The Philippine Statistician, 2000
Vol. 49, Nos. 1-4, pp. 57-71.

Wavelet-based Exact Maximum Likelihood Estimation of ARFIMA(p,d,q) Parameters

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

ABSTRACT

In this paper, we modify wavelets so that wavelet coefficients become statistically independent. Under standard regularity conditions, we determine the distribution of the wavelet coefficients. Using the moving average representation of the ARFIMA(p,d,q) process, we obtain wavelet-based exact maximum likelihood estimators of the ARFIMA(p,d,q) parameters. The corresponding Fisher information matrix is also derived.

KEYWORDS AND PHRASES: Long-memory, ARFIMA(p,d,q), wavelets, maximum likelihood

1. INTRODUCTION

There is ample evidence that the phenomenon of long memory occurs in various areas of human endeavor such as in economics, telecommunications and hydrology. A classic example of a long-memory process is the minimum water level of the Nile River, which is characterized by its slowly decaying autocorrelations. Similar behavior has been observed in many hydrological, geophysical, and climatological records (Beran, 1994). Granger (1966) also noted long-range dependence in economics. Recently, studies have been done on the presence of long-term correlations in telecommunications traffic (see, e.g., Abry and Veitch, 1997). Long-memory processes are also related to other dynamic areas of research such as self-similar processes and fractals (Abry, Veitch and Flandrin, 1997), and unstable processes (Chan and Terrin, 1995). For a comprehensive review on long-memory processes, see the monograph by Beran (1994).

Brockwell (1987) defines a *long-memory process* as a stationary process for which the autocorrelation $\rho(k) \sim Ck^{2d-1}$ as $k \to \infty$, where $C > 0$ and $d < 0.5$. In this case, the autocorrelations decay to zero slowly at a hyperbolic rate. On the other hand, an ARMA process $\{X_t\}$, is considered 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, $\rho(k) \sim Cr^k, k = 1,2,...,$ where $C > 0$ and $0 < r < 1$. For our purpose, we refer to a process as *intermediate memory*, if $d < 0$ and $\sum|\rho(k)| < \infty$, while *long memory* occurs when $0 < d < 0.5$ and $\sum|\rho(k)| = \infty$.

The definition of long memory processes or long-range dependence (LRD) tells us the behavior of the autocorrelations as the lag goes to infinity but not the size of individual correlations. Time series with arbitrarily small autocorrelations that tend to zero very slowly may be considered a

[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

long-memory process. To detect LRD, all autocorrelations must be considered simultaneously, instead of taking them separately. This requires a very lengthy time series for detection of LRD to be reliable. However, LRD allows for more reliable and precise prediction of remote future values of the series (than short-memory processes).

A popular model for long-memory is the ARFIMA(p,d,q) process, a generalization of the well-known ARIMA(p,d,q) model. In this paper, we analyze long memory processes using wavelets (cf. Section 2.2, and Chui 1997).

Wavelets have emerged fairly recently as efficient tools for analyzing long-memory processes. Their use could provide estimators of the LRD parameter which have relatively better statistical properties and are more computationally efficient than some traditional non-wavelet estimators (Abry and Veitch, 1997; Abry, Veitch and Flandrin, 1997).

Jensen (1994) obtained a maximum likelihood estimator of the ARFIMA(0,d,0) parameters. However, his assumption of statistical independence of the wavelet coefficients is quite questionable. To solve this problem, we modify the wavelets so that wavelet coefficients become statistically independent. This allows us to derive maximum likelihood estimators of the ARFIMA(p,d,q) parameters. Here, we use the moving average representation of ARFIMA(p,d,q) process, under the usual regularity assumptions, to determine the distribution of its modified wavelet transform. We then use this to derive the likelihood function for ARFIMA(p,d,q) parameters. This then circumvents the usual estimation problems on ARFIMA(p,d,q) processes (see, e.g., Taqqu, 1985 or Bonzo, 1995).

The organization of this paper is as follows. We firstly present an introduction of ARFIMA(p,d,q) process and wavelets in Section 2. Our main results are given in Section 3. We then give some concluding remarks in Section 4.

2. PRELIMINARY CONCEPTS

In this section we present some concepts and standard results on ARFIMA(p,d,q) processes, wavelets, and wavelet transforms that we use in the succeeding sections of this paper.

2.1 ARFIMA(p,d,q) Process

Long-memory processes are often modeled by means of the *fractionally integrated autoregressive 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

$$\Phi(B)(1-B)^d X_t = \Theta(B)Z_t \tag{1}$$

where Z_t is white noise; B is the backshift operator, i.e., $BX_t = X_{t-1}$; Φ is a p^{th}-order polynomial called the autoregressive operator; Θ is a q^{th}-order polynomial called the moving average operator, and; $(1-B)^d$ is the fractional difference operator. For $0 < d < 0.5$, model (1) defines a

long-memory process with non-summable autocorrelations. If $-0.5 < d < 0$, then $\{X_t\}$ is an intermediate-memory process with summable autocorrelations. If $d = 0$, this simplifies to the ARMA(p,q) model, 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, at least in the usual sense. However, in case $d > 0.5$, appropriate integer differencing can constrain d to satisfy $-0.5 < d < 0.5$. Note that the parameter d determines the long-term behavior, whereas p, q, and the corresponding parameters of $\Phi(B)$ and $\Theta(B)$ allow for more flexible modeling of short-range behavior.

