Nonlinear Processes
in Geophysics
Discussions

Open Access

Efficient Bayesian inference for ARFIMA processes

T. Graves[1], R. B. Gramacy[2], C. L. E. Franzke[3], and N. W. Watkins[4,5,6,7]

[1]URS Corporation, London, UK
[2]The University of Chicago, Booth School of Business, Chicago, USA
[3]Meteorological Institute and Center for Earth System Research and Sustainability (CEN),
University of Hamburg, Hamburg, Germany
[4]Centre for the Analysis of Time Series, London School of Economics and Political Science,
London, UK
[5]Centre for Fusion Space and Astrophysics, University of Warwick, Coventry, UK
[6]Max Planck Institute for the Physics of Complex Systems, Dresden, Germany
[7]Centre for Complexity and Design, Open University, Milton Keynes, UK

Correspondence to: C. Franzke (christian.franzke@uni-hamburg.de)

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

NPGD

2, 573–618, 2015

Bayesian Inference

T. Graves et al.

Title Page

Abstract   Introduction

Conclusions   References

Tables   Figures

◄◄   ►►

◄   ►

Back   Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

Many geophysical quantities, like atmospheric temperature, water levels in rivers, and wind speeds, have shown evidence of long-range dependence (LRD). LRD means that these quantities experience non-trivial temporal memory, which potentially enhances their predictability, but also hampers the detection of externally forced trends. Thus, it is important to reliably identify whether or not a system exhibits LRD. In this paper we present a modern and systematic approach to the inference of LRD. Rather than Mandelbrot's fractional Gaussian noise, we use the more flexible Autoregressive Fractional Integrated Moving Average (ARFIMA) model which is widely used in time series analysis, and of increasing interest in climate science. Unlike most previous work on the inference of LRD, which is frequentist in nature, we provide a systematic treatment of Bayesian inference. In particular, we provide a new approximate likelihood for efficient parameter inference, and show how nuisance parameters (e.g. short memory effects) can be integrated over in order to focus on long memory parameters, and hypothesis testing more directly. We illustrate our new methodology on the Nile water level data, with favorable comparison to the standard estimators.

1 Introduction

Many natural processes are sufficiently complex that a stochastic model is essential, or at the very least an efficient description (Watkins, 2013). Such a process will be specified by several properties, of which a particularly important one is the degree of memory in a time series, often expressed through a characteristic autocorrelation time over which fluctuations will decay in magnitude. In this paper, however, we are concerned with specific types of stochastic processes that are capable of possessing "long memory", or "long-range dependence" (LRD) (Beran, 1994a; Palma, 2007; Beran et al., 2013). Long memory is the notion of there being correlation between the present and *all* points in the past. A standard definition is that a (finite variance, stationary)

process has *long memory* if its autocorrelation function (ACF) has power-law decay: $\rho(\cdot)$ such that $\rho(k) \sim c_\rho \, k^{2d-1}$ as $k \to \infty$, for some non-zero constant c_ρ, and where $0 < d < \frac{1}{2}$. The parameter d is the memory parameter; if $d = 0$ the process does not exhibit long memory, while if $-\frac{1}{2} < d < 0$ the process is said to be anti-persistent.

⁵ The asymptotic power law form of the ACF corresponds to an absence of a characteristic decay timescale, in striking contrast to many standard (stationary) stochastic processes where the effect of each data point decays so fast that it rapidly becomes indistinguishable from noise. An example of the latter is the exponential ACF where the e-folding scale sets a characteristic correlation time. The study of processes that *do*
¹⁰ possess long memory is important because they exhibit unusual properties, because many familiar mathematical results fail to hold, and because of the numerous examples of data sets where LRD is seen.

The study of long memory originated in the 1950s in the field of hydrology, where studies of the levels of the river Nile (Hurst, 1951) demonstrated anomalously fast
¹⁵ growth of the rescaled range of the time series. After protracted debates[1] about whether this was a transient (finite time) effect, the mathematical pioneer Benoît B. Mandelbrot showed that if one retained the assumption of stationarity, novel mathematics would then be essential to sufficiently explain the Hurst effect. In doing so he rigorously defined the concept of long memory (Mandelbrot and Van Ness, 1968;
²⁰ Mandelbrot and Wallis, 1968).

Most research into long memory and its properties has been based on classical statistical methods, spanning parametric, semi-parametric and non-parametric modeling (see Beran et al., 2013, for a review). Very few Bayesian methods have been studied, most probably due to computational difficulties. The earliest works are parametric and
²⁵ include Koop et al. (1997), Pai and Ravishanker (1998), and Hsu and Breidt (2003). If computational challenges could be mitigated, the Bayesian paradigm would offer advantages over classical methods including flexibility in specification of priors (i.e. phys-

---

[1]For a detailed exposition of this period of mathematical history, see Graves et al. (2014).

Bayesian Inference

T. Graves et al.

Full Screen / Esc

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

ical expertise could be used to elicit an informative prior). It would offer the ability to marginalize out aspects of a model apparatus and data, such as short memory or seasonal effects and missing observations, so that statements about long memory effects can be made unconditionally.

Towards easing the computational burden, we focus on the ARFIMA class of processes (Granger and Joyeux, 1980; Hosking, 1981) as the basis of developing a systematic and unifying Bayesian framework for modeling a variety of common time series phenomena, with particular emphasis on detecting potential long memory effects. ARFIMA has become very popular in statistics and econometrics because it is generalizable and its connection to the ARMA family (and to fractional Gaussian noise) is relatively transparent. A key property of ARFIMA is its ability to simultaneously yet separately model long and short memory.

Both Liseo et al. (2001) and Holan et al. (2009) argued, echoing a sentiment in the classical statistics literature, that full parametric long memory models (like ARFIMA) are "too hard" to work with. Furthermore, often d is the only object of real interest, and consideration of a single class of models, such as ARFIMA, is too restrictive. They therefore developed methods which have similarities to classical periodograms.

