BAYESIAN METHODS FOR CHANGE-POINT DETECTION IN LONG-RANGE DEPENDENT PROCESSES

By Bonnie K. Ray and Ruey S. Tsay

IBM T.J.Watson Research Center and University of Chicago

First Version received October 2000

Abstract. We describe a Bayesian method for detecting structural changes in a long-range dependent process. In particular, we focus on changes in the long-range dependence parameter, d, and changes in the process level, μ. Markov chain Monte Carlo (MCMC) methods are used to estimate the posterior probability and size of a change at time t, along with other model parameters. A time-dependent Kalman filter approach is used to evaluate the likelihood of the fractionally integrated ARMA model characterizing the long-range dependence. The method allows for multiple change points and can be extended to the long-memory stochastic volatility case. We apply the method to three examples, to investigate a change in persistence of the yearly Nile River minima, to investigate structural changes in the series of durations between intraday trades of IBM stock on the New York Stock Exchange, and to detect structural breaks in daily stock returns for the Coca Cola Company during the 1990s.

Keywords. Gibbs sampler; Kalman filter; long memory; structural change.

1. INTRODUCTION

Stationary processes exhibiting long-term, persistent fluctuations have been observed in many areas, including hydrology, meteorology, economics, finance and telecommunications. A commonly used model for such processes is the autoregressive fractionally integrated moving-average (ARFIMA) model, introduced by Granger and Joyeux (1980) and Hosking (1981). Accurate estimation of an ARFIMA model often requires a large sample of data taken over a long period of time which, in turn, increases the chance of structural breaks in the process. A structural break may be caused by a change in the physical mechanism that generates the data or by a change in the way that observations are collected over time.

Due to the slowly decaying correlation structure of an ARFIMA process, test statistics commonly used for assessing the stability of the model over time may encounter some difficulties. Kuan and Hsu (1998) show that, although the least-squares estimator of a change-point in mean is consistent for detecting a change if

Corresponding Author: Bonnie K. Ray, IBM Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598 Phone (914) 945 2952 Fax: (914) 954 3434; http://www.research.ibm.com/people/b/bonnie

0143-9782/02/06 687–705 JOURNAL OF TIME SERIES ANALYSIS Vol. 23, No. 6

one exists, it is likely to suggest a spurious change if there is no such change. Wright (1998) shows that the usual sup-Wald and CUSUM tests for structural stability with unknown potential break date falsely indicate a break almost surely when applied to a polynomial regression model with long-range dependent errors. Even structural change tests designed specifically for ARFIMA data with known potential break date may still have large size distortions in small samples (Hidalgo and Robinson, 1996). Beran and Terrin (1996) derive a test that can be used to detect a single change in one of the parameters of an ARFIMA model, but their method is not applicable for detecting a change in mean. Furthermore, most commonly used tests are designed to detect a single change point, although for a long series, several structural breaks may be present.

McCulloch and Tsay (1993) give a Bayesian method for estimating random level and variance shifts in an autoregressive (AR) time series. Their method is based on estimating the probability and size of a shift at each time point, together with other model parameters, using Markov-chain Monte-Carlo (MCMC) techniques. Although only changes in level and variance are discussed in their paper, the method is broadly applicable. For example, it can be used to test for a change in other model parameters, such as an AR parameter. In this paper, we extend the method of McCulloch and Tsay (1993) to detect changes in the parameters of an ARFIMA process. We concentrate on detecting changes in the long memory parameter, d, and in the mean, μ, although as stated above, the method can be used to detect changes in any model parameter. The method is likelihood-based and can detect multiple change points.

The remainder of the paper is organized as follows. Section 2 outlines the method and discusses its implementation. Section 3 presents the results of a small simulation study. Section 4 applies the method to the series of yearly minima of the Nile river to investigate the stability of the long-range dependent parameter, and to two financial series, the duration between trades of IBM stock on the New York Stock Exchange and the absolute returns on Coca Cola stock, to investigate the presence of level shifts and/or changes in persistence. Section 5 discusses the relation of the method presented here to that of other authors. Section 6 concludes.

2. DESCRIPTION OF METHOD

Let $\{y_t\}$ be a long-range dependent process with mean μ. We model $\{y_t\}$ as an ARFIMA(p, d, q) process

$$y_t = \mu + \frac{(1 - \theta_1 B - \ldots - \theta_q B^q)}{(1 - \phi_1 B - \ldots - \phi_p B^p)}(1 - B)^{-d} a_t \tag{1}$$

where a_t is a Gaussian white noise process with variance σ_a^2 and B denotes the backshift operator. The d parameter is a real value governing the amount of persistence in the process. The process is stationary and invertible if all roots of $(1 - \phi_1 z - \cdots - \phi_p z^p)$ and $(1 - \theta_1 B - \cdots - \theta_q B^q)$ lie outside the unit circle and

$-0.5 < d < 0.5$. We assume that the AR and moving-average (MA) polynomials have no common factors. The process is *long-range dependent* when $0 < d < 0.5$. An ARFIMA process has an infinite-order MA representation of the form

$$y_t = \mu + a_t + \sum_{i=1}^{\infty} \psi_i(d, \Phi, \Theta) a_{t-i} \tag{2}$$

where $\psi_i(d, \Phi, \Theta)$ signifies the dependence of the ith MA coefficient on the d and ARMA parameters. See Beran (1994) for additional discussion of ARFIMA models. In this paper, we focus only on stationary long-range dependent models.

2.1. *Random persistence-shift ARFIMA model*

Assume the amount of persistence in the process changes over time as $d_t, 0 < d_t < 0.5$, where the size and time of the shifts in d are unknown. The process may still be expressed in terms of an infinite-order MA representation of the form

$$y_t = \mu + a_t + \sum_{i=1}^{\infty} \psi_i(d_t, \Phi, \Theta) a_{t-i} \tag{3}$$

where the sequence of ψ_i is now different for each different value of d.

