NONINFORMATIVE PRIORS, BAYES ESTIMATION
AND FORECASTING FOR LONG MEMORY PROCESSES

by

Chuanbo Zang and Anirban DasGupta

Technical Report #99-13

Department of Statistics
Purdue University
West Lafayette, IN USA

July 1999

*

Noninformative Priors, Bayes Estimation and Forecasting for Long Memory Processes *

Chuanbo Zang† Anirban DasGupta

Department of Statistics
Purdue University
West Lafayette, IN 47907

Abstract

With the discovery of long range dependence in many types of real data, estimation of the long memory parameter has become an important and challenging problem in time series analysis. Recent work has concentrated on semi and nonparametric estimation and the associated asymptotics.

In this article, we first give a simple closed form approximation to the Jeffrey prior for the parameter vector $\theta = (d, \mu, \sigma)$ of a fractional ARIMA(0,d,0) process. The approximation is based on the familiar Whittle approximation of Σ^{-1} and has a simple interpretation. It is also amenable to further inference and computation. The approximate Jeffrey prior and a natural uniform prior are used to construct Bayes estimates of the long memory parameter d. The estimates are compared to the MLE and the classic Geweke - Porter-Hudak estimate for a range of values of the sample size n. Somewhat surprisingly, we find that except when d is rather small, the Bayes estimates have a consistently smaller mean squared error.

Means of Bayesian predictive distributions developed from these priors are then used to construct forecasts of future observations. We also consider an Empirical BLUP as a classical forecast. They are applied to the Nile River data set as an application, and the forecast based on the uniform prior is seen to be superior, sometimes noticeably so. We give an explanation for why that could be anticipated.

*Research supported by a National Security Agency Grant.

†This is a part of the first author's Ph.D. dissertation at Purdue University

1 Introduction

Time series data with long range dependence has been a very active area of research in recent years. A number of models have been successfully used in various settings; fractional ARIMA processes, fractional Brownian motion, and increments of self-similar processes are among the most common models that have been studied theoretically and used practically. Equivalent definitions of long range dependence via the asymptotic behavior of the autocovariance function at infinity or of the spectral density at the origin are well known; for instance, the requirement on the spectral density is that

$$h(\lambda) \sim L_1(\lambda)|\lambda|^{-2d}, \quad |\lambda| \to 0, \quad d \in (0, \frac{1}{2}), \tag{1.1}$$

where $L_1(.)$ is slowly varying for $\lambda| \to 0$, or equivalently, the autocovariance at lag k, satisfy

$$\gamma(k) \sim L_2(k)|k|^{2d-1}, \quad |k| \to \infty, \quad d \in (0, \frac{1}{2}), \tag{1.2}$$

for some function $L_2(.)$ which is of slow variation for $|k| \to \infty$. See Beran (1994) and Zygmund (1979), in particular.

Estimation of the parameter d, popularly known as the long memory parameter, has recently attracted considerable attention. The problem is important, because accurate estimation of d is of great importance for other inference, for example inference about the stationary variance. It is also important for forecasting and model fitting purposes. Classic literature on estimation of d includes Hurst(1951), Geweke and Porter-Hudak(1983), Yajima(1985), Beran (1986), Dahlhaus(1989), etc. Of late, very substantial literature has accumulated on fine tuning some of these classic estimates and on semiparametric estimation of d; see Agiakloglou et al(1993), Hurvich et al (1995), Robinson (1994,1995), Taqqu and Teverovsky (1997, 1998), Taqqu et al(1995), Koul and Giraitis(1997), Hall et al (1997), among others.

Apart from some of the simulations and empirical studies, much of these developments have focussed on asymptotic performance of the suggested estimates and has been almost exclusively frequentist. The corresponding Bayesian studies have been generally lacking. The purpose of this article is to begin a theoretical foundation for Bayesian estimation of the long memory parameter and forecasting, with an empirical examination of their performance with real and simulated data.

The fractional ARIMA(0,d,0) model has proved to be a useful model for time series data with long memory in a variety of areas of applications; see Beran (1994), for instance. An added advantage of the fractional ARIMA(0,d,0) model is that both the spectral density and the autocovariance function are available in closed form. Indeed, the covariance function is

$$\begin{aligned} \gamma(k) &= \int_{-\pi}^{\pi} h(\lambda, d) e^{ik\lambda} d\lambda \\ &= (-1)^k \frac{\sigma^2 \Gamma(1-2d)}{\Gamma(1+k-d)\Gamma(1-k-d)}, \\ & \quad k = 0, \pm 1, \cdots; \end{aligned} \tag{1.3}$$

and the spectral density is

$$h(\lambda, d) = \frac{\sigma^2}{2\pi}(2 \sin \frac{\lambda}{2})^{-2d}, \quad \lambda \in (-\pi, \pi) \tag{1.4}$$

see Samarov and Taqqu (1988).

The Bayesian calculations that we perform depend on such closed form formulas, and this is the process we consider in this article. Similar calculations seem to be possible for increments of self-similar processes as well, although the expressions would be different. Since we do not focus on the asymptotic behavior of the Bayesian estimates, the exact expressions matter at various places. So it is better to treat these processes separately.

First, in section 2, we derive a closed form approximation to the Jeffrey prior for the ARIMA(0,d,0) process. The stationary mean μ, the scale parameter σ, and the long memory parameter d are all treated as unknown parameters. Thus, the Jeffreys prior is on the vector parameter $\theta = (d, \mu, \sigma)$. The exact Jeffrey prior cannot be written in an analytical form in this problem. It is mostly due to the reason that the Jeffrey prior depends on the form of the likelihood function, which, for the Gaussian case, depends on Σ^{-1}, the inverse covariance matrix. It is well known that except for small n, Σ^{-1} cannot be written in an analytical form in this problem. Therefore, to do any useful inference with realistic values of n, an approximation to the Jeffreys prior is essential. The positive feature of such an approximation is that it is closed form, the final closed form has a nice interpretation, and it is in a form amenable to programming and computation. We show also in section 2 that the Bayes estimate of d under this approximate Jeffreys prior can be obtained by only one dimensional integrations; i.e., the other two parameters μ and σ can in fact be integrated out analytically. This is a major simplification for the empirical evaluations and computations. Also in Section 2, we present a natural "uniform" prior, and the corresponding Bayes estimates. It turned out that the uniform prior estimates actually did better in the simulations and the applications.

Next, in Section 3, we use the approximate Jeffreys prior and the uniform prior for forecasting. Forecasting certainly is one of the paramount goals of time series analysis. In frequentist analysis, the Best Linear Unbiased Predictor (BLUP) has acquired quite universal acceptance and popularity for forecasting purposes. See, for instance, Robinson (1991), Searle, Casella, and McCulloch (1992), Ghosh and Rao (1994), among others. Actually, the expression for BLUP involves the other unknown parameters, and it is common to substitute appropriate estimates for them. The resulting estimates are usually called empirical BLUP. The Bayesian forecasts are the means of the predictive distribution of the future observations given the observed series. The predictive distribution depends on the posterior distribution of the parameter vector , and hence on the prior. As in Section 2, it turns out that the predictive means can be found by only one dimensional integrations, for both the uniform and the Jeffreys prior.

In Section 4, we show a comparative simulation of the mean squared error of five estimates of the long memory parameter d. Two of the five estimates are classical, namely the MLE of d and the Geweke-Porter-Hudak (GPH) estimate; the others are Bayes estimates under the Jeffrey and the uniform prior and an equal mixture of these two priors. The simulations are for a range of values of n. Quite surprisingly, we see that except when d is very small, the Bayes estimates, and especially the Bayes estimate under the uniform prior outperform the MLE. They outperform the GPH estimate as well.

In Section 5, the Bayesian forecasts and the Empirical BLUP developed in Section 3 are applied to the well known Nile river data set. The length of the observed series used to construct the forecasts is n= 20, 50, and 100. Specifically, 20, 50, or 100 years of data prior to the year 1107 AD are used to forecast the minimum level of the Nile River for the years 1107 - 1206. For the Nile River data set, there is some evidence that the first 100 observations or so show white noise variation around a stationary mean; see Beran (1994). That is why we used as our observed series a chunk of observations around the middle of the data set. The conclusions from a comparative evaluation of the forecasts is that the forecast based on the Jeffreys prior and the Empirical BLUP are roughly comparable; but the uniform prior forecast is pretty consistently superior, and sometimes noticeably so. We discuss why that may have been anticipated. The comparisons of the forecasts are presented with tables and plots.

Section 6 gives a concluding summary.

We hope that the formal derivation of the Jeffreys prior and the subsequent comparisons and applications filled a void in the sense that although the frequentist literature on this important problem is growing at a very fast rate, the corresponding Bayesian studies were missing. In addition, the simulations and the application to the Nile river data set show that Bayesian methods seem to have a real potential for success in these problems.

