A Bayesian Approach to Estimating the Long Memory Parameter

Scott Holan*
Tucker McElroy
Sounak Chakraborty*

University of Missouri Columbia*

Statistical Research Division
U.S. Census Bureau
Washington, DC 20233

A Bayesian Approach to Estimating the Long Memory Parameter

Scott Holan[*], Tucker McElroy[†] and Sounak Chakraborty [‡]

University of Missouri-Columbia, U.S. Census Bureau and University of Missouri-Columbia

Abstract

We develop a Bayesian procedure for analyzing stationary long-range dependent processes. Specifically, we consider the fractional exponential model (FEXP) to estimate the memory parameter of a stationary long-memory Gaussian time series. Further, the method we propose is hierarchical and integrates over *all* possible models, thus reducing underestimation of uncertainty at the model-selection stage. Additionally, we establish Bayesian consistency of the memory parameter under mild conditions on the data process. Finally the suggested procedure is investigated on simulated and real data.

Keywords. Adaptive model selection, Bayesian model averaging; FEXP; Hierarchical Bayes; Long-range dependence, Reversible Jump Markov Chain Monte Carlo; Spectral density.

Disclaimer This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the author and not necessarily those of the U.S. Census Bureau.

1 Introduction

Motivated by applications in econometrics, hydrology and other scientific disciplines, fractionally differenced models have been the subject of extensive research over the past two decades. The differencing parameter characterizes the long-memory behavior of a time series and thus has far-reaching utility in areas such as telecommunications, economics, finance, and geophysics among others. Therefore, it is of considerable interest to provide estimators of the memory parameter that have good asymptotic properties, while remaining practical for use in finite samples.

[*](to whom correspondence should be addressed) Department of Statistics, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO, 65211-6100, holans@missouri.edu

[†]Statistical Research Division, U.S. Census Bureau, 4700 Silver Hill Road, Washington, D.C. 20233-9100, tucker.s.mcelroy@census.gov

[‡]Department of Statistics, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO, 65211-6100, chakrabortys@missouri.edu

The long memory process that we consider is Gaussian and has spectral density

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} f^*(\lambda), \ \ \lambda \in (-\pi, \pi). \tag{1}$$

Further, we say that the stationary time series $\{x_t\}$ is *long range dependent* or *long memory* if there exists a $d \in (0, .5)$ such that

$$\lim_{\lambda \to 0^+} \frac{f(\lambda)}{\lambda^{-2d}} = k$$

for some constant $k > 0$. Likewise for $d = 0$ we say that the process is short range dependent or short memory. We assume that the function $f^*(\lambda)$ is even, continuous, and bounded away from zero on $[-\pi, \pi]$. Thus d and f^* respectively govern the long-term and short-term behavior of the process.

Although several methods have been proposed for estimating d, two main methods have emerged. The first method is the so-called parametric approach. One example of the parametric approach was introduced by Fox and Taqqu (1986) where the authors provide a fully parametric maximum likelihood procedure for the case where $\{x_t\}$ is strictly Gaussian. Subsequently this method was developed in greater detail by Giraitis and Taqqu (1999). Geweke and Porter-Hudak (1983) suggested a regression approach to the estimation of the long memory parameter using the spectral domain, while Beran (1993) introduced a procedure based on generalized linear regression. The second method commonly used employs semiparametric estimators and has been the subject of extensive research. Some examples are Robinson (1995a, 1995b) and Giraitis and Surgailis (1990), where both authors justified the use of semiparametric inferential procedures. For a comprehensive discussion on long memory time series and fractionally differenced models see Beran (1994), Robinson (2003), and Palma (2007) and the references therein.

In order to estimate the long memory parameter d we develop a Bayesian approach. Although some research on long memory has taken a Bayesian viewpoint, the literature is very heavily frequentist (see Robinson, 2003 for a discussion). The majority of the Bayesian literature consists of parametric characterizations of long-range dependence based on a class of models. Perhaps the most popular approach is to consider the class of all finite order stationary and invertible ARFIMA (Auto Regressive Fractionally Integrated Moving Average) processes (Hosking, 1981). For example, Ravishanker and Ray (1997), Pai and Ravinshanker (1998), and Ko and Vannucci (2006), all consider Bayesian parametric models.

The first nonparametric Bayesian analysis of long memory data was considered by Petris (1997). This work is posed under the assumption that, for $\lambda \in [0, \pi]$,

$$f(\lambda) \propto \lambda^{-2d} \exp\{g(\lambda)\}, \tag{2}$$

where $g(\lambda)$ is meant to capture the short-range dependence in $\{x_t\}$. Further, at the Fourier frequencies, λ_j, the prior probability measure on $g(\cdot)$ is defined by a centered autoregressive process

of order p with parameters having a Beta prior distribution. As noted in Liseo, Marinucci, and Petrella (2001), in order for the two components λ^{-2d} and $g(\lambda)$ in (2) to be identifiable, appropriate conditions on $g(\lambda)$ need to be imposed.

In this paper we consider an explicitly defined semiparametric estimator of d based on fitting potentially misspecified parametric models. Specifically, the models we fit are from the FEXP class described by Beran (1993, 1994). The FEXP estimator was originally proposed by Janacek (1982) and later discussed by Robinson (1994). The spectral density of the FEXP model has the form of (1) with $f^*(\lambda)$ given by the exponential model of Bloomfield (1973), i.e.,

$$\log f^*(\lambda) = \sum_{k=0}^{m} b_k \cos(k\lambda), \tag{3}$$

where $b_0, \ldots, b_m$ are real constants and m is a positive integer.

The approach we suggest is similar to the Bayesian semiparametric approach of Liseo et al. (2001) to the extent that they consider a spectral method with the short-range dependent portion of the model related (but not equal) to the class of fractional exponential processes introduced by Beran (1993) and to the work of Bloomfield (1973) in the short memory case. However, the method we propose differs from Liseo et al. (2001) in several ways. One distinction between the two methods is that our method is *broad-band* in that the spectral model we impose for the short-range dependent process is modelled on the frequency band $[0, \pi]$; in contrast, the work of Liseo et al. (2001) considers a *narrow-band* estimator of the short memory process by restricting this process to the frequency band $[\lambda^*, \pi]$. This representation induces a model selection problem on the choice of λ^*. The authors suggest treating λ^* as a hyperparameter whose value is ultimately determined by the data. The selection procedure is facilitated via comparison of prior predictive distributions of the observed data set. Further, the method they propose uses a periodogram based Whittle likelihood approximation to the true likelihood (Palma, 2007), while we consider both a "time domain" Whittle likelihood (TD Whittle) as well as the exact Gaussian likelihood. Although the Whittle form we use still constitutes an approximation we denote it as TD Whittle to emphasize the fact that we do not appeal to the periodogram directly. Using the TD Whittle likelihood as well as the exact Gaussian likelihood provides a significant distinction between our models and other models in the FEXP class, as the Toeplitz autocovariance matrix is difficult to deal with computationally. Toward this direction we provide explicit details of algorithms that facilitate its use. Further by avoiding a periodogram based approximation to the likelihood, using both the TD Whittle and the exact Gaussian likelihood, our method can accommodate missing data with relatively little extra computational burden.

Another striking advantage of the method we suggest is that model selection (i.e., the order m in (3)) is embedded in our estimation procedure by making the model order a random parameter to be estimated in the model. This transdimensional approach (Sisson, 2005) provides added flexibility

and reduces uncertainty at the model-selection stage by integrating over *all* possible models from the FEXP class (up to a given finite order). This procedure can be seen as a type of Bayesian model averaging (Hoeting, Madigan, Raferty and Volinsky, 1999). Thus, we consider m a nuisance parameter. Finally, allowing the dimension of the parameter space to change requires a non-trivial extension to the proof of the asymptotic properties.

This paper is organized as follows. In Section 2 we describe the modelling framework, including choice of priors and discussion of implementation of the likelihood calculations. Section 3 provides the details of the Markov Chain Monte Carlo (MCMC) and Reversible Jump Markov Chain Monte Carlo procedures used for estimation. Specifically we present algorithms for both the fixed dimensional and transdimensional cases. Section 4 develops the asymptotic properties of the *a posteriori* estimates of the parameters. In Section 5 our methodology is investigated through Monte Carlo experiments and application to a real data set, the Nile River minima data (Beran, 1994). Finally, Section 6 contains discussion. For convenience of exposition all proofs and derivations are left to the appendix.

2 Hierarchical Bayesian FEXP Model

2.1 Bayesian FEXP

Suppose that the time series of interest follows the model

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d} \exp\left\{ \sum_{l=0}^{m} b_l \cos \lambda l \right\} \tag{4}$$

for $\lambda \in [-\pi, \pi]$. Here f is the spectral density of the stationary process, d is the long memory parameter, and the coefficients b_l parametrize the short memory aspects of the process. Stationarity is maintained by considering $d < 1/2$; since we focus on long memory, we also assume that $d \geq 0$. Note that when $d = 0$, f reduces to the spectral density of an EXP(m) process, which has short memory (Bloomfield, 1973). We start by defining the parameter space $\Theta_m = [0, .5) \times \mathbb{R}^{m+1}$, which explicitly depends on the order m. We use the notation

$$\theta_m = (d, b_0, b_1, \cdots, b_m)$$

for a typical element of Θ_m. Henceforth we use the notation f_{θ_m} to denote the spectral density given in (4) whenever we wish to make the dependence on model parameters explicit. Conditional on m being known, it is simple to write down the Gaussian likelihood for a sample $x^{(n)} = \{x_1, x_2, \cdots, x_n\}$ of size n:

$$p(x^{(n)}|\theta_m) = (2\pi)^{-n/2} \left|\Sigma(f_{\theta_m})\right|^{-1/2} \exp\left\{ -\frac{1}{2} x^{(n)'} \Sigma^{-1}(f_{\theta_m}) x^{(n)} \right\}. \tag{5}$$

We use the following notation for covariance matrices: the Toeplitz covariance matrix of dimension n of a stationary process with spectral density f is written $\Sigma(f)$. Its jk^{th} entry is the autocovariance

function at lag $j - k$ given by

$$\Sigma_{jk}(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) \, e^{i\lambda(j-k)} \, d\lambda.$$

Perhaps one of the most salient features of our approach arises in the choice of the model order m. In a frequentist setting it is most common to choose the order of a FEXP model through order selection criteria (see Hurvich, 2001). Similarly, in the Bayesian context one could maintain the model selection approach and choose m using Bayes factor (Gelman, Carlin, Stern, and Rubin, 2003). However, a more flexible approach takes full advantage of the Bayesian paradigm by putting a prior on m and selecting the number of terms to be included in the model adaptively. Since the main goal of our analysis is estimation of d, we embed the choice of m into the modelling framework. This added layer of flexibility in terms of m enhances the accuracy of the estimate of d by reducing the underestimation of uncertainty at the model-selection stage.