When p=q=0 for the ARFIMA(p,d,q) model, we have a fractional I(d) process. An ARFIMA(p,d,q) process may be viewed as passing a fractional I(d) process through an ARMA(p,q) filter. That is, $X_t = \Phi(B)^{-1}\Theta(B)X_t^*$ 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

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

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 $0 < d < 0.5$.

The following gives the moving average representation of ARFIMA(p,d,q) process:

Theorem 2.1.1. (Brockwell and Davis, 1987) Let $\{X_t: t = 0,1,2,\ldots\}$ be an ARFIMA(p,d,q) process. If $d \in (-0.5, 0.5)$, $\Phi(.)$ and $\Theta(.)$ have no common zeros, and $\Phi(z) \neq 0$ for $|z| \leq 1$, there is a unique stationary solution of equation (1) given by

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \tag{2}$$

where Z_t is a white noise,

$$\sum_{j=0}^{\infty} \psi_j B^j = \nabla^d \frac{\Theta(B)}{\Phi(B)} \text{ and } \sum_{j=0}^{\infty} \psi_j^2 < \infty.$$

Remark: As pointed out in Palma and Chan (1997), ψ_i may be calculated as follows:

$$\psi_i = \sum_{j=0}^{i} \varphi_j \eta_{i-j} \quad \text{for } i \geq 0,$$

where

$$\eta_j = \frac{\Gamma(1-d)}{\Gamma(j+1)\Gamma(1-d-j)}$$

and

$$\varphi_j = \theta_j - \sum_{i=1}^{p} \phi_i \varphi_{j-i}, \quad \varphi_0 = 1, \quad \theta_j = 0 \text{ for } j>q.$$

2.2 Wavelets

A *wavelet* is defined by

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

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 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.

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

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

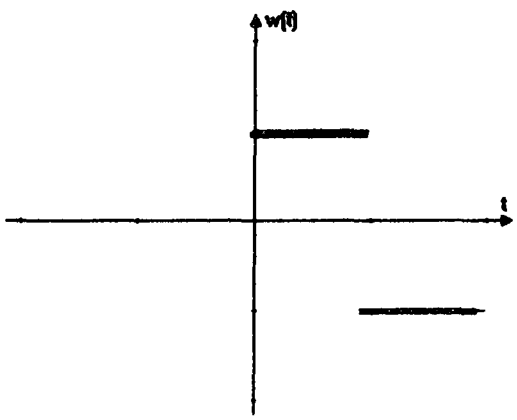

Figure 1 Haar Wavelet

Example 2. *(Mexican Hat Wavelet)* The Mexican hat wavelet (see Figure 2) is represented by

$$\psi(t) = \left(t^2 - 1\right)\exp\left(-\frac{t^2}{2}\right).$$

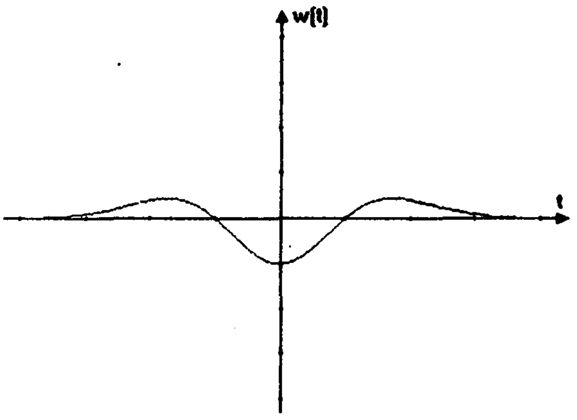

Figure 2 Mexican hat wavelet

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 $C_\psi^{-1} = 2\pi \int |\Psi(\xi)|^2 |\xi|^{-1} d\xi < \infty$. Admissibility $\int \psi(t)dt=0$ is actually implied by the condition $C_\psi^{-1} < \infty$ 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) consists of $\{ d_{m,n}: m \in \{1,2,\ldots,r\}, n(m) \in \{1,2,\ldots,2^{m-1}\} \}$.

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.

We now modify the wavelets so that the wavelet coefficients of X_t are uncorrelated. Define a set of wavelet-like functions $q_{m,n}$ and $q^{m,n}$ such that their Fourier transforms satisfy

$$Q_{m,n}(w) = R_1(w)^{-1/2}\, G_{m,n}(w),$$
$$Q^{m,n}(w) = R_1(w)^{1/2}\, G_{m,n}(w),$$

where $\sigma^2 R_1(w) = R(w)$, $R(.)$ and $G_{m,n}(.)$ are the Fourier transforms of the autocovariance function and the orthonormal wavelet $g_{m,n}(.)$, respectively; σ^2 is the variance of the innovation Z_t.