Given that we do not know *a priori* the time period or size of a shift in d, we model the process allowing for shifts in d of random size at random time periods. In particular, the process $\{y_t\}$ is said to follow a random persistence-shift (RPS) ARFIMA model if d is allowed to change randomly over time as

$$d_t = d_0 + \sum_{j=1}^{t} \delta_j \beta_j = d_{t-1} + \delta_t \beta_t,$$

where the δ_t are independent and identically distributed Bernoulli random variables such that $P(\delta_t = 1) = \epsilon$, and $\{\beta_t\}$ is a sequence of random observations from a known distribution. In this framework, the number of shifts in d is governed by the probability ϵ of the Bernoulli trials. Wang *et al.* (2001) give some examples of processes for which the self-similar behaviour, as characterized by d, might change as the phenomenon evolves. For instance, they find that vertical ocean shear measurements tend to exhibit locally long-range dependent behaviour that changes as a function of ocean depth. Whereas Wang *et al.* (2001) analyse such processes using a semi-parametric wavelet-based approach applied to series blocks, the RPS-ARFIMA model gives a way of formally quantifying both the probability and size of the change at a particular time.

We estimate the parameters of the RPS model in the Bayesian framework, using a Gibbs sampler approach. This entails the derivation of conditional posterior distributions for each of the process parameters. For convenience, we assume that the model of interest contains no ARMA components, i.e., we work with the RPS fractional noise model, $\mathrm{I}(d_t)$. Incorporation of ARMA components is discussed at the end of this section.

Here are the prior distributions used:

- $\frac{v\lambda}{\sigma_a^2} \sim \chi^2(v)$
- $\epsilon \sim \mathrm{Beta}(\gamma_1, \gamma_2)$
- $\mu \sim \mathrm{N}(\mu_0, \sigma_0^2)$
- d_0 and β_j are uniformly distributed over the set of real values such that $0 < d_t < 0.5$ for all t.

Here, v, λ, γ_1, γ_2, μ_0, and σ_0 are specified hyperparameters. The distributions for σ_a^2, ϵ and μ are chosen because they are standard conjugate priors. In the absence of any prior information about the d_t, we assume uniform distributions for d_0 and $\beta_j, j = 1, \ldots, t$ to ensure that each d_t satisfies the stationary, long-memory constraint. Choice of hyperparameters is discussed in the next section.

Using standard Bayesian techniques, we obtain the following conditional posterior distributions for a series of length n:

- $p(\sigma_a^2 | Y, \{\delta_t\}, \{\beta_t\}, d_0, \mu, \epsilon) \sim$ inverted χ^2; i.e., $[v\lambda + s^2]/\sigma_a^2 \sim \chi^2(v+n)$, where $s^2 = \sum_{t=1}^n a_t^2$ and $Y = (y_1, \ldots, y_n)$
- $p(\mu | Y, \{\delta_t\}, \{\beta_t\}, d_0, \sigma_a^2, \epsilon) \sim N(\mu^*, \sigma_*^2)$, where

$$\mu_* = \frac{\sigma_0^2 \sum_{t=1}^n z_t c_{t-1} + \mu_0 \sigma_a^2}{G},$$

$$\sigma_*^2 = \frac{\sigma_a^2 \sigma_0^2}{G}$$

and

$$G = \sigma_0^2 \sum_{t=1}^n c_{t-1}^2 + \sigma_a^2$$

where

$$c_j = 1 - \sum_{i=1}^j \psi_i c_{j-i},$$

and

$$z_t = y_t - \sum_{j=1}^{t-1} \psi_j a_{t-j}.$$

- $p(\epsilon | Y, \mu, \{\delta_t\}, \{\beta_t\}, d_0, \sigma_a^2) \sim \mathrm{Beta}(\gamma_1 + k, \gamma_2 + n - k)$, where k denotes the number of δ_j that are equal to 1.
- The conditional distribution of δ_j is given by

$$p(\delta_j = 1 | Y, \delta_{(-j)}, \{\beta_t\}, d_0, \mu, \epsilon, \sigma_a^2) = \frac{\epsilon P(Y|\delta_j = 1)}{\epsilon P(Y|\delta_j = 1) + (1-\epsilon) P(Y|\delta_j = 0)},$$

where $P(Y|\cdot)$ denotes the likelihood of Y given all parameters and $\{\delta_{(-j)}\}$ denotes the $\{\delta_t\}$ process excluding δ_j.

Both d_0 and β_j have non-standard conditional posteriors. We use a 'griddy' Gibbs procedure to sample values of these parameters. Specifically, to draw d_0, we evaluate $P(Y|d_0, \sigma_a^2, \{\delta_t\}, \{\beta_t\}, \mu, \epsilon)$ over a grid of d_0 values uniformly spaced over a subset of the region $(0.0, 0.5)$ such that $0 < d_t < 0.5$ for all t. The value of d_0 is selected from a multinomial distribution with multinomial weights based on the likelihood values $P(Y|d_0, \sigma_a^2, \{\delta_t\}, \{\beta_t\}, \mu, \epsilon)$. When $\delta_j = 1$, values of β_j are selected similarly, although care is required to find the appropriate region of possible β_j values such that $0 < d_t < 0.5$ for all t. When $\delta_j = 0$, β_j is drawn from a uniform distribution on the appropriate region of possible β_j values.

The likelihood $P(Y|\cdot)$ is evaluated using a time-dependent Kalman filter approach. We use a truncated MA approximation to the ARFIMA likelihood, as discussed by Chan and Palma (1998), with the state-equation matrix modified to allow for changing d. The changes in d are reflected in changes to the MA coefficients. See the Appendix for a state-space representation of the time-dependent ARFIMA model. Using this approach, evaluation of the likelihood is very fast, making the Gibbs sampler approach feasible in practice. Chan and Palma (1998) find that, using a relatively small truncation value, M, for example, $M = 10$ when $n = 1000$, results in accurate parameter estimates for standard ARFIMA models.

ARMA components can easily be accommodated in the current scheme. In this case, the ψ_i coefficients in the truncated MA representation of the model are computed as a function of the (d_t, Φ, Θ) from the relation

$$\Psi_t(B) = \frac{(1 - \theta_1 B - \ldots - \theta_q B^q)}{(1 - \phi_1 B - \ldots - \phi_p B^p)}(1 - B)^{-d_t}$$

To enforce stationarity and invertibility conditions, the AR and MA parameters can be reparameterized in terms of partial autocorrelation coefficients with uniform priors over the region $(-1, 1)$. In this case, the ARMA parameters will have non-standard posterior distributions and can be sampled using, for example, a Metropolis-Hastings algorithm or the 'griddy' Gibbs method. Chib and Greenberg (1994), Marriott *et al.* (1996), and Barnett *et al.* (1997) all discuss various MCMC estimation schemes for ARMA models, while Pai and Ravishanker (1998) and Hsu and Breidt (1999) discuss MCMC estimation of an ARFIMA model with non-zero ARMA terms in the constant d case.