2.2 Likelihood Calculations

Implementation of the Bayesian approach under the exact Gaussian likelihood poses several challenges. First one must form the Toeplitz autocovaraince matrix $\Sigma(f_{\theta_m})$ in (5). This requires calculation of the autocovariances of lag L, $\{c_L\}_{L=0}^{n-1}$, given only knowledge of f_{θ_m}. Further, once the autocovariance matrix is constructed there is still a need to calculate both $|\Sigma(f_{\theta_m})|$ and $x^{(n)'}\Sigma^{-1}(f_{\theta_m})x^{(n)}$. These quantities present computational difficulties because here the autocovariances decline slowly (at a power law rate) and $\Sigma(f_{\theta_m})$ is quite ill-conditioned (Chen, Hurvich, and Lu, 2006).

Calculation of the autocovariances proceeds via a spectral factorization method on the short memory portion of (1) (Pourahmadi, 1983; Hurvich, 2002). Specifically, given $b_0, \ldots, b_n$ one can write down an equivalent $\mathrm{MA}(\infty)$ representation (under mild conditions) using the formula

$$\beta_j = \frac{1}{2j} \sum_{k=1}^{j} k b_k \beta_{j-k},$$

with $\beta_0 \equiv 1$ and β_j ($j = 1, 2, 3, \ldots$) equal to the $\mathrm{MA}(\infty)$ coefficients. Furthermore, the innovation variance, σ_ϵ^2 can be expressed as $\sigma_\epsilon^2 = 2\pi \exp(b_0)$. Next, let $\tilde{c}_h$ denote the theoretical autocovariance associated with a specific short memory $\mathrm{EXP}(m)$ model. Here,

$$\tilde{c}_h = \sigma_\epsilon^2 \sum_{j=h}^{\infty} \beta_j \beta_{j-h} \;\; (h \geq 0).$$

Since β_j decay exponentially fast, a good approximation to $\tilde{c}_h$ can be obtained using the truncated version $\tilde{c}_{h,N}$ for N sufficiently large. Note that $\tilde{c}_{h,N}$ may be efficiently calculated by taking advantage of the FFT (Fast Fourier Transform); see Hurvich (2002) for a comprehensive discussion.

Similarly, let $\{c_s^*\}$ denote the autocovariances of an ARFIMA$(0, d, 0)$ process with unit innovation variance. Again, since $\tilde{c}_h$ decay exponentially fast, c_L can be approximated accurately ($c_{L,Approx.}$) using $\{c_s^*\}$ and $\tilde{c}_{h,N}$ in a "splitting method" (Hurvich, 2002; Bertelli and Caporin, 2002). Additionally, as suggested by Hurvich (2002) the sequence $\{c_{L,Approx.}\}_{L=0}^{n-1}$ can be used to simulate a zero-mean Gaussian realization $x_1, \ldots, x_n$ from the FEXP process $\{x_t\}$ using the Davies-Harte algorithm (Davies and Harte, 1987). The algorithm is exact, in the sense that the autocovariances of the simulated time series exactly match those used as inputs to the algorithm. This algorithm will be used is Section 5 where our methodology is demonstrated on simulated data. Finally, the Toeplitz matrix $\Sigma(f_{\theta_m})$ can be constructed using $\{c_{L,Approx.}\}_{L=0}^{n-1}$.

Ultimately the likelihood will be embedded in an MCMC procedure, Gibbs sampling with Metropolis-Hasting steps (Robert and Casella, 2004). For this purpose it is more convenient to work with the log likelihood rather than the likelihood itself. Thus the quantities needed to calculate $\log p(x^{(n)}|\theta_m)$ are the quadratic form $x^{(n)'}\Sigma^{-1}(f_{\theta_m})x^{(n)}$ and $\log |\Sigma(f_{\theta_m})|$. The first of these quantities, the quadratic form, can be accurately calculated to any desired degree of precision using conjugate gradient (CG) or preconditioned conjugate gradient algorithms (PCG) (Chen et al., 2006). Similarly $\log |\Sigma(f_{\theta_m})|$ can be accurately approximated (numerically) in $O(1)$ operations (Bötcher and Silberman, 1999; Chen et al., 2006). Therefore we can calculate the exact log likelihood with a high degree of precision. For a detailed discussion see Chen et al. (2006).

2.3 Hierarchical Priors and Posteriors

We now fix ideas by setting out some specific hierarchical priors. Let $b = (b_0, b_1, b_2, \cdots)$ and $\sigma^2 = (\sigma_0^2, \sigma_1^2, \sigma_2^2, \cdots)$ be infinite sequences. Then we assign hierarchical priors on the unknown parameters as follows:

$$b_k | m, \sigma_k^2 \sim \begin{cases} \mathcal{N}(0, \sigma_k^2); & \text{if } k \leq m, \\ \Delta_0; & \text{if } k > m, \end{cases}$$

$$d \sim \text{Uniform}(0, 1/2),$$

$$\sigma_k^2 | m \sim \begin{cases} IG(\alpha, \beta); & \text{if } k \leq m, \\ \Delta_0; & \text{if } k > m, \end{cases}$$

$$m \sim \text{Discrete Uniform}\{1, M\}.$$

Alternatively, if one has prior information on the order of the model, this information can be incorporated easily into this framework by choosing a Truncated Poisson$\{\eta, M\}$ prior for m. We denote the point mass at zero via Δ_0, the issue being that b_k and σ_k^2 for k exceeding a given m must be defined to be identically zero. As usual IG denotes the inverse gamma distribution, so

that σ_k^2 has pdf

$$p(\sigma_k^2) \propto \exp\left(-\frac{\beta}{\sigma_k^2}\right)(\sigma_k^2)^{-(\alpha+1)}.$$

Note that the choices of α, β, M, and η (if appropriate) are determined by the practitioner. The selection of priors is usually problem-specific. Ideally, one would like to elicit these priors from past history. However, for most practical applications, such information is unavailable. In such cases it seems reasonable, at least from robustness considerations, to use near-diffuse priors for the hyperparameters of the hierarchical model. Although this would provide a reasonable approach under a fairly broad range of settings we find that FEXP model coefficients tend to be "relatively" small. For this reason one appropriate strategy is to consider a tighter set of priors. Further this choice can be justified by the heuristic argument that typically in long memory settings one requires long time series in order to be able to adequately estimate quantities of interest. Therefore under most prior specifications there is enough data to overcome the effect due to the prior. As we will find in the next section, we have taken "non-informative" priors for b_k, d, σ_k^2, and m, in the sense that on the scale of typical FEXP coefficients the prior provides relatively little information. Finally, if the joint prior is proper, then the posterior is also proper. The present study is based on proper priors; hence, impropriety of the posterior is not an issue in our case. Note that in Section 4 we present asymptotic results under relaxed specification of the the hierarchical priors.

The approach we take is hierarchical Bayes. This type of model has its virtues in that it is richer than a one-stage model, and automatically produces shrinkage estimators of parameters. These shrinkage estimators, by borrowing strength, usually perform better than the regular estimators. Additionally, taking m random we have increased the precision as well as added flexibility to our model by allowing the data to choose the order of the model adaptively given the data. This will be revealed in Section 5, when we illustrate our methodology using both simulated data and an actual data analysis. Below, we consider the joint posterior distribution of the parameters. Noting that b and σ^2 do not depend on d, and the distribution of d does not depend on m, we have

$$\begin{aligned}
p(b, \sigma^2, d, m | x^{(n)}) &\propto p(x^{(n)} | b, \sigma^2, d, m) p(b | \sigma^2, m) p(\sigma^2 | m) p(d) p(m) \\
&\propto (2\pi)^{-n/2} |\Sigma(f_{\theta_m})|^{-1/2} \exp\left\{-\frac{1}{2} x^{(n)'} \Sigma^{-1}(f_{\theta_m}) x^{(n)}\right\} \\
&\quad \times \prod_{k=0}^{m} (\sigma_k^2)^{-1/2} \exp\left\{-\frac{b_k^2}{2\sigma_k^2}\right\} \\
&\quad \times \prod_{k=0}^{m} \exp\left\{-\frac{\beta}{2\sigma_k^2}\right\} (\sigma_k^2)^{-(\alpha+1)} \\
&\quad \times 1_{[0,1/2)}(d) \\
&\quad \times \frac{1}{M}.
\end{aligned} \tag{6}$$

7

Our primary objective is to find the posterior distribution of d given $x^{(n)}$, and use the posterior mean to obtain an estimate for d. The exact posterior distribution $p(d|x^{(n)})$ is given in (14) of Section 4, and the posterior mean is

$$\hat{d} = \int dp(d|x^{(n)}) d\nu(d), \tag{7}$$

where $\nu(dx)$ denotes Lebesque measure on the real line. Asymptotic properties of this estimate are considered in Section 4. In practice, the posterior $p(d|x^{(n)})$ is intractable, as its evaluation requires high dimensional numerical integration. Thus we adopt the popular alternative approach of MCMC, which requires generating samples from the full conditionals of the different parameters given the remaining parameters and the data. Some further details in this direction are provided in Section 3 below.

3 Estimation

3.1 Full Conditionals and Fixed Dimensional MCMC

Some of the full conditionals are not in standard form and thus we have used a Metropolis within Gibbs algorithm (Robert and Casella, 2004). Below we list these full conditionals; for convenience of exposition a full derivation can be found in Appendix B. Our general proposed model makes adaptive selection of m (the model order). However, in this section we take m to be fixed and describe our implementation procedure. Subsequently in Section 3.2 we describe a RJMCMC approach (Green, 1995) for the case in which m is random.