For other results on estimation of the long memory parameter, one may consult Azais and Lang (1993), Beran (1995), Beran and Terrin (1994), Cheung and Diebold (1994), Delgado and Robinson (1994), Graf (1983), Hurvich and Beltrao (1994), Hurvich et al (1995), Hurvich and Ray (1994), Jannacek (1994), Jensen (1999), Kearns and Pagan (1997), Reisen (1994), Veitch and Abry (1999), Giraitis et al (1997), etc.

Forecasting under long range dependence is addressed in Brodsky and Hurvich (1999), Crato and Ray (1996), Eisinga et al (1999),Palma and Chan (1997), Ray(1991, 1993), etc.

Previous literature on Bayesian modelling and analysis of long memory processes includes Haslett and Raftery (1989), Koop et al (1997), Petris (1996),Ravishanker and Ray (1997), etc.

2 Estimation of d

2.1 Jeffreys Prior: First Approximation

Consider now the ARIMA$(0, d, 0)$ model that we are using. Let $\theta = (d, \mu, \sigma)$; then the density of $\underline{X}$ is

$$f(\underline{X}|\theta) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}\sigma^n} e^{-\frac{(\underline{X}-\mu\mathbf{1})'\Sigma^{-1}(\underline{X}-\mu\mathbf{1})}{2\sigma^2}},$$

and the log likelihood function is

$$L(\theta) = logf(\underline{X}|\theta) = -\frac{n}{2}log(2\pi) - \frac{1}{2}log|\Sigma| - nlog\sigma - \frac{(\underline{X}-\mu\underline{\mathbf{1}})'\Sigma^{-1}(\underline{X}-\mu\underline{\mathbf{1}})}{2\sigma^2}. \qquad (2.1)$$

Now, the Jeffreys prior (see Berger, 1986), by definition, is

$$\begin{aligned} \pi(\theta) &= \{det[-E(\frac{\partial^2}{\partial\theta^2}L(\theta))]\}^{\frac{1}{2}} \\ &= n^{\frac{3}{2}}\{det[-E(\frac{\partial^2}{\partial\theta^2}\frac{1}{n}L(\theta))]\}^{\frac{1}{2}}. \end{aligned}$$

The constant term $n^{\frac{3}{2}}$ has no effect on the subsequent Bayes estimates and hence we may proceed with $\{det[-E(\frac{\partial^2}{\partial\theta^2}\frac{1}{n}L(\theta))]\}^{\frac{1}{2}}$ itself, treating this as the Jeffreys prior.

Now from (2.1),

$$\frac{1}{n}L(\theta) = -\frac{1}{2}log(2\pi) - \frac{1}{2n}log|\Sigma| - log\sigma - \frac{(\underline{X}-\mu\underline{\mathbf{1}})'\Sigma^{-1}(\underline{X}-\mu\underline{\mathbf{1}})}{2n\sigma^2}.$$

Grenander and Szegö (1958) proved that,

$$\lim_{n\to\infty}\frac{log|\Sigma|}{n} = \frac{1}{2\pi}\int_{-\pi}^{\pi} log(2\pi h(\lambda, d))d\lambda. \qquad (2.2)$$

Then from (2.2) and Gradshteyn and Ryzhik (1965, Formula 4.22 No.9), we have

$$\lim_{n\to\infty}\frac{log|\Sigma|}{n}=\sigma^2\int_{-\pi}^{\pi}log(2sin\frac{\lambda}{2})^{-2d}d\lambda=0. \qquad (2.3)$$

So if we replace $\frac{1}{n}log|\Sigma|$ by 0, we have the approximation:

$$\frac{1}{n}L(\theta)\approx-\frac{1}{2}log(2\pi)-log\sigma-\frac{(\underline{X}-\mu\underline{\mathbf{1}})'\Sigma^{-1}(\underline{X}-\mu\underline{\mathbf{1}})}{2n\sigma^2}=L_a(\theta)\ (say). \qquad (2.4)$$

We do not provide a precise meaning to the "$\approx$" sign in (2.4). The approximation is a heurism with the goal of obtaining a closed form formula for the prior density that one could use to compute and obtain estimates and forecasts. Now we state the first approximation to the Jeffreys prior.

Theorem 2.1 Based on the approximation (2.4), the Jeffreys prior is

$$\hat{\pi}(\theta)=\{\frac{tr(\Sigma\ddot{\Sigma}^{-1})}{n\sigma^4}-[\frac{tr(\Sigma\dot{\Sigma}^{-1})}{n\sigma^2}]^2\}^{\frac{1}{2}}(\frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{n})^{\frac{1}{2}}, \qquad (2.5)$$

where $\dot{\Sigma}^{-1}$ and $\ddot{\Sigma}^{-1}$ denote the matrices with elements as the first and second derivatives of elements of Σ^{-1} (derivative being with respect to d).

Proof: By straightforward calculation,

$$\begin{aligned}
-E(\frac{\partial^2 L_a(\theta)}{\partial d^2}) &= E\frac{(\underline{X}-\mu\underline{\mathbf{1}})'\ddot{\Sigma}^{-1}(\underline{X}-\mu\underline{\mathbf{1}})}{2n\sigma^2}=\frac{tr(\Sigma\ddot{\Sigma}^{-1})}{2n},\\
-E(\frac{\partial^2 L_a(\theta)}{\partial d\partial\mu}) &= -E\frac{\underline{X}'\dot{\Sigma}^{-1}\underline{\mathbf{1}}-\mu\underline{\mathbf{1}}'\dot{\Sigma}^{-1}\underline{\mathbf{1}}}{n\sigma^2}=0,\\
-E(\frac{\partial^2 L_a(\theta)}{\partial d\partial\sigma}) &= -E\frac{(\underline{X}-\mu\underline{\mathbf{1}})'\dot{\Sigma}^{-1}(\underline{X}-\mu\underline{\mathbf{1}})}{n\sigma^3}=-\frac{tr(\Sigma\dot{\Sigma}^{-1})}{n\sigma},\\
-E(\frac{\partial^2 L_a(\theta)}{\partial\mu^2}) &= \frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{n\sigma^2},\\
-E(\frac{\partial^2 L_a(\theta)}{\partial\mu\partial\sigma}) &= E\frac{2(\underline{\mathbf{1}}'\Sigma^{-1}\underline{X}-\mu\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}})}{n\sigma^3}=0,\\
-E(\frac{\partial^2 L_a(\theta)}{\partial\sigma^2}) &= E\frac{3(\underline{X}-\mu\underline{\mathbf{1}})'\Sigma^{-1}(\underline{X}-\mu\underline{\mathbf{1}})}{n\sigma^4}-\frac{1}{\sigma^2}=\frac{2}{\sigma^2}.
\end{aligned}$$

Hence, $\{det[-E(\frac{\partial^2}{\partial\theta^2}\frac{1}{n}L(\theta))]\}^{\frac{1}{2}}$ is the square root of the determinant of the matrix

$$I(\theta)=\begin{pmatrix}\frac{tr(\Sigma\ddot{\Sigma}^{-1})}{2n} & 0 & -\frac{tr(\Sigma\dot{\Sigma}^{-1})}{n\sigma}\\ 0 & \frac{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}{n\sigma^2} & 0\\ -\frac{tr(\Sigma\dot{\Sigma}^{-1})}{n\sigma} & 0 & \frac{2}{\sigma^2}\end{pmatrix}. \qquad (2.6)$$

This is easily computed and seen to be equal to

$$\{\frac{tr(\Sigma\ddot{\Sigma}^{-1})}{n\sigma^4} - [\frac{tr(\Sigma\dot{\Sigma}^{-1})}{n\sigma^2}]^2\}^{\frac{1}{2}}(\frac{\mathbf{1}'\Sigma^{-1}\mathbf{1}}{n})^{\frac{1}{2}}. \tag{2.7}$$

At this stage, we need an approximation for Σ^{-1} as well to make (2.7) analytically accessible and the Bayes estimates practically computable. We now proceed towards that.

2.2 Whittle's Approximation of Σ^{-1}

Because of the difficulty of the computation of Σ^{-1}, $\dot{\Sigma}^{-1}$ and $\ddot{\Sigma}^{-1}$, Whittle's approximation is used to replace these three matrices.

Let

$$A = [a_{(k-l)}]_{k,l=1,\ldots,n} \tag{2.8}$$

where

$$\begin{aligned} a_{(k-l)} &= (2\pi)^{-2}\int_{-\pi}^{\pi}\frac{1}{h(\lambda,d)}e^{i(k-l)\lambda}d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}(2sin\frac{\lambda}{2})^{2d}e^{i(k-l)\lambda}d\lambda. \end{aligned} \tag{2.9}$$

The matrix A is (asymptotically) the inverse of Σ (see Whittle, 1953; Beran, 1994) and has been widely used as an approximation to Σ^{-1} in the literature.

