WAVELET-BASED BAYESIAN ANALYSIS OF GENERALIZED LONG-MEMORY PROCESS

Alex Gonzaga and Akira Kawanaka

Department of Electrical and Electronics Engineering,
Faculty of Science and Technology, Sophia University
7-1 Kioicho, Chiyoda-ku, Tokyo 102-8554, Japan

ABSTRACT

In this paper we propose a Bayesian approach in estimating the parameters and predicting future values of generalized long-memory process utilizing the approximate likelihood function of discrete wavelet packet coefficients. This approximation does not depend on the length of the signal, but the length of the wavelet filter, which is under the control of the analyst. We illustrate our approach by an example applying simulated data.

1. INTRODUCTION

A digital random signal has long-memory if its autocorrelations decay to zero slowly at a hyperbolic rate. Reference [1] defines a long-memory process as a stationary signal for which the autocorrelation satisfies $\rho(k) \sim Ck^{2d^*-1}$ as $k \to \infty$, where $C > 0$ and $0 < d^* < 0.5$. In this case, the autocorrelations decay to zero slowly at a hyperbolic rate and $\sum |\rho(k)| = \infty$. This phenomenon has been observed in various areas of human endeavor such as in telecommunications, video traffic, economics and hydrology.

A fairly general model of long-memory is the *Gegenbauer autoregressive moving average* (GARMA) process proposed in [2]. It generalizes the definition of long-memory by including a parameter that accounts for persistent cyclic behavior of a random signal. It allows the power spectrum to have pole, not only at 0, but at any frequency in the interval [0,0.5]. This model has been shown to represent well some random signals such as the Bellcore Ethernet trace data [3].

A *GARMA(p,d,u,q) process* is the output of the system function

$$H(z) = \frac{\Theta(z)}{\Phi(z)}(1 - 2uz^{-1} + z^{-2})^{-d} \qquad (1)$$

driven by a stationary white noise input with mean 0 and variance σ^2. The rational function

$$\frac{\Theta(z)}{\Phi(z)} = \frac{1 + \theta_1 z^{-1} + \ldots + \theta_q z^{-q}}{1 - \phi_1 z^{-1} - \ldots - \phi_p z^{-p}}, \quad \theta_q \neq 0 \text{ and } \phi_p \neq 0,$$

is the *autoregressive moving average, ARMA(p,q) system,* such that $\Theta(z)$ and $\Phi(z)$ have no common zeros, and their zeros lie inside the unit circle, which implies that the system is causal and invertible. On the other hand, the *Gegenbauer system*, $(1 - 2uz^{-1} + z^{-2})^{-d}$, $d \neq 0$, can be written as

$$(1 - 2uz^{-1} + z^{-2})^{-d} = \sum_{n=0}^{\infty} C_n^d(u) z^{-n},$$

where $C_n^d(u)$ is the *Gegenbauer polynomial* defined by

$$C_n^d(u) = \sum_{k=0}^{[n/2]} \frac{(-1)^k (2u)^{n-2k} \Gamma((d)-k+n)}{k!(n-2k)!\Gamma(d)},$$

and $[n/2]$ is the largest integer less than or equal to $n/2$. The parameter u provides information about the periodic movement in the random signal. When the input is a stationary white noise, the output is called a *Gegenbauer process*, which is stationary if $d<0.5$ and $|u| <1$ or if $d<0.25$ and $|u|=1$; it is invertible if $-0.5<d$ and $|u| <1$ or $-0.25<d$ and $|u|=1$ [2]. If we set $u=1$ in (1), we obtain the system function of the well-known *autoregressive fractionally integrated moving average (ARFIMA(p,2d,q)) process*, which has a power spectrum with pole at the origin when $0<d<0.5$. It is called a *fractionally differenced process* when $p=q=0$ and $u=1$.

The power spectrum of a GARMA(p,d,u,q) process is given by

$$S_Y(f) = \sigma^2 \left| \frac{\Theta(e^{-i2\pi f})}{\Phi(e^{-i2\pi f})} \right|^2 |2(\cos(2\pi f) - u)|^{-2d} \qquad (2)$$

where $i = \sqrt{-1}$, $f \in (-0.5,0.5]$ and $v = \cos^{-1}(u)/2\pi \in [0,0.5]$ is called the *Gegenbauer frequency* at which the power spectrum becomes unbounded when $0<d<0.5$. It provides an alternative approach in finding the autocovariance function by evaluating the following integral