First, since σ^2 does not depend on d, and the distribution of d does not depend on m, it is easy to see that the full conditional of σ_k^2 is conjugate. As a result, it is straight forward to show that the full conditional of σ_k^2 is IG and is given by

$$p(\sigma_k^2|d, m, b_0, b_1, \ldots, b_k, \ldots, b_m) \quad \sim \quad IG\left\{(\alpha + 1/2), (b_k^2/2 + \beta)\right\}. \tag{8}$$

Although the full conditional of $b = (b_0, \ldots, b_k)$ is not conjugate under this model it is equally straight forward to derive and is given by

$$\begin{aligned}
p(b|m, \sigma^2, d, x^{(n)}) \quad &\propto \quad (2\pi)^{-n/2} |\Sigma(f_{\theta_m})|^{-1/2} \exp\left\{-\frac{1}{2} x^{(n)'} \Sigma^{-1}(f_{\theta_m}) x^{(n)}\right\} \\
&\times \quad \prod_{k=0}^{m} (\sigma_k^2)^{-1/2} \exp\left\{-\frac{b_k^2}{2\sigma_k^2}\right\}.
\end{aligned} \tag{9}$$

Finally, the full conditional of d can be expressed as

$$\begin{aligned}
p(d|\sigma^2, m, b, x^{(n)}) \quad &\propto \quad (2\pi)^{-n/2} |\Sigma(f_{\theta_m})|^{-1/2} \exp\left\{-\frac{1}{2} x^{(n)'} \Sigma^{-1}(f_{\theta_m}) x^{(n)}\right\} \\
&\times \quad 1_{[0,1/2)}(d).
\end{aligned} \tag{10}$$

In the case of fixed model dimension implementation, MCMC is relatively straightforward given the calculated likelihood. Therefore we begin with this case and proceed with m held fixed for any given model. The general algorithm we use is a Gibbs sampler (Gelfand and Smith, 1990), although several Metropolis-Hastings (M-H) steps will be required. The pdf given in (8) is standard and so can be sampled from directly. Unfortunately, generating samples from (9) and (10) is more complicated and requires M-H updates. The algorithm proceeds as follows:

Step 1: Set initial values for all parameter values.

Step 2: Generate samples from (8).

Step 3: Using a Random-Walk, M-H sample the b_k's one at a time from (9) holding all b_j fixed for $j \neq k$. Additionally, since the b_k's are conditionally independent we need only to include the second term of (9) associated with the specific b_k being sampled.

Step 4: Using an Independent M-H step sample from (10).

Step 5: Repeat.

For the implementation of the M-H algorithm, one needs a candidate generating density. Chib and Greenberg (1995) have several proposals in this regard. For specificity, let b_k denote the current state for the parameter b_k; we then draw a candidate value b_k^* using a $\mathcal{N}(b_k, \delta_k^2)$ proposal distribution, where the δ_k are user defined tuning parameters. The algorithm accepts b_k^* as a new value of b_k with acceptance probability

$$\alpha_{b_k} = \min \left\{ 1, \frac{p(b_k^*|m, \sigma^2, d, x^{(n)})}{p(b_k|m, \sigma^2, d, x^{(n)})} \right\},$$

where $p(b_k|m, \sigma^2, d, x^{(n)})$ is given in (9). Similarly for d^* draw a candidate value from a $U(0, 1/2)$ proposal distribution and accept d^* as a new value of d with acceptance probability

$$\alpha_d = \min \left\{ 1, \frac{p(d^*|\sigma^2, m, b, x^{(n)})}{p(d|\sigma^2, m, b, x^{(n)})} \right\},$$

where $p(d|\sigma^2, m, b, x^{(n)})$ is given in (10). Note that in practice it is often easier to compute $\alpha_\bullet$ by making appropriate transformations to the log scale.

In the case where the model dimension is fixed a priori it is natural to estimate models using several different orders (values of m). Then with these models in hand one can choose a final model through the use of Bayes factor or DIC; see Gelman et al. (2003) for further discussion. Since we propose transdimensional modeling of d, where the order m is taken as a model parameter we do not pursue discussion of order selection further.

3.2 Adaptive Model Procedure and RJMCMC

The transdimensional case is substantially more difficult to sample from as we are potentially changing dimensions on each MCMC iteration. Specifically, in this case m is a random parameter to be adaptively estimated from the data. To facilitate model estimation we require the full conditional distribution of m. Thus, for $m = 0, \ldots, M$, the full conditional of m can be written as

$$
\begin{aligned}
p(m|b, \sigma^2, d, x^{(n)}) \quad \propto \quad & (2\pi)^{-n/2} |\Sigma(f_{\theta_m})|^{-1/2} \exp\left\{-\frac{1}{2}x^{(n)\prime}\Sigma^{-1}(f_{\theta_m})x^{(n)}\right\} \\
\times \quad & \prod_{k=0}^{m}(\sigma_k^2)^{-1/2} \exp\left\{-\frac{b_k^2}{2\sigma_k^2}\right\} \\
\times \quad & \prod_{k=0}^{m} \exp\left\{-\frac{\beta}{\sigma_k^2}\right\}(\sigma_k^2)^{-(\alpha+1)} \\
\times \quad & \frac{1}{M}.
\end{aligned}
\tag{11}
$$

Now, for notational convenience, let Θ denote the collection of parameters in our model. To sample from $p(\Theta|x^{(n)})$ we can use the reversible jump algorithm (Green, 1995). Here we outline the steps we employ to traverse the posterior probability surface; see Denison et al. (1998) and Denison et al. (2002) and references therein for a general discussion. We allow three types of moves: BIRTH (B), DEATH (D), and MOVE (S), where these moves are defined by

B: propose with probability p_m^b to increase model dimension by 1,

D: propose with probability p_m^d to decrease model dimension by 1,

S: propose with probability p_m^s to move in the same model dimension.

Note that in the models we consider we require at least one term always in the model (in addition to b_0, i.e. b_0 and b_1). Therefore, the probabilities are given as follows:

$$
\begin{aligned}
p_m^b = p_m^d = p_m^s = 1/3; \;\; \text{for} \;\; m = 2, \ldots, M-1, \\
p_1^b = p_1^s = p_M^d = p_M^s = 1/2, \\
p_M^b = p_1^d = 0.
\end{aligned}
$$

Note that $p_m^b + p_m^d + p_m^s = 1$, for all m. The algorithm proceeds as follows:

1. Set starting values.

2. Generate $u_1 \sim U(0, 1)$

   - if $u_1 < p_m^b$ then goto B step and obtain Θ'
   - otherwise if $p_m^b \leq u_1 \leq p_m^b + p_m^d$ goto D step and obtain Θ'

- otherwise goto S step and obtain Θ'.

3. Suppose we goto B step

   (i.) generate $b_{m+1} \sim \mathcal{N}(0, 1)$

   (ii.) generate entire ($m + 1$ dimensional) σ^2 vector

   (iii.) complete a MCMC iteration as in fixed dimensional case.

3'. Similar steps occur in D and S steps with the appropriate modifications.

4. Generate $u_2 \sim U(0, 1)$

   - if $u_2 < \min\{1, BF(\Theta', \Theta^{(t)}) \times R\}$ then $\Theta^{(t+1)} = \Theta'$

   - else $\Theta^{(t+1)} = \Theta^{(t)}$.

5. Repeat.

Here,

$$
R \;=\; \begin{cases} p^d_{m+1}/p^b_m; & \text{if B step,} \\ p^b_{m-1}/p^d_m; & \text{if D step,} \\ 1; & \text{if S step,} \end{cases}
$$

(12)

and

$$
BF(\Theta', \Theta^{(t)}) = \frac{\int_{\sigma^2} p(\Theta'|x^{(n)})d\sigma^2}{\int_{\sigma^2} p(\Theta^{(t)}|x^{(n)})d\sigma^2}.
$$

(13)

where R and BF denote the proposal ratio and Bayes factor (integrating out the nuisance parameter σ_k^2 for $k = 0, \ldots, m$) respectively. An analytic form for BF and a related discussion regarding its consistency as an order selector in our setting can be found in Appendix C. Note that in the case where a Truncated Poisson$\{\eta, M\}$ prior is imposed on m, R requires slight modification and is given by

$$
R_{TPois} \;=\; \begin{cases} \frac{p^d_{m+1}}{p^b_m} \frac{\eta}{m+1}; & \text{if B step,} \\ \frac{p^b_{m-1}}{p^d_m} \frac{m}{\eta}; & \text{if D step,} \\ 1; & \text{if S step.} \end{cases}
$$

Remark 1 Although our algorithm does not appear to be explicitly sampling from the full conditional of m, it is straight-forward to show that $p(\Theta'|x^{(n)})$ is equivalent to (11). Thus the comparison made in Step 4 is implicitly acting as an M-H step for the parameter m (after integrating out the effects of the nuisance parameter σ^2) with our proposal distribution equal to the birth/death process.

Although all of the estimation in this section was presented under the exact Gaussian likelihood specification, estimation can equally be conducted using the TD Whittle likelihood representation (16), subsequently presented in Section 4. Essentially what this amounts to is calculating the autocovariance structure associated with $f_{-\theta_m}$ (an inverse FEXP model having negative coefficients of the original FEXP model under consideration). This is completely analogous to the first step in our estimation of the exact Gaussian likelihood. The only benefit in using the TD Whittle representation given in (16) is that matrix inversion can be avoided.

3.3 Missing Data

One feature of our approach is that missing data can be handled with relative ease. In previous studies involving the FEXP model there has been a tendency to use Whittle approximations to the exact Gaussian likelihood. Most of these approximations do not facilitate estimation in settings where there is missing data. The reason for this is that usually estimation is based on periodogram estimates obtained from the full data. In both the TD Whittle likelihood and exact Gaussian likelihood that we consider estimation is facilitated via the autocovariance structure and not a periodogram based estimate. Therefore, we can tackle missing data with little added effort.

The method we propose only requires initial (starting) values for the missing data point (in addition to starting values from the algorithm). With these values in hand we can obtain estimates for x_{mis} based on the predictive distribution. Using obvious notation this can be expressed as

$$f(x_{mis}|x_{obs}) = \int_{\Theta} f(x_{mis}|x_{obs})p(\Theta|x_{obs})d\Theta.$$

Although we do not pursue such estimation here, this simply amounts to adding an extra step to the RJMCMC algorithm proposed in Section 3.2. That is, suppose after step t we have a sample $\tilde{\Theta}^{(t)}$ of the parameter vector. Then we may obtain a sample $x_{mis}^{(t)}$ by drawing from the distribution $f(x_{mis}|x_{obs}, \tilde{\Theta}^{(t)})$; see Palma (2007) and Gelman et al. (2003) for further details.

