



Bayesian wavelet analysis of autoregressive fractionally integrated moving-average processes

Kyungduk Ko^a, Marina Vannucci^{b,*}

^a*Department of Mathematics, Boise State University, Boise, ID 83725-1555, USA*

^b*Department of Statistics, Texas A&M University, College Station, TX, 77843-3143, USA*

Received 30 October 2003; accepted 27 January 2005

Available online 22 March 2005

Abstract

Long memory processes are widely used in many scientific fields, such as economics, physics and engineering. In this paper we describe a wavelet-based Bayesian estimation procedure to estimate the parameters of a general Gaussian ARFIMA (p, d, q) , autoregressive fractionally integrated moving average model with unknown autoregressive and moving average parameters. We employ the decorrelation properties of the wavelet transforms to write a relatively simple Bayes model in the wavelet domain. We use an efficient recursive algorithm to compute the variances of the wavelet coefficients. These depend on the unknown characteristic parameters of the model. We use Markov chain Monte Carlo methods and direct numerical integration for inference. Performances are evaluated on simulated data and on real data sets.

© 2005 Elsevier B.V. All rights reserved.

MSC: 62C10; 62F15; 42C40; 65T60; 60G18

Keywords: ARFIMA processes; Bayesian inference; Wavelets

1. Introduction

Long memory processes are widely used in many fields and applications range from financial data to data from biology and hydrology, to mention a few. Fractional ARIMA (p, d, q) , first introduced by Hosking (1981) and Granger and Joyeux (1980), are well

* Corresponding author. Tel.: +1 979 845 0805; fax: +1 979 845 3144.

E-mail address: mvannucci@stat.tamu.edu (M. Vannucci).

known examples of long memory processes. Classical methods for modelling and inference for these processes involve the calculation of the exact likelihood and its maximization with respect to the parameters, see for example [Beran \(1994\)](#). The complexity of such inferential procedures is mainly due to the dense long memory covariance structure that makes the exact likelihood of the data difficult to handle. Exact maximum likelihood estimators, as well as the calculations for posterior distributions in suitable form for inference, are therefore usually impractical for large data sets. Some improvements were achieved by [Pai and Ravishanker \(1996, 1998\)](#) and by [Koop et al. \(1997\)](#), who proposed Bayesian approaches based on the algorithm of [Sowell \(1992a\)](#) to compute the exact likelihood and used importance sampling and Markov chain Monte Carlo methods for a posteriori inference.

Wavelets have proven to be a powerful tool for the analysis and synthesis of data from long memory processes. Wavelets are strongly connected to such processes in that the same shapes repeat at different orders of magnitude, [Wornell \(1996\)](#). The ability of the wavelets to simultaneously localise a process in time and scale domain results in representing many dense matrices in a sparse form. When transforming measurements from a long memory process, wavelet coefficients are approximately uncorrelated, in contrast with the dense long memory covariance structure of the data, see [Tewfik and Kim \(1992\)](#), among others.

Here we propose a Bayesian approach to the wavelet analysis of fractional ARIMA (p, d, q) processes. We first transform the data into wavelet coefficients and use an efficient recursive algorithm from [Vannucci and Corradi \(1999a\)](#) to compute the exact variances and covariances of the wavelet coefficients. We exploit the de-correlation properties of the wavelets to write a simple model in the wavelet domain. The exact variances of the wavelet coefficients are embedded in the model and depend on the parameters of the ARFIMA process. We carry out posterior inference by Markov chain Monte Carlo methods and, in the simpler case of integrated processes, i.e. ARFIMA $(0, d, 0)$, by direct numerical integration. We perform extensive simulation studies and also test our method on benchmark data sets. In all examples we provide comparisons with other methods.

Other work that uses wavelets in the analysis of discrete-time long memory processes can be found in [McCoy and Walden \(1996\)](#), who proposed an approximate wavelet coefficients-based maximum likelihood iterative estimation procedure for ARFIMA $(0, d, 0)$ only. [Jensen \(1999, 2000\)](#) constructed wavelet-based MLE estimators of the long memory parameter for ARFIMA (p, d, q) . His simulations studies showed better overall performances of the wavelet-based estimates with respect to the approximate MLE. Our approach is novel in many respects. We combine Bayesian methods with wavelet-based modelling of long memory processes, for both ARFIMA $(0, d, 0)$ and ARFIMA (p, d, q) models. For ARFIMA (p, d, q) , unlike previous approaches, we also produce estimates of the unknown autoregressive and moving average parameters. Our approach is fairly general and may be applicable to other classes of long memory processes.

The paper is organized as follows: In Section 2, we introduce the necessary mathematical concepts on ARFIMA models and on wavelet methods. In Section 3, we describe the Bayesian model and the posterior inference. We report results from simulations in Section 4 and from applications to GNP data and to the well known Nile river data set in Section 5. Some concluding remarks are given in Section 6.

2. Preliminaries

2.1. Fractional ARIMA processes

A long memory process is characterised by a slow decay of the autocovariance function of the type $\gamma(\tau) \sim C\tau^{-\alpha}$ with $C > 0$ a constant depending on the process, $0 < \alpha < 1$ and τ large. Fractional ARIMA(p, d, q), first introduced by Hosking (1981) and Granger and Joyeux (1980), are well known examples of long memory processes. Let us first define the fractional difference operator $(1 - B)^d$, with $d \in (-.5, .5)$, as the Binomial series expansion:

$$(1 - B)^d \equiv \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j B^j \tag{1}$$

with B the backshift operator and with square summable coefficients:

$$\binom{d}{j} (-1)^j = \frac{\Gamma(d + 1)(-1)^j}{\Gamma(d - j + 1)\Gamma(j + 1)} = \frac{\Gamma(-d + j)}{\Gamma(-d)\Gamma(j + 1)}. \tag{2}$$

Here $\Gamma(\cdot)$ denote the Gamma function. A fractional ARIMA(p, d, q) process with p and q nonnegative integers is defined as the stationary solution of the equation

$$\Phi(B)(1 - B)^d(x_t - \mu) = \Theta(B)\varepsilon_t \tag{3}$$

with polynomials $\Phi(B) = 1 + \phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p$ and $\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ and where μ is the process finite mean and ε_t is a Gaussian white noise process with zero mean and variance σ_ε^2 . Here we assume $\mu = 0$ without loss of generality. Differencing d times the process produces an ARMA(p, q) model. Fractional ARIMA processes are stationary and invertible. They exhibit positive dependency between distant observations for $0 < d < .5$ (long memory), negative dependency for $-.5 < d < 0$ (intermediate memory) and reduce to short memory ARMA(p, q) processes for $d = 0$. A special class of processes is the fractionally integrated obtained for $p = 0$ and $q = 0$, also called fractionally differenced white noise, or $I(d)$, in that differencing d times produces a white noise process.