2.3 Jeffreys Prior: Further Approximation

If in (2.7), $\dot{\Sigma}^{-1}$ and $\ddot{\Sigma}^{-1}$ are replaced by $\dot{A} = [\dot{a}_{(k-l)}]_{k,l=1,\ldots,n}$ and $\ddot{A} = [\ddot{a}_{(k-l)}]_{k,l=1,\ldots,n}$ respectively, then $\tilde{\pi}(\theta)$ of Theorem 2.1 is approximated by

$$\pi^*(\theta) = \{\frac{tr(\Sigma\ddot{A})}{n\sigma^4} - [\frac{tr(\Sigma\dot{A})}{n\sigma^2}]^2\}^{\frac{1}{2}}(\frac{\mathbf{1}'\Sigma^{-1}\mathbf{1}}{n})^{\frac{1}{2}} \tag{2.10}$$

The following penultimate result is the best for practical applications.

Theorem 2.2 For $\pi^*(\theta)$ in (2.10),

$$(1) \quad \lim_{n\to\infty}\frac{\pi^*(\theta)}{(\frac{\mathbf{1}'\Sigma^{-1}\mathbf{1}}{n})^{\frac{1}{2}}} = \frac{\sqrt{C}}{\sigma^2}, \tag{2.11}$$

$$(2) \quad \lim_{n\to\infty}\frac{\pi^*(\theta)}{\frac{1}{n^d}} = \frac{\sqrt{C}}{\sigma^2}\frac{\Gamma(1-d)}{\sqrt{\Gamma(1-2d)\Gamma(2-2d)}}, \tag{2.12}$$

where

$$C = \frac{1}{2\pi} \int_{-\pi}^{\pi} [log(2sin\frac{\lambda}{2})^2]^2 d\lambda \approx 3.29. \tag{2.13}$$

Remark. The usefulness of the above theorem is that we can use $\frac{(\mathbf{1}'\Sigma^{-1}\mathbf{1})^{\frac{1}{2}}}{\sigma^2}$ as a joint default prior for $\theta = (d, \mu, \sigma)$. Similarly, another (and simpler) expression is obtained from (2.12).

To prove this theorem, the following lemma is needed. The two parts of the lemma enable us to handle the two separate terms in (2.10).

Lemma 2.1

(1)

$$\lim_{n\to\infty} \frac{1}{n} tr(\Sigma \ddot{A}) = C,$$

(2)

$$\lim_{n\to\infty} \frac{1}{n} tr(\Sigma \dot{A}) = 0.$$

Proof: Step 1. First, we take part (1) of Lemma 2.1. Let

$$g(\lambda) = [log(2sin\frac{\lambda}{2})^2]^2 (2sin\frac{\lambda}{2})^{2d}; \tag{2.14}$$

then by the dominated convergence theorem,

$$\begin{aligned} \ddot{a}_{(k-l)} &= [\frac{1}{2\pi} \int_{-\pi}^{\pi} (2sin\frac{\lambda}{2})^{2d} e^{i(k-l)\lambda} d\lambda]'' \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} [(2sin\frac{\lambda}{2})^{2d}]'' e^{i(k-l)\lambda} d\lambda \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} [log(2sin\frac{\lambda}{2})^2]^2 (2sin\frac{\lambda}{2})^{2d} e^{i(k-l)\lambda} d\lambda \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) e^{i(k-l)\lambda} d\lambda. \end{aligned} \tag{2.15}$$

Step 2. Now,

$$\begin{aligned} tr(\Sigma \ddot{A}) &= \Sigma_{k=1}^{n} \Sigma_{l=1}^{n} \gamma_{(k-l)} \ddot{a}_{(l-k)} \\ &= n\Sigma_{m=0}^{n-1} \gamma_{(m)} \ddot{a}_{(-m)} - \Sigma_{m=1}^{n-1} m\gamma_{(m)} \ddot{a}_{(-m)} + \Sigma_{m=1}^{n-1} (n-m)\gamma_{(-m)} \ddot{a}_{(m)} \end{aligned}$$

$$\begin{aligned}
&= n\Sigma_{m=1-n}^{n-1}\gamma_{(m)}\ddot{a}_{(-m)} - 2\Sigma_{m=1}^{n-1}m\gamma_{(-m)}\ddot{a}_{(m)} \\
&= n\Sigma_{m=1-n}^{n-1}\gamma_{(m)}\ddot{a}_{(-m)} - 2i\Sigma_{m=1}^{n-1}\hat{a}_{(m)}\gamma_{(-m)},
\end{aligned} \tag{2.16}$$

where

$$\hat{a}_{(m)} = \frac{1}{2\pi}\int_{-\pi}^{\pi}(g(\lambda))'e^{im\lambda}dx = -im\ddot{a}_{(m)}. \tag{2.17}$$

Step 3. Next, of the two terms in (2.16),

$$\begin{aligned}
\Sigma_{m=1}^{n-1}\hat{a}_{(m)}\gamma_{(-m)} &= \Sigma_{m=1}^{n-1}\frac{1}{2\pi}\int_{-\pi}^{\pi}g'(x)e^{imx}dx\frac{1}{2\pi}\int_{-\pi}^{\pi}h(y,d)e^{-imy}dy \\
&= \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\Sigma_{m=1}^{n-1}e^{im(x-y)}g'(x)h(y,d)dxdy \\
&= \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{e^{i(n-1)(x-y)}-1}{1-e^{-i(x-y)}}g'(x)h(y,d)dxdy \\
&= \frac{i}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{sin\frac{n(x-y)}{2}sin\frac{(n-1)(x-y)}{2}}{sin\frac{x-y}{2}}g'(x)h(y,d)dxdy.
\end{aligned} \tag{2.18}$$

Step 4. The need for this step will be clear in step 5. Now observe that

$$\begin{aligned}
&\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}sin\frac{n(x-y)}{2}g'(x)h(y,d)dxdy \\
&= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}sin\frac{nx}{2}cos\frac{ny}{2}g'(x)h(y,d)dxdy - \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}cos\frac{nx}{2}sin\frac{ny}{2}g'(x)h(y,d)dxdy \\
&= I_1 + I_2 \ (say).
\end{aligned} \tag{2.19}$$

The Riemann-Lebesgue Lemma says that if a function $f(x)$ is integrable on (a,b), then

$$\int_a^b f(x)sin(nx)dx \to 0, \quad as\ n \to \infty. \tag{2.20}$$

Let us check the integrability of $g'(x)$: from (2.14),

$$\begin{aligned}
g'(x) &= \{[log(2sin\frac{x}{2})^2]^2(2sin\frac{x}{2})^{2d}\}' \\
&= 2log(2sin\frac{x}{2})^2\frac{cos\frac{x}{2}}{(2sin\frac{x}{2})^{1-2d}} + 2d[log(2sin\frac{x}{2})^2]^2(2sin\frac{x}{2})^{2d-1}cos\frac{x}{2} \\
&\sim \frac{logx}{x^{1-2d}} + \frac{(logx)^2}{x^{1-2d}}, \quad as\ x \to 0.
\end{aligned}$$

So $g'(x)$ is integrable on $(-\pi,\pi)$. Therefore,

$$\int_{-\pi}^{\pi}sin\frac{nx}{2}g'(x)dx \to 0, \ as\ n \to \infty. \tag{2.21}$$

Similarly, $h(y,d)$ is integrable on $(-\pi,\pi)$ and

$$\int_{-\pi}^{\pi} sin\frac{ny}{2}h(y,d)dy \to 0, \ as \ n \to \infty. \tag{2.22}$$

Hence,

$$I_1 \to 0, \ and \ \ I_2 \to 0, \tag{2.23}$$

because $cos\frac{nx}{2}$ and $cos\frac{ny}{2}$ are uniformly bounded.

Step 5. On the other hand, by induction on n, we can prove that

$$|sin\frac{(n-1)(x-y)}{2}| \leq (n-1)|sin\frac{x-y}{2}|.$$

So

$$|\frac{sin\frac{(n-1)(x-y)}{2}}{n \ sin\frac{x-y}{2}}| < 1. \tag{2.24}$$

By (2.18), (2.19), (2.23) and (2.24)

$$\lim_{n\to\infty} \frac{1}{n}\Sigma_{m=1}^{n-1}\hat{a}_{(m)}\gamma_{(-m)} = 0 \tag{2.25}$$

Step 6. Now we come to the first term in (2.16). By Parseval's formula, as $n \to \infty$,

$$\begin{aligned} \Sigma_{m=1-n}^{n-1}\gamma_{(m)}\ddot{a}_{(-m)} &\to \frac{1}{2\pi}\int_{-\pi}^{\pi} g(x)h(x,d)dx \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} [log(2sin\frac{x}{2})^2]^2 dx \\ &= C \\ &\approx 3.29. \end{aligned} \tag{2.26}$$

Therefore, from (2.16) it follows that

$$\lim_{n\to\infty} \frac{1}{n} tr(\Sigma\ddot{A}) = C,$$

completing part (1) of our lemma. We will now proceed to part (2) of Lemma 2.1.