4 Asymptotic Properties

Here we assume the general setup of Section 2.1, so that the priors on the parameters are allowed to be more general than the specific choices of Section 2.2. The prior on m is some discrete distribution that is compactly supported, denoted by $p(m)$. We also have priors $p(b_k|\sigma^2, m)$ and $p(\sigma_k^2|m)$ for all

$k \geq 0$, but with these distributions equalling the point mass at zero if $k > m$. From these we define the joint priors $p(b|\sigma^2, m)$ and $p(\sigma^2|m)$. We also have a prior $p(d)$ on d, but this does not depend on m (this can be relaxed in the proof). So the posterior distribution of all the parameters is

$$p(b, \sigma^2, d, m|x^{(n)}) \propto p(x^{(n)}|b, \sigma^2, d, m)p(b|\sigma^2, m)p(\sigma^2|m)p(d)p(m).$$

The last four quantities on the right hand side are known priors, and the first quantity is just the likelihood function. The posterior distribution for d is given by

$$p(d|x^{(n)}) = \frac{\sum_m p(d) \int_{\mathbb{R}^{2m+2}} p(x^{(n)}|b, \sigma^2, d, m)p(b|\sigma^2, m)p(\sigma^2|m)d\nu_{m+1}(b)d\nu_{m+1}(\sigma^2)p(m)}{\sum_m \int_{\mathbb{R}^{2m+3}} p(d)p(x^{(n)}|b, \sigma^2, d, m)p(b|\sigma^2, m)p(\sigma^2|m)d\nu_1(d)d\nu_{m+1}(b)d\nu_{m+1}(\sigma^2)p(m)}. \tag{14}$$

Here $d\nu_j(x)$ denotes j-dimensional Lebesque measure. For notational convenience – since $p(b|\sigma^2, m) \times p(\sigma^2|m) = p(b, \sigma^2|m)$ anyways – let $\phi = (b, \sigma^2)$ denote the joint variable, with distribution $p(\phi|m)$. Conditional on m, ϕ takes values in $\mathbb{R}^{m+1} \times \mathbb{R}^{m+1}$ (though the variance parameters σ^2 are always non-negative). Denote this space by Φ_m, so that

$$p(d|x^{(n)}) \propto \sum_m p(d) \int_{\Phi_m} p(x^{(n)}|\phi, d, m)p(\phi|m)d\nu_{2m+2}(\phi)p(m). \tag{15}$$

As mentioned in Section 2, the posterior mean is $\hat{d} = \int_{\mathbb{R}} dp(d|x^{(n)})d\nu_1(d)$. We next present a theoretical result for the asymptotics of $\hat{d}$. In the discussion below, we replace the Gaussian likelihood $p(x^{(n)}|\theta_m)$ by the Whittle likelihood $p_w(x^{(n)}|\theta_m)$, which is an approximation. The formula is given by

$$p_w(x^{(n)}|\theta_m) = \exp\{-\frac{n}{2}\left(\log 2\pi + \frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_{\theta_m}(\lambda)d\lambda + \frac{1}{n}x^{(n)\prime}\Sigma(1/f_{\theta_m})x^{(n)}\right)\}. \tag{16}$$

The data itself is generated from a Gaussian process with spectral density given by (4), with fixed parameters $\tilde{m}$ and $\tilde{\theta}_{\tilde{m}}$ that are unknown to the practitioner (here $\tilde{\theta}_m = (\tilde{d}, \tilde{b}_0, \tilde{b}_1, \cdots, \tilde{b}_m)$ for any m). We next establish asymptotic consistency of $\hat{d}$ for the true parameter $\tilde{d}$; our method follows that of Liseo, Marinucci, and Petrella (2001), appropriately generalized to our hierarchical model.

Theorem 1 *Under the conditions discussed in this section, $\hat{d}$ converges to $\tilde{d}$ almost surely.*

Remark 2 The proof shows that $\hat{d}$ is consistent even when the true model order $\tilde{m}$ lies outside the support of the prior on m. This is because m and $(b_0, b_1, \cdots, b_m)$ are essentially nuisance parameters that have a more limited effect on the distribution of d.

5   Case Studies

In this section we illustrate our proposed Bayesian FEXP model for estimation of the long memory parameter d. We first begin with fixed dimension estimation for both simulated data and the

benchmark Nile River Minima data (Beran, 1994). Note that for the purpose of this illustration we use the exact Gaussian likelihood (5), results for the TD Whittle likelihood are similar in terms of estimation for d and are thus not pursued here. For model fitting with the simulated data we run a MCMC chain for 10,000 iterations, while for the Nile River data we run a MCMC chain for 6325 iterations, in both cases discarding the first half for burn in. The exact starting parameters and tuning parameters are detailed below. Convergence is assessed through careful monitoring of the chains along with Geweke's diagnostic as implemented in the coda contributed package in R (2007); for convenience we only display the trace plots for d. In the case of the transdimensional model, convergence is extremely difficult to assess (Sisson, 2005). Therefore we assess convergence through the trace plot of d, as the other parameters are considered nuisance parameters. Lastly, since the posterior expectation (7) is approximated by MCMC method, in actuality, we report a MCMC approximation of (15). In particular, the posterior expectation is approximated by $\hat{d} = R^{-1} \sum_{i=1}^{R} d_{(i)}$, where $d_{(i)}$ are generated from the ith MCMC sample, and R denotes the number of such samples used to estimate the parameters.

5.1 Simulation Study

To illustrate the performance of our methodology we consider a simulated example. For this simulated example we generated data from a FEXP(3) ($n = 663$) model using the Davies-Harte algorithm (Davies and Harte, 1987). The parameter values chosen for this model were obtained by taking the values of a FEXP(3) fit to the Nile River data (to be analyzed in Section 5.2) using log periodogram regression (Hurvich, 2002). The values of the parameters along with their estimated values can be found in Table 1. In conducting this analysis we chose starting values for $b_0, \ldots, b_3$ by generating from $\mathcal{N}(0,1)$ random variables, the initial value d was taken to be .25, and the initial value for all σ_k^2 was equal to 1. Finally the parameters α and β from the IG were fixed at 2.333 and 1.333 respectively. As can be seen from Table 1 and Figure 1 the estimates appear quite good. Although the point estimate for d is experiencing a slight departure from the input value to the simulation the 95% highest posterior density (hpd) covers the "true" value (Figure 1). Similarly even though our estimates for $b_0, \ldots, b_3$ are good, we emphasize that our objective here is effective estimation of d. Graphically, using the hpd - Figure 2, we see that our model is performing satisfactorily concentrating mass around the *psuedo-true* value. Furthermore, the trace plots - Figure 2, are consistent with the MCMC chains converging to their target distributions. Finally, in both the fixed and transdimensional cases we thinned the MCMC sequence, retaining every 5th value.

5.2 Nile River Data

We now illustrate our proposed methodology with the bench mark Nile River minima data (Beran, 1994). This data set contains yearly minimal water levels of the Nile river for the years 622 to 1281 ($n = 663$). Detailed description about the data is obtained from Beran (1994). The first model we estimate has m fixed a priori equal to 3. Further, the starting values for $b_0, \ldots, b_3$ and d are taken as the *psuedo-true* values and are equal to the input values of the simulated data (i.e., the log periodogram regression coefficients). The remaining initial parameters were set identical to the initial parameters from the simulated example. This analysis produces interesting results. That is, when we generate a process using coefficients equal to the log periodogram coefficients (Section 5.1) (via the Davies-Harte algorithm) we find very close agreement between $\hat{d}$ and d, with similar results for the other model parameters. However, the parameter estimates obtained in our analysis, especially for d, differ from those obtained from log periodogram regression. This leads us to believe that, for this data set (with $m = 3$), the estimate of d based on a FEXP model via log periodogram regression is perhaps overstated. However, it should be noted that this conclusion is not entirely clear cut as the hpd's from both analysis overlap slightly. In short, this phenomenon may due in part to having fixed the model order a priori. One way around this peculiarity is to let m be a model parameter to be estimated from the data and estimate d via our transdimensional approach (Section 3.2).

In order to demonstrate the utility of our transdimensional approach we re-analyzed the Nile River data using the reversible jump algorithm of Section 3.2 with $p(m) = DU(1, 6)$. As starting parameters we set $m = 3$ and let all other parameters be the same as in the fixed dimensional case (i.e. log periodogram regression). In terms of estimating d, the results obtained from this analysis are consistent with the results from the fixed dimension case, see Table 3. Furthermore, even though we chose a non-informative prior on m, we obtained parsimonious models from our MCMC samples. This aspect of our approach is notable as it exemplifies the fact that the data is driving the model rather than the prior distribution. However, in the setting where it is advantageous to incorporate prior knowledge about m into the model we can choose the Truncated Poisson prior for m. This prior imposes a penalty for selecting higher order models. Thus, our method is quite flexible and can be implemented on a broad range of long range dependent processes.

6 Discussion

The model we propose extends the current long memory literature in several ways. First we provide a Bayesian estimate of the FEXP model based on the exact Gaussian likelihood. Additionally we provide the details for algorithms that facilitate quick computation. Moreover, our approach uses a Whittle formulation based on a time domain estimate of the quadratic form in the likelihood, thus

dispensing with the need for calculating periodogram estimates. Further, we allow the dimension of our parameter space to take the form of a random parameter to be estimated in the model. Therefore, the method we propose is hierarchical and integrates over *all* possible models, thus reducing underestimation of uncertainty at the model-selection stage.

The hierarchical structure we propose can be implemented with two different goals in mind. First, if the goal of the analysis is strictly estimation of d, our method can be viewed as Bayesian model averaging (Hoeting et al. 1999). In this case our estimate of d results in a weighted average of the estimates of d from our adaptive model, with the weights proportional to the probability of occurrence of the different selected model orders. This approach improves the precision of our estimate in contexts where the "true" model order is unknown a priori. Second, if the goal of the analysis is to produce spectral model then one can appeal to the FEXP($\widehat{m}$) model, where $\widehat{m}$ denotes the posterior mode of m. With this model one can also obtain an equivalent AR representation using the formulas in Hurvich (2002). Since, the goal of our procedure is mainly estimation of d, the long memory parameter, we do not pursue detailed discussion of this model usage here. However, both of these uses constitute novel contributions to the current state of FEXP modelling.