We think ARFIMA deserves another look – that many of the above drawbacks, to ARFIMA in particular and Bayesian computation more generally, can be addressed with a careful treatment. We provide a new approximate likelihood for ARFIMA processes that can be computed quickly for repeated evaluation on large time series, and which underpins an efficient MCMC scheme for Bayesian inference. Our sampling scheme can be best described as a modernization of a blocked MCMC scheme proposed by Pai and Ravishanker (1998) – adapting it to the approximate likelihood and extending it to handle a richer form of (known) short memory effects. We then further extend the analysis to the case where the short memory form is unknown, which requires transdimensional MCMC. This aspect is similar to the work of Ehlers and Brooks (2008) who considered the simpler ARIMA model class, and to Holan et al. (2009) who worked with a nonparametric long memory process. Our contribution has aspects in common

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

with Eğrioğlu and Günay (2010) who presented a more limited method focused on model selection rather than averaging. The advantage of averaging is that the unknown form of short memory effects can be integrated out, focusing on long-memory without conditioning on nuisance parameters.

The aim of this paper is to introduce an efficient Bayesian algorithm for the inference of the parameters of the ARFIMA(p, d, q) model, with particular emphasis on the LRD parameter d. Our Bayesian inference algorithm has been designed in a flexible fashion so that, for instance, the innovations can come from a wide class of different distributions; e.g. α-stable or t distribution (to be published in a companion paper). The remainder of the paper is organized as follows. Section 2 summarizes the ARFIMA model required for our purposes. Section 3 discusses the important numerical calculation of likelihoods, representing a hybrid between earlier classical statistical methods, and our new contributions towards a full Bayesian approach. Section 4 describes our proposed Bayesian framework and methodology in detail, focusing on long-memory only. Then, in Sect. 5, we consider extensions for additional short memory. Empirical illustration and comparison of all methods is provided in Sect. 6. The paper concludes with a discussion in Sect. 7 focused on the potential for further extension.

2   Time series definitions and the ARFIMA model

Because ARFIMA models have not yet been very widely used in the geosciences we provide here a brief review of them. Readers familiar with ARFIMA models can skip this section.

We define an autocovariance ACV $\gamma(\cdot)$ of a weakly stationary process as $\gamma(k) = \mathrm{Cov}(X_t, X_{t+k})$, where k is referred to as the (time) "lag". The (normalized) autocorrelation function ACF $\rho(\cdot)$ is defined as: $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$. Another useful time domain tool is the "backshift" operator $\mathcal{B}$, where $\mathcal{B}X_t = X_{t-1}$, and powers of $\mathcal{B}$ are defined iteratively: $\mathcal{B}^k X_t = \mathcal{B}^{k-1}(\mathcal{B}X_t) = \mathcal{B}^{k-1}X_{t-1} = \cdots = X_{t-k}$. A stationary process $\{X_t\}$ is said to

be *causal* if there exists a sequence of coefficients $\{\psi_k\}$, with finite total mean square $\sum_{k=0}^{\infty}\psi_k^2 < \infty$ such that for all t, a given member of the process can be expanded as a power series in the backshift operator acting on the "innovations", $\{\varepsilon_t\}$:

$$X_t = \Psi(\mathcal{B})\varepsilon_t, \quad \text{where} \quad \Psi(z) = \sum_{k=0}^{\infty} \psi_k z^k. \tag{1}$$

5   The innovations are a white (i.e. stationary, zero mean, iid) noise process with variance σ^2. Causality specifies that for every t, X_t can only depend on the past and present values of the innovations $\{\varepsilon_t\}$.

    A process $\{X_t\}$ is said to be an *auto-regressive process of order p*, AR(p), if for all t:

$$\Phi(\mathcal{B})X_t = \varepsilon_t, \quad \text{where} \quad \Phi(z) = 1 + \sum_{k=1}^{p} \phi_k z^k, \quad \text{and} \quad (\phi_1, \ldots, \phi_p) \in \mathbb{R}^p. \tag{2}$$

10   AR(p) processes are invertible, stationary and causal if and only if $\Phi(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. $\{X_t\}$ is said to be a *moving average process of order q*, MA(q), if

$$X_t = \Theta(\mathcal{B})\varepsilon_t, \quad \text{where} \quad \Theta(z) = 1 + \sum_{k=1}^{q} \theta_k z^k, \quad \text{and} \quad (\theta_1, \ldots, \theta_p) \in \mathbb{R}^q, \tag{3}$$

for all t.[2] MA(q) processes are stationary and causal, and are invertible if and only if $\Theta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.

15   A natural extension of the AR and MA classes arises by combining them (Box and Jenkins, 1970). The process $\{X_t\}$ is said to be an *auto-regressive moving average (ARMA) process* process of orders p and q, ARMA(p, q), if for all t:

$$\Phi(\mathcal{B})X_t = \Theta(\mathcal{B})\varepsilon_t. \tag{4}$$

---

[2]Many authors define $\Phi(z) = 1 - \sum \phi_k z^k$. Our version emphasises connections between Φ and Eqs. (2) and (3).

Although there is no simple closed form for the ACV of an ARMA process with arbitrary p and q, so long as the process is causal and invertible, then $|\rho(k)| \leq Cr^k$, for $k > 0$, i.e. it decays exponentially fast. In other words, although correlation between nearby points may be high, dependence between distant points is negligible.

Before turning to "long memory", we require one further result. Under some extra conditions, stationary processes with ACV $\gamma(\cdot)$ possess a spectral density function (SDF) $f(\cdot)$ defined such that: $\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) \, d\lambda$, $\forall k \in \mathbb{Z}$. This can be inverted to obtain an explicit expression for the SDF (e.g. Brockwell and Davis, 1991, Sect. 4.3): $f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\lambda}$, where $-\pi \leq \lambda \leq \pi$.[3] Finally, the SDF of an ARMA process is

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\Theta(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2}, \qquad 0 \leq \lambda \leq \pi. \tag{5}$$

The restriction $|d| < \frac{1}{2}$ is necessary to ensure stationarity; clearly if $|d| \geq \frac{1}{2}$ the ACF would not decay. The continuity between stationary and non-stationary processes around $|d| = \frac{1}{2}$ is similar to that which occurs for AR(1) process with $|\phi_1| \to 1$ (such processes are stationary for $|\phi_1| < 1$, but the case $|\phi_1| = 1$ is the non-stationary random-walk).