$$\gamma(k) = 2\int_0^{0.5} S_Y(f)\cos(2\pi kf)df\,, \qquad (3)$$

which can be computed using any software that allows singularities in the integrand. From [4], if $u \in (-1,1)$ the autocovariance function at lag k of a GARMA(0,d,u,0) is given by

$$\gamma(k) = \frac{\sigma_z^2}{2\sqrt{\pi}}\Gamma(1-2d)[2\sin(2\pi v)]^{0.5-2d} \cdot\left[P_{k-0.5}^{2d-0.5}(u) + (-1)^k P_{k-0.5}^{2d-0.5}(-u)\right], \qquad (4)$$

where $P_a^b(x)$ are the *associated Legendre functions of the first kind.* If $|u|=1$, $|\gamma(k)|$ is the absolute autocovariance of a fractionally differenced process given in [1, p. 523]. The lack of simple expression for the autocovariance function, particularly for $u \in (-1,1)$, makes the analysis of Gegenbauer process difficult. It could be simplified by applying a decorrelating transformation such as a wavelet transform.

In this paper, we deal with GARMA(p,d,u,q) process for which the Gegenbauer component is stationary, invertible and d>0, the ARMA component is causal and invertible, and $\Theta(z)$ and $\Phi(z)$ are coprime. We call this process a *generalized long-memory process.* We propose a Bayesian approach in estimating the parameters and predicting future values of this process by utilizing the approximate likelihood function of discrete wavelet packet coefficients. This provides an alternative method incorporating prior information about the parameters and utilizing the decorrelating property of wavelet transform. We illustrate our approach by an example applying simulated data.

2. DECORRELATING THE SIGNAL

The standard discrete wavelet transform (DWT) has been shown to be an effective within-and across-scale decorrelator of fractionally differenced process [5] and fractional Brownian motion [6]. However, for other processes including the generalized long-memory process, within-scale autocorrelations may not be negligible for some scales and for any length of the filter. To see this let $\{d_{jt} | j=1,2\ldots,J,\ t=0,\ldots,N2^{-j}-1\}$ be the nonboundary DWT coefficients of the random signal $\{Y_t\}_{t=0}^{N-1}$, where $N=2^J$ for some integer J. From [7, p.348], the autocovariance function could be written as

$$Cov(d_{jt}, d_{j't'}) = \int_{-1/2}^{1/2} e^{i^* 2\pi f(2^{j'}(t'+1)-2^j(t+1))} H_j(f) H_{j'}^*(f) S_Y(f) df\,, \qquad (5)$$

where $H_{j,L}(f)$ is the Fourier transform of the level j Daubechies wavelet filter $\{h_{j,l}\}$ and * denotes the complex conjugation operator. By setting $j=j'$ and $t'=t+s$ in (5) the within-scale covariance is given by

$$Cov(d_{jt}, d_{j(t+s)}) = \int_{-1/2}^{1/2} e^{i^* 2\pi f 2^j s} |H_j(f)|^2 S_Y(f) df\,. \qquad (6)$$

Given the wavelet filter $\{h_{\ell,1}\}$, the scaling filter $\{g_{\ell,1}\}$ is defined by $g_{\ell,1} = (-1)^{\ell+1} h_{L-\ell-1,1}$ for $\ell = 0,\ldots, L$-1. For j>1, wavelet and scaling filters $\{h_{\ell,j}\}$ and $\{g_{\ell,j}\}$ are of the same length $L_j = (2^j-1)(L-1)+1$ and denoted by $H_{j,L}(f)$ and $G_{j,L}(f)$. From [5, Lemma 4.1]

$$\left|H_{j,L}(f)\right|^2 \to \begin{cases} 2^j, & |f| \in (2^{-j-1}, 2^{-j}) \\ 2^{j-1}, & |f| = 2^{-j-1} \text{ or } 2^{-j} \\ 0, & \text{otherwise} \end{cases}$$

Hence for large L, (5) is approximately zero if $j \neq j'$. This is not true for within-scale correlations, which could be relatively large even for large values of L. From [8], Daubechies' wavelet filters converge monotonically to an ideal high-pass filter and Daubechies' scaling filters converge monotonically to an ideal low-pass filter as the filter length L increases without bound. Thus, for an appropriate choice of large L, from (6) we may write the autocovariance in the form

$$Cov(d_{j,t}, d_{j,t+s}) \approx 2^{j+1}\int_{2^{-(j+1)}}^{2^{-j}} \cos(2^{j+1}\pi sf) S_Y(f) df\,,$$

which is relatively large when $v \in [2^{-j-1}, 2^{-j}]$ and $0<d<0.5$. However, as v approaches zero the effect of singularity diminishes due to the decreasing length of the interval $[2^{-j-1}, 2^{-j}]$. Hence, for appropriate choice of filter length L, within-scale correlations of wavelet coefficients of a fractionally differenced process, where v=0, are approximately zero. This approximation is not valid for long-memory processes with pole not close to the zero frequency. Hence, statistical analysis of a generalized long-memory process utilizing approximate likelihood function with a diagonal autocovariance matrix may not be valid when the orthonormal discrete wavelet transform (DWT) is used. This problem may be solved by applying a transformation that generalizes the partitioning scheme of the DWT.