2.2. Covariance structure of ARFIMA models

If the roots of $\Phi(B)$ are outside the unit circle, $\Phi(B)$ can be written as

$$\Phi(x) = \prod_{j=1}^p (1 - \rho_j x), \tag{4}$$

where $|\rho_j| < 1$ for $j = 1, 2, \dots, p$. If, in addition, the roots of $\Phi(x)$ are simple, Sowell (1992a) has showed that the autocovariance function of a Gaussian fractional ARIMA (p, d, q) process is

$$\gamma(\tau) = \sigma_\varepsilon^2 \sum_{l=-q}^q \sum_{j=1}^p \psi(l)\zeta_j C(d, p - \tau + l, \rho_j), \tag{5}$$

where

$$\psi(l) = \sum_{\tau=\max[0,l]}^{\min[q,q-l]} \theta_\tau \theta_{\tau-l}, \tag{6}$$

$$\zeta_j = \left[\rho_j \prod_{i=1}^p (1 - \rho_i \rho_j) \prod_{m \neq j} (\rho_j - \rho_m) \right]^{-1} \tag{7}$$

and where

$$C(d, h, \rho) = \frac{\Gamma(1 - 2d)\Gamma(d + h)}{\Gamma(1 - d + h)\Gamma(1 - d)\Gamma(d)} [\rho^{2p} F(d + h, 1; 1 - d + h; \rho) + F(d - h, 1; 1 - d - h; \rho) - 1]. \tag{8}$$

Here $F(a, 1; c; \rho)$ is the hypergeometric function, satisfying the recursive relationship $F(a, 1; c; \rho) = (c - 1)/\rho(a - 1)[F(a - 1, 1; c - 1; \rho) - 1]$. Formulas (5)–(8) will be used to simulate data from ARFIMA models in Section 4.1.

The autocovariance function of an $I(d)$ process simplifies into the form

$$\gamma(\tau) = \sigma_\varepsilon^2 \frac{\Gamma(1 - 2d)\Gamma(d + \tau)}{\Gamma(1 - d + \tau)\Gamma(1 - d)\Gamma(d)}. \tag{9}$$

2.3. Discrete wavelet transforms

Suppose we observe a time series as a realization of a random process and let us indicate the data vector as $X = (x_1, \dots, x_n)$ with $n = 2^J$ and J a positive integer denoting the scale of the data. Using 2^J points, with J integer, is not a real restriction and methods exist to overcome the limitation allowing wavelet transforms to be applied to any length of data.

The standard wavelet transform, as proposed by Mallat (1989), is essentially a filtering operation that operates on octave bands of frequency. It starts by applying to the data the filters:

$$x_{J-1,k} = \sum_m h_{m-2k} x_{J,m}, \quad d_{J-1,k} = \sum_m g_{m-2k} x_{J,m}. \tag{10}$$

Eqs. (10) are convolutions followed by a downsampling operation and provide a coarser approximation of the data (the vector with components $x_{J-1,k}$, the so called scaling coefficients) and a set of details (the vector of wavelet coefficients $d_{J-1,k}$) at the scale $J - 1$. Scaling and wavelet coefficient vectors have length 2^{J-1} . Coefficients h_l define a low-pass filter and vary according to the wavelet family. Here we are concerned with Daubechies (1992) wavelets. These wavelets have compact support, implying filters with a finite number of nonzero coefficients h_l . Filter coefficients g_l define a high-pass filter and are commonly defined as $g_l = (-1)^l h_{1-l}$.

The discrete wavelet transform (DWT) proceeds by keeping the wavelet coefficients $d_{J-1,k}$ and applying Eqs. (10) to the scaling coefficients $x_{J-1,k}$. The procedure is repeated until a desired scale is reached, say $J - r$. Although it operates via recursive applications

of filters, for practical purposes the DWT is often represented in matrix form as $Z = WX$, with W an orthogonal matrix of the form

$$W = [W_{J-1}^T, W_{J-2}^T, \dots, W_{J-r}^T, V_{J-r}^T]^T \tag{11}$$

that decomposes the data into sets of coefficients

$$Z = [d_{J-1}^T, d_{J-2}^T, \dots, d_{J-r}^T, x_{J-r}^T]^T \tag{12}$$

with $d_{J-j} = W_{J-j}X$ and $x_{J-r} = V_{J-r}X$. At scale $\tau_j = 2^{j-1}$, or level j , the $n/2^j$ wavelet coefficients are associated with changes in averages of the data on a scale τ_j at a set of location times. This means that each wavelet coefficient at that level tells us how much a weighted average of the data changes from a particular time period of effective length τ_j to the next one. Scaling coefficients of the wavelet transform are instead weighted averages of the data with bandwidth τ_{J+1} over a particular time period of effective length τ_{J+1} . The wavelet transform is a cumulative measure of the variations in the data over regions proportional to the wavelet scales, with coefficients at coarser and coarser level describing features at lower frequency ranges and larger time periods.

An inverse transformation exists to reconstruct a set of data from its wavelet decomposition.

2.4. Variances and covariances of wavelet coefficients

Exact variances and covariances of the wavelet and scaling coefficients can be computed as follows. If the data were generated from a random process with autocovariance function $\gamma(\tau)$, we can write the variance-covariance matrix of the vector X as

$$\Sigma_X(i, j) = [\gamma(|i - j|)]. \tag{13}$$

Using the matrix notation of the DWT it is therefore straightforward to compute the variance-covariance matrix Σ_Z of the wavelet and scaling coefficients vector Z as

$$\Sigma_Z = W \Sigma_X W^T. \tag{14}$$

Although expressions for W are available, see for example [McCoy and Walden \(1996\)](#), computation of Σ_Z through the matrix product of Eq. (14) is not efficient. A faster and computationally less expensive way is to use the recursive filters of the DWT. This is the idea behind the work of [Vannucci and Corradi \(1999a\)](#), who have proposed a recursive way of computing covariances of coefficients by using the recursive filters of the DWT, see Proposition 1 on page 974 of their paper. At a generic step the filters are used to get

$$\text{cov}[d_{j-1,k}, d_{j'-1,k'}] = \sum_m \sum_n g_{m-2k} g_{n-2k'} \text{cov}[x_{j,m}, x_{j',n}] \tag{15}$$

and similarly for $\text{cov}[x_{j-1,k}, x_{j'-1,k'}]$ and $\text{cov}[x_{j-1,k}, d_{j'-1,k'}]$ with j, j' and k, k' integers. At the first step these filters are applied to the elements of the matrix Σ_X and give within-scale variances and covariances of coefficients that belong to the coarser scale $J - 1$ and across-scales variances and covariances of coefficients that belong to scales $J - 2$ and $J - 1$.

The filters are then applied to the variances and covariances of scaling coefficients at level $J - 1$, and so on until scale $J - r$ to get all elements of Σ_Z .

The Vannucci and Corradi algorithm has an interesting link to the two-dimensional discrete wavelet transform (DWT2) that makes computations simple. In the context of this paper, the matrix Σ_Z in Eq. (14) can be computed by first applying the DWT2 to the matrix Σ_X . The diagonal blocks of the resulting matrix will provide the within-scale variances and covariances at the different levels. One can then apply the one-dimensional DWT to the rows of the non diagonal blocks to obtain the across-scale variances and covariances. The boundary conditions of the DWT will affect some of the variances. In the case of a stationary process wavelet coefficients have constant variance at a given scale and it is a good practice to replace the affected variances with those unaffected (Jensen, personal communication).

3. Bayesian modelling

Let now $X = (x_1, \dots, x_n)$ be a vector of observations from a Gaussian ARFIMA(p, d, q) process. We model wavelet coefficients, rather than the original data. Long memory data have, in fact, a dense covariance structure that makes the exact likelihood of the data difficult to handle, see for example Beran (1994). On the contrary, simpler models can be used for wavelet coefficients.