Now,

$$\{ q_{m,n}(t);\ m,n \in Z\} = \{q_m(t-2^{-m}n);\ m,n \in Z\},$$
$$\{ q^{m,n}(t);\ m,n \in Z\} = \{q^m(t-2^{-m}n);\ m,n \in Z\},$$

and $\{q_{m,n}(t),\ q^{m,n}(t)\}$ is a biorthogonal sequence (See Lemma 13.4 of Walter, 1994). The corresponding scaling-like functions induce a multiresolution-like analysis. If the spectral density of X(t) is strictly positive, then it has the expansion

$$X(t) = \sum_m \sum_n d^*_{m,n}\, q^m(t - 2^{-m} n). \qquad (3)$$

with all the expansion coefficients uncorrelated, and convergence is in the mean square sense (Theorem 13.5, Walter(1994)). Note that

$$d^*_{m,n} = \langle X(t),\ q_m(t-2^{-m}n) \rangle,$$

where the inner product is defined in the L^2 sense.

3. MAIN RESULTS

In this section, we derive maximum likelihood estimators of the parameters of an ARFIMA(p,d,q) process. If $d \in (-0.5, 0.5)$, $\Phi(.)$ and $\Theta(.)$ have no common zeros, and $\Phi(z) \neq 0$ for $|z| \leq 1$, by Theorem 2.1.1, the moving average representation of ARFIMA(p,d,q), X_t, is given by

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j},$$

where $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Hence, $|\psi_r| \to 0$ as $r \to \infty$. For sufficiently large r, $\psi_r \approx 0$. Thus, if the expansion is truncated after r terms, where r is sufficiently large,

$$X_t^{(r)} = \sum_{j=0}^{r} \psi_j Z_{t-j}.$$

(4)

Note that (4) converges to X_t almost surely. We use this truncated expression to avoid some technicalities in determining the distribution of the wavelet coefficients.

In the following lemma and theorem we determine the distribution of $d^*_{m,n}$ of a long-memory ARFIMA(p,d,q) process and we derive the wavelet-based maximum likelihood estimator of ARFIMA(p,d,q) parameters.

Lemma 3.1. Let X_t be an ARFIMA(p,d,q) process with $d\in(0,0.5)$, $\Phi(.)$ and $\Theta(.)$ have no common zeros, $\Phi(z) \neq 0$ for $|z| \leq 1$, $\Theta(z) \neq 0$ for $|z| \leq 1$, and the innovations $Z_t \sim IIDN(0,\sigma^2)$. Then the wavelet-like coefficients $d^*_{m,n}$ are independent $N(0, (2\pi)^{-1}\sigma^2)$.

Proof.

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

$$R(w) = \frac{\sigma^2 \,|\,\Theta(e^{-iw})\,|^2}{2\pi\,|\,\Phi(e^{-iw})\,|^2}\,|\,1-e^{-iw}\,|^{-2d}.$$

Since $|\,1-e^{-iw}\,| = \left|2\sin\left(\frac{1}{2}w\right)\right| = \left(2-2\cos w\right)^{1/2}$, $d\in(0,0.5)$, and $\Theta\,(z) \neq 0$ for $|z| \leq 1$, the spectral density is strictly positive. Hence, Equation (3) applies.

Now,

$$E\,(d^*_{m,n}) = E(<X_t,\, q_m(t-2^{-m}n)>) = \int_R E(X_t)\, q_m(t-2^{-m}n)\,dt = 0.$$

The variance is given by

$$\begin{aligned} E<X_t,\, q_m(t-2^{-m}n)>^2 &= E\,\{\,\int X_t q_m(t-2^{-m}n)dt \;\; \int X_s q_m(s-2^{-m}n)ds\,\} \\ &= \int\int E(X_t X_s) q_m(t-2^{-m}n) q_m(s-2^{-m}n)dtds \\ &= \int\int r(t-s) q_m(t-2^{-m}n) q_m(s-2^{-m}n)dtds. \end{aligned}$$

By Parseval's identity, we have

$$E<X_t,\, q_m(t-2^{-m}n)>^2 = \int \frac{1}{2\pi}\int R(w)e^{-isw}\,\overline{Q_{mn}(w)e^{-i2^{-m}nw}}\, q_m(s-2^{-m}n)dwds.$$

By Fubini's theorem, we get

$$\begin{aligned} E<X_t,q_m(t-2^{-m}n)>^2 &= \frac{1}{2\pi}\int R(w)\overline{Q_{mn}(w)}e^{i2^{-m}nw}\{\int q_m(s-2^{-m}n)e^{-i(s-2^{-m}n)w}ds\}e^{-i2^{-m}nw}dw \\ &= \frac{1}{2\pi}\int R(w)\overline{Q_{mn}(w)}Q_{mn}(w)dw = \frac{1}{2\pi}\int R(w)\,|Q_{mn}(w)|^2\,dw \\ &= \frac{1}{2\pi}\int \sigma^2 R_1(w)\,|Q_{mn}(w)|^2\,dw = \frac{1}{2\pi}\sigma^2\int |G_{mn}^{\cdot}(w)|^2\,dw \\ &= \frac{1}{2\pi}\sigma^2\,\|\,G_{mn}(w)\,\|^2 = \frac{1}{2\pi}\sigma^2. \end{aligned}$$