2.2. *Random level-shift ARFIMA model*

Recent work by Granger and Hyung (1999) indicates that undetected level shifts in a time series may spuriously indicate the presence of long-range dependence, similar to the result that a large mean shift can lead to over-differencing in ARIMA modelling. On the other hand, a typical stationary long-range dependent series exhibits features such as apparent local trends and/or cycles and a level that seems to change over time. Because of this typical behaviour, it may be difficult to distinguish between level shifts and persistence subjectively. However, by explicitly accounting for level shifts in the ARFIMA model, we can estimate the probability and size of a shift at a particular point, together with the other ARFIMA model parameters. A random level-shift (RLS) ARFIMA model satisfies

$$y_t = \mu_t + a_t + \sum_{i=1}^{\infty} \psi_i(d, \Phi, \Theta) a_{t-i} \tag{4}$$

where $\mu_t = \mu_{t-1} + \delta_t \beta_t$, analogous to the random-level shift AR (RLAR) model of Chen and Tiao (1990) and McCulloch and Tsay (1993).

Implementation of the Gibbs sampler for estimating the parameters of a RLS-ARFIMA model is the same as that for the RPS-model for σ_a^2, ϵ, and $\{\delta_t\}$. The values of d are chosen using a Metropolis–Hastings algorithm with a $\mathrm{N}(0, \sigma_d^2)$ proposal. Rejection sampling is used to enforce $0 < d < 0.5$. A normal prior, $\mathrm{N}(0, \sigma_\beta^2)$, is used to sample β_j. If $\delta_j = 0$, y_t contains no information on β_j and a new value is drawn from the prior. If $\delta_j = 1$, $\{y_t\}$ contains information on β_j for all $t \geqslant j$. Let

$$a_t = (y_t - \mu_t^*) - \sum_{j=1}^{M} \psi_j a_{t-j},$$

where $\mu_t^* = \mu_t, t < j$ and $\mu_t^* = \mu_t - \delta_j \beta_j, t \geqslant j$. Then $a_t \sim \mathrm{N}(c_{t-j}\beta_j, \sigma_a^2)$, where

$$c_0 = 1,$$
$$c_j = 1 - \sum_{i=1}^{\min(j,M)} \psi_i c_{j-i}, \qquad j > 0$$

The conditional posterior distribution of β_j is normal with mean β_j^* and variance σ_j^2 given by

$$\beta_j^* = \frac{\sigma_\beta^2 G}{\sigma_a^2 + \sigma_\beta^2 D} \qquad \text{and} \qquad \sigma_j^2 = \frac{\sigma_\beta^2 \sigma_a^2}{\sigma_a^2 + \sigma_\beta^2 D}.$$

Here,

$$G = \sum_{t=j}^{n} c_{t-j} a_t \qquad \text{and} \qquad D = \sum_{t=j}^{n} c_{t-j}^2.$$

This is a standard result for a conjugate normal posterior.

2.3. *Implementation issues*

Implementation of the Gibbs sampler requires choice of the hyperparameters specifying the prior distributions. Here, we used $\gamma_1 = 1, \gamma_2 = 20$ in the prior for ϵ.

This choice places most of the mass of ϵ on small values, indicating that we do not expect the value of $d(\mu)$ to jump too often. In the RLS model, we then let σ_β^2 be a multiple (e.g., 3.0) of some initial estimate of σ_a^2. We set $\lambda = 0.5$, $v = 1$ to give a diffuse prior on σ_a^2. Values of μ_0 and σ_0^2 were also chosen to give a diffuse prior on μ, for example, $\mu_0 = 0$, $\sigma_0^2 = 3\hat{\sigma}_a^2$. The variance of the Gaussian proposal for d in the RLS model was set to a small value, 0.001 in our examples. See McCulloch and Tsay (1993) for a more detailed discussion on choice of hyperparameters in similar models.

In the RPS model, we used a grid of 20 values to select d_0 and β_j using the griddy Gibbs procedure. Ten steps of the Hastings algorithm were used within each Gibbs iteration to sample values of d for the RLS-ARFIMA model. The method of Raftery and Lewis (1992) was used to find the minimum burn-in, N_0, and the number of iterations, N, required to estimate the posterior cumulative distribution function of the 0.025 and 0.975 quantiles of the parameters to within ± 0.01 with probability 0.95. The sampler was run for N iterations, with only the last $N - N_0$ iterations used for analysis.

Initial values for d_0, σ_a^2, and μ were obtained using standard maximum likelihood estimation for a fixed parameter ARFIMA model. Initial values for $\{\delta_t\}$ and $\{\beta_t\}$ were obtained by sampling from their respective priors. We let $\epsilon = 0.05$ initially, giving little prior weight to a parameter change.

Due to the inherent long-range correlation properties of the ARFIMA model, a reasonable number of observations is needed to obtain adequate parameter estimates. In our implementation, we allowed shifts only at time periods $t = b + 1, 2b + 1, \ldots, n - b + 1$, where b denotes a block size. Thus if $\delta_{b+1} = 1$, d_0 in the RPS-ARFIMA model is estimated using only the first b observations. In the RLS-ARFIMA model, we set $\hat{\mu} = \bar{x}$ initially and allowed an additional shift at $t = 1$. If there is no mean change, this provides an accurate initial estimate of μ. If a mean change exists, then $\delta_1 = 1$ with high probability and β_1 gives an estimate of the initial mean based on the first b observations.

Note that a shift occurring in the middle of a block will be detected at the closest block end points, unless the shift persists less than b time periods. In other words, our method will be unable to detect single or short runs of additive outliers, but can detect long-lasting or permanent shifts in process structure. Other methods, such as the sup-Wald test, are also unable to detect temporary changes. We investigate the effect of different block sizes on estimation results in the next section.

Bayes factors can be used in our framework, for instance to compare a model that allows d to change and a model having constant d, or to compare the RPS-ARFIMA model to the RLS-ARFIMA model for a given data set. The Bayes factor for model M_r relative to model M_s is defined as

$$B_{rs} = \frac{m(Y|M_r)}{m(Y|M_s)} \tag{5}$$

where $m(Y|M_r)$ is the marginal likelihood of model M_r and

$$m(Y|M_r) = \int f(Y|M_r, \Omega)\pi(\Omega)\mathrm{d}\Omega, \tag{6}$$

where Ω denotes the vector of parameters for model M_r. Kass and Raftery (1995) review different methods for computing the marginal likelihood under many different types of model formulations and discuss interpretation of the computed Bayes factor value. In our framework, we use the harmonic mean of the likelihood values evaluated over a subset of the sampled parameter values as an estimate of the marginal likelihood for each model.