We explore decorrelation properties of the wavelets using plots as those displayed in Figs. 1 and 2. Plots (a) in the two figures show the covariance matrices Σ_X in (14) for $p = 1, q = 1$ with $d = .2, \phi = .5, \theta = -.8$ and $d = .4, \phi = .1, \theta = -.8$, respectively. Plots (b) show the corresponding matrices Σ_Z . Plot (b) of Fig. 1 has been obtained with Haar wavelets, while plot (b) of Fig. 2 uses Daubechies wavelets with seven vanishing moments. These are the two wavelet families we will later use in the simulation study. The horizontal bars below each plot indicate the color scales. Plots (b) show the whitening properties of the wavelets. Matrices of both figures are in fact close to diagonal. They also show that wavelets with a higher number of vanishing moments have greater decorrelation power.

For further insights, we can approximate Σ_Z as diagonal and use the inverse transform in Eq. (14) to compute a reconstructed $\tilde{\Sigma}_X$ from the diagonal Σ_Z . Such reconstructions are shown in plots (c) of Figs. 1 and 2. We notice that a diagonal structure in the wavelet domain does not imply uncorrelated observations. On the contrary, due to the good compression properties of the wavelets, most of the original covariance structures is retained. Summary measures of the difference between the original Σ_X matrices and their reconstructions $\tilde{\Sigma}_X$ can be computed as mean squares and mean absolute deviations:

$$D_1 = \frac{1}{m} \sum_{i,j} [\Sigma_X(i, j) - \tilde{\Sigma}_X(i, j)]^2 \quad \text{and} \quad D_2 = \frac{1}{m} \sum_{i,j} |\Sigma_X(i, j) - \tilde{\Sigma}_X(i, j)|. \quad (16)$$

These give $D_1 = .0038$ and $D_2 = 5.3 \times 10^{-5}$ for the case shown in Fig. 1 and $D_1 = .0231$ and $D_2 = 6.6 \times 10^{-4}$ for Fig. 2. Similar results hold for different values of the d, θ, ϕ parameters and $p = 0, 1$ and/or $q = 0, 1$. Overall, reconstructions from diagonal structures in the wavelet domain are reasonably close to the original ones, the only exceptions being

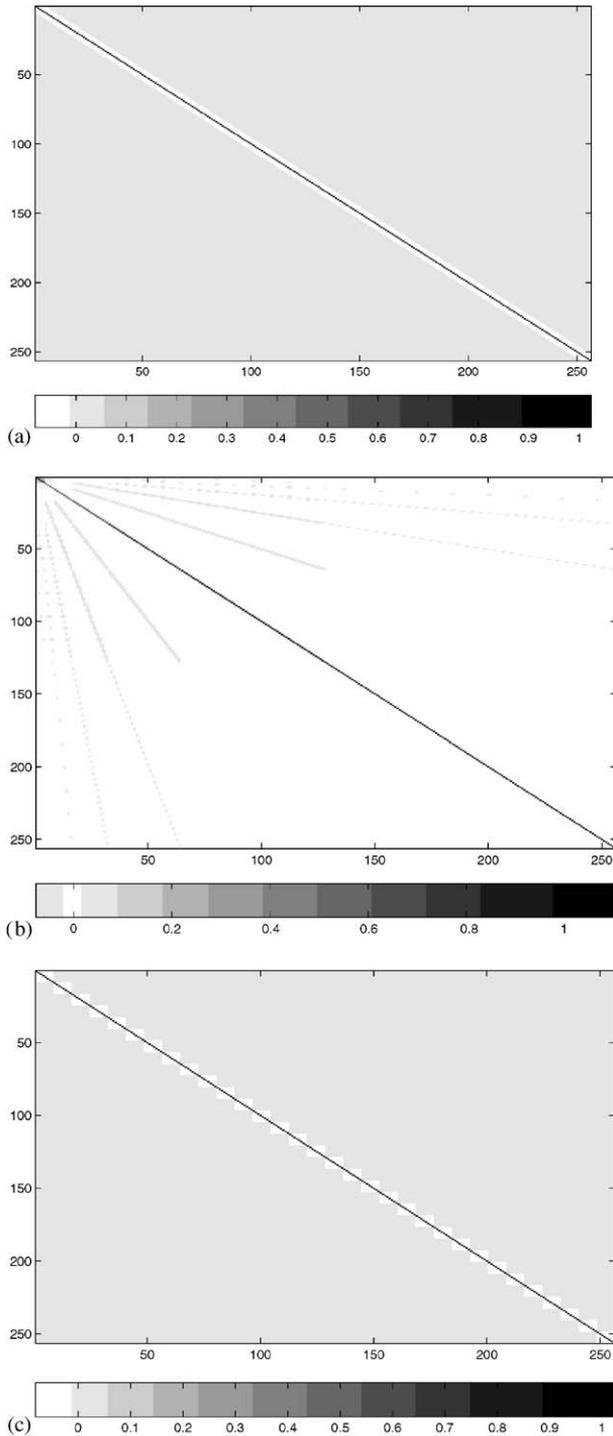


Fig. 1. ARFIMA(1,2,1), $\phi = .5$, $\theta = -.8$. Haar wavelets.

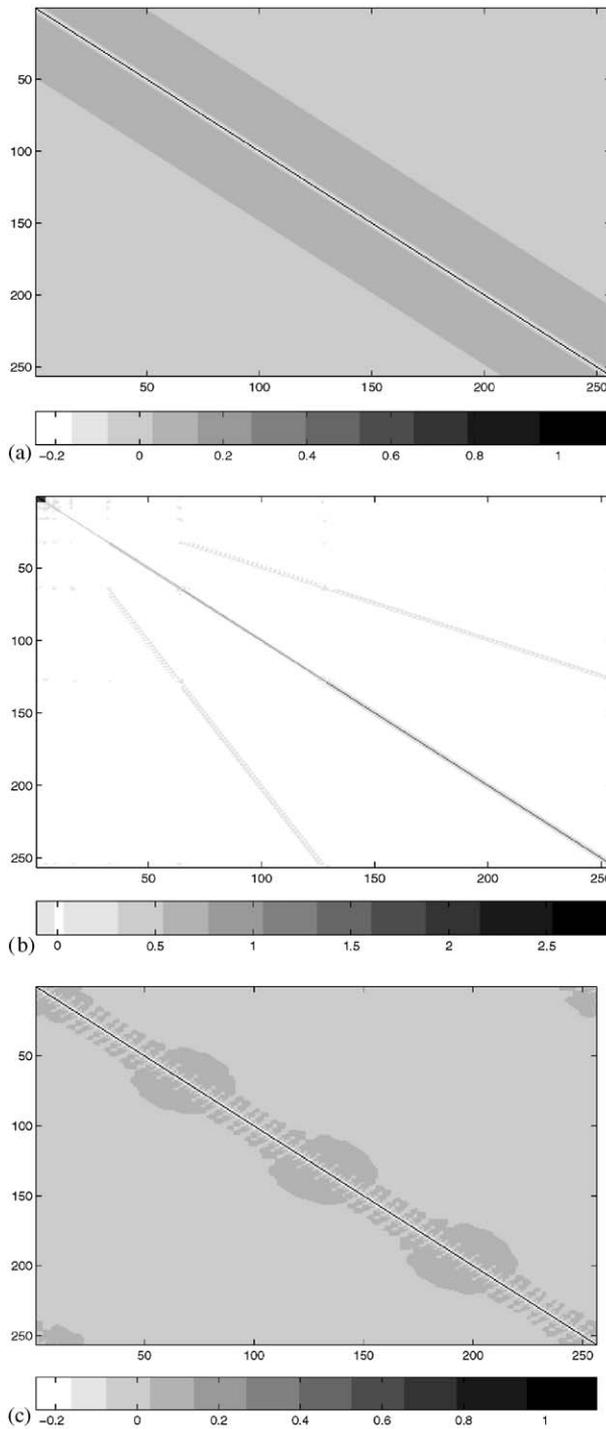


Fig. 2. ARFIMA(1,4,1), $\phi = .1$, $\theta = -.8$. MP(7) wavelets.

extreme cases where d is close to .5 and the MA and AR parameters are close to the unitary root.