Another striking distinction between our FEXP models and most of the methods currently being implemented for estimation of the long memory parameter is that by using a Bayesian approach we gain more information about d, since we can obtain the whole posterior distribution rather than just a point estimate and standard error. This is something log periodogram regression and GPH (Geweke, and Porter-Hudak, 1983) don't achieve. This distinction is fundamental and is often one of the advantages of appealing to the Bayesian paradigm. One illustration of the benefits of having a posterior density of d arises in Section 5.2 were we analyze the Nile River minima data. In this setting, examining the posterior density of d (Figure 1), we notice that although we started from a noninformative prior on d we are able to extract the embedded information about d from the likelihood. That is, by taking advantage of the Bayesian machinary we ended up with a highly informative posterior. Therefore having this posterior density definitely provides an added boost to the performance of our estimate of d.

Additionally, we establish Bayesian consistency of the memory parameter under mild conditions on the data process. This poses a non-trivial extension to the proof presented in Liseo et al. (2001). Although our asymptotics rely on the Whittle formulation (16), this quantity can be calculated with the same level of effort as the exact Gaussian likelihood and provides comparable results. Therefore, we do not consider this a draw-back to our method. Rather, we view this as a choice to be made by the practitioner, weighing the benefits of having an *exact* likelihood estimate versus an estimate with proven optimal asymptotic behavior. We believe that this should ultimately depend on the goals of the analysis.

In summary, we have developed a flexible modeling approach for Gaussian long range dependent processes. The algorithms we provide are computationally straight forward and relieve the

practitioner from the burden of order selection when the main focus is estimation of the memory parameter. Lastly, we develop asymptotic properties of our estimator under mild conditions on the data process.

Appendix A - Asymptotics

Proof of Theorem 1. First note that we have

$$\hat{d} - \tilde{d} = \frac{\sum_m \int_{\mathbb{R} \times \Phi_m} (d - \tilde{d}) p_w(x^{(n)} | \phi, d, m) p(\phi|m) p(d) d\nu_{2m+2}(\phi) d\nu_1(d) p(m)}{\sum_m \int_{\mathbb{R} \times \Phi_m} p_w(x^{(n)} | \phi, d, m) p(\phi|m) p(d) d\nu_{2m+2}(\phi) d\nu_1(d) p(m)}.$$

Note that we have substituted the Whittle likelihood into this expression. The quadratic form in the Whittle likelihood can be written as

$$\frac{1}{n} x^{(n)'} \Sigma(1/f_{\theta_m}) x^{(n)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_n(\lambda)}{f_{\theta_m}(\lambda)} \, d\lambda,$$

where $I_n(\lambda) = n^{-1} |\sum_{t=1}^{n} x_t e^{-i\lambda t}|^2$ is the periodogram. Adapting Lemma 1 of Fox and Taqqu (1986) to handle the case of an uncentered periodogram, we have almost sure convergence of the quadratic form to

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_{\tilde{\theta}_{\tilde{m}}}(\lambda)}{f_{\theta_m}(\lambda)} \, d\lambda$$

uniformly in θ_m (for each $m \geq 1$ fixed). In order for this result to be true, we require that $f_{\theta_{\tilde{m}}}(\lambda)$ be continuous at all $(\lambda, \theta_{\tilde{m}})$, $\lambda \neq 0$; $1/f_{\theta_{\tilde{m}}}(\lambda)$ is continuous at all $(\lambda, \theta_{\tilde{m}})$; $f_{\theta_{\tilde{m}}}(\lambda) = O(|\lambda|^{-(\alpha(\theta_{\tilde{m}})+\delta)})$ as $\lambda \to 0$, where $0 < \alpha(\theta_{\tilde{m}}) < 1$ and $\delta > 0$; $\frac{\partial}{\partial \lambda} f_{\theta_{\tilde{m}}}(\lambda)$ is continuous at all $(\lambda, \theta_{\tilde{m}})$, $\lambda \neq 0$; $\frac{\partial}{\partial \lambda} f_{\theta_{\tilde{m}}}(\lambda) = O(|\lambda|^{-(\alpha(\theta_{\tilde{m}})+1+\delta)})$ as $\lambda \to 0$; and $1/f_{\theta_m}(\lambda)$ is continuous on $[-\pi, \pi] \times \Theta_m$ for each $m \geq 1$. It is straightforward to see that all of these conditions are met by our specification of f in (4). Next, let

$$K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_{\tilde{\theta}_{\tilde{m}}}(\lambda)}{f_{\theta_m}(\lambda)} - \log \frac{f_{\tilde{\theta}_{\tilde{m}}}(\lambda)}{f_{\theta_m}(\lambda)} \right) \, d\lambda.$$

Then we have

$$p_w(x^{(n)} | \theta_m) = \exp\{-\frac{n}{2} \left(\log 2\pi - \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_{\tilde{\theta}_{\tilde{m}}}(\lambda) \, d\lambda + K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) + R_n(\theta_m) \right)\},$$

where $R_n(\theta_m)$ is a remainder term that tends to zero a.s., uniformly in θ_m. Now the first two terms above do not depend upon the variable θ_m, nor upon m either. Thus

$$\hat{d} - \tilde{d} = \frac{\sum_m \int_{\mathbb{R} \times \Phi_m} (d - \tilde{d}) \exp\{-\frac{n}{2} \left(K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) + R_n(\theta_m) \right)\} p(\phi|m) p(d) d\nu_{2m+2}(\phi) d\nu_1(d) p(m)}{\sum_m \int_{\mathbb{R} \times \Phi_m} \exp\{-\frac{n}{2} \left(K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) + R_n(\theta_m) \right)\} p(\phi|m) p(d) d\nu_{2m+2}(\phi) d\nu_1(d) p(m)},$$

noting that $\int_{-\pi}^{\pi} \log f_{\tilde{\theta}_{\tilde{m}}}(\lambda) \, d\lambda$ does not depend upon m. Since the exponential integrands do not depend on σ^2, we can effectively integrate out these nuisance parameters so that $p(\phi|m)$ reduces

17

to $p(b|m)$. Note that $p(b|m)$ cannot be computed in an implementation – since only $p(b|\sigma^2, m)$ is defined – but for the purposes of our proof, this reduction simplifies the notation. Letting $p(b|m)p(d) = p(\theta_m)$, we obtain

$$\hat{d} - \tilde{d} = \frac{\sum_m \int_{\Theta_m} (d - \tilde{d}) \exp\{-\frac{n}{2}\left(K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) + R_n(\theta_m)\right)\} p(\theta_m) d\nu_{m+2}(\theta_m) p(m)}{\sum_m \int_{\Theta_m} \exp\{-\frac{n}{2}\left(K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) + R_n(\theta_m)\right)\} p(\theta_m) d\nu_{m+2}(\theta_m) p(m)}.$$

Now let $\epsilon > 0$ and for each $m \geq 1$ let $\epsilon_m = \epsilon 2^{-m}$, and consider the set $\{\theta_m \in \Theta_m : \|\theta_m - \tilde{\theta}_m\| < \epsilon_m\}$. This is an ϵ_m-neighborhood – with respect to the Euclidean norm $\|\cdot\|$ – of $\tilde{\theta}_m$, which is the vector consisting of the first $m + 2$ correct parameters (but if $m > \tilde{m}$, pad out with zeroes). Denote this set by $N_{\epsilon_m}(\tilde{\theta}_m)$; we split the numerator integral in the expression for $\hat{d} - \tilde{d}$ into an integration over $N_{\epsilon_m}(\tilde{\theta}_m)$ and its complement. In the former term, we obtain the strict upper bound of $\sum_{m \geq 1} \epsilon_m = \epsilon$, using $|d - \tilde{d}| \leq \|\theta_m - \tilde{\theta}_m\| \leq \epsilon_m$. The latter term is bounded by

$$\frac{\sum_m \int_{N^c_{\epsilon_m}(\tilde{\theta}_m)} (d - \tilde{d}) \exp\{-\frac{n}{2}\left(K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) + R_n(\theta_m)\right)\} p(\theta_m) d\nu_{m+2}(\theta_m) p(m)}{\sum_m \int_{N_{\epsilon_m/2}(\tilde{\theta}_m)} \exp\{-\frac{n}{2}\left(K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) + R_n(\theta_m)\right)\} p(\theta_m) d\nu_{m+2}(\theta_m) p(m)}. \tag{A.1}$$

Here we have made the denominator smaller by restricting the region of integration. The function $K_{\tilde{\theta}_{\tilde{m}}}$ is interpreted as a sequence of functions depending on m; given m, it is a real-valued function defined on Θ_m. On this domain it is strictly convex, with a minimum value of $1/2$ (misstated in Liseo et al. (2001)) occurring at $\tilde{\theta}_m$ for each m. To demonstrate this, fix m, and let $g(y) = y - \log y$. This function satisfies $\dot{g}(y) = 1 - 1/y$ and $\ddot{g}(y) = 1/y^2$, so that $y = 1$ is the global minimizer. Clearly $K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) = (2\pi)^{-1} \int_{-\pi}^{\pi} g\left(f_{\tilde{\theta}_{\tilde{m}}}(\lambda)/f_{\theta_m}(\lambda)\right) d\lambda$, and its derivatives are

$$\frac{\partial}{\partial \theta_m^j} K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \dot{g}\left(\frac{f_{\tilde{\theta}_{\tilde{m}}}(\lambda)}{f_{\theta_m}(\lambda)}\right) \left[-\frac{f_{\tilde{\theta}_{\tilde{m}}}(\lambda)}{f_{\theta_m}^2(\lambda)} \frac{\partial}{\partial \theta_m^j} f_{\theta_m}(\lambda)\right] d\lambda$$

$$\frac{\partial^2}{\partial \theta_m^j \partial \theta_m^k} K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ddot{g}\left(\frac{f_{\tilde{\theta}_{\tilde{m}}}(\lambda)}{f_{\theta_m}(\lambda)}\right) \left[-\frac{f_{\tilde{\theta}_{\tilde{m}}}(\lambda)}{f_{\theta_m}^2(\lambda)}\right]^2 \frac{\partial}{\partial \theta_m^j} f_{\theta_m}(\lambda) \frac{\partial}{\partial \theta_m^k} f_{\theta_m}(\lambda) d\lambda$$

$$+ \frac{1}{2\pi} \int_{-\pi}^{\pi} \dot{g}\left(\frac{f_{\tilde{\theta}_{\tilde{m}}}(\lambda)}{f_{\theta_m}(\lambda)}\right) \left[\frac{2f_{\tilde{\theta}_{\tilde{m}}}(\lambda)}{f_{\theta_m}^3(\lambda)} \left(\frac{\partial}{\partial \theta_m^j} f_{\theta_m}(\lambda) \frac{\partial}{\partial \theta_m^k} f_{\theta_m}(\lambda)\right) - \frac{f_{\tilde{\theta}_{\tilde{m}}}(\lambda)}{f_{\theta_m}^2(\lambda)} \frac{\partial^2}{\partial \theta_m^j \partial \theta_m^k} f_{\theta_m}(\lambda)\right] d\lambda.$$

