2), we have

$$|cov\,(log_2 \hat{v}^2(\lambda_1), log_2 \hat{v}^2(\lambda_2))| = |\sum_{k=m}^{\infty} \sum_{r=m}^{\infty} \frac{a_k}{k!} \frac{a_r}{r!} E[H_k(X_{\lambda_1}) H_r(X_{\lambda_2})]|.$$

$$\le \sum_{k=m}^{\infty} \frac{a_k^2}{k!^2} k! [|\operatorname{cov}(\hat{v}^2(\lambda_1), \hat{v}^2(\lambda_2))|]^k = 0$$

for $\lambda_1 \neq \lambda_2$. Let $\lambda_1 = 2^j$ and $\lambda_2 = 2^k$. Therefore, if $j \neq k$

$$cov\,(log_2 \hat{v}^2(2^j), log_2 \hat{v}^2(2^k)) \approx 0.$$

The following theorem gives us the weighted-least-square estimator of the long-memory parameter d.

Theorem 5.3. Let Y_t be an ARFIMA(p,d,q) process, where $d \in (-1/2, 1/2)$. A wavelet-based weighted-least-square estimator of the long-memory parameter d is given by

$$\hat{d} = \frac{1}{2} \left[\frac{\sum_{j=0}^{j^*} u_j j y_j - (\sum_{j=0}^{j^*} u_j y_j)(\sum_{j=0}^{j^*} u_j j)}{\sum_{j=0}^{j^*} u_j j^2 - (\sum_{j=0}^{j^*} u_j j)^2} + 1 \right]$$

where $\hat{v}^2(2^j)$ is the maximal-overlap estimator of the wavelet variance at scale 2^j, $y_j = log_2 \hat{v}^2(2^j)$, $v_j = \{\phi(s/2)\}^{-1}$, $\phi(.)$ is the trigamma function and

$$u_j = \frac{v_j}{\sum_{j=0}^{j^*} v_j}.$$

Proof.

The spectral density of an ARFIMA(p,d,q) process Y_t at the origin is given by

$$R(w) \sim R_{ARMA}(0)\, |w|^{-2d}.$$

Hence, the wavelet variance of Y_t is

$$v_Y^2(\lambda) \approx C\lambda^{2d-1}.$$

where $C = 2f_{ARMA}(0)$ and $-\frac{1}{2} < d < \frac{1}{2}$.

Now, let

$$\varepsilon_j = \{ \log_2 \hat{v}^2(2^j) - E(\log_2 \hat{v}^2(2^j))\},$$

and

$$y_j = \log_2 \hat{v}^2(2^j),$$

where $\lambda = 2^j$. Clearly, by Lemma 5.1,

$$E(\varepsilon_j) \approx 0 \text{ and } var(\varepsilon_j) \approx \phi(s/2)/\ln 2.$$

(Note that p is a function of j for $\lambda = 2^j$.) By Lemma 5.2, the error terms ε_j are approximately uncorrelated with respect to scale. Hence, we have the regression equation

$$y_j = (2d-1)j + \log_2 C + \varepsilon_j^3, \quad j=0,1,2,3,\ldots,j^*.$$

Let

$$v_j = \{\phi(s/2)\}^{-1}, \text{ and } u_j = \frac{v_j}{\sum_{j=0}^{j^*} v_j}.$$

Performing a weighted least squares fit between y_j and j (Bickel and Doksum, 1977) with weights u_j yields the following estimator of the long-memory parameter d

$$\hat{d} = \frac{1}{2}\left[\frac{\sum_{j=0}^{j^*} u_j j y_j - (\sum_{j=0}^{j^*} u_j y_j)(\sum_{j=0}^{j^*} u_j j)}{\sum_{j=0}^{j^*} u_j j^2 - (\sum_{j=0}^{j^*} u_j j)^2} + 1\right].$$

Q.E.D.

From the maximal-overlap estimator, the upper limit is N and the lower limit is

$$L_\lambda = (2\lambda-1)(L-1)+1,$$

which is an increasing function of λ. Hence, for $\lambda = 2^j$, the maximum possible value of j in the summation is the largest integer satisfying $N - L_\lambda \geq 0$.

This implies that

$$\lambda \leq \frac{N+L-2}{2L-2}, \; L>1.$$

Thus, the choice of j^* should not exceed $\left\| \left\lfloor \log_2\left(\frac{N+L-2}{2L-2}\right) \right\rfloor \right\|$.