Decorrelation properties of the wavelets for long memory processes are well documented in the literature. [Tewfik and Kim \(1992\)](#) and [Dijkerman and Mazumdar \(1994\)](#) proved that the correlation between wavelet coefficients decreases exponentially fast across scales and hyperbolically fast along time. [Jensen \(1999, 2000\)](#) provides evidence that these rates of decay allow the DWTs to do a credible job at decorrelating the highly autocorrelated long-memory processes. These results, of course, strictly depend on the long-memory structure and do not apply to other processes. [Percival et al. \(2000\)](#) suggest to look at wavelet packets as a way to decorrelate processes for which the standard DWTs fail, such as for short-memory processes. See also [Gabbanini et al. \(2004\)](#).

3.1. Model in the wavelet domain

The DWT is a linear and orthogonal transformation and wavelet coefficients therefore inherit the distribution of the data, specifically they are zero mean Gaussian. Let $\Psi = (\phi, \theta, d, \sigma_\varepsilon^2)$ and $\Psi_0 = (\phi, \theta, d)$ where $\phi = (\phi_1, \phi_2, \dots, \phi_p)$ and $\theta = (\theta_1, \theta_2, \dots, \theta_q)$. We can write

$$[Z_i | \Psi] \sim N(0, \sigma_{Z_i}^2(\Psi)) \tag{17}$$

independently for $i=1, \dots, n$. The variances $\sigma_{Z_i}^2$ of the wavelet coefficients can be computed using the Vannucci and Corradi algorithm as previously described. Notice that, since we are interested in the variances only, computations simplify considerably. We simply have to apply the DWT2 to the matrix Σ_X and the diagonal elements of the resulting matrix will be the variances $\sigma_{Z_i}^2$. Moreover, because of the form of the autocovariance function (5) of an ARFIMA model we have

$$\sigma_{Z_i}^2(\Psi) = \sigma_\varepsilon^2 \sigma_{Z_i}^2(\Psi_0), \tag{18}$$

where $\sigma_{Z_i}^2(\Psi_0)$ depends only on the long memory parameter and the moving average and autoregressive parameters. Eq. (17) becomes then

$$[Z_i | \Psi] \sim N(0, \sigma_\varepsilon^2 \sigma_{Z_i}^2(\Psi_0)) \tag{19}$$

and inference on the parameters can now be carried out.

We first need to specify priors for the unknowns, i.e. the long memory parameter, d , the autoregressive coefficients, $\phi_1, \phi_2, \dots, \phi_p$, the moving average coefficients, $\theta_1, \theta_2, \dots, \theta_q$, and the nuisance scale parameter, σ_ε^2 . We assume that $\pi(\Psi) = \pi(\phi)\pi(\theta)\pi(d)\pi(\sigma_\varepsilon)$. A natural prior for σ_ε^2 is an inverse gamma $\mathcal{IG}(\alpha/2, \beta/2)$

$$\pi(\sigma_\varepsilon^2) = \frac{(\alpha/2)^{\beta/2}}{\Gamma(\beta/2)} (\sigma_\varepsilon^2)^{-(\beta+2)/2} \exp(-\alpha/(2\sigma_\varepsilon^2)). \tag{20}$$

As for the long memory parameter, a sensible choice is a mixture of two Beta distributions on $-1/2 < d < 0$ and $0 < d < 1/2$ with a point mass at $d=0$. However, in all our simulations we found that uniform priors in $(-1/2, 1/2)$ give good results. As for the priors of ϕ s and

θ s we use uniform distributions in $(-1, 1)$ to satisfy the causality and invertibility of the ARMA process.

3.2. Posterior analysis

The posterior distribution of Ψ may be written as

$$\pi(\Psi|Z) \propto (\sigma_\varepsilon^2)^{-(n/2+1)} \left(\prod_{i=1}^n \sigma_{z_i}^2(\Psi_0) \right)^{-1/2} \exp \left[-\frac{\sum_{i=1}^n (z_i^2 / \sigma_{z_i}^2(\Psi_0))}{2\sigma_\varepsilon^2} \right] \pi(\Psi). \tag{21}$$

We treat σ_ε^2 as a nuisance parameter and obtain the marginal posterior distribution of Ψ_0 by integrating out σ_ε^2 in Eq. (21) as

$$\pi(\Psi_0|Z) \propto \left(\frac{1}{\prod_{i=1}^n \sigma_{z_i}^2(\Psi_0)} \right)^{1/2} \left[\sum_{i=1}^n \left(\frac{z_i^2}{\sigma_{z_i}^2(\Psi_0)} \right) \right]^{-n/2} \pi(\Psi_0). \tag{22}$$

Moreover, by suitable integrations we obtain the following inverse Gamma distribution associated with σ_ε^2 :

$$\pi(\sigma_\varepsilon^2 | \Psi_0, Z) \sim IG \left(\frac{n}{2}, \frac{2}{\sum_{i=1}^n (z_i^2 / \sigma_{z_i}^2(\Psi_0))} \right) \tag{23}$$

and the marginal posterior distribution of σ_ε^2 , given the data, is therefore

$$\pi(\sigma_\varepsilon^2 | Z) = \int_{\Psi_0} \pi(\sigma_\varepsilon^2 | \Psi_0, Z) \pi(\Psi_0 | Z) d\Psi_0. \tag{24}$$

Samples can now be drawn from Eq. (22) using the Metropolis algorithm, [Metropolis et al. \(1950\)](#). If of interest, inference on σ_ε^2 can be carried out based on the MCMC samples on Ψ_0 using a Rao–Blackwellization procedure, i.e. by sampling a σ_ε^2 from Eq. (23) for each sample on Ψ_0 .