We approximately decorrelate a generalized long-memory process across and within scale by utilizing the discrete wavelet packet transform (DWPT). This has been proposed in [9] to decorrelate a process when DWT fails. Unlike DWT, which has only one particular filtering sequence, DWPT executes all possible low-pass and high-pass filtering combinations. From [7, p. 215], for $j=1,\ldots,J_0 (J_0 \leq J)$, we can write the DWPT coefficients $\{d_{jnt} | j=1,2\ldots,J, n=0,\ldots,2^j-1, t=0,\ldots,N2^{-j}-1\}$ in the form

$$d_{jnt} = \sum_{l=0}^{L_j-1} u_{jnl} Y_{2^j[t+1]-1-l \bmod M}\,, \quad t=0,\ldots,M_j-1,$$

where the jth level filter $\{u_{jnl}\}$ has length

$L_j = (2^j\text{-}1)(L\text{-}1)+1$ and computed from $\{h_{\ell,1}\}$ and $\{g_{\ell,1}\}$. The within-scale autocovariance function of nonboundary DWPT coefficients is given by

$$Cov(d_{jnt}, d_{jn(t+s)}) = \int_{-1/2}^{1/2} e^{i2\pi f 2^j s} |U_{jn}(f)|^2 S_Y(f) df ,$$

where the transfer function $U_{jn}(f)$ depends only on $G(\cdot)$ and $H(\cdot)$ such that the squared gain function $|U_{jn}(f)|^2$ is nominally band-pass over the frequency interval $I_{jn} = \left(\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}} \right]$. Although $S_Y(f)$ has a pole at some frequency between 0 and 0.5, $|U_{jn}(f)|^2$ can have its pass-band on any interval I_{jn}, $n=0,...,2^j$. This allows for decorrelation by choosing the wavelet basis functions $\{u_{jnl}\}$ appropriately for some index n. For this purpose we consider the algorithm proposed in [9]. It successively selects the DWPT coefficients at each level by performing a white noise test on wavelet packet subbands at each level.

3. ESTIMATION AND PREDICTION

In this section, we propose a Bayesian approach in estimating parameters and predicting future values of generalized long-memory process utilizing the independence assumption across and within scales on appropriately selected DWPT coefficients. This formulation simplifies the likelihood into a univariate density.

Let $D = (D'_{1n_1}, D'_{2n_2}, ..., D'_{Jn_J})'$ be the $(2^J\text{-}1)$x1 vector containing the nonboundary DWPT coefficients of $Y_t (t = 0, 1, \ldots, 2^J\text{-}1)$, where $n_1, n_2, ..., n_k$ are selected using the approach in [9]. The vector $D_{jn_k} = (D_{jn_k 1}, ..., D_{jn_k 2^{J-j}})$, has DWPT coefficients at scale j as its components. Consider the re-scaled approximate bandpass variance

$$\sigma^2 \sigma_j^2(\tau) = 2^{j+1} \int_{n2^{-(j+1)}}^{(n+1)2^{-(j+1)}} S_Y(f) df .$$

Assuming independence across and within scale, and Gaussian white noise innovations, the approximate likelihood function is given by $f(D|\Psi) =$

$$\prod_{j=1}^{J} (2\pi\sigma^2\sigma_j^2)^{-1/2} \cdot exp\left[-\tfrac{1}{2} \sum_{j=1}^{J} \sigma^{-2}\sigma_j^{-2} \sum_{t=0}^{N2^{-j}-1} D_{jn_k t}^2 \right].$$

This approximation does not depend on the length of the signal, but the length of the wavelet filter, which is under the control of the analyst.