Now, we have

$$\begin{aligned} <Z_{t-j},\, q_m(t-2^{-m}n)> &= \int Z_{t-j} q_m(t-2^{-m}n)dt \\ &= \lim_{r\to\infty}\sum_{i=0}^{r-1} Z_{t_{ri}-j}\, q_m(t_{ri}-2_r^{-m}n)(t_{r,i+1}-t_{ri}), \end{aligned}$$

where

$$\sup_i (t_{r,i+1}-t_{ri}) \to 0 \;\; as\;\, r\to\infty$$

and lim indicates the convergence in quadratic mean. Hence, $<Z_{t-j}, q_m(t-2^{-m}n)>$ is a limit of sums of independent normal random variables, and has mean zero and finite variance. By the Continuity Theorem, it must be normally distributed. Moreover,

$$Cov\left(<Z_{t-j}, q_m(t-2^{-m}n)>, <Z_{s-j}, q_m(s-2^{-m}n)>\right)$$
$$= \iint E(Z_{t-j}Z_{s-j})q_m(t-2^{-m}n)q_m(s-2^{-m}n)dtds,$$

which is 0 if $t \neq s$. Hence, the inner products $<Z_{t-j}, q_m(t-2^{-m}n)>$ are independent.

Consider $X_t^{(r)}$ in (4). By linearity of inner product, we have

$$<X_t^{(r)}, q_m(t-2^{-m}n)> = <\sum_{j=0}^{r} \psi_j Z_{t-j}, q_m(t-2^{-m}n)> = \sum_{j=0}^{r} \psi_j <Z_{t-j}, q_m(t-2^{-m}n)>.$$

Now, since $X_t^{(r)}$ converges to X_t with probability 1, the distribution of $X_t^{(r)}$ converges to that of X_t. Since the inner product is a continuous function, the distribution of $<X_t^{(r)}, q_m(t-2^{-m}n)>$ converges to the distribution of $<X_t, q_m(t-2^{-m}n)>$. Since $<X_t, q_m(t-2^{-m}n)>$ is a limit sum of independent normal random variables, it must be normally distributed. Therefore, $<X_t, q_m(t-2^{-m}n)>$ is distributed as $N(0, (2\pi)^{-1}\sigma^2)$.

Independence of the wavelet-like coefficients d^*_{mn} follows directly from the fact that d^*_{mn} are uncorrelated. QED.

In the preceding lemma, the conditions $\Phi(z) \neq 0$ for $|z| \leq 1$ and $\Theta(z) \neq 0$ for $|z| \leq 1$ are actually conditions for causality and invertibility of the process.

Theorem 3.2 Let $\{X_1, X_2, \ldots, X_{2^r}\}$ be a realization of the ARFIMA(p,d,q) process under the assumptions of Lemma 3.1. The maximum likelihood estimators of the unknown parameters $(\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q,d,\sigma)=(\underline{\Phi}, \Theta, d, \sigma)$ are $(\hat{\phi}_1,\ldots,\hat{\phi}_p,\hat{\theta}_1,\ldots,\hat{\theta}_p,\hat{d},\hat{\sigma}) = (\hat{\Phi},\hat{\Theta},\hat{d},\hat{\sigma})$, where

$$\hat{\sigma}^2 = \frac{2\pi S(\hat{\Phi},\hat{\Theta},\bar{d})}{N^*},$$

N^* is the number of wavelet-like coefficients and $S(\hat{\Phi},\hat{\Theta},\hat{d})$ minimizes

$$S(\Phi,\Theta,d) = \sum_m \sum_n {d^*_{mn}}^2 .$$

Proof.

Let X_t be an ARFIMA(p,d,q) process under the assumptions of Lemma 3.1. For a time series of length $N = 2^r$, the set of wavelet-like coefficients consists of

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

Let $M = \{1,2,\ldots,r\}$ and $N(m) = \{1,2,\ldots,2^{m-1}\}$. Since $\{d^*_{m,n}: m,n \in Z\}$ are independent $N(0, (2\pi)^{-1}\sigma^2)$, the likelihood function is

$$L(\Phi,\Theta,d,\sigma) = \prod_{m\in M} \prod_{n\in N(m)} \frac{1}{\sigma} \exp\left[\frac{-\pi \cdot d^{*\,2}_{m,n}}{\sigma^2}\right].$$

Hence, the loglikelihood is

$$\ln L(\Phi,\Theta,d,\sigma) = -N^* \ln\sigma - \frac{\pi \sum_m \sum_n d^{*\,2}_{mn}}{\sigma^2}.$$