We sample from the posterior distribution (22) using a Gaussian proposal centered at the maximum likelihood estimates of the parameters and with covariance matrix given by the observed Fisher information. A transformation method can be used in order to take into account the restrictions on d and on θ and ϕ . In the case of $p = 0, 1$ and/or $q = 0, 1$, that is $-1/2 < d < 1/2$ and $-1 < \theta, \phi < 1$, we can use a transformation of the type

$$d^* = \frac{1 - 1 + e^d}{2 \ 1 + e^d}, \quad \theta^* = \frac{-1 + e^\theta}{1 + e^\theta} \quad \text{and} \quad \phi^* = \frac{-1 + e^\phi}{1 + e^\phi}, \tag{25}$$

where $-\infty < \theta, \phi, d < +\infty$ are the values sampled from the Gaussian proposal. The transformed posterior density is then

$$\pi(\Psi^* | Z) = \pi(\Psi | Z) |J|, \tag{26}$$

where $|J| = |\partial\Psi/\partial\Psi^*|$ is given by the expression

$$\frac{\exp(d) \exp(\phi_1 + \phi_2 + \dots + \phi_p) \exp(\theta_1 + \theta_2 + \dots + \theta_q)}{(1 + \exp(d))^2 (1 + \exp(\phi_1))^2 \dots (1 + \exp(\phi_p))^2 (1 + \exp(\theta_1))^2 \dots (1 + \exp(\theta_q))^2}$$

Similar transformations can be used for higher order models. Alternatively, sampling with rejection can be used as a simpler but less efficient computational method.

3.3. Inference for $I(d)$ processes

The simpler ARFIMA(0, d , 0), or $I(d)$ process, is characterized by two parameters only, the long memory parameter of interest, d , and the variance parameter, σ_ε^2 . We use the same priors on d and σ_ε^2 as previously described and treat σ_ε^2 as a nuisance parameter. The posterior inference simplifies considerably and we can avoid the use of MCMC methods. The nuisance variance parameter can in fact be integrated out and inference on the long memory parameter can be carried out by numerical integration methods. We have

$$p(d|Z) = \int p(Z|d, \sigma_\varepsilon^2) \pi(\sigma_\varepsilon^2) \pi(d) d\sigma_\varepsilon^2 = \pi(d) p(Z|d) \tag{27}$$

with $p(Z|d)$ a Student t -distribution

$$p(Z|d) \sim \frac{1}{\left[\prod \sigma_{Z_i}^2(d)\right]^{1/2}} \left[\alpha + \sum Z_i^2 / \sigma_{Z_i}^2(d)\right]^{-(\beta+n)/2} \tag{28}$$

with degrees of freedom β , O’Hagan (1994). Inference on d can now be carried out by univariate numerical integration. The marginal (27) can be plotted for a grid G of values in the range of d and a summary point estimate can be computed as $\hat{d} = \sum_{d_i \in G} d_i p(d_i|Z)$ with normalized $\{p(d_i|Z), d_i \in G\}$. Numerical credibility intervals can be also computed. Some preliminary work on a similar Bayesian wavelet approach to $I(d)$ processes, that uses MCMC methods rather than numerical integration, is briefly described in Vannucci and Corradi (1999b).

4. Simulation studies

There are a number of ways to generate a time series that exhibits long memory properties. A computationally simple one was proposed by McLeod and Hipel (1978) and involves the Cholesky decomposition of the correlation matrix $R_X(i, j) = [\rho(|i - j|)]$. Given $R_X = MM'$ with $M = [m_{i,j}]$ a lower triangular matrix, if $\varepsilon_t, t = 1, \dots, n$ is a Gaussian white noise series with zero mean and unit variance, then the series

$$X_t = \gamma(0)^{1/2} \sum_{i=1}^t m_{t,i} \varepsilon_i \tag{29}$$

will have autocorrelation $\rho(\tau)$.

Table 1
 ARFIMA(1, d , 0): estimates of d and ϕ from our wavelet Bayes method with MP(7) wavelets, MLE and the Geweke and Porter-Hudak (1983) method, respectively

(ϕ, d)		$\phi = .5$	$d = .2$	$\phi = -.8$	$d = .4$
2^7	KV	.5474(.0702)	.1502(.0274)	-.6669(.0601)	.3801(.0281)
	MLE	.6466	.0015	-.6373	.3290
	GPH		.0550		.4993
2^9	KV	.5599(.0458)	.1585(.0247)	-.7245(.0332)	.3729(.0256)
	MLE	.6015	.1252	-.7198	.3671
	GPH		.2529		.5018

Numbers in parentheses are std's.

4.1. Fractional ARIMA processes

We used the McLeod and Hipel method to simulate data with $\rho(\tau)$ computed from (5) with $\sigma_\epsilon^2 = 1$. Parameters included in the study were the autoregressive and moving average parameters and the long memory parameter. We report here results on ARFIMA(1, d , 0), ARFIMA(0, d , 1) and ARFIMA(1, d , 1). In order to check robustness of the estimates we simulated data for different values of the parameters and of the sample size n . We used uniform priors, as previously described. In all simulations we used Daubechies (1992) minimum phase wavelets MP(7). See next paragraph for sensitivity to the choice of the wavelet family.

We adopted similar settings to those of Pai and Ravishanker (1998). For each combination of the parameters under investigation we based our wavelet-based Bayes estimates on ten MCMC chains. We used the maximum likelihood estimates as initial values for one MCMC sampler and then perturbed these estimates to obtain overdispersed values to be used to initialize the ten independent parallel MCMC chains. All chains ran for 1000 iterations after a burnin time of 1000. Estimates were computed as posterior means, together with posterior std's, obtained from the pooled MCMC sample. Tables 1–3 summarize the numerical results of the study. For comparison we also report the values of the MLE and of the classical estimator of Geweke and Porter-Hudak (1983), based on a regression on periodogram ordinates. In the case of ARFIMA(1, d , 0) we notice that our wavelet-based Bayes estimates are always better than MLE and GPH estimates, for both small and large sample sizes. As expected, MLE estimates improve considerably for large sample sizes. For ARFIMA(0, d , 1) Bayes estimates are almost always better, with the noticeable exception of the case $\theta = .5$, for which our method does not improve on MLE. For ARFIMA(1, d , 1) Bayes estimates for the long memory parameter improve on the other two methods, while MLE do a better job at estimating the parameter ϕ , although Bayes estimates are very close to the MLE values. Overall, these results seem to indicate a better overall performance of our method with respect to the other two methods. See also the simulation study on $I(d)$ processes for more comparisons with the Geweke and Porter-Hudak method.

Table 2

ARFIMA(0, d , 1): estimates of d and θ from our wavelet Bayes method with MP(7) wavelets, MLE and the Geweke and Porter-Hudak (1983) method, respectively

(d, θ)		$d = .2$	$\theta = .5$	$d = .4$	$\theta = -.8$
2^7	KV	.2059(.0659)	.3484(.2051)	.4104(.0170)	-.6968(.0816)
	MLE	.0699	.4671	.1367	-.4029
	GPH	.4237		.2913	
2^9	KV	.1948(.0253)	.4029(.0797)	.3192(.0147)	-.7155(.0313)
	MLE	.1559	.5547	.3024	-.7189
	GPH	.2316		.3580	

Numbers in parentheses are std's.

Table 3

ARFIMA(1, d , 1): estimates of d , ϕ and θ from our wavelet Bayes method with MP(7) wavelets, MLE and the Geweke and Porter-Hudak (1983) method, respectively

(ϕ, d, θ)	$\phi = .1$	$d = .4$	$\theta = .5$	$\phi = -.1$	$d = .4$	$\theta = -.5$	
2^7	KV	.1703(.1579)	.2617(.0042)	.6650(.2324)	-.2568(.1709)	.3755(.0226)	-.2995(.1558)
	MLE	.0847	.2588	.7020	-.2147	.2447	-.2344
	GPH		.4777			.3145	
2^9	KV	.0459(.1346)	.3915(.0075)	.5577(.2136)	-.0529(.0982)	.2834(.0261)	-.4183(.0921)
	MLE	.0698	.3859	.5646	-.0386	.2783	-.4374
	GPH		.5134			.1962	

Numbers in parentheses are std's.