3. SIMULATION RESULTS

As a check on our method, we applied it to some simulated examples in which the presence and size of structural breaks were known.

3.1. *RPS-ARFIMA*

We conducted a simulation study using series containing two shifts in d. We generated 50 series of length $n = 1000$ from a mean-zero $\mathrm{I}(d_t)$ process having $d_t = 0.10, 1 \leqslant t < 330, d_t = 0.40, 330 \leqslant t < 585, d_t = 0.25, 585 \leqslant t < 1000$. The series were generated using a truncated MA of Gaussian white noise variables with MA coefficients as in (3). Each series was initially generated using $\sigma_a^2 = 1.0$. The series was then rescaled to have sample variance 1.0. We estimated the parameters of a RPS-ARFIMA model using block size $b = 20, 50, 100$. Table I gives the average posterior means and standard deviations of $\hat{\theta} = (\hat{\mu}, \hat{d}_0, \hat{\sigma}_a^2, \hat{\epsilon})$ for each block size over the 50 replications, while Figures 1 and 2 show the average posterior mean of d and the average estimated probability of a change over time.

From Table 1, we see that μ and d_0 are accurately estimated no matter the block size, while the posterior mean of ϵ is larger on average than the prior mean of 0.05. Figure 1 shows that the general trend in d is captured quite well. However, from Figure 2, the estimated probability of a change is at most around 0.60 at $t = 300$

TABLE I

MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) OF ESTIMATED PARAMETERS OVER 50 REPLICATIONS OF AN $\mathrm{I}(d_t)$ PROCESS OF LENGTH $n = 1000$ WITH $d_0 = d_t = 0.10, 1 \leqslant t < 330; d_t = 0.40, 330 \leqslant t < 585; d_t = 0.25, 585 \leqslant t \leqslant 1000$

b	μ	d_0	σ_a^2	ϵ
20	0.0265 (0.1401)	0.1469 (0.0602)	0.7646 (0.1395)	0.1431(0.0602)
50	0.0295 (0.1454)	0.1309 (0.0543)	0.7857 (0.1227)	0.1104 (0.0385)
100	0.0307 (0.1512)	0.1329 (0.0567)	0.7789 (0.1279)	0.1019 (0.0310)

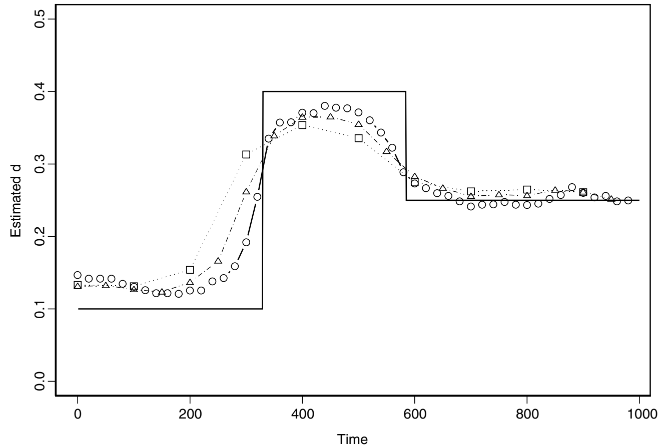

FIGURE 1. Average estimated posterior mean of d over time for 50 simulated $\mathrm{I}(d_t)$ processes; solid line indicates true value of d, while dashed lines indicate estimated d using block sizes $b = 20$ (circles), $b = 50$ (triangles), and $b = 100$ (squares).

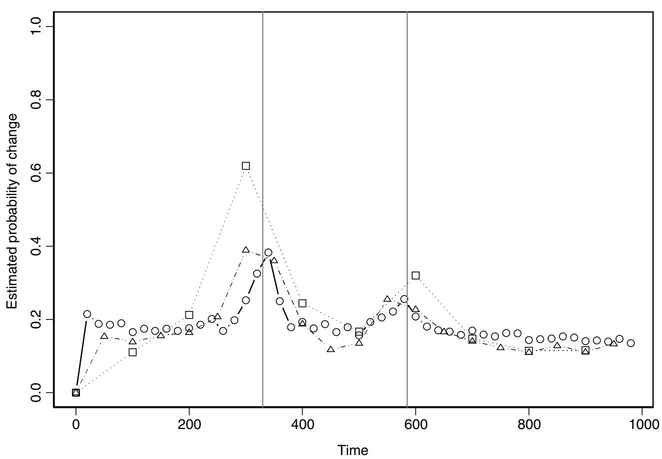

FIGURE 2. Average estimated probability of change in d at block end points for 50 replicated $\mathrm{I}(d_t)$ processes; vertical lines indicate true change points, while dashed lines indicate estimated probability of change using block sizes $b = 20$ (circles), $b = 50$ (triangles), and $b = 100$ (squares).

when $b = 100$, corresponding to the change from $d = 0.10$ to $d = 0.40$ at $t = 330$. As the changes in d do not occur at block end points in this example, the estimated probability of a change may be diffused across several block end points. To investigate further, we conducted another small simulation study using only ten series in which $P(A_t \cup A_{t+b})$ was computed for $t = b + 1, 2b + 1, n - 2b + 1$, where A_t denotes the event a change point is detected at time t. If a change in d

occurs at some time t' between t and $t + b$, we expect to find evidence of that change at time t, time $t + b$, or possibly both. We found that, for both $b = 20$ and $b = 50$, the indicated probability of change around the break points increased from about 0.35 to about 0.50. This suggests that our method can be reliably used to detect both the sizes and approximate times of changes in d. To more precisely estimate the point of change, the MCMC algorithm could be run for a second time, with change points only allowed at a subset of time periods selected based on the initial results.

To check that our method does not find spurious changes in d, we simulated a single realization of a zero-mean Gaussian I(0.4) process of length $n = 1000$ having white noise variance $\sigma_a^2 = 1.0$ using the method discussed in Hosking (1984). The algorithm was implemented exactly as in the previous example using $b = 50$. A posterior mean of 0.4690 was obtained for d_0, comparable to the estimate of 0.4407 obtained using Splus. The largest estimated posterior probability for $\delta_t = 1$ was 0.0133 at $t = 51$, with corresponding mean $\beta_j = -0.0193$.