Step 7. Let us introduce the function

$$w(x) = log(2sin\frac{x}{2})^2 \ (2sin\frac{x}{2})^{2d}. \tag{2.27}$$

Similar to (2.15),

$$\begin{aligned}
\dot{a}_{(k-l)} &= [\frac{1}{2\pi}\int_{-\pi}^{\pi}(2sin\frac{\lambda}{2})^{2d}e^{i(k-l)\lambda}d\lambda]' \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} log(2sin\frac{\lambda}{2})^{2}(2sin\frac{\lambda}{2})^{2d}e^{i(k-l)\lambda}d\lambda \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} w(\lambda)e^{i(k-l)\lambda}d\lambda. \qquad (2.28)
\end{aligned}$$

Now, as in (2.16),

$$\begin{aligned}
tr(\Sigma\dot{A}) &= \Sigma_{k=1}^{n}\Sigma_{l=1}^{n}\gamma_{(k-l)}\dot{a}_{(l-k)} \\
&= n\Sigma_{m=1-n}^{n-1}\gamma_{(m)}\dot{a}_{(-m)} - 2i\Sigma_{m=1}^{n-1}a^{*}_{(m)}\gamma_{(-m)}, \qquad (2.29)
\end{aligned}$$

where

$$a^{*}_{(m)} = -im\dot{a}_{(m)}.$$

Step 8. Following the steps of part (1) of this lemma, it can be proved that

$$\lim_{n\to\infty}\frac{1}{n}\Sigma_{m=1}^{n-1}a^{*}_{(m)}\gamma_{(-m)} = 0. \qquad (2.30)$$

Step 9. And, again, by Parseval's formula, as $n \to \infty$,

$$\begin{aligned}
\Sigma_{m=1-n}^{n-1}\gamma_{(m)}\dot{a}_{(-m)} &\to \int_{-\pi}^{\pi} w(x)h(x,d)dx \\
&= \int_{-\pi}^{\pi} log(2sin\frac{x}{2})^{2}dx \\
&= 0 \ (See\ (2.3)). \qquad (2.31)
\end{aligned}$$

Step 10. By (2.29), (2.30) and (2.31),

$$\lim_{n\to\infty}\frac{1}{n}tr(\Sigma\dot{A}) = 0.$$

This completes the proof of both parts of Lemma 2.1.

Proof of Theorem 2.2:

(1) By (2.10) and lemma 2.1, part (1) is obtained.

(2) By part (1) of this theorem and the fact (see Beran, 1994)

$$\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}} \sim \frac{\Gamma(1-2d)\Gamma(2-2d)}{\Gamma^2(1-d)} n^{2d-1},$$

part (2) is obtained.

By Theorem 2.2, we can use $\frac{1}{\sigma^2}(\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}})^{\frac{1}{2}}$, or even better, $\frac{1}{\sigma^2}\frac{\Gamma(1-d)}{\sqrt{\Gamma(1-2d)\Gamma(2-2d)}}n^{-d}$ as an approximation of the Jeffreys prior for this problem. In Figure 1, we have superimposed these two functions on the (correctly scaled) exact formula of $\pi^*(\theta)$. Recall that this exact formula was given in equation (2.10). It is clear that the above approximations are quite good.

Discussion. The preceding calculations provide **only an approximation** to the true Jeffreys prior for $\theta = (d, \mu, \sigma)$. We had to use approximations for $log|\Sigma|$, Σ^{-1}, $\dot{\Sigma}^{-1}$ and $\ddot{\Sigma}^{-1}$ in order to finally provide an analytical, clear, and useful expression that one can use. The exact Jefferys prior does not have such an useful expression. Furthermore, the approximation $\frac{1}{\sigma^2}(\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}})^{\frac{1}{2}}$ has a nice interpretation; it says that one should use the uniform prior for μ, the familiar $\frac{1}{\sigma^2}$ prior for σ, and $\sqrt{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}}$ for d.

2.4 Estimation of d

Having obtained an approximation to the Jeffreys prior, we will now address the very important problem of finding Bayesian estimates of d and then comparing them, on the basis of mean squared error, to certain other estimates of d. In addition to the approximate Jeffreys prior, the prior $\frac{2}{\sigma^2}$ jointly for $\theta = (d, \mu, \sigma)$ is also used; casually speaking, according to this latter prior, μ and d have uniform priors and σ has the density $\frac{1}{\sigma^2}$. For Jeffreys prior and the uniform prior, the Bayes estimates of d are discussed below.

(1) First let us consider the approximate Jeffreys prior

$$\pi_J(\theta) \sim \frac{1}{\sigma^2}(\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}})^{\frac{1}{2}}. \tag{2.32}$$

Then the Bayes estimate of d is

$$\hat{d}_J = \frac{\int_0^{\frac{1}{2}} \int_0^{\infty} \int_{-\infty}^{\infty} d\ f(\underline{x}|\theta) \frac{(\underline{1}'\Sigma^{-1}\underline{1})^{\frac{1}{2}}}{\sigma^2} d\mu d\sigma dd}{\int_0^{\frac{1}{2}} \int_0^{\infty} \int_{-\infty}^{\infty} f(\underline{x}|\theta) \frac{(\underline{1}'\Sigma^{-1}\underline{1})^{\frac{1}{2}}}{\sigma^2} d\mu d\sigma dd}. \tag{2.33}$$

The denominator in (2.32) is proportional to

$$\begin{aligned}
&\int_0^{\frac{1}{2}} \int_0^{\infty} \int_{-\infty}^{\infty} \frac{(\underline{1}'\Sigma^{-1}\underline{1})^{\frac{1}{2}}}{\sigma^{n+2}|\Sigma|^{\frac{1}{2}}} e^{-\frac{(\underline{X}-\mu\underline{1})'\Sigma^{-1}(\underline{X}-\mu\underline{1})}{2\sigma^2}} d\mu d\sigma dd \\
= &\int_0^{\frac{1}{2}} \int_0^{\infty} \int_{-\infty}^{\infty} \frac{(\underline{1}'\Sigma^{-1}\underline{1})^{\frac{1}{2}}}{\sigma^{n+2}|\Sigma|^{\frac{1}{2}}} e^{-\frac{\underline{X}'\Sigma^{-1}\underline{X} - (\underline{1}'\Sigma^{-1}\underline{X})^2/\underline{1}'\Sigma^{-1}\underline{1}}{2\sigma^2}} \times \\
&\qquad e^{-\frac{\underline{1}'\Sigma^{-1}\underline{1}(\mu - \underline{1}'\Sigma^{-1}\underline{X}/\underline{1}'\Sigma^{-1}\underline{1})^2}{2\sigma^2}} d\mu d\sigma dd \\
\propto &\int_0^{\frac{1}{2}} \int_0^{\infty} \frac{|\Sigma|^{-\frac{1}{2}}}{\sigma^{n+1}} e^{-\frac{\underline{X}'\Sigma^{-1}\underline{X} - (\underline{1}'\Sigma^{-1}\underline{X})^2/\underline{1}'\Sigma^{-1}\underline{1}}{2\sigma^2}} d\sigma dd \\
\propto &\int_0^{\frac{1}{2}} |\Sigma|^{-\frac{1}{2}} \Gamma(\frac{n}{2}) (\underline{X}'\Sigma^{-1}\underline{X} - \frac{(\underline{1}'\Sigma^{-1}\underline{X})^2}{\underline{1}'\Sigma^{-1}\underline{1}})^{-\frac{n}{2}} dd.
\end{aligned} \tag{2.34}$$

Similarly, the numerator in (2.32) is proportional to

$$\int_0^{\frac{1}{2}} d|\Sigma|^{-\frac{1}{2}} \Gamma(\frac{n}{2}) (\underline{X}'\Sigma^{-1}\underline{X} - \frac{(\underline{1}'\Sigma^{-1}\underline{X})^2}{\underline{1}'\Sigma^{-1}\underline{1}})^{-\frac{n}{2}} dd. \tag{2.35}$$

Computationally, (2.33) and (2.34) are not useful because they again include the infeasible Σ^{-1}. If we replace Σ^{-1} by A, then we get the pseudo-Bayes estimate:

$$\hat{d}_J = \frac{\int_0^{\frac{1}{2}} d|\Sigma|^{-\frac{1}{2}} (\underline{X}'A\underline{X} - \frac{(\underline{1}'A\underline{X})^2}{\underline{1}'A\underline{1}})^{-\frac{n}{2}} dd}{\int_0^{\frac{1}{2}} |\Sigma|^{-\frac{1}{2}} (\underline{X}'A\underline{X} - \frac{(\underline{1}'A\underline{X})^2}{\underline{1}'A\underline{1}})^{-\frac{n}{2}} dd}. \tag{2.36}$$

This is computable. A small clarification is in order: although an exact expression for $\underline{1}'\Sigma^{-1}\underline{1}$ is available, we still used $\underline{1}'A\underline{1}$ instead. The reason is that otherwise, the quantity $\underline{X}'A\underline{X} - \frac{(\underline{1}'A\underline{X})^2}{\underline{1}'\Sigma^{-1}\underline{1}}$ can be negative. Note that we have retained the same notation $\hat{d}_J$ in (2.36).