4.2. Integrated processes

We performed a second simulation study on $I(d)$ processes using the McLeod and Hipel method to simulate data with $\rho(\tau)$ computed from (9) with $\sigma_\epsilon^2 = 1$. Computations are considerably faster for this class of models, in that the posterior inference does not require the use of MCMC methods. Also, we have a single parameter, d , to include in the study. We therefore assessed performances of our estimation procedure by computing Monte Carlo mean squared error (MSE) and bias of the estimates on 1000 simulated time series. In order to check robustness of the estimates to parameter values and sample size, we repeated the procedure for different values of the long memory parameter d and of the sample size n . We set uniform priors on σ_ϵ^2 and d . We also repeated simulations with different wavelet families. We used Daubechies (1992, p. 195) minimum phase wavelets MP(1), i.e. Haar wavelets, and MP(7). Wavelets with a higher number of vanishing moments ensure wavelet coefficients approximately uncorrelated. On the other hand the support of the wavelets increases with the regularity and boundary effects may arise in the DWT, so that a trade-off is often necessary. Daubechies (1992, p. 198) wavelets “least asymmetric”, LA(N), with N coefficients gave very similar results (not reported) to MP(N) wavelets.

Table 4

ARFIMA(0, d , 0): bias and MSE of estimates of d for our wavelet Bayes method with MP(1) and MP(7) wavelets, the McCoy and Walden (1996) method with MP(7) wavelets and the Geweke and Porter-Hudak (1983) method, respectively

d		$d = .05$		$d = .2$		$d = .4$		$d = .45$	
		BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
2^7	KV-MP1	.0341	.0023	-.0396	.0045	-.0781	.0129	-.0927	.0127
	KV-MP7	.0306	.0030	-.0411	.0049	-.0822	.0099	-.1026	.0129
	MW-MP7	.0235	.0021	-.0701	.0098	-.1024	.0141	-.1186	.0171
	GPH	-.0100	.0790	.0135	.0850	.0187	.0773	.0227	.0723
2^8	KV-MP1	.0173	.0019	-.0203	.0037	-.0569	.0053	-.0523	.0049
	KV-MP7	.0127	.0028	-.0225	.0030	-.0519	.0043	-.0671	.0058
	MW-MP7	-.0243	.0016	-.0488	.0049	-.0755	.0076	-.0861	.0091
	GPH	.0004	.0465	.0120	.0419	.0150	.0430	.0187	.0427
2^9	KV-MP1	.0085	.0006	-.0144	.0016	-.0282	.0021	-.0357	.0029
	KV-MP7	.0076	.0015	-.0121	.0014	-.0313	.0019	-.0423	.0025
	MW-MP7	-.0239	.0014	-.0353	.0025	-.0586	.0044	-.0662	.0052
	GPH	.0018	.0276	.0011	.0274	.0083	.0298	.0111	.0297
2^{10}	KV-MP1	.0061	.0008	-.0091	.0008	-.0184	.0009	-.0277	.0012
	KV-MP7	.0038	.0006	-.0065	.0006	-.0158	.0007	-.0261	.0011
	MW-MP7	.0223	.0011	-.0274	.0013	-.0463	.0027	-.0549	.0035
	GPH	.0034	.0170	.0016	.0175	.0084	.0208	.0084	.0177

For $I(d)$ processes we had available for comparison the approximate maximum likelihood wavelet estimation procedure of McCoy and Walden (1996) (we used MP(7) wavelets). We also computed the Geweke and Porter-Hudak estimates. Table 4 summarizes the numerical results of the simulation study. Wavelets with higher degrees of regularity produce slightly better results than Haar wavelets, for larger sample sizes. There is no evident sensitivity of the estimates to the different values of d , showing a good robustness to the different values of the long memory parameter. In comparing our Bayes method with that of Geweke and Porter-Hudak one may notice that the bias of the Bayes estimates is often worse. However, this is compensated by the large improvement in the MSE, that is the Bayes estimates have lower variance than the GPH estimates. Also, notice how, for a fixed value of d , the Bayes estimates tend to improve when n increases, unlike the GPH estimates. When comparing our Bayes method with the McCoy and Walden approximate MLE method with MP(7) wavelets, we can notice that both procedures produce very good results, the Bayes estimates being almost always better than the approximate MLEs. Bias and MSE are in fact lower in almost all cases. The bias is almost always negative, for both Bayes and approximate MLE estimators, suggesting that both procedures tend to underestimate d .

In order to investigate the effect of the approximately uncorrelated property used to specify the matrix Σ_W we looked into results on the estimation of the long memory parameter

Table 5
ARFIMA(0, d , 0). Estimation results under different correlation structures

		Diagonal	Block-diagonal	Full
2^7	$d = 0.05$	0.0411(−0.0089)	0.0471(−0.0029)	0.0460(−0.0040)
	$d = 0.20$	0.1971(−0.0029)	0.1456(−0.0544)	0.1532(−0.0468)
	$d = 0.45$	0.4410(−0.0090)	0.4419(−0.0081)	0.4282(−0.0218)
2^9	$d = 0.05$	0.0487(−0.0013)	0.0551(0.0051)	0.0542(0.0042)
	$d = 0.20$	0.2286(0.0286)	0.2105(0.0105)	0.2113 (0.0113)
	$d = 0.45$	0.4659(0.0159)	0.4525(0.0025)	0.4451(−0.0049)

Numbers in parentheses denote biases.

d under different correlation structures. Table 5 reports estimates and biases for different values of the long memory parameter d and different sample sizes n . Daubechies wavelets MP(7) were used. In the table the term “Diagonal” means that only the diagonal elements of $\Sigma_W = W \Sigma_X W^T$ are used in the estimation, that is the wavelet coefficients are assumed independent, while “Block-Diagonal” denotes estimates obtained by specifying Σ_W as block diagonal with blocks corresponding to within-scale variances and covariances. The term “Full” indicates that the exact form of Σ_W is used, with no approximation. Results confirm that the approximation to uncorrelated wavelet coefficients is reasonable with data from a long memory process.

5. Examples

We finally illustrate our method on some real data sets, specifically on GNP data and on the well known Nile river data set.

5.1. The GNP data

We examined post-war quarterly data on the logarithms of seasonally adjusted US real GNP from 1947 to 1991. First differences are shown in Fig. 3. This series has been analysed by several authors. A slightly shorter time series was fitted by Sowell (1992b) with both ARMA and ARFIMA models. The AIC criterion indicated the ARFIMA(3, d , 2) as the best model. Sowell also reported MLE estimates for the long memory and the autoregressive and moving average parameters of all models. As for Bayesian approaches, Pai and Ravishanker (1996) showed evidence for the ARFIMA(0, d , 0) model without mean as the best fit, while Koop et al. (1997) reported ARFIMA(1, d , 0) as best model. For some models there appears to be some discrepancy between the parameter estimates they report and those of Sowell.