3.2. *RLS-ARFIMA*

To validate the method for detecting level shifts, we conducted a simulation study involving 100 replicated Gaussian I(0.40) series of length $n = 500$ having a mean shift from zero to 2.77 at $t = 211$. We set white noise variance, σ_a^2, to $\Gamma(1-d)^2/\Gamma(1-2d) = 0.4381$, resulting in process variance $\sigma_y^2 = 1.0$, so that the size of the mean shift was almost $3\sigma_y^2$. Table II gives the average posterior means of $\hat{d}$, $\hat{\sigma}_a^2$ and $\hat{\epsilon}$ over the 100 replications for $b = 20, 50, 100$. The presence of a mean shift does not seem to adversely affect the estimation of d. The average value of ϵ increased to about 0.13 from its initial value 0.05. A mean shift of average size 2.28 was detected at $t = 200$ with probability 1.0 when $b = 100$. The maximum average probability of a mean shift at other allowed time periods was 0.23, with average size only 0.37. A mean shift at $t = 200$ was indicated with average probability 0.98 and average size 2.00 when $b = 50$, while a mean shift at $t = 250$ was indicated with average probability 0.35 and average size 0.70. The maximum average probability at other allowed time periods was less than 0.15. When $b = 20$, a mean shift was indicated with average probability close to 0.60 at

TABLE II

Means and Standard Deviations (in parentheses) of Estimated Parameters over 100 Replications of an I(0.4) Model of Length $n = 500$ with $\mu_t = 0.0, 1 \leqslant t < 211, \mu_t = 2.7, 211 \leqslant t \leqslant 500$

b	d	σ_a^2	ϵ
20	0.3757 (0.0351)	0.6994 (0.0854)	0.1127 (0.0200)
50	0.3897 (0.0335)	0.7106 (0.0853)	0.1358 (0.0204)
100	0.3966 (0.0339)	0.7167 (0.0849)	0.1339 (0.0170)

TABLE III

MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) OF ESTIMATED PARAMETERS OVER 100 REPLICATIONS OF AN I(0.4) PROCESS OF LENGTH $n = 500$ WITH NO LEVEL CHANGES

				Maximum probability of change				
b	d_0	σ_a^2	ϵ	$p > 0.50$	$p > 0.60$	$p > 0.70$	$p > 0.80$	$p > 0.90$
20	0.365 (0.038)	0.675 (0.077)	0.041 (0.024)	22	17	13	10	8
50	0.369 (0.036)	0.677 (0.076)	0.049 (0.024)	20	17	14	12	11
100	0.375 (0.035)	0.680 (0.074)	0.048 (0.017)	10	7	7	5	5

both $t = 200$ and $t = 220$ with average size approximately 1.3 in each case. This is reasonable, as the change actually occurs in the middle of the block, at time $t = 211$. A change in level is indicated at other time periods with probability less than 0.10. Note that, in their RLAR model, McCulloch and Tsay (1993) estimate β_t at every time period, i.e., $b = 1$. This is not computationally feasible for the ARFIMA model.

To verify that our method does not find spurious changes in μ, we applied it to 100 replications of a zero-mean Gaussian I(0.4) process of length $n = 500$. Table III shows the estimated d_0, σ_a^2, and ϵ parameters, along with the number of times (out of 100) that the maximum estimated probability of a change was greater than p, where $p = 0.50, 0.60, 0.70, 0.80, 0.90$. We see that using blocks of size $b = 100$ minimizes the chance of finding spurious level changes, while results of the previous simulation show that this is sufficient to detect a change if one exists. When $b = 20$ or $b = 50$, there is about a 17% chance of spuriously detecting a level change using $p = 0.6$ as the cut-off. The value 0.60 was the estimated average probability of a change at the time periods bracketing the true change in the previous example when $b = 20$. When $b = 50$, the previous simulation results indicate that a larger cut-off could be used, giving smaller chance of spuriously detecting a mean shift. The simulations of Wright (1998) indicate that when a CUSUM or sup-Wald test is applied to a long-range dependent process having $d = 0.40$, a spurious mean change is indicated in about 90% of cases. In our simulations, the average size of the shift detected at the time period having maximum probability of a shift ranged from 0.85 to 1.00 across different probability thresholds and different block sizes.

4. APPLICATIONS

4.1. *Yearly minima of Nile river*

The series of yearly minima of the Nile river has been extensively studied in relation to long-range dependence, and has typically been found to have a long-range dependence parameter d between 0.35 and 0.40; see, for example, Beran (1994). For the series of yearly Nile minima based on measurements near Cairo for the years 622–1284 AD ($n = 663$)[1], Beran and Terrin (1996) find evidence of a change in the amount of long-range dependence around the year 722 AD, with $\hat{d} = 0.04$ for the first 100 years, and $\hat{d} = 0.38$ thereafter. Various historical studies – e.g. Balek (1977) – indicate that a new type of device was introduced for taking measurements around the year 715 AD. A change in the measuring device may be the source of the change in long-range dependent structure, although Whitcher *et al.* (2002) attribute this change to a change in the underlying measurement variability. Using the method of Section 2.1, we estimated the probability of a shift in d for the sample-mean corrected series using $b = 20$. A change in d from 0.05 to 0.45 was indicated at year 722 AD with probability 1.0, analogous to the results of Beran and Terrin (1996). Changes in d at other time periods were indicated with probability less than 0.04.

In theory, to distinguish between changes in σ_a^2 and changes in d, one could allow both random persistence shifts and random variance shifts, discussed in McCulloch and Tsay (1993). We hypothesize that very large amounts of data would be required to distinguish these two possibilities using the Bayesian method.

4.2. *Trade durations for IBM stock*

Our second example pertains to the series of durations between trades of IBM stock on the NYSE from November 1990 to January 1991, a total of 63 trading days ($n = 19022$). Durations are measured in seconds and are time intervals between two consecutive trades. There are no durations between trading days. The durations distribution has a large positive skew, thus we analyse the logarithm of durations.

As an initial step, we applied the spectral regression method of Geweke and Porter-Hudak (1983) with $m = [n^{.6}]$ Fourier frequencies to estimate d for moving blocks of size $n = 1000$. We also computed the sample average of each moving block. The estimated d values ranged from -0.02 to 0.66, while the sample averages ranged from 2.5 to 4.1, suggesting that d, μ, or both may change over time. Given the large sample size, we focus on only one week of durations at a time to determine whether a level or persistence shift occured during that time frame. Figure 3 shows the logarithm of durations for days 48–59, corresponding approximately to the second and third weeks of January 1991. A vertical line marks the beginning of day 55. The durations appear to have changes in both level and persistence over the first period of $n = 2734$ transactions, while they are relatively stable over the second period, consisting of $n = 2316$ transactions.