There are a number of alternative definitions of LRD, one of which is particularly useful, as it considers the frequency domain: A stationary process has long memory when its SDF follows $f(\lambda) \sim c_f \lambda^{-2d}$, as $\lambda \to 0^+$ for some positive constant c_f, and where $0 < d < \frac{1}{2}$.

The simplest way of *creating* a process which exhibits long memory is through the SDF. Consider $f(\lambda) = |1 - e^{i\lambda}|^{-2d}$, where $0 < |d| < \frac{1}{2}$. By simple algebraic manipulation, this is equivalently $f(\lambda) = \left(2 \sin \frac{\lambda}{2}\right)^{-2d}$, from which we deduce that $f(\lambda) \sim \lambda^{-2d}$ as $\lambda \to 0^+$. Therefore, assuming stationarity, the process which has this SDF (or any scalar

---

[3]Since ACV of a stationary process is an even function of lag, the above equation implies that the associated SDF is an even function. One therefore only needs to be interested positive arguments: $0 \leq \lambda \leq \pi$.

multiple of it) is a long memory process. More generally, a process having spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| 1 - e^{i\lambda} \right|^{-2d}, \qquad 0 < \lambda \le \pi \tag{6}$$

is called *fractionally integrated* with memory parameter d, FI(d) (Barnes and Allan, 1966; Adenstedt, 1974). The full trichotomy of negative, short, and long memory is determined solely by d.

In practice this model is of limited appeal to time series analysts because the entire memory structure is determined by just one parameter, d. One often therefore generalizes it by taking any short memory SDF $f^*(\cdot)$, and defining a new SDF: $f(\lambda) = f^*(\lambda) \left| 1 - e^{i\lambda} \right|^{-2d}$, $0 \le \lambda \le \pi$. An obvious class of short memory processes to use this way is ARMA. Taking f^* from Eq. (5) yields so-called auto-regressive fractionally integrated moving average process with parameter d, and orders p and q (ARFIMA(p, d, q)), having SDF:

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\Theta(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2} |1 - e^{i\lambda}|^{-2d}, \qquad 0 \le \lambda \le \pi. \tag{7}$$

Choosing $p = q = 0$ recovers FI(d) $\equiv$ ARFIMA($0, d, 0$)

Practical utility from the perspective of (Bayesian) inference demands finding a representation in the temporal domain. To obtain this, consider the operator $(1 - \mathcal{B})^d$ for real $d > -1$, which is formally defined using the generalized form of the binomial expansion (Brockwell and Davis, 1991, Eq. 13.2.2):

$$(1 - \mathcal{B})^d =: \sum_{k=0}^{\infty} \pi_k^{(d)} \mathcal{B}^k, \quad \text{where} \quad \pi_k^{(d)} = (-1)^k \frac{1}{\Gamma(k + 1)} \frac{\Gamma(d + 1)}{\Gamma(d - k + 1)}. \tag{8}$$

From this observation, one can show that $X_t = (1 - \mathcal{B})^{-d} Z_t$, where $\{Z_t\}$ is an ARMA process, has SDF shown in Eq. (7). The operator $(1 - \mathcal{B})^d$ is called the "fractional differencing" operator since it allows a degree of differencing between zeroth and first

order. The process $\{X_t\}$ is fractionally "inverse-differenced", i.e. it is an "integrated" process. The operator is used to redefine both the ARFIMA$(0, d, 0)$ and more general ARFIMA(p, d, q) processes in the time domain. A process $\{X_t\}$ is an ARFIMA$(0, d, 0)$ process if for all t: $(1 - \mathcal{B})^d X_t = \varepsilon_t$. Likewise, a process $\{X_t\}$ is an ARFIMA(p, d, q) pro-

5 cess if for all t: $\Phi(\mathcal{B})(1 - \mathcal{B})^d X_t = \Theta(\mathcal{B})\varepsilon_t$, where Φ and Θ are given in Eqs. (2) and (3) respectively.

Finally, to connect back to our first definition of long memory, consider the ACV of the ARFIMA$(0, d, 0)$ process. By using the definition of spectral density to directly integrate $f(\lambda)$ in Eq. (6), and an alternative expression for $\pi_k^{(d)}$ in Eq. (8)

10
$$\pi_k^{(d)} = \frac{1}{\Gamma(k+1)} \frac{\Gamma(k-d)}{\Gamma(-d)}, \tag{9}$$

one can obtain the following representation of the ACV of the ARFIMA$(0, d, 0)$ process:

$$\gamma_d(k; \sigma) = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} \frac{\Gamma(k+d)}{\Gamma(1+k-d)}. \tag{10}$$

Because the parameter σ^2 is just a scalar multiplier, we may simplify notation by defining $\gamma_d(k) = \gamma_d(k; \sigma)/\sigma^2$, whereby $\gamma_d(\cdot) \equiv \gamma_d(\cdot; 1)$. Then the ACF is:

15
$$\rho_d(k) = \frac{\Gamma(1-d)}{\Gamma(d)} \frac{\Gamma(k+d)}{\Gamma(1+k-d)}, \tag{11}$$

from which Stirling's approximation gives $\rho_d(k) \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1}$, confirming a power-law relationship for the ACF. Finally, note that Eq. (9) can be used to represent ARFIMA$(0, d, 0)$ as an AR(∞) process, as $X_t + \sum_{k=1}^{\infty} \pi_k^{(d)} X_{t-k} = \varepsilon_t$. And noting that $\psi_k^{(d)} = \pi_k^{(-d)}$, leads to the following MA(∞) analog: $X_t = \sum_{k=0}^{\infty} \frac{1}{\Gamma(k+1)} \frac{\Gamma(k+d)}{\Gamma(d)} \varepsilon_{t-k}$.

Bayesian Inference

T. Graves et al.

3 Likelihood evaluation for Bayesian inference

For now we restrict our attention to (a Bayesian) analysis of an ARFIMA$(0,d,0)$ process, having no short-ranged ARMA components, placing emphasis squarely on the memory parameter d.