Now, let

[3] The ε_j at least approximately satisfy $E(\varepsilon_j) = 0$ and $cov(\varepsilon_i, \varepsilon_j) = 0$, $1 \leq i < j \leq k$. (Bickel and Doksum(1977), p.95)

$$a_j = u_j j - u_j \sum_{j=0}^{j*} u_j j$$

and

$$B = \sum_{j=0}^{j*} u_j j^2 - \left(\sum_{j=0}^{j*} u_j j \right)^2 .$$

Hence,

$$\hat{d} = \frac{1}{2} \left[\frac{\sum_{j=0}^{j*} a_i y_i}{B} + 1 \right].$$

Note that $\sum_{j=0}^{j^*} u_j = 1$, then $\sum_{j=0}^{j^*} a_j = 0$ and $\sum_{j=0}^{j^*} a_j j = B$. The variance could be written as

$$Var\ \hat{d} = var \left[\frac{\sum_{j=0}^{j*} a_j y_i}{2B} + \frac{1}{2} \right] = (2B)^{-2} \left\{ \sum_{j=0}^{j*} a_j^{\ 2} \operatorname{var}(y_j) + \sum_{i \neq j=0}^{j*} a_i a_j \operatorname{cov}(y_i, y_j) \right\}.$$

By Lemma 5.2, $cov(y_i,y_j) = 0$ for $i \neq j$. Hence,

$$Var\ \hat{d} = (2B)^{-2} \left\{ \sum_{j=0}^{j*} a_j^{\ 2}\ var(y_j) \right\}.$$

By Lemma 5.1, the large sample variance of y_i is given by

$$Var(\log_2 \hat{v}^2(\lambda)) = \frac{A_{w_\lambda} \log_2^2 e}{2\lambda^2 N_{W_\lambda} v^4(\lambda)}$$

where $\lambda = 2^j$. Hence, var$\hat{d}$ can be made arbitrarily small by small choice of j^* for large N. Moreover, for large N, $y_j = (2d-1)j + log_2C + \varepsilon_j$ satisfies the properties of a generalized linear model. Hence, the resulting estimator $\hat{d}$ must be asymptotically unbiased for d. However, consistency of $\hat{d}$ can only be assured if $var\hat{d} = o(1)$. Clearly, this is satisfied by choosing a fixed and relatively small j^*. Thus, the choice of j^* provides a trade-off between bias and variance.

The preceding weighted-least-square estimator allows us to estimate the long-memory parameter d of an ARFIMA(p,d,q) process without the knowledge of p and q. It could be applied to any process that behaves as ARFIMA(p,d,q) at the pole. It may be expressed directly in terms of the wavelet-transform estimator of the wavelet variance by replacing the corresponding maximal-overlap estimator. The wavelet-transform estimator is more computationally efficient but less statistically efficient than the maximal-overlap estimator (Percival, 1995).

The following theorem shows that the weighted-least-square estimator based on the maximal-overlap estimator is more statistically efficient that the one based on the wavelet-transform estimator.

Theorem 5.4 Let $\hat{d}_W$ and $\hat{d}_v$ be the weighted-least-square estimators of the long-memory parameter d based on maximal-overlap and wavelet-transform estimators of the wavelet variance, respectively. Under the assumptions of Theorem 3.1, the asymptotic relative efficiency of $\hat{d}_W$ with respect to $\hat{d}_v$ satisfies

$$ARE(\hat{d}_W, \hat{d}_v) < c_{j*} < 1,$$

where $j^* \leq \left\lfloor \log_2\left(\frac{N+L-2}{2L-2}\right)\right\rfloor$ and c_{j*} is an increasing function of j*.

Proof.

From Lemma 5.1, y_j is asymptotically normal. From Lemma 5.2, the y_j are independent. Now, we can write the estimator $\hat{d}_W$ in the form

$$\hat{d}_W = \sum_{j=0}^{j*} y_j \left[\frac{(u_j j - u_j(\sum_{j=0}^{j*} u_j j))}{2\sum_{j=0}^{j*} u_j j^2 - 2(\sum_{j=0}^{j*} u_j j)^2}\right] + \frac{1}{2}.$$

Hence, we have

$$\hat{d}_W \xrightarrow{d} N(d, \operatorname{var}\hat{d}_W)$$

where

$$Var\ \hat{d}_W = var\left[\frac{\sum_{j=0}^{j*} a_j y_j}{2B} + \frac{1}{2}\right]$$

$$= (2B)^{-2}\left\{\sum_{j=0}^{j*} a_j^2 \operatorname{var}(y_j) + \sum_{i \neq j=0}^{j*} a_i a_j \operatorname{cov}(y_i, y_j)\right\},$$

where $\sum_{j=0}^{j*} a_j j = B$. By Lemma 5.2, $cov(y_i,y_j) = 0$ for $i \neq j$. Hence,

$$var\ \hat{d}_W = (2B)^{-2}\left\{\sum_{j=0}^{j*} a_j^2\ var(y_j)\right\}.$$

Thus,

$$var\ \hat{d}_W = (2B)^{-2}(\log_2^2 e)\left\{\sum_{j=0}^{j*} \frac{a_j^2 A_{W_{2^j}}}{2^{2j+1} v^4(2^j) N_{W_{2^j}}}\right\}.$$