Let $\Psi = (\Phi, \Theta, d, u, \sigma^2)$ and $\Psi' = (\Phi, \Theta, d, u)$ denote the vectors of parameters, where $\Phi = (\phi_1, ..., \phi_p)$ and $\Theta = (\theta_1, ..., \theta_q)$. For fixed p and q, we consider the prior structure $\pi(\Psi) = \pi(\sigma^2)\ \pi(d,u)\ \pi(\Phi,\Theta)$. We adopt a uniform prior for Φ and Θ over C_p and C_q, respectively; an Inverse-Gamma $(0.5v_0, 0.5\delta_0)$ for σ^2 and a Beta(α,β) for d and $2v$, where $\alpha, \beta \geq 1$. Note that the choice $\alpha = \beta = 1$ gives the standard uniform prior on [0,1]. For a GARMA(p,d,u,q) process, the given prior for the long-memory parameter d does not impose stationarity apriori, and so applies to both explosive and nonexplosive cases. The posterior density of Ψ is given by $\pi(\Psi|D) \propto f(D|\Psi)\ \pi(\Psi)$. We make inferences about the parameters $\Psi = (\Phi, \Theta, d, \sigma^2)$ by sampling from $\pi(\Psi|D)$ using Markov chain Monte Carlo (MCMC) algorithm. To do this, we propose an MCMC sampler that utilizes the Metropolis-Hastings algorithm.

First we block the parameters Ψ' into two groups $A_1 = (d,u)$ and $A_2 = (\Phi,\Theta)$. The sampling algorithm can then be summarized as follows:

1) Initialize Ψ by taking $\hat{\Psi}$ = argmax log($f(D|\Psi)$.

2) Sample σ^2 routinely from

$$Inverse-Gamma\left(\frac{v_0 + J}{2}, \frac{\delta_0 + \sum_{i=1}^{J} \sigma_j^{-2} \sum_{t=0}^{N2^{-j}-1} D_{jn_k t}^2}{2} \right).$$

3) Propose a value of the kth block conditioned on the previous value of the kth block and σ^2, and the other values of the other blocks A_{-k},

$A'_k \sim q_k(A_k^{(r-1)}, \cdot | Y_t, A_{-k}, \sigma^2)$.

We define the proposal distribution by

$$q_k(A_k^{(r-1)}, A | Y_t) \propto exp\left\{ -0.5(A - A^{(r-1)})'(V)^{-1}(A - A^{(r-1)}) \right\},$$

where V is the Fisher information matrix of $\hat{\Psi}$.

4) Compute the acceptance probability

$$\alpha\left(A_k^{r-1}, A'_k | Y_t, A_{-k}, \sigma^2\right) = \min\left\{ 1, \frac{\pi\left(A'_k | Y_t, A_{-k}, \sigma^2\right) q_k(A'_k, A_k^{(r-1)} | Y_t, A_{-k}, \sigma^2)}{\pi\left(A_k^{(r-1)} | Y_t, A_{-k}, \sigma^2\right) q_k(A_k^{(r-1)}, A'_k | Y_t, A_{-k}, \sigma^2)} \right\}.$$

5) Sample U from Uniform(0,1). Update the kth block as follows:

$$A_k^{(r)} = \begin{cases} A'_k & if \quad u \leq \alpha\left(A_k^{r-1}, A'_k | Y_t, A_{-k}, \sigma^2\right) \\ A_k^{(r-1)} & otherwise \end{cases}.$$

Repeat steps 2-5 for r= 1,2,…,N+M and for k = 1,2. Estimate the parameters from the $\{\Psi^{(N+1)}, \Psi^{(N+2)}, ..., \Psi^{(N+M)}\}$ beyond a burn-in of N iterations.

For prediction purposes we use the original observations and the parameter values obtained through the preceding sampling algorithm. This provides a simpler approach with no need to reconstruct the signal.

Let $Y_{N+F} = (y_0,\ldots,y_{N-1}, y_N,\ldots,y_{N+F-1})$, where Y_N is the vector representing the observed values and $Y_N^F = (y_N,\ldots,y_{N+F-1})$ is the vector of F future unknown realizations. A GARMA(p,d,u,q) process can be written as $\Phi(B)Y_t = \Theta(B)X_t$, where $X_t = \left(1-2uB+B^2\right)^{-d} Z_t$ is the associated GARMA(0,d,u,0) process. Hence, we have

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + X_t - \sum_{i=1}^{q} \theta_i X_{t-i} . \qquad (7)$$

Now, the joint density of Y_N^F is

$$f(Y_N^F \mid Y_N, \Psi) = \prod_{j=0}^{F-1} f_j(y_{N+j} \mid y_{N+j-1}, \Psi) , \qquad (8)$$

where each f_j(j=0,2,…,F-1) is univariate normal. Using (7), the mean and variance of f_j are given by

$$E(Y_{N+j} \mid Y_{N+j-1}, \Psi) = \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \theta_i x_{t-i}$$

and $\mathrm{Var}(Y_{N+j} \mid Y_{N+j-1}, \Psi) = \mathrm{Var}(X_{N+j})$, respectively. $\mathrm{Var}(X_{N+j})$ is the variance of the associated GARMA(0,d,u,0) process given by (4) when k=0.