We fitted ARFIMA(p , d , q) models with $p, q = 0, 1$. We used a circulant filter by padding the series with replicas and truncating the wavelet transform. Estimates and standard deviations for the parameters of the different models are reported in Table 6. In the case of ARFIMA(1, d , 0), ARFIMA(0, d , 1) and ARFIMA(1, d , 1), estimates are based on ten independent parallel MCMC chains with 1000 iterations each and burnin times of 1000

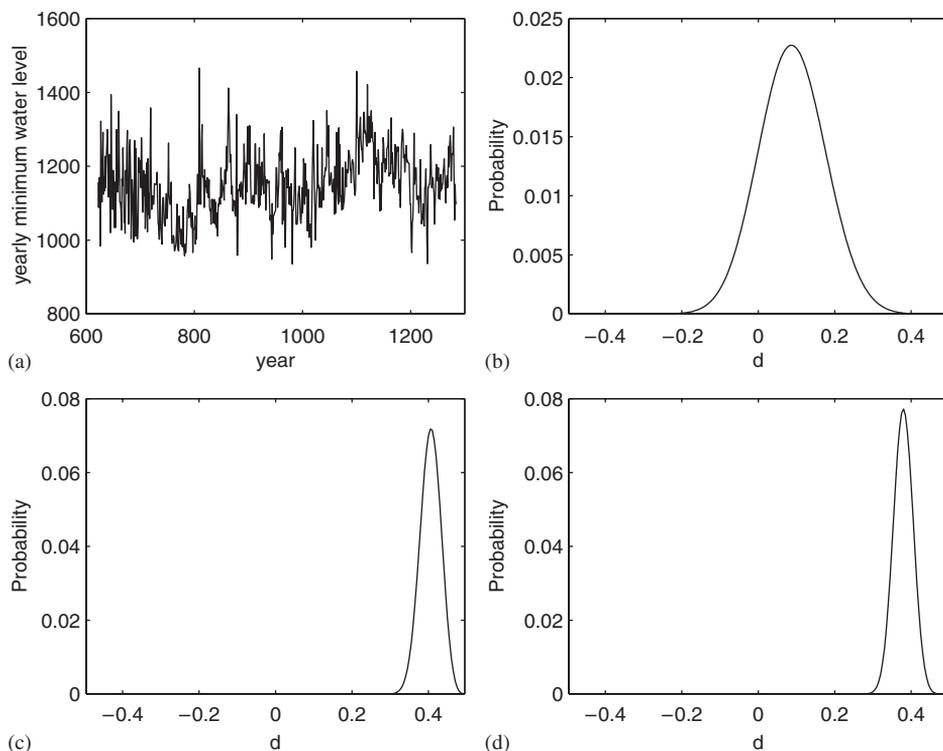


Fig. 3. First differences of GNP data, in plot (a), and kernel estimates of the posterior density of d for (b) ARFIMA(1, d , 0), (c) ARFIMA(0, d , 1) and (d) ARFIMA(1, d , 1).

Table 6
Wavelet-based Bayes estimates for US GNP data

Model	d	ϕ	θ
(0, d , 0)	.2528(.0024)		
(1, d , 0)	-.4499(.0027)	-.6910(.0376)	
(0, d , 1)	.1925(.0303)		.1705(.0521)
(1, d , 1)	-.3628(.0124)	-.6991(.0203)	.0456(.0361)

iterations. With respect to previous works, these estimates appear to be in better agreement with those reported by Sowell (1992b). For example, Koop et al. (1997) report estimates of the long memory parameter of $d = -.29, .23, -.16$, while MLE estimates are $d = -.45, .16, -.38$, for ARFIMA(1, d , 0), ARFIMA(0, d , 1) and ARFIMA(1, d , 1), respectively. Fig. 3 shows kernel density estimates of the posterior distributions of the long memory parameter for the three different ARFIMA models. The estimate of the long memory parameter for the ARFIMA(0, d , 0) model was instead based on the marginal (27)

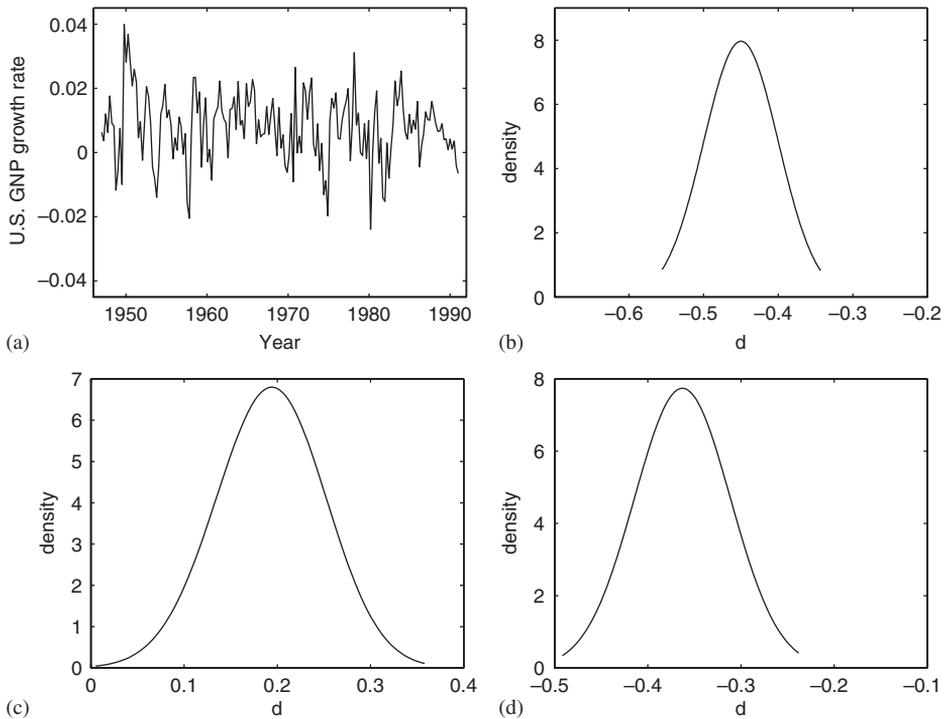


Fig. 4. Yearly minimum water level of the Nile River at the Roda Gauge (622–1284 AD), in plot (a), and marginal posterior distributions of d for (b) first 100, (c) subsequent 500, (d) all observations.

computed on a grid of 500 values in the range of the parameter. This value appears to be slightly lower than the estimate reported by Sowell ($d = .29$) and by Koop et al. (1997) and Pai and Ravishanker (1996) ($d = .32$).

5.2. The Nile river data set

Probably the most well known example of a time series that exhibits long memory behaviour is the Nile river minimum water levels data set, Toussoun (1925). Whitcher et al. (2002) provide a recent analysis of this data set, although they concentrate more on the problem of detecting a change in the variance of the time series. Data consist of 663 yearly values, from 622 AD to 1284 AD, see Fig. 4. Beran (1994, p. 117–118), showed that this time series can be modelled as fractionally differenced $I(d)$ and obtained an estimate of $d = .4$ using the Whittle’s approximate maximum likelihood approach. However he suspected a change of the long memory parameter in the time series. Looking at the plot of the data he noticed that the first part of the series seemed to fluctuate much more independently than the subsequent measurements. He partitioned the first 600 measurements into two subsets of 100 and 500 observations, respectively. His maximum likelihood estimates of the long memory parameter were indeed different, $d = .04$ and $d = .38$ respectively, see

Beran (1994, Section 10.3). Whitcher et al. (2002) used the approximate relationship between a wavelet estimate of the variance of the data and the autocovariance function of the long memory process to compute estimates of d based on a simple regression model. They got $d = .38$, $.42$ and $d = -.07$ for the whole time series, the last 563 observations and the first 100 observations, respectively.