In the case that $\tilde{m} \leq m$, then we have $\tilde{\theta}_m = (\tilde{\theta}_{\tilde{m}}, 0, \cdots, 0)$ where there are $m - \tilde{m}$ zeroes. Then $f_{\tilde{\theta}_{\tilde{m}}}(\lambda) = f_{\tilde{\theta}_m}(\lambda)$ and the first derivatives of $K_{\tilde{\theta}_{\tilde{m}}}$ are zero at $\tilde{\theta}_m$; moreover the second derivatives satisfy

$$\frac{\partial^2}{\partial^2 \theta_m^j} K_{\tilde{\theta}_{\tilde{m}}}(\tilde{\theta}_m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{\tilde{\theta}_m}^{-2}(\lambda) \left(\frac{\partial}{\partial \theta_m^j} f_{\tilde{\theta}_m}(\lambda)\right) d\lambda,$$

which is strictly positive. Thus $\tilde{\theta}_m$ is a local minimum. More generally (say when m may be less than $\tilde{m}$), the function $K_{\tilde{\theta}_{\tilde{m}}}(\theta_m)$ is convex, which is shown as follows. Using the form (4) for f, we

18

have

$$K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\exp\{-(\tilde{d}-d)\log|1-e^{-i\lambda}|^2 + \sum_{l=0}^{m}(\tilde{b}_l-b_l)\cos\lambda l + \sum_{l=m+1}^{\tilde{m}}\tilde{b}_l\cos\lambda l\}$$
$$+ \left((\tilde{d}-d)\log|1-e^{-i\lambda}|^2 - \sum_{l=0}^{m}(\tilde{b}_l-b_l)\cos\lambda l - \sum_{l=m+1}^{\tilde{m}}\tilde{b}_l\cos\lambda l\right)d\lambda.$$

This assumes that $\tilde{m} > m$, but the formula is easily adapted to handle $\tilde{m} \leq m$ in an obvious manner. Now in the Taylor Series expansion of $\log(1-e^{-i\lambda})$ there are no constant terms, and thus the integral is zero; hence we have

$$K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\exp\{-(\tilde{d}-d)\log|1-e^{-i\lambda}|^2 + \sum_{l=0}^{m}(\tilde{b}_l-b_l)\cos\lambda l + \sum_{l=m+1}^{\tilde{m}}\tilde{b}_l\cos\lambda l\} - \left(\tilde{b}_0 - b_0\right)d\lambda.$$

Using the convexity of the exponential function, it is now straightforward to verify the convexity of $K_{\tilde{\theta}_{\tilde{m}}}$. Now returning to the case that $\tilde{m} \leq m$, we know that the minimum value of $K_{\tilde{\theta}_{\tilde{m}}}$ is achieved at $\tilde{\theta}_m$, and the value is $1/2 > 0$; therefore we can find positive constants $c(\epsilon_m/2)$ and $C(\epsilon_m)$ such that

$$c(\epsilon_m/2) = \sup_{\theta_m\in N_{\epsilon_m/2}(\tilde{\theta}_m)} K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) < \inf_{\theta_m\in N^c_{\epsilon_m}(\tilde{\theta}_m)} K_{\tilde{\theta}_{\tilde{m}}}(\theta_m) = C(\epsilon_m). \tag{A.2}$$

But when $\tilde{m} > m$, we have

$$K_{\tilde{\theta}_{\tilde{m}}}(\tilde{\theta}_m) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\exp\{\sum_{l=m+1}^{\tilde{m}}\tilde{b}_l\cos\lambda l\}\,d\lambda,$$

which is strictly positive. Hence, using the strict convexity, in this case as well we can find positive constants $c(\epsilon_m/2)$ and $C(\epsilon_m)$ satisfying (A.2), making ϵ smaller if necessary (we may need to contract each ϵ_m in order to ensure that $c(\epsilon_m/2) > 0$, but there are only finitely many such m's to consider).

Next we consider the remainder terms $R_n(\theta_m)$. For each m fixed, let Ω_m be the subset of the full probability space Ω such that $R_n(\theta_m)(\omega) \to 0$ uniformly in θ_m for every $\omega \in \Omega_m$, with $\mathbb{P}[\Omega_m] = 1$. Now for any ω, we have $R_n(\theta_{m+1})(\omega) \to 0$ implies $R_n(\theta_m)(\omega) \to 0$, since in the former expression we can just set $g_{m+1} = 0$ to obtain the latter. Thus $\Omega_{m+1} \subset \Omega_m$, and $\mathbb{P}[\bigcap_{m\geq 0}\Omega_m] = \lim_{m\to\infty}\mathbb{P}[\Omega_m] = 1$, using continuity from below of the probability measure $\mathbb{P}$. So on $\bigcap_{m\geq 0}\Omega_m$, a set of measure one, we have $R_n(\theta_m) \to 0$ for all m. Then for any $\delta > 0$, there exists N such that $|R_n(\theta_m)| < \delta$ for all m and every $n \geq N$. Applying these results to (A.1), we obtain the bound of

$$\frac{\sum_m \int_{N^c_{\epsilon_m}(\tilde{\theta}_m)}(d-\tilde{d})p(\theta_m)d\nu_{m+2}(\theta_m)\exp\{-\frac{n}{2}(C(\epsilon_m)-\delta)\}p(m)}{\sum_m \int_{N_{\epsilon_m/2}(\tilde{\theta}_m)}p(\theta_m)d\nu_{m+2}(\theta_m)\exp\{-\frac{n}{2}(c(\epsilon_m/2)+\delta)\}p(m)}$$
$$\leq \mathbb{E}|d-\tilde{d}|\frac{\sum_m \exp\{-\frac{n}{2}(C(\epsilon_m)-\delta)\}p(m)}{\sum_m \int_{N_{\epsilon_m/2}(\tilde{\theta}_m)}p(\theta_m)d\nu_{m+2}(\theta_m)\exp\{-\frac{n}{2}(c(\epsilon_m/2)-\delta)\}p(m)}.$$

Since only finitely-many values of m have $p(m) > 0$, we can find an $\epsilon^* < \epsilon_m$ for all such m's, such that $C(\epsilon^*) < C(\epsilon_m)$ and $c(\epsilon^*/2) > c(\epsilon_m/2)$ but with $c(\epsilon^*/2) < C(\epsilon^*)$. Then the bound becomes

$$\mathbb{E}|d - \tilde{d}| \frac{\sum_m p(m) \exp\{-\frac{n}{2}\left(C(\epsilon^*) - c(\epsilon^*/2) - 2\delta\right)\}}{\sum_m \int_{N_{\epsilon_m/2}(\tilde{\theta}_m)} p(\theta_m) d\nu_{m+2}(\theta_m) p(m)}.$$

So we choose $\delta < (C(\epsilon^*) - c(\epsilon^*/2))/2$. The denominator in the above formula does not depend upon n, so the bound tends to zero as n tends to ∞ on the set $\bigcap_{m \geq 0} \Omega_m$, which establishes strong consistency. $\quad \square$

Appendix B - Full Conditionals

$$\begin{aligned}
p(m|b, \sigma^2, d, x^{(n)}) \;&\propto\; p(x^{(n)}|m, b, \sigma^2, d)p(m|b, \sigma^2, d) \\
&\propto\; p(x^{(n)}|m, b, \sigma^2, d)p(b|m, \sigma^2, d)p(m|\sigma^2, d) \\
&\propto\; p(x^{(n)}|m, b, \sigma^2, d)p(b|m, \sigma^2)p(\sigma^2|m, d)p(m|d) \\
&\propto\; p(x^{(n)}|m, b, \sigma^2, d)p(b|m, \sigma^2)p(\sigma^2|m)p(m) \\
&\propto\; (2\pi)^{-n/2} |\Sigma(f_{\theta_m})|^{-1/2} \exp\left\{-\frac{1}{2}x^{(n)\prime}\Sigma^{-1}(f_{\theta_m})x^{(n)}\right\} \\
&\times\; \prod_{k=0}^{m}(\sigma_k^2)^{-1/2} \exp\left\{-\frac{b_k^2}{2\sigma_k^2}\right\} \\
&\times\; \prod_{k=0}^{m} \exp\left\{-\frac{\beta}{\sigma_k^2}\right\} (\sigma_k^2)^{-(\alpha+1)} \\
&\times\; \frac{1}{M}.
\end{aligned}$$

$$\begin{aligned}
p(\sigma_k^2|d, m, b_0, b_1, \ldots, b_k, \ldots, b_m) \;&\propto\; p(b_k|\sigma_k^2)p(\sigma_k^2) \\
&\propto\; (\sigma_k^2)^{-1/2} \exp\left\{-\frac{b_k^2}{2\sigma_k^2}\right\} \exp\left\{-\frac{\beta}{\sigma_k^2}\right\} (\sigma_k^2)^{-(\alpha+1)} \\
&=\; (\sigma_k^2)^{-(\alpha+1/2+1)} \exp\left\{-\frac{b_k^2}{2\sigma_k^2} - \frac{\beta}{\sigma_k^2}\right\} \\
&=\; (\sigma_k^2)^{-(\alpha+1/2+1)} \exp\left\{-\frac{(b_k^2/2 + \beta)}{\sigma_k^2}\right\} \\
&\sim\; IG\left\{(\alpha + 1/2), (b_k^2/2 + \beta)\right\}.
\end{aligned}$$

$$\begin{aligned}
p(b|m, \sigma^2, d, x^{(n)}) \;&\propto\; p(x^{(n)}|b, m, \sigma^2, d)p(b|m, \sigma^2, d) \\
&\propto\; (2\pi)^{-n/2} |\Sigma(f_{\theta_m})|^{-1/2} \exp\left\{-\frac{1}{2}x^{(n)\prime}\Sigma^{-1}(f_{\theta_m})x^{(n)}\right\} \\
&\times\; \prod_{k=0}^{m}(\sigma_k^2)^{-1/2} \exp\left\{-\frac{b_k^2}{2\sigma_k^2}\right\}.
\end{aligned}$$

$$
\begin{aligned}
p(d|\sigma^2, m, b, x^{(n)}) &\propto p(x^{(n)}|b, m, \sigma^2, d)p(d|\sigma^2, m, b) \\
&\propto (2\pi)^{-n/2} |\Sigma(f_{\theta_m})|^{-1/2} \exp\left\{ -\frac{1}{2} x^{(n)\prime} \Sigma^{-1}(f_{\theta_m}) x^{(n)} \right\} \\
&\times 1_{[0,1/2)}(d).
\end{aligned}
$$