Similarly,

$$\hat{d}_V \xrightarrow{d} N(d, \operatorname{var}\hat{d}_V)$$

Thus, by Theorem 5.2.1 (Lehmann (1983), p.345), we obtain

$$ARE(\hat{d}_W, \hat{d}_v) = \frac{\sum_{j=0}^{j*} \frac{a_j^2 A_{W_\lambda}}{\lambda^2 N_{W_\lambda} v^4(\lambda)}}{\sum_{j=0}^{j*} \frac{a_j^2 A_{V_\lambda}}{\lambda^2 N_{V_\lambda} v^4(\lambda)}},$$

where $\lambda = 2^j$. From Percival (1995)

$$\frac{A_{W_\lambda}}{2A_{V_\lambda}} = c_j < 1,$$

where c_j is an increasing function of $j < \infty$, and $\lambda = 2^j$. Hence,

$$ARE(\hat{d}_W, \hat{d}_v) = \frac{\sum_{j=0}^{j^*} \frac{a_j^2 2c_j A_{V_\lambda}}{\lambda^2 2N_{V_\lambda} v^4(\lambda)}}{\sum_{j=0}^{j^*} \frac{a_j^2 A_{V_\lambda}}{\lambda^2 N_{V_\lambda} v^4(\lambda)}},$$

Thus, for

$$j^* \le \left\lfloor \log_2\left(\frac{N+L-2}{2L-2}\right)\right\rfloor$$

we have

$$ARE(\hat{d}_W, \hat{d}_v) < c_{j^*} < 1. \qquad \text{QED.}$$

In the following theorem, we derive an asymptotically uniformly most powerful test for testing H_o: $d \le 0$ versus H_1: $d > 0$. Note that if $d < 0$ the process has intermediate memory, if $d = 0$ the process has short memory (ARMA(p,q)), and if $d \in (0,1/2)$ the process has long memory.

Theorem 5.5. Let Y_t be a ARFIMA(p,d,q) process with $d \in (-0.5,0.5)$ under the assumptions of Lemma 3.1. An asymptotically uniformly most powerful test for testing H_o: $d \le 0$ versus H_1: $d>0$
is given by

$$\phi(x) = \begin{cases} 1 & \hat{d} > k \\ 0 & otherwise \end{cases}$$

where $k = Z_\alpha \sqrt{Var\ \hat{d}}$, $\sum_{j=0}^{j^*} a_j j = B$, and $var\ \hat{d} = (2B)^{-2} (\log_2^2 e)\left\{\sum_{j=0}^{j^*} \frac{a_j^2 A_{W_{2^j}}}{2^{2j+1} v^4(2^j) N_{W_{2^j}}}\right\}$.

Proof.

From the proof of Theorem 5.4, we have

$$\hat{d} \xrightarrow{d} N(d, Var\ \hat{d}),$$

where

$$var\ \hat{d} = (2B)^{-2} (\log_2^2 e)\left\{\sum_{j=0}^{j^*} \frac{a_j^2 A_{W_{2^j}}}{2^{2j+1} v^4(2^j) N_{W_{2^j}}}\right\}.$$

Thus,

$$P_d\left(\frac{\hat{d}-d}{\sqrt{Var\ \hat{d}}} > Z_\alpha\right) = \alpha,$$

where Z_α is the $1-\alpha$ quantile of the standard normal distribution. If $d=0$ we have

$$P_0\left(\hat{d} > Z_\alpha \sqrt{Var\ \hat{d}}\right) = \alpha.$$

Hence, an asymptotically uniformly most powerful test for testing H_o: $d \le 0$ versus H_1: $d>0$ is given by

$$\phi(x) = \begin{cases} 1 & \hat{d} > k \\ 0 & otherwise \end{cases}$$

where $k = Z_\alpha \sqrt{Var\, \hat{d}}$, $Var\ \hat{d} = (2B)^{-2} (\log_2^2 e)\left\{\sum_{j=0}^{j^*} \frac{a_j^{\,2} A_{W_{\lambda_j}}}{2^{2j+1} v^4 (2^j) N_{W_{\lambda_j}}}\right\}$, and $\sum_{j=0}^{j^*} a_j j = B$.QED.

The preceding test may be expressed directly in terms of the wavelet-transform estimator of the wavelet variance, which is more computationally efficient but less statistically efficient than the maximal-overlap estimator.

6. Applications

We apply the wavelet-transform estimator and maximal-overlap estimator of the wavelet variance in estimating and testing for the long memory parameter of the Nile river data – yearly minimum water levels of the Nile for the years 622-1133 A.D. measured at the Roda Gauge near Cairo. This is a well-known example of a long-memory process, which is usually used to assess the performance of estimators of the long-memory parameter.