[1]The data were obtained from http://www.stat.cmu.edu/S/beran

In any given day, two to three hundred durations are recorded. We analysed the data using $b = 220$, i.e. allowing approximately one shift per day. Figure 4 shows the estimated values of d_t and μ_t for both the RPS-ARFIMA and RLS-ARFIMA models during the first period of interest. During this time, three major shifts in d were indicated with probabilities 0.74, 0.45, and 0.79, respectively, when a RPS-ARFIMA model with constant mean was fitted. When a RLS-ARFIMA model with constant d was fitted, two major shifts in μ were indicated, both with probability one. The computed Bayes factor for the RLS-ARFIMA model

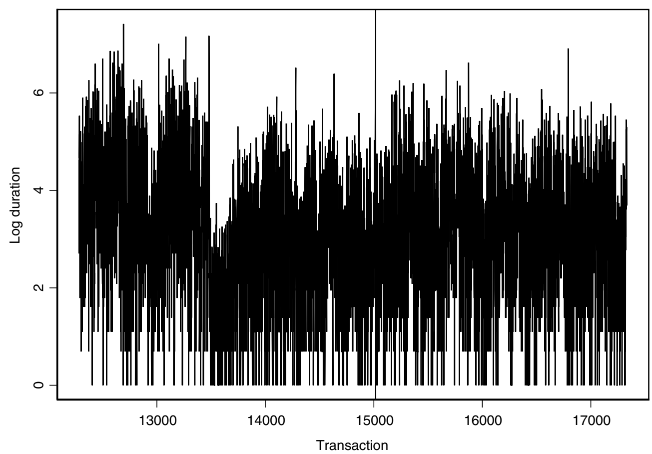

FIGURE 3. Logarithm of time intervals between IBM stock trades for days 48–59 of the period November 1990–January 1991; vertical line indicates start of day 55.

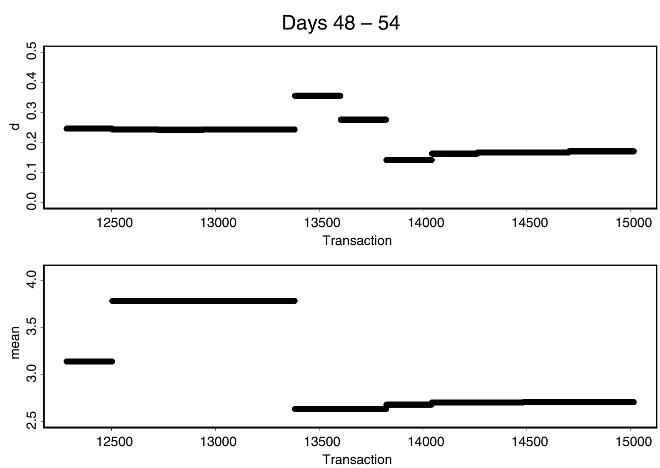

FIGURE 4. Estimated d and μ for the logarithm of time intervals between IBM stock trades for Days 48 through 54 of the period November 1990 through January 1991.

relative to the RPS-ARFIMA model was estimated as 23.66 (on the $\log_{10}$ scale), indicating decisively that changes in the structure of the durations process were due to shifts in the level of the durations rather than shifts in the persistence of durations. Changes in both d and μ were indicated with very low probability during the second set of days.

4.3. *Returns on Coca Cola Stock*

As a third example, we analysed the volatility of daily Coca Cola stock returns, as measured by the absolute value of the actual returns, from 2 January 1990 through 3 December 1999 ($n = 2528$). Recent empirical evidence indicates that, although stock return *levels* have little predictability, variations in stock returns exhibit persistent correlation over time (Lobato and Savin, 1997; Ray and Tsay, 2000). Other authors suggest that apparent long memory may actually be due to unaccounted for changes in the levels of stock return volatilities (Granger and Hyung, 1999; Diebold and Inoue, 2001). An analysis of the first 50 sample autocorrelations of the absolute returns showed that they were all positive and decayed to zero slowly, indicating possible long-range dependent behaviour. However, a time series plot of the data suggests that there may be structural changes in the series. A standard I(d) model fit using approximate MLE methods gave $\hat{d} = 0.143$.

We first fit a RLS-ARFIMA model to the data allowing changes in μ every 63 trading days, corresponding roughly to once a quarter. Columns 2–3 of Table IV give the posterior means and standard deviations of the model parameters. An estimated d value of 0.11 suggests weak persistence in volatility. The solid line in the top plot of Figure 5 shows the estimated mean over time. The analysis indicates a jump in the level of volatility early in 1997.

Fitting a RPS-ARFIMA model with constant mean gave posterior means and standard deviations for model parameters shown in columns 4–5 of Table IV. The solid line of the bottom plot of Figure 5 shows the estimated persistence over time. The value of d decreased from its initial value of 0.168 to 0.12 and remained fairly stable until early 1998. During the first quarter of 1998, the estimated value of d jumped first to 0.181 with probability 0.40 and then to 0.42 in the late second

TABLE IV

POSTERIOR MEANS AND STANDARD DEVIATIONS OF RLS-ARFIMA AND RPS-ARFIMA MODELS FOR ABSOLUTE VALUE OF DAILY STOCK RETURNS FOR THE COCA COLA COMPANY USING $b = 63$

	RLS-ARFIMA		RPS-ARFIMA		RLS-ARFIMA (DC)		RPS-ARFIMA (MC)	
	Mean	Std dev.	Mean	Std dev.	Mean	Std dev.	Mean	Std dev.
μ	–	–	1.175	0.036	–	–	–	–
d_0	0.112	0.016	0.168	0.061	–	–	0.130	0.044
σ_a^2	1.014	0.020	1.025	0.020	1.017	0.020	1.004	0.084
ϵ	0.052	0.028	0.149	0.077	0.049	0.030	0.021	0.049