For comparison with previous results, we have used our wavelet Bayes method for $I(d)$ processes to compute estimates of the parameter d for different subsets of the data. We used the first 100, the subsequent 500 and all observations. A circulant filter was used by padding the series with replicas and then truncating the wavelet transform. Fig. 4 shows the marginal posterior distributions of d for the different sub-series. Estimates of d were $\hat{d} = .0891$ with a 95% credibility interval of $(-.083, .257)$ (first 100 observations), $\hat{d} = .4052$ with a 95% credibility interval of $(.347, .453)$ (subsequent 500 observations) and $\hat{d} = .3793$ with a 95% credibility interval of $(.327, .427)$ (all observations). Our results well agree with Beran's results.

6. Commentary

We have proposed a wavelet-based Bayesian approach to the analysis of long memory processes, specifically Gaussian ARFIMA(p, d, q), autoregressive fractionally integrated moving average models with unknown autoregressive and moving average parameters. We have used the decorrelation nature of the wavelet transform and shown how the variances of the wavelet coefficients depend on the characteristic parameters of the process generating the data. We have carried out Bayesian posterior inference on the parameters by MCMC methods and direct numerical integration. The proposed strategy is quite general and may be applied to different classes of processes. Simulation studies and real examples have demonstrated the usefulness of wavelet methods and Bayesian methodologies in the analysis of data from long memory processes.

We have provided evidence for the whitening properties of the wavelets suggesting that the approximation to uncorrelated coefficients can be reasonable. This was also confirmed by the overall good performance of our method in the simulation studies. In future work we plan on investigating dependence modeling that will allow us to incorporate some of the nondiagonal structure we see in plots (b) of Figs. 1 and 2.

For inference we have mainly focused on the long memory parameter and the autoregressive and moving average parameters. The inferential procedure, however, can be easily generalized to include other model parameters, such as the noise variance and the process mean, either by embedding the Metropolis into a Gibbs sampler that employs the full conditional distributions of the additional parameters or by using a Rao–Blackwellization procedure. More interesting, inferential procedures could be generalized to detect structural changes in the long memory process, such as changes in the long memory parameter d . This is currently under investigation.

Finally, although our main focus has been on the estimation of the parameters of the model, forecasting is another important feature of estimation procedures for time series. On this regard, we notice that prediction of future values may not be intuitive if done in the wavelet domain. However, it can be easily carried out in the time domain in the same

manner as typically done in previous Bayesian approaches to ARFIMA processes, i.e. via approximating the predictive distribution by Monte Carlo integration using the output from the MCMC sampler, see, for example, [Pai and Ravishanker \(1996, 1998\)](#).

Acknowledgements

MV is partially supported by National Science Foundation, CAREER award number DMS-0093208.

References

- Beran, J., 1994. *Statistics for Long-Memory Processes*. Chapman & Hall, New York.
- Daubechies, I., 1992. *Ten Lectures on Wavelets*, vol. 61. SIAM, CBMS-NSF Conference Series. SIAM, Philadelphia, PA.
- Dijkerman, R.W., Mazumdar, R.R., 1994. On the correlation structure of the wavelet coefficients of fractional Brownian motion. *IEEE Trans. Inform. Theory* 40 (5), 1609–1612.
- Gabbanini, F., Vannucci, M., Bartoli, G., Moro, A., 2004. Wavelet packet methods for the analysis of variance of time series with application to crack widths on the Brunelleschi dome. *J. Comput. Graphical Statist.* 13(3), 639–658.
- Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. *J. Time Series Anal.* 4, 221–238.
- Granger, C.W., Joyeux, R., 1980. An introduction to long memory time series models and fractional differencing. *J. Time Series Anal.* 1, 15–29.
- Hosking, J.R.M., 1981. Fractional differencing. *Biometrika* 68, 165–176.
- Jensen, M.J., 1999. Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. *J. Forecasting* 18, 17–32.
- Jensen, M., 2000. An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *J. Econom. Dyn. Control* 24 (3), 361–386.
- Koop, G., Ley, E., Osiewalski, J., Steel, M.F.J., 1997. Bayesian analysis of long memory and persistence using ARFIMA models. *J. Econometrics* 76, 149–169.
- Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (7), 674–693.
- McCoy, E.J., Walden, A.T., 1996. Wavelet analysis and synthesis of stationary long-memory processes. *J. Comput. Graphical Statist.* 5 (1), 26–56.
- McLeod, A.I., Hipel, K.W., 1978. Preservation of the rescaled adjusted range parts, 1, 2 and 3. *Water Resour. Res.* 14, 491–512.
- Metropolis, N., Rosenblut, A.W., Rosenblut, M.N., Teller, A.H., Teller, E., 1950. Equations of state calculations by fast computing machines. *J. Chem. Phys.* 21, 1087–1092.
- O'Hagan, A., 1994. *Kendall's Advanced Theory of Statistics*, vol. 2B. University Press, Cambridge.
- Pai, J.S., Ravishanker, N., 1996. Bayesian modelling of ARFIMA processes by Markov chain Monte Carlo methods. *J. Forecasting* 15, 63–82.
- Pai, J.S., Ravishanker, N., 1998. Bayesian analysis of autoregressive fractionally integrated moving-average processes. *J. Time Ser. Anal.* 19 (1), 99–112.
- Percival, D.P., Sardy, S., Davison, A.C., 2000. Wavestrapping time series: adaptive wavelet-based bootstrapping. In: Fitzgerald, W.J. et al. (Eds.), *Nonlinear and Nonstationary Signal Processing*. Cambridge University Press, Cambridge.
- Sowell, F., 1992a. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *J. Econometrics* 53, 165–188.
- Sowell, F., 1992b. Modeling long-run behavior with the fractional arima model. *J. Monetary Econom.* 29, 277–302.

- Tewfik, A.H., Kim, M., 1992. Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Trans. Inform. Theory* 38 (2), 904–909.
- Toussoun, O., 1925. Mémoire sur l'histoire du Nil. *Mémoires a l'Institut d'Égypte* 18, 366–404.
- Vannucci, M., Corradi, F., 1999a. Covariance structure of wavelet coefficients: theory and models in a Bayesian perspective. *J. Roy. Statist. Soc. Ser. B* 61 (4), 971–986.
- Vannucci, M., Corradi, F., 1999b. Modeling dependence in the wavelet domain. In: Vidakovic, B. et al., Müller, P. (Eds.), *Bayesian Inference in Wavelet-based Models*. Springer, Berlin, pp. 173–186.
- Whitcher, B., Byers, S.D., Guttorp, P., Percival, D.B., 2002. Testing for homogeneity of variance in time series: long memory wavelets and the Nile river. *Water Resour. Res.* 38 (5), 65–88.
- Wornell, G.W., 1996. *Signal Processing with Fractals: A Wavelet-based Approach*. Prentice-Hall, Englewood Cliffs, NJ.