Inferences about Y_N^F can be made on the basis of $f(Y_N^F \mid Y_N, \Psi)$. The density of future unknown data is given by

$$f(Y_N^F \mid Y_N) = \int f(Y_N^F \mid Y_N, \Psi)\pi(\Psi \mid Y_N) d\Psi , \qquad (9)$$

where $\pi(\Psi \mid Y_N)$ is the prior density of the vector of parameters Ψ. Using the output Ψ_r (r = 1,…,M) from the sampler, (9) can be approximated by Monte Carlo integration: $\hat{f}(Y_N^F \mid Y_N) = \frac{1}{M}\sum_{r=1}^{M} f(Y_N^F \mid Y_N, \Psi_r)$, where $f(Y_N^F \mid Y_N, \Psi)$ is defined by (8).

4. EXAMPLE

We simulate a GARMA(p,d,u,q) process as follows.

Step 1. Generate the autocovariances R(0), R(1),…,R(n) of GARMA(0,d,u,0) by evaluating (3) numerically for k = 0,1,…,n.

Step 2. Simulate X_t, a stationary Gaussian GARMA(0,d,u,0) time series of length n, using the method proposed in [10], which is used for any stationary process with given autocovariances.

Step 3. Simulate an ARMA(p,q) process, using X_t as innovations. The result is the simulated GARMA(p,d,u,q) process.

Using this procedure, we constructed a data set with n=512 for a GARMA(1,0.25,0.6,1) process, where θ=0.3, ϕ= 0.2, and σ^2=1. The posterior means and the posterior standard deviations (in parentheses) for the parameters are $\hat{\theta}$=0.283 (0.036), $\hat{\phi}$=0.192 (0.032), $\hat{d}$=0.271 (0.027), $\hat{u}$=0.612 (0.029), and $\hat{\sigma}$=1.01 (0.005). Convergence of the sampler was monitored by estimating the potential scale reduction factor based on 1000 iterations and 10 replications. The kernel density estimates for the parameters were unimodal and symmetric about the posterior means.

5. CONCLUSION

Discrete wavelet packet transform simplifies the variance-covariance matrix of a generalized long-memory process by approximately decorrelating wavelet coefficients within and across scales. This approximation does not depend on the length of the signal, but the length of the wavelet filter, which is under the control of the analyst. It allows for a computationally efficient sampling from the posterior density of p+q+2 parameters.

6. REFERENCES

[1] P. Brockwell and R. Davis, *Time Series: Theory and Methods*, Springer-Verlag, 1991

[2] H.L. Gray, N.F. Zhang, and W. Woodward, "On Generalized Fractional Processes," *Journal of Time Series Analysis,* Vol. 10, pp. 233-257, 1989.

[3] R. Ramachandran and V.N. Bhethanabotla, "Generalized Autoregressive Moving Average Modeling of the Bellcore Data," *Procedings of the 25th Annual IEEE Conference on Local Computer Networks,* pp. 654-661, 2000

[4] C. Chung, "Estimating a generalized long-memory process," *Journal of Econometrics*, Vol. 73, pp. 237-259, 1996.

[5] Y. Fan, On the Approximate Decorrelation Property of Discrete Wavelet Transform for Fractionally Differenced Process, *IEEE Transactions on Information Theory*, Vol. 49 pp. 516-521, 2003.

[6] A. Tewfik and M. Kim, "Correlation Structure of the Discrete Wavelet Coefficients of Fraction Brownian Motion", *IEEE Transactions on Information Theory*, Vol. 38, No. 2, pp.904-909, 1992

[7] D.B. Percival and A.T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, 2000.

[8] M.J. Lai, "On the Digital Filter Associated with Daubechies' Wavelets," *IEEE Transactions on Signal Processing*, Vol. 43, pp. 2203-2205, 1995.

[9] D.B. Percival, S. Sardy and A.C. Davison, "Wavestrapping time series: adaptive wavelet-based bootstrapping," *Nonlinear and Nonstationary Signal Processing,* Cambridge University Press, 442-471, 2000

[10] R.B. Davies, S. Harte, "Tests for Hurst Effect," *Biometrika*, Vol. 74(1), pp. 95-101, 1987.