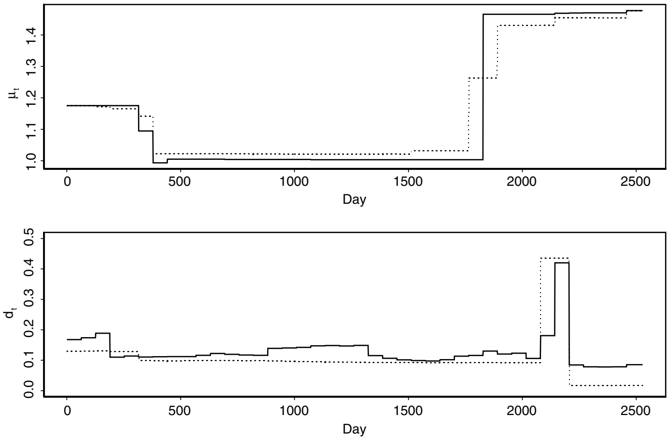

FIGURE 5. Top plot: estimated mean over time for RLS-ARFIMA model of Coke return volatilities: the solid line denotes the estimates when d is held fixed; the dotted line denotes the estimates when $(1-B)^{d_t}$ is used to prefilter the data. Bottom plot: estimated d over time for RPS-ARFIMA models of Coke return volatilities: the solid line denotes the estimates when μ is held fixed; the dotted line denotes the estimates when the data is corrected for a changing mean.

quarter, with probability 0.77. The value of d decreased sharply, to 0.14 with probability 1.0, at the end of the third quarter and remained stable throughout 1999.

It is possible that the changes in persistence indicated by the fitted RPS-ARFIMA model may actually be due to changes in level or vice versa. To investigate further, we fit a RLS-ARFIMA model in which the long-range dependence parameter was not updated, but specified to take values d_t according to the previously estimated RPS-ARFIMA model. Posterior means and standard deviations of estimated model parameters are given in columns 6–7 of Table IV, while the dotted line in the top plot of Figure 5 shows the estimated mean over time. A jump in the level of volatility is still indicated, even after allowance for changing levels of persistence. We next used the estimated mean process from the fitted RLS-ARFIMA model to mean-correct the absolute returns and fit a RPS-ARFIMA model to the mean-corrected data. Posterior means and standard deviations of estimated model parameters are given in columns 8–9 of Table IV, while the dotted line in the bottom plot of Figure 5 shows the estimated d parameter over time. The estimated d value is a little more stable over time than when the mean is held fixed, but there is still a significant jump and subsequent decrease in persistence during 1998. Taken together with the previous results, the analysis suggests that the level of volatility in the returns of Coca Cola stock increased after 1998 and the long-run predictability decreased. Historical data show that the price of Coca Cola stock experienced a significant drop in the second quarter of 1997, but recovered to reach a new high in the second quarter of 1998. The price dropped even more dramatically at the end of the second quarter of 1998, from about \$85 to about \$55 per share, and has not fully recovered ever since. Various reasons ranging from global economic slowdown to company specific problems caused the price to fluctuate, leading to the appointment of a new CEO in the beginning of year 2000. These big price changes may account for the structural changes detected by the proposed method.

5. RELATION TO OTHER WORK

Liu and Kao (1999) also use a Bayesian framework to identify multiple changes of the long-memory parameter in an ARFIMA model. Their method requires that an upper bound on the number of changes in d is determined in advance. They then apply the reversible jump MCMC technique to obtain posterior distributions for the number of changes, the times of occurrence of the changes, and the values of d at each change point. They also allow the random errors in the ARFIMA model to have changing conditional variance, which they parameterize using a GARCH(1,1) model. The ARFIMA likelihood is approximated using the truncated AR representation of the model and d is allowed to range over the interval (0,1). In their examples, the maximum number of allowed change points, k, is kept small, for example, $k = 3$. In contrast, our method does not *a priori* restrict the number of possible change points. Additionally, our use of a truncated MA representation for likelihood evaluation provides a more accurate approximation, as the MA coefficients decay more rapidly than the AR coefficients (Chan and Palma, 1998).

Chib (1998) uses a Bayesian MCMC method to estimate a multiple change-point model in the general setting. His method requires *a prior* specification of the number of change points, but allows the probability of a change to be non-constant.

6. DISCUSSION

We have shown how the Bayesian methodology of McCulloch and Tsay (1993) can be extended in a natural way to allow estimation of the probability of parameter changes in an ARFIMA model. In fact, the method may be easily applied to any time series model having a linear state–space representation. The state–space framework also allows for straightforward extension of the methods of Section 2 to the long-memory stochastic volatility (LMSV) framework, discussed in Breidt *et al.* (1998). Specifically, the distribution of the disturbance terms ϵ_t in the LMSV model can be approximated using a mixture Gaussian distribution. See Hsu and Breidt (2001) for details concerning MCMC estimation of an LMSV model in the constant d case.

Although not considered in this paper, the method of appending a probit model to the ARFIMA model to estimate the probability and size of a shift given a set of explanatory variables can also be implemented in a manner analogous to that of McCulloch and Tsay (1993). We note that, as for the method of Beran and Terrin (1996) for detecting a change in d, model misspecification may result in failure to detect structural changes. The issue of ARFIMA model selection in the presence of structural breaks is a subject for future research.

APPENDIX

Let $y'_t = y_t - \mu$ be a zero mean $\mathrm{I}(d_t)$ process, where

$$d_t = d_0 + \sum_{j=1}^{t} \delta_j \beta_j = d_{t-1} + \delta_t \beta_t.$$

For d fixed, y'_t has an infinite MA representation with MA coefficients that depend on d, i.e.,

$$y'_t = a_t + \sum_{i=1}^{\infty} \psi_i(d) a_{t-i}.$$

Following Chan and Palma (1998), we approximate y'_t as an MA(M) process, where M is chosen dependent on the length of the modelled series, and write the MA model in a state–space representation. The only difference for the RPS-ARIMA model is that the coefficients of the MA representation change over time as d changes. A state–space representation that incorporates these changes is obtained as

$$\tilde{X}_{t+1} = F\tilde{X}_t + T_{t+1}\tilde{Z}_{t+1} \tag{7}$$

$$y'_t = G\tilde{X}_t \tag{8}$$

where $\tilde{X}_t$ is a vector of length $(M+1)$ having ith element

$$X_{t,i} = \sum_{k=i}^{M+1} \psi_{k-1}(d_t) G\tilde{Z}_{t+i-k}$$

Here, $G = [1, 0, \ldots, 0]$ has length $(M+1)$, $\tilde{Z}'_t = [a_t, \ldots, a_{t-M}]$, $F_{i,j} = [\delta_{i+1,j}]_{i,j=1}^{M+1}$ and

$$T_{t+1} = \begin{bmatrix} 1 & \Delta_1(\beta_{t+1}\delta_{t+1}) & \cdots & \Delta_M(\beta_{t+1}\delta_{t+1}) \\ \psi_1(d_{t+1}) & \Delta_2(\beta_{t+1}\delta_{t+1}) & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \psi_M(d_{t+1}) & 0 & \cdots & 0 \end{bmatrix} \tag{9}$$

where $\Delta_i(\beta_{t+1}\delta_{t+1}) = \psi_i(d_{t+1}) - \psi_i(d_t)$.