(2) Next we consider the prior

$$\pi_U(\theta) = \frac{2}{\sigma^2}. \tag{2.37}$$

Analogus to (2.36), we now get an alternative Bayesian estimate:

$$\hat{d}_U = \frac{\int_0^{\frac{1}{2}} \int_0^{\infty} \int_{-\infty}^{\infty} d\ f(\underline{x}|\theta) \frac{2}{\sigma^2} d\mu d\sigma dd}{\int_0^{\frac{1}{2}} \int_0^{\infty} \int_{-\infty}^{\infty} f(\underline{x}|\theta) \frac{2}{\sigma^2} d\mu d\sigma dd}$$

$$= \frac{\int_0^{\frac{1}{2}} d\, |\Sigma|^{-\frac{1}{2}}(\underline{1}'\Sigma^{-1}\underline{1})^{-\frac{1}{2}}(\underline{X}'A\underline{X} - \frac{(\underline{1}'A\underline{X})^2}{\underline{1}'A\underline{1}})^{-\frac{n}{2}}dd}{\int_0^{\frac{1}{2}} |\Sigma|^{-\frac{1}{2}}(\underline{1}'\Sigma^{-1}\underline{1})^{-\frac{1}{2}}(\underline{X}'A\underline{X} - \frac{(\underline{1}'A\underline{X})^2}{\underline{1}'A\underline{1}})^{-\frac{n}{2}}dd}. \tag{2.38}$$

This is also computable.

Example 2.1 For the Nile River data, the estimates of d using the full series (622-1281 AD) are as follows:

R/S	0.436
MLE	0.393
GPH	0.384
Uniform prior	0.404
Jeffreys prior	0.394

So with the exception of the R/S estimate, the numerical values of the other estimates are quite comparable. However, over repeated simulated series, the Bayes estimates seem to be generally better as we will see in section 4.

3 Forecasting

3.1 Introduction

Forecasting is one of the primary goals of time series analysis. In many areas of statistics, the Best Linear Unbiased Predictor (BLUP) has acquired popularity for forecasting purposes. See, for instance, Robinson (1991), Searle, Casella and McCulloch (1992), and Ghosh and Rao (1994). Bayesian forecasting of a future observation would, on the other hand, use the predictive distribution of the future observation given the observed data, and then use the mean of that predictive distribution as the forecast.

In reality, however, the problem is a little more involved. The simple reason is that the BLUP will actually involve, in its formula, parameters of the model that really are not known. Likewise, the Bayesian predictive distribution involves, by definition, a prior on all the unknown parameters of the model. Frequentists, typically, get around this dilemma by simply plugging estimates of the unknown parameters in the BLUP expression. This is known as the Empirical BLUP. There is some choice in which estimates are going to be plugged in for writing the Empirical BLUP. It depends on simplicity of computation, for example. The Bayesian predictive distribution is easily written, in principle. But evidently, there is an integration involved in getting rid of the parameters. And this integration, in all but the most convenient models, may be impossible to perform in closed form. So there is some work involved in either the frequentist or the Bayesian domain in order to make the forecasts correspond to realism.

Interestingly enough, long memory is actually good for forecasting purposes. The stronger the dependence between the observations, the more the future observations look like the past, and easier is forecasting. Fine predictions at large lags may be obtained under long range dependence, provided a good long series of data are available to estimate the long memory parameter. For the Bayesian forecasts, we will use the two default priors described in section 2, namely, the approximate Jeffreys prior $\pi_J(\theta) = \frac{1}{\sigma^2}(\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}})^{\frac{1}{2}}$, and the "uniform" prior $\pi_U(\theta) = \frac{2}{\sigma^2}$. In the Empirical BLUPS , we will use the MLE of d to estimate d, and the BLUE to estimate μ. Note that the BLUE itself has the parameter d in its expression, but there too d is replaced by its MLE.

3.2 Forecasting Formulas

For an ARIMA$(0, d, 0)$ process, suppose we have observations $\underline{X} = (X_1, ..., X_n)'$, and we want to predict the future value X_{n+k} at lag k. Denote

$$\xi = (\gamma(n+k-1), \gamma(n+k-2), ..., \gamma(k))', \tag{3.1}$$

the vector of covariances (apart from a multiplier of σ^2) between the components of $\underline{X}$ and the future observation X_{n+k}. Let also

$$\underline{X}^+ = (X_1, ..., X_n, X_{n+k})' \tag{3.2}$$

and

$$\Sigma^+ = \begin{pmatrix} \Sigma & \xi \\ \xi' & a \end{pmatrix}; \tag{3.3}$$

here, $a = \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}$. Given $\underline{X}$, the Bayes predictor of X_{n+k} is

$$\hat{X}_{n+k} = E^{\theta|\underline{X}} E^{X_{n+k}|\underline{X},\theta}(X_{n+k}). \tag{3.4}$$

Since we have a Gaussian time series, we have that

$$X_{n+k}|\underline{X}, \theta \sim N(\mu + \xi'\Sigma^{-1}(\underline{X} - \mu\underline{\mathbf{1}}), \sigma^2(a - \xi'\Sigma^{-1}\xi)). \tag{3.5}$$

In particular,

$$\begin{aligned} E(X_{n+k}|\underline{X}, \theta) &= \mu + \xi'\Sigma^{-1}(\underline{X} - \mu\underline{\mathbf{1}}) \\ &= \mu(1 - \xi'\Sigma^{-1}\underline{\mathbf{1}}) + \xi'\Sigma^{-1}\underline{X}. \end{aligned} \tag{3.6}$$

So for the uniform prior in (2.37), the Bayes predictor of X_{n+k} is

$$\begin{aligned} \hat{X}^U_{n+k} &= E^{\theta|\underline{X}} E^{X_{n+k}|\underline{X},\theta}(X_{n+k}) \\ &= E^{\theta|\underline{X}}(\mu(1 - \xi'\Sigma^{-1}\underline{\mathbf{1}}) + \xi'\Sigma^{-1}\underline{X}) \\ &= \frac{\int\int\int \frac{\mu(1-\xi'\Sigma^{-1}\underline{\mathbf{1}})+\xi'\Sigma^{-1}\underline{X}}{\sigma^{n+2}|\Sigma|^{\frac{1}{2}}} e^{-\frac{(\underline{X}-\mu\underline{\mathbf{1}})'\Sigma^{-1}(\underline{X}-\mu\underline{\mathbf{1}})}{2\sigma^2}} d\mu d\sigma dd}{\int\int\int \frac{1}{\sigma^{n+2}|\Sigma|^{\frac{1}{2}}} e^{-\frac{(\underline{X}-\mu\underline{\mathbf{1}})'\Sigma^{-1}(\underline{X}-\mu\underline{\mathbf{1}})}{2\sigma^2}} d\mu d\sigma dd} \\ &= \frac{\int \frac{|\Sigma|^{-\frac{1}{2}}}{(\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}})^{\frac{1}{2}}}[\xi'\Sigma^{-1}\underline{X} + \frac{(1-\xi'\Sigma^{-1}\underline{\mathbf{1}})\underline{\mathbf{1}}'\Sigma^{-1}\underline{X}}{(\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}})}](\underline{X}'\Sigma^{-1}\underline{X} - \frac{(\underline{\mathbf{1}}'\Sigma^{-1}\underline{X})^2}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}})^{-\frac{n}{2}} dd}{\int \frac{|\Sigma|^{-\frac{1}{2}}}{(\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}})^{\frac{1}{2}}}(\underline{X}'\Sigma^{-1}\underline{X} - \frac{(\underline{\mathbf{1}}'\Sigma^{-1}\underline{X})^2}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}})^{-\frac{n}{2}} dd}, \end{aligned} \tag{3.7}$$

on integrating out μ and σ. The integrations with respect to μ and σ, fortunately, can be done in closed form.

Similarly, for the Jeffreys prior in (2.32),

$$\hat{X}^J_{n+k} = \frac{\int |\Sigma|^{-\frac{1}{2}}[\xi'\Sigma^{-1}\underline{X} + \frac{(1-\xi'\Sigma^{-1}\underline{\mathbf{1}})\underline{\mathbf{1}}'\Sigma^{-1}\underline{X}}{(\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}})}](\underline{X}'\Sigma^{-1}\underline{X} - \frac{(\underline{\mathbf{1}}'\Sigma^{-1}\underline{X})^2}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}})^{-\frac{n}{2}} dd}{\int |\Sigma|^{-\frac{1}{2}}(\underline{X}'\Sigma^{-1}\underline{X} - \frac{(\underline{\mathbf{1}}'\Sigma^{-1}\underline{X})^2}{\underline{\mathbf{1}}'\Sigma^{-1}\underline{\mathbf{1}}})^{-\frac{n}{2}} dd}. \tag{3.8}$$

Now we turn to the classical forecasts. The exact formula for the BLUP of X_{n+k} is

$$\mu(1 - \xi'\Sigma^{-1}\underline{\mathbf{1}}) + \xi'\Sigma^{-1}\underline{X}.$$

If in this formula, we write the BLUE of μ in place of μ, we get

$$BLUP = (\underline{1}'\Sigma^{-1}\underline{1})^{-1}\underline{1}'\Sigma^{-1}\underline{X}(1 - \xi'\Sigma^{-1}\underline{1}) + \xi'\Sigma^{-1}\underline{X}. \tag{3.9}$$

Finally, if we replace d by its MLE based on the observed series $\underline{X} = (X_1, ..., X_n)$, we get the Empirical BLUP.