Here we develop an efficient and new scheme for evaluating the (log) likelihood, via approximation. This scheme is very flexible in the sense that it seamlessly allows to use different noise distributions (like a t distribution instead of a Gaussian; this will be reported elsewhere). Throughout, suppose that we have observed the vector $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ as a realization of a stationary, causal and invertible ARFIMA$(0,d,0)$ process $\{X_t\}$ with mean $\mu \in \mathbb{R}$. The innovations will be assumed to be independent, and taken from a zero-mean *location-scale probability density* $f(\cdot; 0, \sigma, \boldsymbol{\lambda})$, which means the density can be written as $f(x; \delta, \sigma, \boldsymbol{\lambda}) \equiv \frac{1}{\sigma} f\left(\frac{x-\delta}{\sigma}; 0, 1, \boldsymbol{\lambda}\right)$. The parameters δ and σ are called the "location" and "scale" parameters respectively. The m dimensional $\boldsymbol{\lambda}$ is a "shape" parameter (if it exists, i.e. $m > 0$). An common example is the Gaussian $\mathcal{N}(\mu, \sigma^2)$, where $\delta \equiv \mu$ and there is $\boldsymbol{\lambda}$. We classify the four parameters μ, σ, $\boldsymbol{\lambda}$, and d, into three distinct classes: (1) the mean of process, μ; (2) innovation distribution parameters, $\boldsymbol{\upsilon} = (\sigma, \boldsymbol{\lambda})$; and (3) memory structure, d. Together, $\boldsymbol{\psi} = (\mu, \boldsymbol{\upsilon}, \boldsymbol{\omega})$, where $\boldsymbol{\omega}$ will later encompass the short-range parameters p and q.

Our proposed likelihood approximation uses a truncated AR(∞) approximation (cf. Haslett and Raftery, 1989). We first re-write the AR(∞) approximation of ARFIMA$(0,d,0)$ to incorporate the unknown parameter μ, and drop the (d) superscript for convenience: $X_t - \mu = \varepsilon_t - \sum_{k=1}^{\infty} \pi_k (X_{t-k} - \mu)$. Then we truncate this AR(∞) representation to obtain an AR(P) one, with P large enough to retain low frequency effects, e.g. $P = n$.

We denote: $\Pi_P = \sum_{k=0}^{P} \pi_k$ and, with $\pi_0 = 1$, rearrange terms to obtain the following modified model:

$$X_t = \varepsilon_t + \Pi_P \mu - \sum_{k=1}^{P} \pi_k X_{t-k}. \tag{12}$$

It is now possible to write down a *conditional* likelihood. For convenience the notation $x_k = (x_1, \ldots, x_k)^\top$ for $k = 1, \ldots, n$ will be used (and x_0 is interpreted as appropriate where necessary). Denote the unobserved P vector of random variables $(x_{1-P}, \ldots, x_{-1}, x_0)^\top$ by x_A (in the Bayesian context these will be "auxiliary", hence "A").

Consider the likelihood $L(x|\psi)$ as a joint density which can be factorized as a product of conditionals. Writing $g_t(x_t|x_{t-1}, \psi)$ for the density of X_t conditional on x_{t-1}, we obtain $L(x|\psi) = \prod_{t=1}^{n} g_t(x_t|x_{t-1}, \psi)$.

This is still of little use because the g_t may have a complicated form. However by further conditioning on x_A, and writing $h_t(x_t|x_A, x_{t-1}, \psi)$ for the density of X_t conditional on x_{t-1} *and* x_A, we obtain: $L(x|\psi, x_A) = \prod_{t=1}^{n} h_t(x_t|x_A, x_{t-1}, \psi)$. Returning to Eq. (12) observe that, conditional on both the observed and *un*observed past values, X_t is simply distributed according to the innovations' density f with a suitable change in location: $X_t|x_{t-1}, x_A \sim f\left(\cdot; \left[\Pi_P \mu - \sum_{k=1}^{P} \pi_k x_{t-k}\right], \sigma, \lambda\right)$. Then using location-scale representation:

$$h_t(x_t|x_A, x_{t-1}, \psi) \approx f\left(x_t; \left[\Pi_P \mu - \sum_{k=1}^{P} \pi_k x_{t-k}\right], \sigma, \lambda\right) \tag{13}$$

$$\equiv \frac{1}{\sigma} f\left(\frac{c_t - \Pi_P \mu}{\sigma}; 0, 1, \lambda\right), \quad \text{where} \quad c_t = \sum_{k=0}^{P} \pi_k x_{t-k}, \qquad t = 1, \ldots, n.$$

Therefore, $L(x|\psi, x_A) \approx \sigma^{-n} \prod_{t=1}^{n} f\left(\frac{c_t - \Pi_P \mu}{\sigma}; \lambda\right)$, or equivalently:

$$\ell(x|\psi, x_A) \approx -n \log \sigma + \sum_{t=1}^{n} \log\left\{f\left(\frac{c_t - \Pi_P \mu}{\sigma}; \lambda\right)\right\}. \tag{14}$$

Evaluating this expression efficiently depends upon efficient calculation of $c = (c_1, \ldots, c_n)^t$ and $\log f(\cdot)$. From Eq. (13), c is a convolution of the augmented data, (x, x_A), and coefficients depending on d, which can be evaluated quickly in $\mathbb{R}$ via

`convolve` via FFT. Consequently, evaluation of the *conditional* likelihood in the Gaussian case costs only $\mathcal{O}(n \log n)$ – a clear improvement over the "exact" method. Obtaining the *un*conditional likelihood requires marginalization over $\boldsymbol{x}_A$, which is analytically infeasible. However this conditional form will suffice in the context of our Bayesian in-
ferential scheme, presented below.

4 A Bayesian approach to long memory inference

We are now ready to consider Bayesian inference for ARFIMA$(0, d, 0)$ processes. Our method can be succinctly described as a modernization of the blocked MCMC method of Pai and Ravishanker (1998). Isolating parameters by blocking provides significant scope for modularization which helps accommodate our extensions for short memory. Pairing with efficient likelihood evaluations allows much longer time series to be entertained than ever before. Our description begins with appropriate specification of priors which are more general than previous choices, yet still encourages tractable inference. We then provide the relevant updating calculations for all parameters, including those for auxiliary parameters $\boldsymbol{x}_A$.