Maximizing this with respect to the unknown parameters, the maximum likelihood estimators are given by $(\hat{\phi}_1,...,\hat{\phi}_p,\hat{\theta}_1,...,\hat{\theta}_p,\hat{d},\hat{\sigma}) = (\hat{\Phi},\hat{\Theta},\hat{d},\hat{\sigma})$, where

$$\hat{\sigma}^2 = \frac{2\pi S(\hat{\Phi},\hat{\Theta},\hat{d})}{N^*},$$

and N* is the number of wavelet-like coefficients. Substituting $\frac{2\pi S(\Phi,\Theta,d)}{N^*}$ to σ^2 in $\ln(\Phi,\Theta,d,\sigma)$, we have the following

$$\ln(\Phi,\Theta,d,\sigma) = \frac{-N^*}{2}\ln\left(\frac{2\pi}{N^*}\right) - \frac{N^*}{2}\ln\left(\frac{S(\Phi,\Theta,d)}{N^*}\right) - \frac{N^*}{2}.$$

Thus, $S(\hat{\Phi},\hat{\Theta},\hat{d})$ minimizes

$$S(\Phi,\Theta,d) = \sum_m \sum_n d^{*\,2}_{mn}.$$ QED

The following lemma will be used to derive the Fisher information matrix of the wavelet-like coefficient $d^*_{m,n}$ of ARFIMA(0,d,0), $d \in (0,0.5)$.

Lemma 3.3 If $d^*_{m,n}$ is the wavelet-like coefficient in Lemma 3.1 of ARFIMA(0,d,0) process, then

i) $E\left(d^*_{mn} \frac{\partial}{\partial d} d^*_{mn}\right) = 0;$

ii) $E\left(d^*_{mn} \frac{\partial^2}{\partial d^2} d^*_{mn}\right) = 0.$

Proof.

To show (i),

$$E\left(d^*_{mn} \frac{\partial}{\partial d} d^*_{mn}\right) = E \int X_t q_{mn}(t-2^{-m}n)dt \; \frac{\partial}{\partial d} \int X_s q_{mn}(s-2^{-m}n)ds$$

By Leibnitz's rule of differentiating an integral we have

$$E\left(d^*_{mn} \frac{\partial}{\partial d} d^*_{mn}\right) = E \int X_t q_{mn}(t-2^{-m}n)dt \; \int X_s \frac{\partial}{\partial d} q_{mn}(s-2^{-m}n)ds$$

$$= \int\int E(X_t X_s) q_m(t-2^{-m}n) \frac{\partial}{\partial d} q_m(s-2^{-m}n) dt ds$$

$$= \int\int r(t-s) q_m(t-2^{-m}n) \frac{\partial}{\partial d} q_m(s-2^{-m}n) dt ds .$$

By Parseval's identity, we have

$$E\left(d^{\bullet}_{mn} \frac{\partial}{\partial d} d^{\bullet}_{mn}\right) = \int \frac{1}{2\pi} \int R(w) e^{-isw} \overline{Q_{mn}(w) e^{-i2^{-m}nw}} \frac{\partial}{\partial d} q_m(s-2^{-m}n) dw ds .$$

By Fubini's theorem, we get

$$E\left(d^{\bullet}_{mn} \frac{\partial}{\partial d} d^{\bullet}_{mn}\right)$$

$$= \frac{1}{2\pi} \int R(w) \overline{Q_{mn}(w)} e^{i2^{-m}nw} \left\{ \frac{\partial}{\partial d} \int q_m(s-2^{-m}n) e^{-i(s-2^{-m}n)w} ds \right\} e^{-i2^{-m}nw} dw$$

$$= \frac{1}{2\pi} \int R(w) \overline{Q_{mn}(w)} \frac{\partial}{\partial d} Q_{mn}(w) dw$$

$$= \frac{1}{2\pi} \int R(w) \overline{Q_{mn}(w)} \frac{\partial}{\partial d} R_1^{-1/2}(w) G_{mn}(w) dw$$

$$= \frac{1}{2\pi} \int R(w) |Q_{mn}(w)|^2 \ln|2\sin(w/2)| dw$$

$$= \frac{1}{2\pi} \int \sigma^2 R_1(w) |Q_{mn}(w)|^2 \ln|2\sin(w/2)| dw$$

$$= \frac{1}{2\pi} \sigma^2 \int |G_{mn}(w)|^2 \ln|2\sin(w/2)| dw$$

Integrating by parts and noting that $\int |G_{mn}(w)|^2 dw = 1$, we have

$$E\left(d^{\bullet}_{mn} \frac{\partial}{\partial d} d^{\bullet}_{mn}\right) = \frac{\sigma^2}{2\pi}\left(\ln|2\sin(w/2)|\big|_A - \ln|\sin(w/2)|\big|_A\right) = \frac{\sigma^2}{2\pi}\left(\ln 2\big|_A\right) = 0,$$

where A is the passband of $G_{mn}(.)$.