4 Simulation for estimation of d

In this section, we provide a comparative simulation of the mean squared error of the following 5 estimates of d:

i. $\hat{d}_U$;

ii. $\hat{d}_J$;

iii. The MLE of d;

iv. The Geweke-Porter-Hudak (GPH) estimate;

v. The Bayes estimate with respect to the prior $\frac{1}{2}\pi_U(\theta) + \frac{1}{2}\pi_J(\theta)$.

The simulation results for the five estimates are listed in Table 1 for n =20,50, 100, 200 and d =0.01, 0.1, 0.25, 0.4, 0.49. For $n = 50$ and $n = 200$, the mean squared errors are additionally displayed in Figure 2. From these results, we can see that, only for small d, the MLE is the best. As d becomes larger, the Bayes estimates based on uniform prior, Jeffreys prior and the mixed prior are substantially better than the MLE and GPH. The evidence from Table 1 is that the performance of the Bayes estimates is definitely encouraging and the derivation of the approximations given in Theorem 2.2 was worthwhile.

Table 1. Simulated MSE for estimators of d

$n = 20$

d	Uniform	Jeffreys	MLE	GPH	Mixed
0.01	0.035596	0.017922	0.004471	0.059235	0.031542
0.10	0.013685	0.003756	0.008147	0.050953	0.011283
0.25	0.003943	0.007955	0.035887	0.042557	0.004336
0.40	0.021669	0.042638	0.078513	0.037652	0.024826
0.49	0.038851	0.069552	0.084226	0.030318	0.043068

$n = 50$

d	Uniform	Jeffreys	MLE	GPH	Mixed
0.01	0.022175	0.012583	0.007157	0.024291	0.018773
0.10	0.006397	0.002047	0.006649	0.016553	0.004671
0.25	0.007243	0.009199	0.021278	0.024925	0.007812
0.40	0.012328	0.023663	0.028967	0.025021	0.014844
0.49	0.020665	0.038455	0.033576	0.020106	0.023906

Table 1. Simulated MSE for Estimators of d (continued)

$n = 100$

d	Uniform	Jeffreys	MLE	GPH	Mixed
0.01	0.008592	0.005438	0.002041	0.012043	0.008122
0.10	0.002633	0.001208	0.004232	0.008091	0.002438
0.25	0.004496	0.005283	0.007241	0.013985	0.004566
0.40	0.003415	0.006972	0.006817	0.006642	0.003568
0.49	0.007983	0.015058	0.011992	0.010873	0.008361

$n = 200$

d	Uniform	Jeffreys	MLE	GPH	Mixed
0.01	0.002472	0.001704	0.000578	0.002910	0.002138
0.10	0.001874	0.001535	0.003348	0.004646	0.001758
0.25	0.004037	0.004821	0.005121	0.006232	0.004302
0.40	0.002648	0.004179	0.004389	0.006269	0.002833
0.49	0.003043	0.005764	0.004455	0.003362	0.003193

5 Application

We now apply the forecasts developed in section 3 to the Nile River data set. Beran (1994), Beran and Terrin (1996) have shown that for the Nile River data, there is some evidence that the first 100 observations (622 - 721 AD) follow a pattern of white noise variation around a stationary mean. For instance, the MLE of d for the first 100 observations is only about 0.04. In subsequent applications, Beran (1994) uses the observations for the years 1007 - 1106 AD, to evaluate forecasts for the next 100 years. We will use the same convention. Actually, we will also consider the cases when the available data range over 20 years (1087 - 1106 AD), and 50 years (1057 - 1106 AD). Then the empirical BLUPS will be compared with the Bayesian forecasts. The conclusion will be that the empirical BLUP is largely comparable to the forecast based on the Jeffreys prior, but the forecast based on the uniform prior is more accurate than these two. We will discuss why that might have been anticipated.

The empirical BLUP as well as the two Bayesian forecasts for the Nile River minima for the years 1107 - 1206 are presented in Table 2. Some selected lags are presented for ease of presentation. These forecasts use 50 years of observed data, i.e., data for the years 1057 - 1106.

Table 3 lists the forecast errors, respectively, for 100, 50, and 20 years of observed data.

Examination of Table 3 shows the empirical BLUP and the Bayesian forecast based on the Jeffreys prior to be roughly of the same quality. But the forecast based on the uniform prior is a bit superior. As an illustration, let us take the forecast errors based on 50 years of observations. With the exception of the forecast for the years 1111, 1146, 1161, 1181 and 1201 (lags 5, 40, 55, 75 and 95), the forecast based on the uniform prior is the best. Sometimes, relatively speaking, the superiority of the uniform prior forecast is substantial. For instance, for the year 1108, the uniform prior forecast is 16% better than the empirical BLUP.

On hindsight, one can explain why this ought to be the case. Figure 2 illustrates that the uniform prior Bayes estimate of d is superior to the MLE when the true d is moderate or large. For the Nile River data, the true value of d seems to be fairly high. One may therefore expect that forecasts that use the uniform prior Bayes estimate of d will outperform the empirical BLUP in some overall sense. That is what we are seeing in Table 3.

In Figure 3, we plot the cumulative mean absolute error against the years, for all three forecasts. For a given series of forecast values, cumulative mean absolute error at year j is defined as

$$\textit{Cumulative Mean Absolute Error} \;=\; \frac{1}{j}\Sigma_{k=1}^{j}|X_{n+k} - \hat{X}_{n+k}|. \tag{5.1}$$

Figure 3 is for the case n=50, i.e., 50 years of data are used to construct the three forecasts. Again we see that the empirical BLUP and the Jeffreys prior forecast are of about the same quality, and the forecast based on the uniform prior is noticeably superior.

6 Concluding Summary

Estimation of the dependence parameter under long range dependence has recently become a very active area of research. The literature has focussed on semi and nonparametric estimation and the corresponding asymptotics. Theoretical Bayesian studies and comparative evaluation had been lacking. We consider the fractional ARIMA(0,d,0) Gaussian process and give a simple and closed form approximation to the Jeffrey prior density for the parameter vector $\theta = (d, \mu, \sigma)$. The approximated prior has a nice interpretation and is amenable to further inference and computation for essentially any n, not just small n. A comparative simulation shows that except when d is rather small, the Jeffrey prior estimate and the uniform prior estimate each outperforms the MLE of d. They also outperform the classic Geweke - Porter-Hudak estimate.

It seemed natural to develop Bayesian forecasts for ARIMA(0,d,0) processes using these priors. Fortunately, the means of the predictive distributions involve only one dimensional integration. The formulas and an Empirical BLUP are applied to the well known Nile river minima data set. Interestingly, the forecasts based on the uniform prior are seen to be the best quite consistently.

Our results lay the foundation for filling a void in this very important area. The results in this article suggest that Bayesian methods have a really good potential for success in other long memory processes.

Table 2. Forecasts using the Nile River data between the years 1057-1106

			forecast		
year	k=lag	observations	uniform	Jeffreys	classical
1107	1	1333	1247.5	1239.1	1240.6
1108	2	1270	1234.5	1226.8	1227.5
1109	3	1245	1227.6	1220.5	1220.8
1110	4	1245	1223.0	1216.3	1216.4
1111	5	1211	1219.6	1213.3	1213.2
1116	10	1295	1210.1	1205.0	1204.6
1121	15	1268	1205.4	1201.0	1200.5
1126	20	1319	1202.5	1198.6	1198.0
1131	25	1263	1200.5	1197.0	1196.4
1136	30	1223	1199.1	1195.8	1195.2
1141	35	1205	1198.0	1195.0	1194.4
1146	40	1128	1197.1	1194.3	1193.7
1151	45	1209	1196.4	1193.7	1193.1
1156	50	1247	1195.8	1193.3	1192.7
1161	55	1182	1195.3	1192.9	1192.3
1166	60	1236	1194.9	1192.6	1192.0
1171	65	1259	1194.6	1192.3	1191.8
1176	70	1252	1194.3	1192.1	1191.6
1181	75	1088	1194.0	1191.8	1191.4
1186	80	1286	1193.8	1191.7	1191.2
1191	85	1250	1193.5	1191.5	1191.1
1196	90	1209	1193.4	1191.4	1190.9
1201	95	989	1193.2	1191.2	1190.8
1206	100	1290	1193.0	1191.1	1190.7