We follow earlier work (Koop et al., 1997; Pai and Ravishanker, 1998) and assume a priori independence for components of ψ. Each component will leverage familiar prior forms with diffuse versions as limiting cases. Specifically, we use a diffuse Gaussian prior on μ: $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$, with σ_0 large. The improper flat prior is obtained as the limiting distribution when $\sigma_0 \to \infty$: $p_\mu(\mu) \propto 1$. We place a gamma prior on the precision $\tau = \sigma^{-2}$ implying a *Root-Inverse Gamma* distribution $\mathcal{R}(\alpha_0, \beta_0)$ for σ, with density $f(\sigma) = \frac{2}{\Gamma(\alpha)} \beta_0^{\alpha_0} \sigma^{-(2\alpha_0 + 1)} \exp\left(-\frac{\beta_0}{y^2}\right)$, $\sigma > 0$. A diffuse/improper prior is obtained as the limiting distribution when $\alpha_0, \beta_0 \to 0$: $p_\sigma(\sigma) \propto \sigma^{-1}$. Finally, we specify $d \sim \mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$.

Updating μ: following Pai and Ravishanker (1998), we use a symmetric random walk (RW) MH update with proposals $\xi_\mu \sim \mathcal{N}(\mu, \sigma_\mu^2)$, for some σ_μ^2. The acceptance ratio is

$$A_\mu(\mu, \xi_\mu) = \sum_{t=1}^{n} \log \left\{ f\left(\frac{c_t - \Pi_P \xi_\mu}{\sigma}; \lambda\right) \right\} - \sum_{t=1}^{n} \log \left\{ f\left(\frac{c_t - \Pi_P \mu}{\sigma}; \lambda\right) \right\}$$
$$+ \log \left[\frac{p_\mu(\xi_\mu)}{p_\mu(\mu)}\right] \qquad (15)$$

under the approximate likelihood.

Updating σ: we diverge from Pai and Ravishanker (1998) here, who suggest independent MH with moment-matched inverse gamma proposals, finding poor performance under poor moment estimates. We instead prefer a Random Walk (RW) Metropolis-Hastings (MH) approach, which we conduct in log space since the domain is $\mathbb{R}^+$. Specifically, set: $\log \xi_\sigma = \log \sigma + \upsilon$, where $\upsilon \sim \mathcal{N}(0, \sigma_\sigma^2)$ for some σ_σ^2. $\xi_\sigma|\sigma$ is lognormal and we obtain: $\frac{q(\sigma; \xi_\sigma)}{q(\xi_\sigma; \sigma)} = \frac{\xi_\sigma}{\sigma}$. Recalling Eq. (15) the MH acceptance ratio under the approximate likelihood is

$$A_\sigma(\sigma, \xi_\sigma) = \sum_{t=1}^{n} \log \left\{ f\left(\frac{c_t - \Pi_P \mu}{\xi_\sigma}; \lambda\right) \right\} - \sum_{t=1}^{n} \log \left\{ f\left(\frac{c_t - \Pi_P \mu}{\sigma}; \lambda\right) \right\}$$
$$+ \log \left[\frac{p_\sigma(\xi_\sigma)}{p_\sigma(\sigma)}\right] + (n-1)\log\left[\frac{\sigma}{\xi_\sigma}\right].$$

The MH algorithm, applied alternately in a Metropolis-within-Gibbs fashion to the parameters μ and σ, works well. However *actual* Gibbs sampling is an efficient alternative in this two-parameter case (i.e. for known d, see Graves, 2013).

Update of d: updating the memory parameter d is far less straightforward than either μ or σ. Regardless of the innovations' distribution, the conditional posterior $\pi_{d|\psi_{-d}}(d|\boldsymbol{\psi}_{-d}, \boldsymbol{x})$ is not amenable to Gibbs sampling. We use RW proposals from truncated Gaussian $\xi_d \sim \mathcal{N}^{(a,b)}(\mu, \sigma^2)$, with density

$$f(x; \mu, \sigma, a, b) = \frac{1}{\sigma} \frac{\phi[(x-\mu)/\sigma]}{\Phi[(b-\mu)/\sigma] - \Phi[(a-\mu)/\sigma]}, \qquad a < x < b. \qquad (16)$$

In particular, we use $\xi_d|d \sim \mathcal{N}^{(-1/2,1/2)}(d,\sigma_d^2)$ via rejection sampling from $\mathcal{N}(d,\sigma_d^2)$ until $\xi_d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. Although this may seem inefficient, it is perfectly acceptable: as an example, if $\sigma_d = 0.5$ the expected number of required variates is still less than 2, regardless of d. More refined methods of directly sampling from truncated normal distributions exist – see for example Robert (1995) – but we find little added benefit in our context.

A useful cancellation in $q(d; \xi_d)/q(\xi_d; d)$ obtained from Eq. (16) yields

$$A_d = \ell(\boldsymbol{x}|\xi_d, \boldsymbol{\psi}_{-d}) - \ell(\boldsymbol{x}|d, \boldsymbol{\psi}_{-d}) + \log\left[\frac{p_d(\xi_d)}{p_d(d)}\right]$$

$$+ \log\left\{\frac{\Phi\left[\left(\frac{1}{2} - d\right)/\sigma_d\right] - \Phi\left[\left(-\frac{1}{2} - d\right)/\sigma_d\right]}{\Phi\left[\left(\frac{1}{2} - \xi_d\right)/\sigma_d\right] - \Phi\left[\left(-\frac{1}{2} - \xi_d\right)/\sigma_d\right]}\right\}.$$