To show (ii), by an analogous argument to the proof of (i), we have

$$E\left(d^{\bullet}_{mn} \frac{\partial^2}{\partial d^2} d^{\bullet}_{mn}\right) = \frac{1}{2\pi} \sigma^2 \int_A |G_{mn}(w)|^2 \ln^2|2\sin(w/2)| dw$$

Integrating by parts and noting that $\int |G_{mn}(w)|^2 dw = 1$, we have

$$E\left(d^{\bullet}_{mn} \frac{\partial^2}{\partial d^2} d^{\bullet}_{mn}\right) = \frac{\sigma^2}{2\pi}\left\{\ln^2|2\sin(w/2)|\big|_A - \int_A \cot(w/2)\ln|2\sin(w/2)| dw\right\}$$

Integrating the second term by parts, we get

$$E\left(d^{\bullet}_{mn} \frac{\partial^2}{\partial d^2} d^{\bullet}_{mn}\right) = \frac{\sigma^2}{2\pi}\left(\ln^2 2\big|_A\right) = 0. \qquad \text{QED.}$$

Theorem 3.4 If $d^*_{m,n}$ is the wavelet-like coefficient in Lemma 3.1 of ARFIMA(0,d,0) process, then the Fisher information matrix of $(d_{11}, d_{12}, \ldots, d_{2^{r-1},r})$ is given by

$$I(d,\sigma) = \sum_{m \in M} \sum_{n \in N(m)} \begin{bmatrix} \Lambda_{mn} & 0 \\ 0 & 2/\sigma^2 \end{bmatrix}$$

where Λ_{mn} is the inverse Fourier transform of $\{G_{mn}(w)\}^2 \ln^2(2\sin(w/2))$ at $t = 2^{-m+1}n$.

Proof.

The loglikelihood function of d^*_{mn} is given by

$$\ln L(d,\sigma) = -\ln\sigma - \frac{\pi}{\sigma^2} d^{*\,2}_{mn}.$$

Hence,

$$E\left(\frac{\partial \ln L(d,\sigma)}{\partial \sigma}\right) = \frac{-1}{\sigma} + \frac{2\pi}{\sigma^3} E\left(d^{*2}_{mn}\right) = \frac{-1}{\sigma} + \frac{1}{\sigma} = 0.$$

Moreover, by Lemma 3.3

$$E\left(\frac{\partial \ln L(d,\sigma)}{\partial d}\right) = \frac{-\pi}{\sigma^2} 2E\left(d^*_{mn} \frac{\partial}{\partial d} d^*_{mn}\right) = 0.$$

Now,

$$E\left(-\frac{\partial^2}{\partial\sigma\partial d} \ln L(d,\sigma)\right) = -E\left(\frac{\partial^2}{\partial\sigma\partial d}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn}\right]\right) = E\left(\frac{\partial}{\partial\sigma}\left[\frac{\pi}{\sigma^2}\frac{\partial}{\partial d} d^{*2}_{mn}\right]\right)$$

$$= \frac{-4\pi}{\sigma^3} E\left(d^*_{mn} \frac{\partial}{\partial d} d^*_{mn}\right)$$

By Lemma 3.3, we have

$$E\left(-\frac{\partial^2}{\partial\sigma\partial d} \ln L(d,\sigma)\right) = 0.$$

Similarly,

$$E\left(-\frac{\partial^2}{\partial d\partial\sigma} \ln L(d,\sigma)\right) = 0.$$

Since d^*_{mn} and $\frac{\partial}{\partial d} d^*_{mn}$ are Gaussian, from (i), d^*_{mn} and $\frac{\partial}{\partial d} d^*_{mn}$ are independent. Hence,

$$E\left(d^{*3}_{mn} \frac{\partial}{\partial d} d^*_{mn}\right) = 0.$$

This and Lemma 3.3 imply that

$$E\left(\frac{\partial}{\partial\sigma}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn}\right]\frac{\partial}{\partial d}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn}\right]\right)$$

$$= E\left(\left[\frac{-1}{\sigma} + \frac{2\pi}{\sigma^3} d^{*2}_{mn}\right]\left[\frac{-\pi}{\sigma^2}\frac{\partial}{\partial d} d^{*2}_{mn}\right]\right)$$

$$= \frac{2\pi}{\sigma^3} E\left(d^*_{mn} \frac{\partial}{\partial d} d^*_{mn} \right) - \frac{4\pi^2}{\sigma^5} E\left(d^{*3}_{mn} \frac{\partial}{\partial d} d^*_{mn} \right) = 0.$$

Thus,

$$E\left(-\frac{\partial^2}{\partial\sigma\partial d} \ln L(d,\sigma) \right) = E\left(\frac{\partial}{\partial\sigma}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn} \right] \frac{\partial}{\partial d}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn} \right] \right) = 0$$

and

$$E\left(-\frac{\partial^2}{\partial d\partial\sigma} \ln L(d,\sigma) \right) = E\left(\frac{\partial}{\partial d}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn} \right] \frac{\partial}{\partial\sigma}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn} \right] \right) = 0.$$

Now,

$$E\left(-\frac{\partial^2}{\partial\sigma\partial\sigma} \ln L(d,\sigma) \right) = -E\left(\frac{\partial}{\partial\sigma}\left[-\frac{1}{\sigma} + \frac{2\pi}{\sigma^3} d^{*2}_{mn} \right] \right) = -E\left(\left[\frac{1}{\sigma^2} - \frac{6\pi}{\sigma^4} d^{*2}_{mn} \right] \right)$$

$$= -\left(\frac{1}{\sigma^2} - \frac{6\pi}{\sigma^4} E\left(d^{*2}_{mn}\right) \right) = \frac{2}{\sigma^2}.$$