Table 3. Absolute errors for predicting the Nile River minimum

(a) Based on the observations between the years 1007 and 1106

year	k=lag	uniform prior	Jeffreys prior	classical prediction
1107	1	77.386	82.807	80.322
1108	2	29.327	34.840	32.621
1109	3	12.764	18.178	16.167
1110	4	18.604	23.894	22.046
1111	5	10.946	5.788	7.503
1116	10	86.241	90.853	89.566
1121	15	66.322	70.528	69.489
1126	20	121.953	125.845	124.973
1131	25	69.283	72.922	72.171
1136	30	31.820	35.250	34.592
1141	35	15.828	19.082	18.497
1146	40	59.534	56.432	56.957
1151	45	22.829	25.799	25.324
1156	50	61.985	64.839	64.404
1161	55	2.021	0.728	0.330
1166	60	52.843	55.500	55.131
1171	65	76.602	79.175	78.833
1176	70	59.274	61.771	61.452
1181	75	93.126	90.699	90.997
1186	80	105.413	107.776	107.497
1191	85	69.900	72.205	71.941
1196	90	29.342	31.592	31.343
1201	95	190.255	188.055	188.291
1206	100	111.114	113.267	113.044

Table 3. Absolute errors for predicting the Nile River minimum (continued)

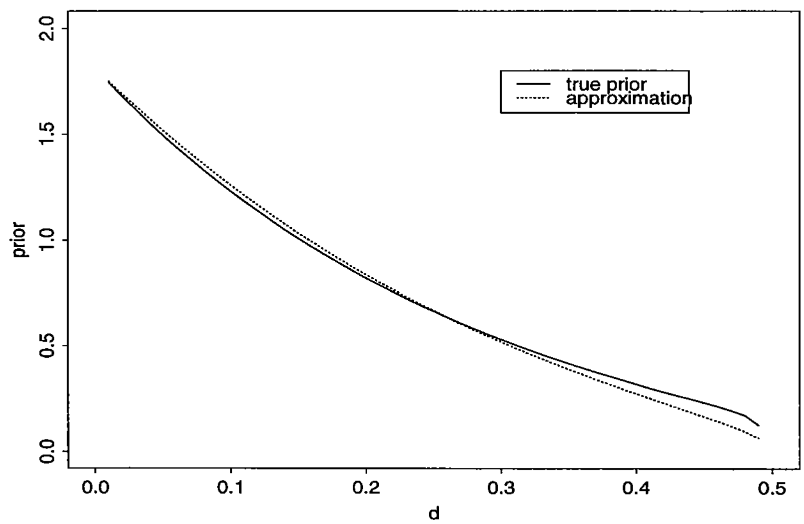
(b) Based on the observations between the years 1057 and 1106

year	k=lag	uniform prior	Jeffreys prior	classical prediction
1107	1	85.525	93.935	92.364
1108	2	35.508	43.191	42.482
1109	3	17.402	24.512	24.205
1110	4	21.997	28.656	28.589
1111	5	8.607	2.319	2.223
1116	10	84.904	89.978	90.432
1121	15	62.612	66.971	67.534
1126	20	116.509	120.382	120.981
1131	25	62.490	66.006	66.612
1136	30	23.935	27.175	27.777
1141	35	7.037	10,058	10.648
1146	40	69.095	66.253	65.677
1151	45	12.606	15.298	15.860
1156	50	51.184	53.750	54.296
1161	55	13.332	10.875	10.343
1166	60	41.080	43.442	43.959
1171	65	64.433	66.713	67.216
1176	70	46.740	48.947	49.437
1181	75	105.991	103.849	103.372
1186	80	92.247	94.330	94.796
1191	85	56.458	58.488	58.943
1196	90	15.646	17.629	18.073
1201	95	204.184	202.245	201.810
1206	100	96.969	98.869	99.294

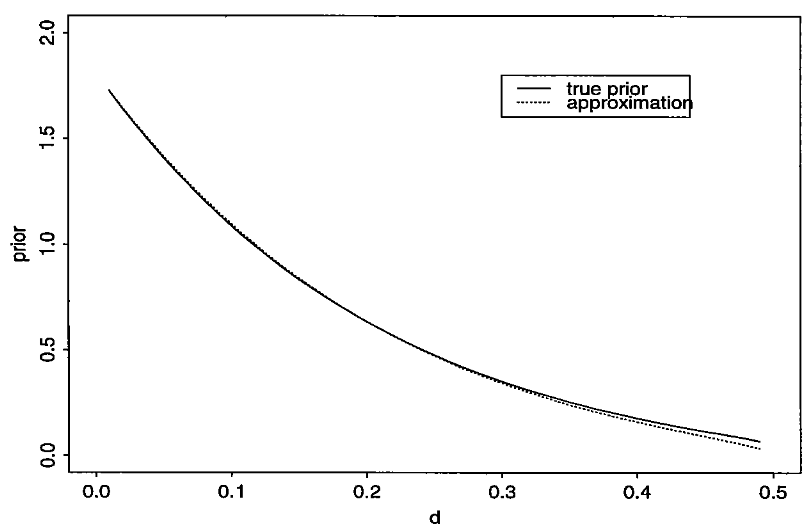
Table 3. Absolute errors for predicting the Nile River minimum (continued)

(c) Based on the observations between the years 1087 and 1106

year	k=lag	uniform prior	Jeffreys prior	classical prediction
1107	1	70.404	77.591	72.317
1108	2	16.684	22.855	19.022
1109	3	3.670	1.840	1.253
1110	4	0.715	4.323	1.699
1111	5	32.618	27.937	30.299
1116	10	56.856	60.473	59.015
1121	15	32.294	35.370	34.260
1126	20	84.690	87.430	86.507
1131	25	29.592	32.103	31.293
1136	30	9.778	7.435	8.170
1141	35	27.314	25.099	25.782
1146	40	103.959	101.845	102.491
1151	45	22.680	20.646	21.265
1156	50	15.546	17.512	16.915
1161	55	49.268	47.359	47.939
1166	60	4.887	6.749	6.181
1171	65	28.019	29.839	29.282
1176	70	10.132	11.916	11.368
1181	75	142.770	141.017	141.558
1186	80	55.315	57.041	56.505
1191	85	19.391	21.091	20.561
1196	90	21.542	19.864	20.390
1201	95	241.482	239.824	240.347
1206	100	59.571	61.211	60.691

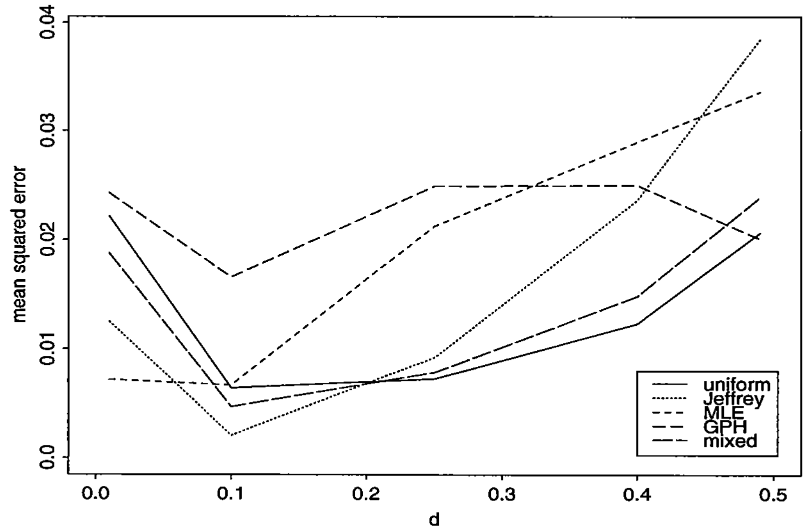


(a) $n = 50$

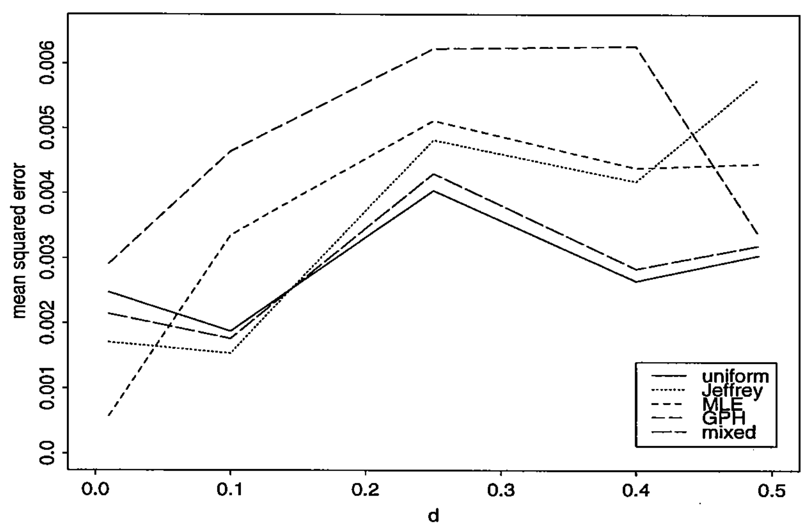


(b) $n = 200$

Figure 1: Jeffreys prior $\pi^*(\theta)$ and its approximation $\frac{\sqrt{C}}{\sigma^2}\frac{\Gamma(1-d)}{\sqrt{\Gamma(1-2d)\Gamma(2-2d)}}n^{-d}$, here $\sigma^2 = 1$