For given $\mu, d_0, \sigma_a^2, \{\delta_t\}$ and $\{\beta_t\}$, the Gaussian log-likelihood function of y_t is given by

$$-0.5\ln(2\pi) + \sum_{i=1}^{n} \ln(s_t) + \sum_{i=1}^{n} \frac{(y_t - y'_t)}{s_t} \tag{10}$$

where s_t and y'_t are evaluated using standard Kalman recursion equations; see, for example, (Brockwell and Davis, 1987, ch. 12). To start the recursions, values of $\tilde{X}_1$, the mean of the process, and $\tilde{\Sigma}_1$, the covariance matrix of the process, are needed. For the RPS-ARFIMA model, the filter is initialized by letting $\tilde{X}_1 = \mu$ and computing $\tilde{\sigma}_1 = \mathrm{E}(X_1 X_1^{\mathrm{T}}) - \mathrm{E}(\tilde{X}_1 \tilde{X}_1^{\mathrm{T}})$ assuming $d = d_0$.

ACKNOWLEDGEMENTS

The research of Bonnie K. Ray was supported, in part, by the National Science Foundation. That of Ruey S. Tsay was supported by the National Science Foundation and the Graduate School of Business, University of Chicago. The authors thank the editor and two anonymous referees for insightful comments that helped to improve the paper.

REFERENCES

BALEK, J. (1977) Hydrology and water resources in tropical Africa. In *Developments in Water Science*, 8, New York: Elsevier.

BARNETT, G., KOHN, R. and SHEATHER, C. (1997) Robust Bayesian estimation of auto-regression moving average models. *Journal of Time Series Analysis* 18, 11–28.

BERAN, J. (1994) *Statistics for Long Memory Processes*. New York: Chapman & Hall.

——— and TERRIN, N. (1996) Testing for a change of the long-memory parameter. *Biometrika* 83, 627–38.

BREIDT, F. J., CRATO, N., and de LIMA, P. (1998) On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* 83, 325–48.

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

CHAN, N. H. and PALMA, W. (1998) State space modeling of long-memory processes. *Annals of Statistics* 26, 719–39.

CHEN, C. and TIAO, G. C. (1990) Random level shift time series models, ARMA approximations, and level-shift detection. *Journal of the Business and Economic Statistics* 8, 83–97.

CHIB, S. (1998) Estimation and comparison of multiple change-point models. *Journal of Econometrics* 86, 221–41.

——— and GREENBERG, E. (1994) Bayes inference in regression models with ARMA(p, q) errors. *Journal of Econometrics* 64, 183–206.

DIEBOLD, F. and INOUE, A. (2001) Long memory and regime switching, *Journal of Econometics* 105, 131–159.

GEWEKE, J. and PORTER-HUDAK, S. (1983) The estimation and application of long-memory time series models. *Journal of Time Series Analysis* 4, 221–38.

GRANGER, C. W. J. and HYUNG, N. (1999) Occasional structural breaks and long memory, University of California, San Diego Department of Economics Discussion Paper 99-14, June, 1999.

——— and JOYEUX, R. (1980) An introduction to long-memory time series models and fractional differencing. *J Time Series Analysis* 1, 15–29.

HIDALGO, J. and ROBINSON, P. (1996) Testing for structural change in a long-memory environment. *Journal of Econometrics* 70, 159–74.

HOSKING, J. R. M. (1981) Fractional differencing. *Biometrika* 68, 165–76.

——— (1984) Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research* 20, 1898–908.

HSU, N. -J. and BREIDT, F. J. (2001) Bayesian analysis of fractionally integrated ARMA with additive noise, *Journal of Forecasting*, to appear.

KASS, R. and RAFTERY, A. (1995) Bayes factors. *Journal of the American Statistical Association* 90, 773–95.

KUAN, C.-M. and HSU, C.-C. (1998) Change-point estimation of fractionally integrated processes. *Journal of Time Series Analysis* 19, 693–708.

Liu, S. -I. and Kao, M. -H. (1999) Bayesian analysis for multiple changes of the long memory parameter, Graduate Institute of Statistics, National Central University, Chung-Li, Taiwan.

Lobato, I. N. and Savin, N. E. (1997) Real and spurious long memory properties of stock market data. *Journal of Business and Economic Statistics* 16, 261–7.

Mariott, J., Ravishanker, N., Gelfand, A. and Pai, J. (1996) Bayesian analysis of ARMA processes: Complete sampling based inference under full likelihoods. In *Bayesian Statistics and Econometrics: Essays in Honor of Arnold Zellner* (eds D. Berry, K. Chaloner and J. Geweke), New York: Wiley.

McCulloch, R. and Tsay, R. (1993) Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the Amererican Statistical Association* 88, 968–78.

Pai, J. S. and Ravishanker, N. (1998) Bayesian analysis of autoregressive fractionally integrated moving average processes. *Journal of Time Series Analysis* 19, 99–112.

Raftery, A. E. and Lewis, S. M. (1992) Comment on "The Gibbs sample and Markov chain Monte Carlo", One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science* 7, 493–7.

Ray, B. K. and Tsay, R. S. (2000) Size effects on common long-range dependence in stock volatilities. *Journal of Business and Economic Statistics* 18, 254–62.

Wang, Y., Cavanaugh, J. E. and Song, C. (2001) Self-similarity index estimation via wavelets for locally self-similar processes. *Journal of Statistical Planning and Inferences*. 99, 91–110.

Whitcher, B., Byers, S. D., Guttorp, P. and Pericval, D. B. (2002) Testing for homogeneity of variance in time series: Long memory, wavelets, and the Nile River, *Water Resources Research* 38(5), 10.1029/2001WR0005.93.

Wright, J. (1998) Testing for a structural break at unknown data with long-memory disturbances. *Journal of Time Series Analysis* 19, 369–76.