Appendix C - Bayes Factor and BIC

Here we provide justification, in our context, for the reversible jump algorithm as a consistent method of model selection within each repetition of the chain. First let Q denote $BF(\Theta', \Theta^{(}t)) \times R$ where R and BF are defined similar to (12) and (13) respectively. First we consider the numerator of (13)

$$
\int_{\sigma^2} p(\Theta'|x^{(n)})d\sigma^2 = L(\Theta'|x^{(n)}) \frac{\prod_{i=1}^{m'} \Gamma(\alpha + 1/2)}{\prod_{i=1}^{m'} (\frac{b_i^2}{2} + \beta)^{\alpha+1/2}}. \tag{C.1}
$$

Similarly it follows that

$$
\int_{\sigma^2} p(\Theta^{(t)}|x^{(n)})d\sigma^2 = L(\Theta^{(t)}|x^{(n)}) \frac{\prod_{i=1}^{m} \Gamma(\alpha + 1/2)}{\prod_{i=1}^{m} (\frac{b_i^2}{2} + \beta)^{\alpha+1/2}}. \tag{C.2}
$$

Taking the logarithms of (C.1) yields the following expression

$$
\begin{aligned}
\log\left\{ \int_{\sigma^2} p(\Theta'|x^{(n)})d\sigma^2 \right\} &= \log\left\{ L(\Theta'|x^{(n)}) \right\} + m' \log\left\{ \Gamma(\alpha + 1/2) \right\} \\
&- (\alpha + 1/2) \sum_{i=1}^{m'} \log\left(\frac{b_i^2}{2} + \beta \right),
\end{aligned} \tag{C.3}
$$

with a similar expression obtained for (C.2). Therefore $\log\{BF(\Theta', \Theta^{(t)})\}$ can be expressed as

$$
\begin{aligned}
\log\{BF(\Theta', \Theta^{(t)})\} &= \left[\log\left\{ L(\Theta'|x^{(n)}) \right\} - \log\left\{ L(\Theta'|x^{(n)}) \right\} \right] + (m' - m) \log \Gamma(\alpha + 1/2) \\
&- (\alpha - 1/2) \left\{ \sum_{i+1}^{m'} \log\left(\frac{b_i^2}{2} + \beta \right) - \sum_{i+1}^{m} \log\left(\frac{b_i^2}{2} + \beta \right) \right\}.
\end{aligned} \tag{C.4}
$$

Now consider the Bayesian information criterion (BIC) of Θ' and $\Theta^{(t)}$ respectively,

$$
\text{BIC}(\Theta') = -2\log\left\{ L(\Theta'|x^{(n)}) \right\} + 2p' \log(n),
$$

and

$$
\text{BIC}(\Theta^{(t)}) = -2\log\left\{ L(\Theta^{(t)}|x^{(n)}) \right\} + 2p \log(n),
$$

where $p' = 2m' + 1$ and $p = 2m + 1$, the number of parameters in the respective models. Now from the expressions above it follows that

$$\text{BIC}(\Theta') - \text{BIC}(\Theta^{(t)}) \approx -2\log\left\{BF(\Theta', \Theta^{(t)})\right\}.$$

Thus,

$$
\begin{aligned}
\text{BIC}(\Theta') - \text{BIC}(\Theta^{(t)}) &= -2\log\left\{L(\Theta^{(t)}|x^{(n)})\right\} + 2p'\log(n) + 2\log\left\{L(\Theta^{(t)}|x^{(n)})\right\} - 2p\log(n) \\
&= -2\left[\log\left\{BF(\Theta', \Theta^{(t)})\right\} - (m' - m)\log\Gamma(\alpha + 1/2)\right. \\
&\quad + (\alpha + 1/2)\left\{\sum_{i=1}^{m'}\log\left(\frac{b_i^2}{2} + \beta\right) - \sum_{i=1}^{m}\log\left(\frac{b_i^2}{2} + \beta\right)\right\} \\
&\quad \left. - 2(m' - m)\log\int_{\sigma^2} p(\Theta'|x^{(n)})d\sigma^2\right].
\end{aligned}
$$

Now by the Laplace approximation (Robert and Casella, 2004) it follows that

$$
\begin{aligned}
\log\left\{\int_{\Theta_{-m}} p(\Theta'|x^{(n)})d\Theta\right\} &\simeq L(\Theta'|x^{(n)})\frac{\prod_{i=1}^{m'}\Gamma(\alpha + 1/2)}{\prod_{i=1}^{m'}(\frac{b_i^2}{2} + \beta)^{\alpha + 1/2}} \\
&= \log L(\Theta'|\hat{\theta}, m) - \frac{p'}{2}\log(n) + O(1).
\end{aligned}
$$

Similarly

$$
\begin{aligned}
\log\left\{\int_{\Theta_{-m}} p(\Theta^{(t)}|x^{(n)})d\Theta\right\} &\simeq L(\Theta^{(t)}|x^{(n)})\frac{\prod_{i=1}^{m}\Gamma(\alpha + 1/2)}{\prod_{i=1}^{m}(\frac{b_i^2}{2} + \beta)^{\alpha + 1/2}} \\
&= \log L(\Theta^{(t)}|\hat{\theta}^t, m) - \frac{p}{2}\log(n) + O(1),
\end{aligned}
$$

where $\hat{\theta}$ and $\hat{\theta}^t$ are maximum likelihood estimators. Thus we have

$$-2\log\left\{BF(\Theta', \Theta^{(t)})\right\} = -2\log\left[\int_{\sigma^2} p(\Theta'|x^{(n)})d\sigma^2 - \int_{\sigma^2} p(\Theta^{(t)}|x^{(n)})d\sigma^2\right] \approx \text{BIC}(\Theta') - \text{BIC}(\Theta^{(t)}).$$

Since BIC is asymptotically consistent as an order selection criterion so are the BF's.

References

[1] Beran, J. (1993) Fitting long-memory models by generalized linear regression. *Biometrika*, **80**, 817-822.

[2] Beran, J. (1994) *Statistics for Long Memory Processes*. New York: Chapman and Hall.

[3] Bertelli, S. and Caporin, M. (2002) A note on calculating autocovariances of long-memory processes. *Journal of Time Series Analysis*, **23**, 503-508.

[4] Bloomfield, P. (1973) An exponential model for the spectrum of a scalar time series. *Biometrika*, **60**, 217-226.

[5] hen, W., Hurvich, C., Lu, Y. (2006) On the correlation matrix of the discrete Fourier transform and fast solution of large Toeplitz systems for long-memory time series. *Journal of the American Statistical Association*, **101** 812-822

[6] Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings Algorithm. *American Statistician*, **49**, 327–335.

[7] Davies, R. B., Harte, D. S. (1987) Tests for Hurst effect. *Biometrika*, **74**, 95-101.

[8] Denison, D.G.T., Mallick, B.K., and Smith, A.F.M. (1998) Automatic Bayesian curve fitting. *Journal of the Royal Statistical Society B*, **60**, 333-350.

[9] Denison, D.G.T., Mallick, B.K., Holmes, C.C., and Smith, A.F.M. (2002) *Bayesian Methods for Nonlinear Classification and Regression.* New York: Wiley.

[10] Fox, R. and Taqqu, M. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, **14**, 517-532.

[11] Gelfand, A. and Smith, A. F. M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 385-409

[12] Gelman, A., Carlin, J.B., Stern, H., and Rubin, D. (2003) *Bayesian Data Analysis: Second Ed.* New York: Chapman & Hall/CRC

[13] Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221-238

[14] Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221-238.

[15] Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (ed JM Bernardo, JO Berger, AP Dawid, and AFM Smith). Clarendon Press, Oxford, U.K.

[16] Gilks, W. R. and Wild, P. (1992) Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, **41**, 337-348.

[17] Giraitis, L.. and Surgailis, D. (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields*, **86**, 87-104.

[18] Giraitis, L. and Taqqu, M. (1999) Whittle estimator for finite-variance non-Gaussian time series with long memory. *The Annals of Statistics*, **27**, 178-203.

[19] Green, P.J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711-732.

[20] Hoeting, J.A., Madigan, D., Rafferty, A.E., and Volinsky, C.T. (1999) Bayesian model averaging: a tutorial (with Discussion). *Statistical Science*, **14**, 382-417.

[21] Hosking, J.R.M. (1981) Fractional differencing. *Biometrika*, **68**, 165-176.

[22] Hurvich, C. (2001) Model selection for broadband semiparametric estimation of long memory time series. *Journal of Time Series Analysis*, **22**, 679-709.

[23] Hurvich, C. (2002) Multistep Forecasting of long memory series using fractional exponential models. *International Journal of Forecasting*, **18**, 167-179.

[24] Janacek, G. J. (1982) Determining the degree of differencing for time series via the log spectrum. *Journal of Time Series Analysis*, **3**, 177-183.