(a) $n = 50$



(b) $n = 200$

Figure 2: Mean squared errors for estimators of d

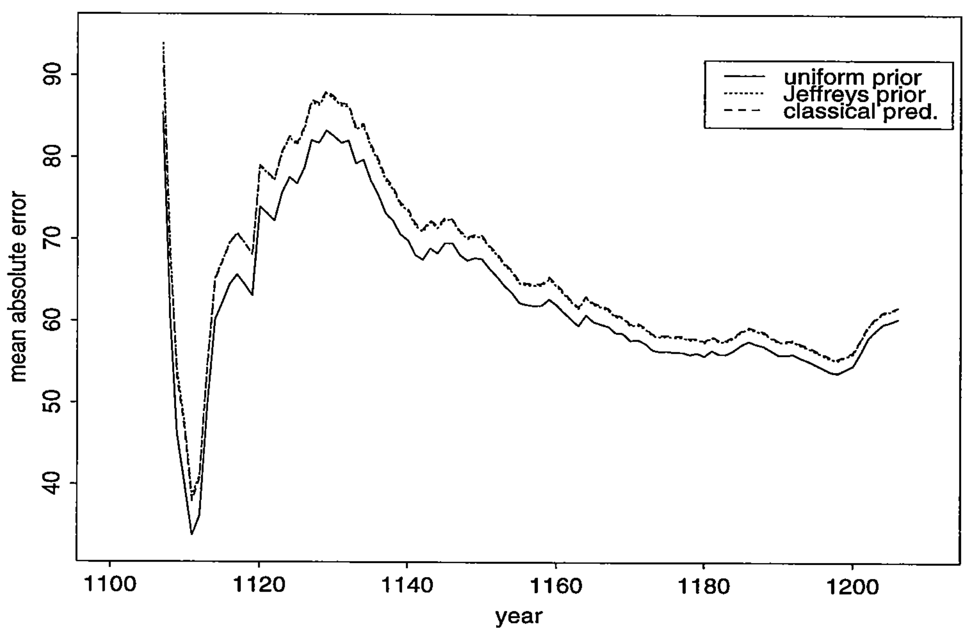


Figure 3: Cumulative mean absolute error for forecasts based on the previous 50 years observations

References

[1] Agiakloglou, C., Newbold, P., and Wohar, M. (1993). Bias in an estimator of the fractional difference parameter. *J. Time Ser. Anal.,* **14**, 235-246.

[2] Azais, J. M. and Lang, G. (1993). Nonparametric-estimation of the long-range dependence exponent (French). *Comp. Rendus. Academ. Sci. Ser. I-Math.*, **316**, 611-614.

[3] Beran, J. (1986). Estimation, testing and prediction for self-similar and related processes. *Ph.D. thesis, ETH Zürich.*

[4] Beran, J. (1994). *Statistics for Long-Memory Processes.* Chapman & Hall, New York.

[5] Beran, J. (1995). Maximum-likelihood-estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. Roy. Statist. Soc. B*, **57**, 659-672.

[6] Beran, J. and Terrin, N. (1994). Estimation of the long-memory parameter, based on a multivariate central limit theorem. *J. Time Ser. Anal.*, **15**, 269-84.

[7] Berger, J. O. (1986). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York.

[8] Brodsky, J. and, Hurvich, C. M. (1999). Multi-step forecasting for long-memory processes. *J. Forecasting*, **18**, 59-75.

[9] Dahlhaus, R. (1989). Efficient Parameter Estimation for Self-similar Processes. *Ann. Statist.*, **17**, 1749-1766.

[10] Delgado, M. A. and Robinson, P. M. (1994). New methods for the analysis of long-memory time series: Application to Spanish inflation. *J. Forecasting*, **13**, 97-107.

[11] Eisinga, R., Franses, P. H. and Ooms, M. (1999). Forecasting long memory left-right political orientations. *Int. J. Forecasting*, **15**, 185-199.

[12] Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.*, **4**, 221-237.

[13] Ghosh, M. and Rao, J. N. K. (1994). Small Area Estimation : An Appraisal. *Statist. Sci.*, **9**, 55 - 93.

[14] Giraitis, L., Robinson, P. M. and Samarov, A. (1997). Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long-range dependence. *J. Time Ser. Anal.*, **18**, 49-60.

[15] Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Tables of Integrals, Series and Products.* Academic Press, London.

[16] Graf, H. P. (1983). Long-range correlations and estimation of the self-similarity parameter. *Ph.D. thesis, ETH Zürich.*

[17] Grenander, U. and Szegö, G. (1958). *Toeplitz forms and their application.* Univ. of California Press, Berkeley.

[18] Hall, P., Koul, H. L. and Turlach, B. A. (1997). Note on convergence rates of semiparametric estimators of dependent index. *Ann. Statist.*, **25**, 1725-1739.

[19] Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. Invited paper with discussion. *J. Appl. Statist.,* **38**, 1-50.

[20] Hurvich, C. M. and Beltrao, K. I. (1994). Automatic semiparametric estimation of the memory parameter of a long-memory time series. *J. Time Ser. Anal.*, **15**. 285-302.

[21] Hurvich, C. M., Deo, R. and Brodsky, J. (1995). The mean squared error of Geweke and Porter-Hudaks's estimator of the memory parameter of a long memory time series. *Preprint.*

[22] Hurvich, C. M. and Ray, B. K. (1994). Estimation of the long-memory parameter for nonstationary or noninvertible fractionally differenced integrated processes. *Preprint.*

[23] Jannacek, G. (1994). Estimating d for long and short-memory time-series. *Environmetrics*, **5**, 255-271.

[24] Jensen, M. J. (1999). Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. *J. Forecasting*, **18**, 17-32.

[25] Kearns, P and Pagan, A. (1997). Estimating the density tail index for financial time series. *Rev. Econ. Statist.*, **79**, 171-175.

[26] Koop, G., Ley, E., Osiewalski, J. and Steel, M. F. J. (1997). Bayesian analysis of long memory and persistence using ARFIMA models. *J. Econometrics*, **76**, 149-169.

[27] Koul, H. L. and Giraitis, L. (1997). Estimation of the dependence parameter in linear regression with long-range dependent error. *Stochastic Process.*, **71**, 207-224.

[28] Palma, W. and Chan, N. H. (1997). Estimation and forecasting of long-memory processes with missing values. *J. Forecast.*, **16**, 395-410.

[29] Petris, G. (1996). Bayesian spectral analysis of long memory time series. *ASA Proceedings of the Section on Bayesian Statistical Science, AmerStat: VA,* 18-23.

[30] Ravishanker, N. and Ray, B. K. (1997). Bayesian analysis of vector ARFIMA processes. *Aust. J. Statist.*, **39**, 295-311.

[31] Ray, B. K. (1991). Forecasting long memory processes under misspecification. *ASA Proc. Bus. Econ. Statist. Sect.*, pp. 237-242.

[32] Ray, B. K. (1993). Modelling long-memory processes for optimal long-range prediction. *J. Time Ser. Anal.*, **18**, 511-526.

[33] Reisen, V. A. (1994). Estimation of the fractional difference parameter in the ARIMA(p, d, q) model using the smoothed periodogram. *J. Time Ser. Anal.*, **15**, 335-350.

[34] Robinson, G. K. (1991). The Estimation of Random Effects. *Statist. Sci.*, **6**, 15-51.

[35] Robinson, P. M. (1994). Semiparametric analysis of long-memory time series. *Ann. Statist.*, **22**, 515-539.

[36] Robinson, P. M. (1995). Gaussian Semiparametric estimation of long range dependence. *Ann. Statist.*, **23**, 1630-1661.

[37] Samarov, A. and Taqqu, M. S. (1988). On the efficiency of the sample mean in long-memory noise. *J. Time Ser. Anal.*, **9**, 191-200.

[38] Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components.* John Wiley, New York.

[39] Taqqu, M. S. and Teverovsky, V. (1998). On estimating the intensity of long-range dependence in finite and infinite variance series. In *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (R. Adler, R. Feldman and M. S. Taqqu, eds.), 177-217.

[40] Taqqu, M.S., Teverovsky, V. and Willinger, W. (1995). Estimators for long-range dependence: An empirical study. *Fractals*, **3**, 785-798.

[41] Veitch, D. and Abry, P. (1999). A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Trans. Inform. Th.*, **45**, 878-897.

[42] Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.*, **2**, 423-434.

[43] Yajima, Y. (1985). On estimation of long-memory time series models. *Austral. J. Statist.*, **27**, 303-320.

[44] Zygmund, A. (1979). *Trigonometric series.* Cambridge University Press, London.