Denote $\xi_{c_t} = \sum_{k=0}^{P} \xi_{\pi_k} x_{t-k}$ for $t = 1, \dots, n$, where $\{\xi_{\pi_k}\}$ are the proposed coefficients $\{\pi_k^{(\xi_d)}\}$; $\pi_k^{(d)} = \frac{1}{\Gamma(k+1)}\frac{\Gamma(k-d)}{\Gamma(-d)}$. Denote $\xi_{\Pi_P} = \sum_{k=0}^{P} \xi_{\pi_k}$. Then in the approximate case:

$$A_d = \sum_{t=1}^{n} \log\left\{f\left(\frac{\xi_{c_t} - \xi_{\Pi_P}\mu}{\sigma}; \boldsymbol{\lambda}\right)\right\} - \sum_{t=1}^{n} \log\left\{f\left(\frac{c_t - \Pi_P\mu}{\sigma}; \boldsymbol{\lambda}\right)\right\}$$

$$+ \log\left[\frac{p_d(\xi_d)}{p_d(d)}\right] + \log\left\{\frac{\Phi\left[\left(\frac{1}{2} - d\right)/\sigma_d\right] - \Phi\left[\left(-\frac{1}{2} - d\right)/\sigma_d\right]}{\Phi\left[\left(\frac{1}{2} - \xi_d\right)/\sigma_d\right] - \Phi\left[\left(-\frac{1}{2} - \xi_d\right)/\sigma_d\right]}\right\}. \tag{17}$$

Optional update of x_A: when using the approximate likelihood method, one must account for the auxiliary variables $\boldsymbol{x}_A$, a P vector (where $P = n$ is sensible). We find that, in practice, it is not necessary to update all the auxiliary parameters at each iteration. In fact the method can be shown to work perfectly well, empirically, if we *never* update them, provided they are given a sensible initial value (such as the sample mean of the observed data $\bar{x}$). This is not an uncommon tactic in the long memory (big-n) context (e.g. Beran, 1994b); for further discussion refer to Graves (2013, Appendix C).

|◀ | ▶|

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

For a full MH approach, we recommend an independence sampler to "backward project" the observed time series. Specifically, first relabel the observed data: $y_{-i} = x_{i+1}$, $i = 0, \ldots, n-1$. Then use the vector $(y_{-(n-1)}, \ldots, y_{-1}, y_0)^t$ to generate a new vector of length n, $(Y_1, \ldots, Y_n)^t$ where Y_t via Eq. (12): $Y_t = \varepsilon_t + \Pi_P \mu - \sum_{k=1}^{n} \pi_k Y_{t-k}$, where the coefficients $\{\pi\}$ are determined by the current value of the memory parameter(s). Then take the proposed x_A, denoted ξ_{x_A}, as the reverse sequence: $\xi_{x_{-i}} = y_{i+1}$, $i = 0, \ldots, n-1$. Since this is an independence sampler, calculation of the acceptance probability is straightforward. It is only necessary to evaluate the proposal density $q(\xi_{x_A} | x, \psi)$. But this is easy using the results from Sect. 3. For simplicity, we prefer uniform prior for x_A.

Besides simplicity, justification for this approach lies primarily in is preservation of the auto-correlation structure – this is clear since the ACF is symmetric in time. The proposed vector has a low acceptance rate, and the potential remedies (e.g. multiple-try methods) seem unnecessarily complicated given the success of the simpler method.

5 Extensions to accommodate short memory

Simple ARFIMA$(0, d, 0)$ are mathematically convenient but have limited practical applicability because the entire memory structure is determined by just one parameter, d. Although d is often of primary interest, it may be unrealistic to assume no short memory effects. This issue is often implicitly acknowledged since semi-parametric estimation methods, such as those used as comparators in Sect. 6.1, are motivated by a desire to circumvent the problem of specifying precisely (and inferring) the form of short memory (i.e. the values of p and q in an ARIMA model). Full parametric Bayesian modeling of ARFIMA(p, d, q) processes represents an essentially untried alternative, primarily due to computational challenges. Related, more discrete, alternatives show potential. Pai and Ravishanker (1998) considered all four models with $p, q \leq 1$, whereas Koop et al. (1997) considered sixteen with $p, q \leq 3$.

Such approaches, especially ones allowing larger p, q, can be computationally burdensome as much effort is spent modeling unsuitable processes towards a goal (in-

ferring p, q) which is not of primary interest (d is). To develop an efficient, fully-parametric, Bayesian method of inference that properly accounts for varying models, and to marginalize out these nuisance quantities, we use reversible-jump (RJ) MCMC (Green, 1995). We extend the parameter space to include the set of models (p and q), with chains moving *between* and within models, and focus on the marginal posterior distribution of d obtained by (Monte Carlo) integration over all models and parameters therein. RJ methods have previously been applied to both auto-regressive models (Vermaak et al., 2004), and full ARMA models (Ehlers and Brooks, 2006, 2008). In the long memory context, Holan et al. (2009) applied RJ to FEXP processes. However for ARFIMA, the only related work we are aware of is by Eğrioğlu and Günay (2010) who demonstrated a promising if limited alternative.

Below we show how the likelihood may be calculated with extra short-memory components when p and q are known, and subsequently how Bayesian inference can be applied in this case. Then, the more general case of unknown p and q via RJ is described.

5.1 Likelihood derivation and inference for known short memory

Recall that short memory components of an ARFIMA process are defined by the AR and MA polynomials, Φ and Θ respectively, (see Sect. 2). Here, we distinguish between the polynomial, Φ, and the vector of its coefficients, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$. When the polynomial degree is required explicitly, bracketed superscripts will be used; $\Phi^{(p)}$, $\boldsymbol{\phi}^{(p)}$, $\Theta^{(p)}$, $\boldsymbol{\theta}^{(p)}$, respectively.

We combine the short memory parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ with d to create a single "memory" parameter, $\boldsymbol{\omega} = (\boldsymbol{\phi}, \boldsymbol{\theta}, d)$. For a given unit-variance ARFIMA(p, d, q) process, we denote its ACV by $\gamma_{\boldsymbol{\omega}}(\cdot)$, with $\gamma_d(\cdot)$ and $\gamma_{\boldsymbol{\phi}, \boldsymbol{\theta}}(\cdot)$ those of the relevant unit-variance ARFIMA($0, d, 0$) and ARMA(p, q) processes respectively. The SDF of the unit-variance ARFIMA(p, d, q) process is written as $f_{\boldsymbol{\omega}}(\cdot)$, and its covariance matrix is $\Sigma_{\boldsymbol{\omega}}$.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

An "exact" likelihood evaluation requires an explicit calculation of the ACV $\gamma_{\boldsymbol{\omega}}(\cdot)$, however there is no simple closed form for arbitrary ARFIMA processes. Fortunately, our proposed approximate likelihood method of Sect. 3 can be ported over directly. Given the coefficients $\{\pi_k^{(d)}\}$ and polynomials Φ and Θ, it is trivial to calculate the $\{\pi_k^{(\boldsymbol{\omega})}\}$ co-
5   efficients required by again applying the numerical methods of Brockwell and Davis (1991, Sect. 3.3).