We first compute the weighted-least-square estimator of d using the wavelet-transform estimator of the wavelet variance, which could be obtained directly from the discrete wavelet transform of the process. Using WAVELAB 0.701, a library of MATLAB routines for wavelet analysis, we compute the wavelet coefficients at different scales after subtracting the mean. The corresponding weights are computed using the polygamma function of MATHEMATICA. Hence, the weighted-least-square estimator of the long-memory parameter based on the wavelet-transform estimator of the wavelet variance for $N=2^9$ is

$$\hat{d}_V = 0.322786,$$

which shows that the Nile river data, indeed, represents a long-memory process.

We also compute the weighted-least-square estimator of d using the maximal-overlap estimator of the wavelet variance. We use the fact that since

$$W_{t,\lambda} = \sum_{i=0}^{L_\lambda - 1} h_{i,\lambda} Y_{t-i},$$

the spectral density of $W_{t,\lambda}$ satisfies

$$R_W(\lambda) = |H(\lambda)|^2 R_Y(\lambda).$$

Hence, given a realization of the process Y_t and the transfer function $H(\lambda)$ of the wavelet filters, the values of $W_{t,\lambda}$ in the definition of maximal-overlap estimator may be computed by taking the inverse transform of R_W. For $N = 2^9$, the weighted-least-square estimator of the long-memory parameter d based on the maximal-overlap estimator of the wavelet variance is

$$\hat{d}_W = 0.310858,$$

which shows that the Nile river data, indeed, represents a long-memory process.

We summarize the values of $\hat{d}$ for the two estimators of the wavelet variance, and for $N = 2^9$ and $N = 2^8$ as follows:

	$N = 2^9$	$N = 2^8$
d_v (wavelet transform)	0.322786	0.254552
d_w (maximal overlap)	0.310858	0.29227

Clearly, the value of $\hat{d}$ for $N = 2^8$ based on the maximal-overlap estimator is much closer to its value at $N=2^9$ than $\hat{d}$ based on the wavelet-transform estimator. This demonstrates the asymptotic relative efficiency of the two estimators in Theorem 5.4.

For the testing procedure in Theorem 5.5, the value of k for $\alpha = 0.05$ is *0.239589*. Thus, for both estimates of *d*, we reject the null hypothesis and conclude that the Nile data represents a long-memory process.

7. Concluding Remarks

The weighted-least-square estimator based on the maximal-overlap estimator of the wavelet variance is shown to be more statistically efficient, but less computationally efficient, than the one based on the wavelet-transform estimator. An algorithm may be designed to compute the maximal-overlap estimator without applying spectral analysis. Moreover, simulations of long-memory processes may implemented to fully assess the efficiency of the estimator and test in making inferences about the long-memory parameter.

The preceding procedures have the advantage of being able to make inferences about d without the knowledge of p and q. However, just like other least-square estimation procedures on d, we just exploited the simple form of the pole of the spectral density at the origin. They do not tell us about the short-term properties of the process.

ACKNOWLEDGMENTS

We wish to thank the editors of the Philippine Statistician and the anonymous referee for their helpful comments and suggestions.

References

ABRY, P., VEITCH, D.(1997). Wavelet Analysis of Long-range Dependent Traffic, submitted for publication.

ABRY, P., et. al.(1997) Long-range Dependence: Revisiting Aggregation with Wavelets, submitted for publication.

BERAN, J. (1994). Statistics for Long-memory Processes, Chapman and Hall, New York.

BICKEL, P., DOKSUM, K. (1977). Mathematical Statistics, Holden-Day, Inc., San Francisco.

BINSOL, A. and BONZO, D. (1996). On the Application of ARFIMA Model to the Philippine Stock Exchange Oil Index, The Philippine Statistician, Vol. 44-45, 33-48.

BROCKWELL, P., and DAVIS, R. (1987). Time Series:Theory and Methods, Springer- Verlag, New York.

CHAN, N., TERRIN, N.(1995) Inference for Unstable Long-memory Processes with Applications to Fractional Unit Root Autoregressions, Annals of Statistics,23,5, pp. 1662-1683.

DAUBECHIES, I.(1992). Ten Lectures on Wavelets, SIAM, Philadelphia.

GRAGER, C.W.J.(1966). The Typical Spectral Shape of an Economic Variable, Econometrica, 34, 150-161.

JENSEN, M.J.(1995). Ordinary Least Square Estimate of the Fractional Differencing Parameter Using Wavelets as Derived from Smoothing Kernels, Working Paper, Southern Illinois University.

LEHMANN, E.L. (1983). Theory of Point Estimation, John Wiley and Sons, New York.

LEHMANN, E.L.(1986). Testing Statistical Hypotheses, John Wiley and Sons, New York.

PERCIVAL, D. (1995). On Estimation of the Wavelet Variance, Biometrika, 82, 3, pp. 619-31.