Similarly,

$$E\left(\frac{\partial}{\partial\sigma}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn} \right] \frac{\partial}{\partial\sigma}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn} \right] \right) = E\left(\left[\frac{-1}{\sigma} + \frac{2\pi}{\sigma^3} d^{*2}_{mn} \right]^2 \right)$$

$$= E\left(\frac{1}{\sigma^2} - \frac{4\pi}{\sigma^4} d^{*2}_{mn} + \frac{4\pi^2}{\sigma^6} d^{*4}_{mn} \right).$$

Since $E\left(d^{*2}_{mn}\right) = \frac{\sigma^2}{2\pi}$ and $E\left(d^{*4}_{mn}\right) = 3\left(\frac{\sigma^2}{2\pi}\right)^2$, then

$$E\left(\frac{\partial}{\partial\sigma}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn} \right] \frac{\partial}{\partial\sigma}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn} \right] \right) = \frac{2}{\sigma^2}.$$

Thus,

$$E\left(-\frac{\partial^2}{\partial\sigma\partial\sigma} \ln L(d,\sigma) \right) = E\left(\frac{\partial}{\partial\sigma}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn} \right] \frac{\partial}{\partial\sigma}\left[-\ln\sigma - \frac{\pi}{\sigma^2} d^{*2}_{mn} \right] \right) = \frac{2}{\sigma^2}.$$

Now,

$$\frac{\partial}{\partial d} d^*_{mn} = \frac{\partial}{\partial d} \int X_t q_{mn}(t - 2^{-m} n) dt$$

By Parseval's identity, we have

$$\frac{\partial}{\partial d} d^*_{mn} = \frac{1}{2\pi} \frac{\partial}{\partial d} \int X(w) \overline{Q_{mn}(w) e^{-i2^{-m} nw}} dw = \frac{1}{2\pi} \int X(w) \frac{\partial}{\partial d} \overline{Q_{mn}(w) e^{-i2^{-m} nw}} dw$$

$$= \frac{1}{2\pi} \int X(w) \overline{Q_{mn}(w)} e^{i2^{-m} nw} \ln|2\sin(w/2)|\, dw$$

Let

$$\overline{B(w)} = \overline{Q_{mn}(w)}e^{i2^{-m}nw} \ln|2\sin(w/2)|$$

and

$$\overline{b(t)} = \int \overline{Q_{mn}(w)}e^{i2^{-m}nw} \ln|2\sin(w/2)|e^{iwt}dw.$$

Hence,

$$\frac{\partial}{\partial d}d^{\bullet}_{mn} = \int X_t \overline{b(t)}dt.$$

Thus,

$$E\left(\frac{\partial}{\partial d}d^{\bullet}_{mn}\right)^2 = E\iint X_t X_s \overline{b(t)b(s)}dtds = \iint E(X_t X_s)\overline{b(t)b(s)}dtds$$
$$= \iint R(t-s)\overline{b(t)b(s)}dtds.$$

By Parseval's identity and Fubini's theorem, we have

$$E\left(\frac{\partial}{\partial d}d^{\bullet}_{mn}\right)^2 = \frac{1}{2\pi}\int R(w)\overline{B(w)}\left[\int \overline{b(s)}e^{-iws}ds\right]dw$$
$$= \frac{1}{2\pi}\int R(w)\{\overline{B(w)}\}^2 dw = \frac{\sigma^2}{2\pi}\int R_1(w)\{\overline{B(w)}\}^2 dw$$
$$= \frac{\sigma^2}{2\pi}\int R_1(w)\{\overline{Q_{mn}(w)}\}^2 \ln^2|2\sin(w/2)|e^{itw}dw$$
$$= \frac{\sigma^2}{2\pi}\int \{G_{mn}(w)\}^2 \ln^2|2\sin(w/2)|e^{itw}dw = \frac{\sigma^2}{2\pi}\Lambda_{mn}$$

where Λ_{mn} is the inverse Fourier transform of $\{G_{mn}(w)\}^2 \ln^2(2\sin(w/2))$ at $t = 2^{-m+1}n$. Thus,

$$E\left(\frac{\partial}{\partial d}d^{\bullet}_{mn}\right)^2 = \frac{\sigma^2}{2\pi}\Lambda_{mn}.$$

We then have

$$E\left(\frac{\partial}{\partial d}\ln L(d,\sigma)\frac{\partial}{\partial d}\ln L(d,\sigma)\right) = E\left(\frac{-\pi}{\sigma^2}\frac{\partial}{\partial d}d^{\bullet 2}_{mn}.\frac{-\pi}{\sigma^2}\frac{\partial}{\partial d}d^{\bullet 2}_{mn}\right).$$