To focus the exposition, consider the simple, yet useful, ARFIMA$(1, d, 0)$ model where the full memory parameter is $\boldsymbol{\omega} = (d, \phi_1)$. Because the parameter spaces of d and ϕ_1 are independent, it is simplest to update each of these parameters separately; d with
10 the methods of Sect. 4 and ϕ_1 similarly: $\xi_{\phi_1}|\phi_1 \sim \mathcal{N}^{(-1,1)}(\phi_1, \sigma_{\phi_1}^2)$, for some $\sigma_{\phi_1}^2$. In practice however, the posteriors of d and ϕ_1 typically exhibit significant correlation so independent proposals are inefficient. One solution would be to parametrize to some d^* and orthogonal ϕ_2^*, but the interpretation of d^* would not be clear. An alternative to explicit reparametrisation is to update the parameters jointly, but in such a way that
15 proposals are aligned with the correlation structure. This will ensure a reasonable acceptance rate and mixing.

To propose parameters in the manner described above, a two-dimensional, suitably truncated Gaussian random walk, with covariance matrix aligned with the posterior covariance, is required. To make proposals of this sort, and indeed for arbitrary $\boldsymbol{\omega}$ in larger
20 p and q cases, requires sampling from a *hypercuboid*-truncated MVN $\mathcal{N}_r^{(a,b)}(\boldsymbol{\omega}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}})$, where $(\boldsymbol{a}, \boldsymbol{b})$ describe the coordinates of the hypercube. We find that rejection sampling based unconstrained similarly parameterized MVNs samples (e.g. using `mvtnorm`, Genz et al., 2012) works well, because in the RW setup the mode of the distribution always lies inside the hypercuboid. Returning to the specific ARFIMA$(1, d, 0)$ case,
25 clearly $r = 2$, $\boldsymbol{b} = (0.5, 1)$ and $\boldsymbol{a} = -\boldsymbol{b}$, is appropriate. Calculation of the MH acceptance ratio $A_{\boldsymbol{\omega}}(\boldsymbol{\omega}, \boldsymbol{\xi}_{\boldsymbol{\omega}})$ is trivial; it simply requires numerical evaluation of $\Phi_r(\cdot; \cdot, \boldsymbol{\Sigma}_{\boldsymbol{\omega}})$, e.g. via `mvtnorm`, since the ratios of hypercuboid normalization terms would cancel. We find that initial $\phi^{[0]}$ chosen uniformly in $\mathcal{C}_1$, i.e. the interval $(-1, 1)$, and $d^{[0]}$ are systemat-

ically from $\{-0.4, -0.2, 0, 0.2, 0.4\}$ work well. Any choice of prior for $\boldsymbol{\omega}$ can be made, although we prefer flat (proper) priors.

The only technical difficulty is the choice of proposal covariance matrix $\Sigma_{\boldsymbol{\omega}}$. Ideally, it would be aligned with the posterior covariance – however this is not a priori known. We find that running a "pilot" chain with independent proposals via $\mathcal{N}_r^{(a,b)}(\boldsymbol{\omega}, \sigma_{\boldsymbol{\omega}}^2 \mathbf{I}_r)$ can help choose a $\Sigma_{\boldsymbol{\omega}}$. A rescaled version of the sample covariance matrix from the pilot posterior chain, following Roberts and Rosenthal (2001), works well (see Sect. 6.2).

5.2 Unknown short memory form

We now expand the parameter space to include models $M \in \mathcal{M}$, the set of ARFIMA models with p and q short memory parameters, indexing the size of the parameter space $\Psi^{(M)}$. For our "transdimensional moves", we only consider adjacent models, on which we will be more specific later. For now, note that the choice of bijective function mapping between models spaces (whose Jacobian term appears in the acceptance ratio), is crucial to the success of the sampler. To illustrate, consider transforming from $\Phi^{(p+1)} \in \mathcal{C}_{p+1}$ down to $\Phi^{(p)} \in \mathcal{C}_p$. This turns out to be a non-trivial problem however because, for $p > 1$, $\mathcal{C}_p$ has a very complicated shape. The most natural map would be: $(\phi_1, \ldots, \phi_p, \phi_{p+1}) \longmapsto (\phi_1, \ldots, \phi_p)$. However there is no guarantee that the image will lie in $\mathcal{C}_p$. Even if the model dimension is fixed, difficulties are still encountered; a natural proposal method would be to update each component of $\boldsymbol{\phi}$ separately but, because of the awkward shape of $\mathcal{C}_p$, the "allowable" values for each component are a complicated function of the others. Nontrivial proposals are required.

A potential approach is to parametrize in terms of the inverse roots (poles) of Φ, as advocated by Ehlers and Brooks (2006, 2008): By writing $\Phi(z) = \prod_{i=1}^{p}(1 - \alpha_i z)$, we have that $\boldsymbol{\phi}^{(p)} \in \mathcal{C}_p |\alpha_i| < 1$ for all i. This looks attractive because it transforms $\mathcal{C}_p$ into $D^p = D \times \cdots \times D$ (p times) where D is the open unit disc, which is easy to sample from. But this method has serious drawbacks when we consider the RJ step. To decrease dimension, the natural map would be to remove one of the roots from the polynomial.

But because it is assumed that Φ has real coefficients (otherwise the model has no realistic interpretation), any complex α_i must appear as conjugate pairs. There is then no obvious way to remove a root; a contrived method might be to remove the conjugate pair and replace it with a real root with the same modulus, however it is unclear how this new polynomial is related to the original, and to other aspects of the process, like ACV.