[25] Ko, K., Vannucci, M. (2006) Bayesian wavelet analysis of autregressive fractionally integrated moving-average processes. *Journal of Statistical Planning and Inference*, **136**, 3415-3434

[26] Liseo, B., Marinucci, D., and Petrella, L. (2001) Bayesian semiparametric inference on long-range dependence. *Biometrika*, **88**, 1089-1104

[27] Pourahmadi, M. (1983) Exact factorization of the spectral density and its application to forecasting and time series analysis. *Communications in Statistics: Theory and Methods*, **12**, 2085-2094.

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

[29] Palma, W. (2007) *Long-Memory Time Series*, New York: John Wiley and Sons.

[30] Petris, G. (1997) *Bayesian Analysis of Long Memory Time Series*, Unpublished Ph.D. Thesis

[31] R Development Core Team (2007) R: A Language and Environment for Statistical Computing, http://www.R-project.org.

[32] Ravishanker, N., and Ray, B. K. (1997) Bayesian analysis of vector ARFIMA processes. *Australian Journal of Statistics*, **39**, 295-311.

[33] Robert, C., and Casella, G. (2004) *Monte Carlo Statistical Methods*, New York: Springer-Verlag.

[34] Robinson, P. M. (1994) Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association*, **89**, 1420-1437 .

[35] Robinson, P. M. (1995a) Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, **23**, 1047-1072 .

[36] Robinson, P. M. (1995b) Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, **23**, 1630-1661 .

[37] Robinson, P. M. (Ed.) (2003) *Time Series With Long Memory*, Oxford: Oxford University Press.

[38] Sisson, S. A. (2005) Transdimensional Markov chains: A decade of progress and future perspectives *Journal of the American Statistical Association*, **100**, 1077-1089.

Posterior Results: Simulated Data - Thin = 5 (simulated)					
Param.	"Truth"	Mean	Med.	Stdev	95% HPD
b_0	6.640	6.615	6.615	.061	(6.504, 6.741)
b_1	-.121	-.057	-.059	.094	(-.230,.135)
b_2	-.232	-.156	-.157	.079	(-.301,.000)
b_3	-.044	.016	.015	.079	(-.141,.157)
d	.479	.453	.455	.026	(.406,.498)

Table 1: Posterior for simulated data, thinning imposed = 5. Thinning with 10 and no thinning were also explored with no appreciable differences. Note $n = 663$ and number of MCMC iterations equaled 10,000 with first half being allocated toward burn in. m was chosen a priori equal to 3 for the purpose of illustration.

Posterior Results: Nile River Data - Thin = 5				
Param.	Mean	Med.	Stdev	95% HPD
b_0	6.659	6.655	.060	(6.542, 6.781)
b_1	.096	-.088	.149	(-.153,.392)
b_2	-.070	-.072	.097	(-.266,.114)
b_3	.008	.003	.088	(-.140,.186)
d	.374	.380	.061	(.248,.469)

Table 2: Posterior for Nile River Data, thinning imposed = 5. Thinning with 10 and no thinning were also explored with no appreciable differences. Note $n = 663$ and number of MCMC iterations equaled 6325 with first half being allocated toward burn in. m was chosen a priori equal to 3 for the purpose of illustration.
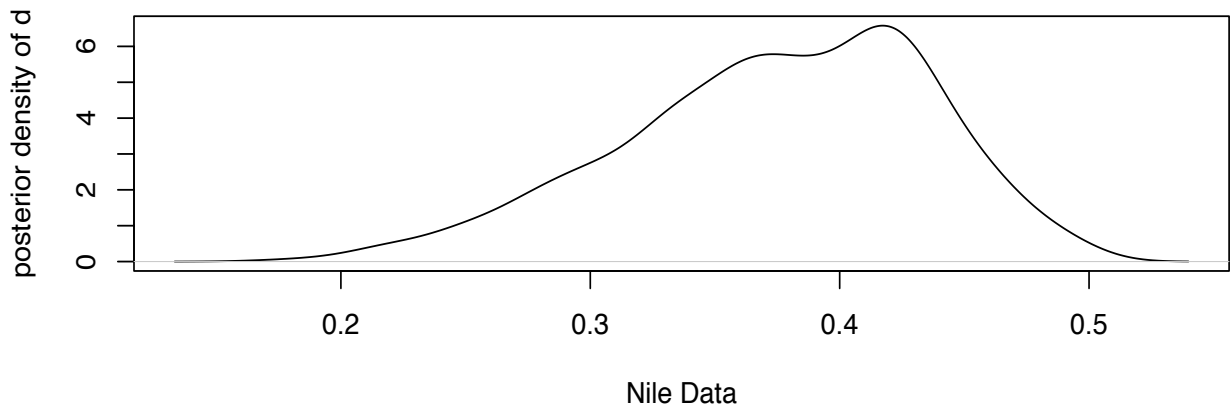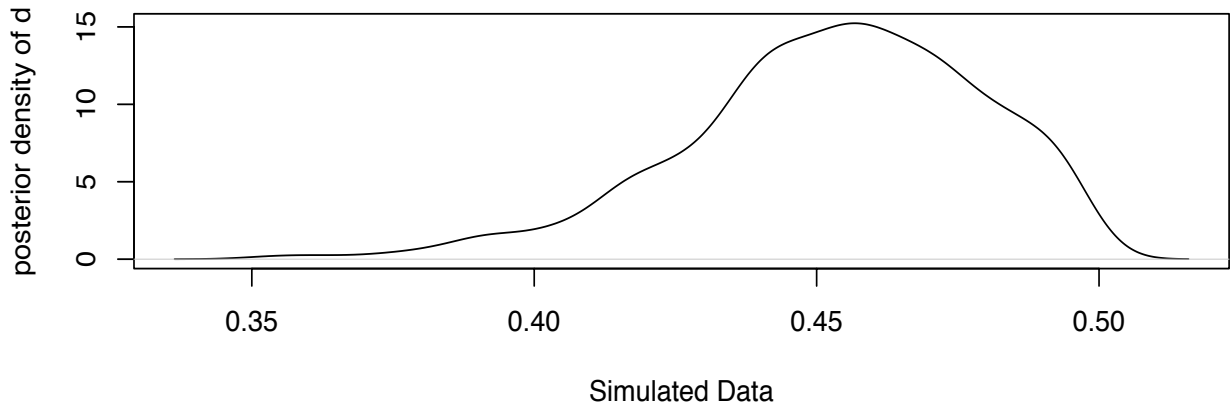
Figure 1: 95% HPD - d for both the simulated data and Nile River analysis. Note that this plot was constructed using a default kernel smoother with the bandwidth determined by Nadaraya Watson
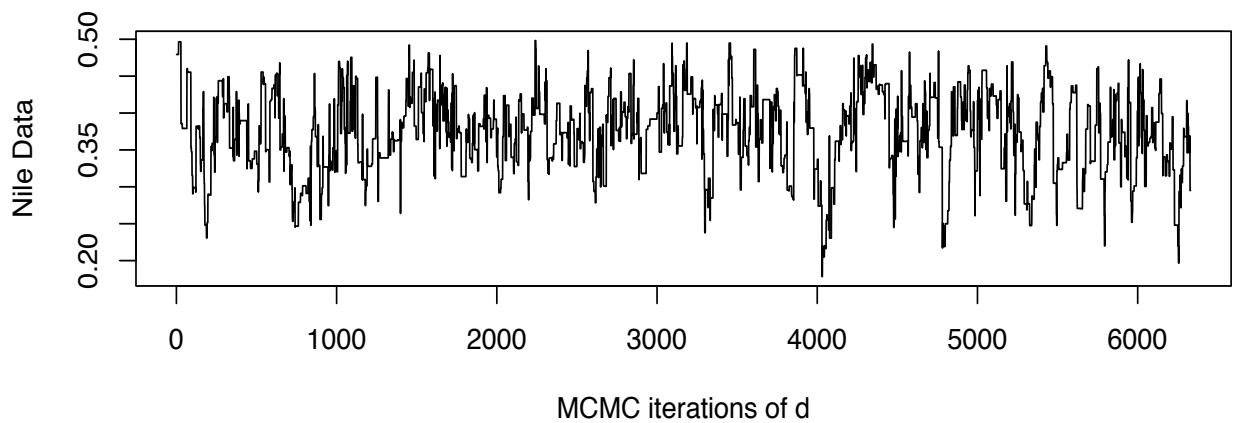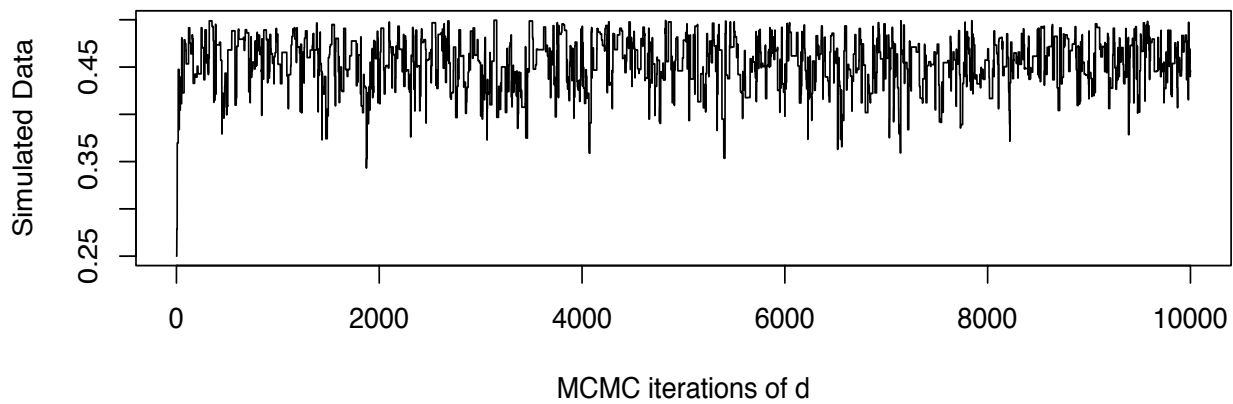
Figure 2: Trace Plots - d for both the simulated data and Nile River analysis. Note that these plots are consistent with the MCMC chains converging to their target stationary distributions.