By Lemma 3.3

$$E\left(\frac{\partial}{\partial d}\ln L(d,\sigma)\frac{\partial}{\partial d}\ln L(d,\sigma)\right) = \frac{4\pi^2}{\sigma^4}E\left(d^{\bullet 2}_{mn}\right)E\left(\frac{\partial}{\partial d}d^{\bullet}_{mn}\right)^2$$
$$= \frac{2\pi}{\sigma^2}E\left(\frac{\partial}{\partial d}d^{\bullet}_{mn}\right)^2 = \Lambda_{mn}.$$

Moreover,

$$E\left(-\frac{\partial^2}{\partial d^2}\ln L(d,\sigma)\right) = \frac{2\pi}{\sigma^2}E\left(d^{\bullet}_{mn}\frac{\partial}{\partial d}\frac{\partial}{\partial d}d^{\bullet}_{mn} + \frac{\partial}{\partial d}d^{\bullet}_{mn}\frac{\partial}{\partial d}d^{\bullet}_{mn}\right).$$

By Lemma 3.3

$$E\left(-\frac{\partial^2}{\partial d^2}\ln L(d,\sigma)\right) = \frac{2\pi}{\sigma^2}\left(E\left[d^{\bullet}_{mn}\frac{\partial}{\partial d}\frac{\partial}{\partial d}d^{\bullet}_{mn}\right] + E\left[\frac{\partial}{\partial d}d^{\bullet}_{mn}\right]^2\right)$$

$$= \frac{2\pi}{\sigma^2} E\left(\frac{\partial}{\partial d} d^{*}_{mn}\right)^2 = \Lambda_{mn}.$$

Hence,

$$E\left(\frac{\partial}{\partial d}\ln L(d,\sigma)\frac{\partial}{\partial d}\ln L(d,\sigma)\right) = E\left(-\frac{\partial^2}{\partial d^2}\ln L(d,\sigma)\right) = \Lambda_{mn}.$$

Thus, the Fisher information matrix of $d^{*}_{m,n}$ is

$$I(d,\sigma) = \begin{bmatrix} \Lambda_{mn} & 0 \\ 0 & 2/\sigma^2 \end{bmatrix}.$$

By independence of $d^{*}_{m,n}$, the Fisher information matrix of $(d_{11}, d_{12}, \ldots, d_{2^{r-1},r})$ is given by

$$I(d,\sigma) = \sum_{m\in M}\sum_{n\in N(m)} \begin{bmatrix} \Lambda_{mn} & 0 \\ 0 & 2/\sigma^2 \end{bmatrix}.$$ QED.

This Fisher information matrix may be used to assess the statistical efficiency of the maximum likelihood estimators of ARFIMA(0,d,0) parameters.

4. CONCLUDING REMARKS

In this paper, we modified the wavelets and the wavelet coefficients. Hence, Mallat's algorithm in computing the wavelet coefficients d_{mn} is not directly applicable to the computation of the wavelet-like coefficients $d^{*}_{m,n}$. An algorithm for this purpose may be designed in future research. Nevertheless, we proposed an exact maximum likelihood estimation procedure in estimating both the short- and long-memory parameters of a long-memory process, which is not dependent on the choice of the wavelet basis functions. This proposed method would also allow us to exploit the independence of the wavelet-like coefficients in investigating other statistical properties of the estimators.

ACKNOWLEDGMENTS

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

References

ABRY, P.,VEITCH, D. (1997a). Wavelet Analysis of Long Range Dependence Traffic (submitted for publication).

ABRY, P., VEITCH, D., FLANDRIN, P.(1997b) Long-Range Dependence: Revisiting Aggregation with Wavelets (submitted for publication).

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

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

BONZO, D.I.B.C. (1995). On a Nonparametric Approach, Diagnosting Testing and Model Building of Nonlinear Time Series, Working Paper, Statistical Center, University of the Philippines.

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

CHAN, N.H., TERRIN, N. (1995) Inference for Unstable Long-Memory Processes with Applications to Fractional Unit Root Autoregressions. Annals of Statistics, 23, No. 5, 1662-1683.

CHUI, C.(1997). Wavelets: A Mathematical Tool for Signal Analysis. SIAM, Philadelphia.

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

JENSEN, M.J.(1994). Wavelet Analysis of Fractionally Integrated Processes, Working Paper, Southern Illinois University.

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.

NASON, G.(1994). Wavelet Regression by Cross-validation, Working Paper, University of Bristol, UK.

NASON, G., and SILVERMAN, B. (1994) . The Stationary Wavelet Transform and some Statistical Applications, Working Paper, University of Bristol, UK.

PALMA, W., and CHAN, A.N.H. (1997). Estimation and Forecasting of Long-memory Processes with Missing Values, Journal of Forecasting, Vol. 16, 385-410.

PEEBLES, P.(1987).Probability, Random Variables, and Random Signal Principles. McGraw-Hill, Singapore.

TAQQU, M.S.(1985). A Bibliographical Guide to Self-similar Processes and Long-range Dependence, In Dependence in Probability and Statistics, Ed. by Eberlein, E. and Taqqu, M.S., 137-161.

WALTER, G. (1994). Wavelets and Other Orthogonal Systems With Applications, CRC Press, Inc., Florida, USA.