5.2.1 Reparametrisation of Φ and Θ

We therefore propose reparametrisation Φ (and Θ) using the bijection between C_p and $(-1, 1)^p$ advocated by various authors, e.g. Marriott et al. (1995) and Vermaak et al. (2004). To our knowledge, these methods have not previously been deployed towards integrating out short memory components in Bayesian analysis of ARFIMA processes.

Monahan (1984) defined a mapping $\boldsymbol{\phi}^{(p)} \longleftrightarrow \boldsymbol{\varphi}^{(p)}$ recursively as follows:

$$\phi_i^{(k-1)} = \frac{\phi_i^{(k)} - \phi_k^{(k)}\phi_{k-i}^{(k)}}{1 - \left(\phi_k^{(k)}\right)^2}, \qquad k = p, \ldots, 2, \qquad i = 1, \ldots, k-1. \tag{18}$$

Then set $\varphi_k^{(p)} = \phi_k^{(k)}$ for $k = 1, \ldots, p$. The reverse recursion is given by:

$$\phi_i^{(k)} = \begin{cases} \varphi_k^{(p)} & \text{for } i = k \qquad\qquad k = 1, \ldots, p \\ \phi_i^{(k-1)} + \varphi_k^{(p)}\phi_{k-i}^{(k-1)} & \text{for } i = 1, \ldots, k-1 \quad k = 2, \ldots, p \end{cases}.$$

Note that $\phi_p^{(p)} = \varphi_p^{(p)}$. Moreover, if $p = 1$, the two parameterizations are the same, i.e. $\phi_1 = \varphi_1$ (consequently the brief study of ARFIMA$(1, d, 0)$ in Sect. 5.1 fits in this framework). The equivalent parametrized form for $\boldsymbol{\theta}$ is $\boldsymbol{\vartheta}$. The full memory parameter $\boldsymbol{\omega}$ is parametrized as $\bar{\Omega} = (-1/2, 1/2) \times$ (the image of $C_{p,q}$). However recall that in practice, $C_{p,q}$ will be assumed equivalent to $C_p \times C_q$, so the parameter space is effectively: $\bar{\Omega} = (-1/2, 1/2) \times (-1, 1)^{p+q}$.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀◀ | ▶▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Besides mathematical convenience, this bijection has a very useful property (cf. Kay and Marple, 1981) which helps motivate its use in defining RJ maps. In other words, if $d = q = 0$, using this parametrization for $\boldsymbol{\varphi}$ when moving between different values of p allows one to automatically choose processes that have very closely matching ACFs at low lags. In the MCMC context this is useful because it allows the chain to propose models that have a similar correlation structure to the current one. Although this property is nice, it may be of limited value for full ARFIMA models, since the proof of the main result does not easily lend itself to the inclusion of either a MA or long memory component. Nevertheless, our empirical results similarly indicate a "near-match" for a full ARFIMA(p, d, q) model.

5.2.2 Application of RJ MCMC to ARFIMA(p, d, q) processes

We now use this reparametrisation to efficiently propose new parameter values. Firstly, it is necessary to propose a new memory parameter ϖ whilst keeping the model fixed. Attempts at updating each component individually suffer from the same problems of excessive posterior correlation that were encountered in Sect. 5.1. Therefore the simultaneous update of the entire $r = (p + q + 1)$-dimensional parameter ϖ is performed using the hypercuboid-truncated Gaussian distribution from definition $\xi_{\varpi} | \varpi \sim \mathcal{N}_r^{\mathcal{H}_r}(\varpi, \Sigma_{\varpi})$, where $\mathcal{H}_r$ defines the r dimensional rectangle. The covariance matrix Σ_{ϖ} is discussed in some detail below. The choice of prior $p_{\varpi}(\cdot)$ is arbitrary. Pai and Ravishanker (1998) used a uniform prior for $\boldsymbol{\omega}$ which has an explicit expression in the ϖ parametrization (Monahan, 1984). However, their expression is unnecessarily complicated since a uniform prior over Ω holds no special interpretation. We therefore prefer uniform prior over $\bar{\Omega}$: $p_{\varpi}(\varpi) \propto 1$, $\varpi \in \bar{\Omega}$.

Now consider the "between-models" transition. We must first choose a model prior $p_{\mathcal{M}}(\cdot)$. A variety of priors are possible; the simplest option would be to have a uniform prior over $\mathcal{M}$, but this would of course be improper. We may in practice want to restrict the possible values of p, q to $0 \le p \le P$ and $0 \le q \le Q$ for some P, Q (say 5), which would render the uniform prior proper. However even in this formulation, a lot of

prior weight is being put onto complicated models which, in the interests of parsimony, might be undesired. We prefer a truncated joint Poisson distribution with parameter λ:
$$p_{\mathcal{M}}(p,q) \propto \frac{\lambda^{p+q}}{p!q!} \mathbb{I}(p \le P, q \le Q).$$

Now, denote the probability of jumping from model $M_{p,q}$ to model $M_{p',q'}$ by $U_{(p,q),(p',q')}$. U could allocate non-zero probability for every model pair, but for convenience we severely restrict the possible jumps (whilst retaining irreducibility) using a two-dimensional bounded birth and death process. Consider the subgraph of $\mathbb{Z}^2$: $G = \{(p,q) : 0 \le p \le P, 0 \le q \le Q\}$, and allocate uniform non-zero probability only to neighboring values, i.e. if and only if $|p - p'| + |q - q'| = 1$. Each point in the "body" of G has four neighbors; each point on the "line boundaries" has three; and each of the four "corner points" has only two neighbors. Therefore the model transition probabilities $U_{(p,q),(p',q')}$ are either 1/4, 1/3, 1/2, or 0.

Now suppose the current $(p + q + 3)$-dimensional parameter is $\boldsymbol{\psi}^{(p,q)}$, given by $\boldsymbol{\psi}^{(p,q)} = (\mu, \sigma, d, \boldsymbol{\varphi}^{(p)}, \boldsymbol{\vartheta}^{(q)})$, using a slight abuse of notation. Because the mathematical detail of the AR and MA components are almost identical, we consider only the case of de/increasing p by 1 here; all of the following remains valid if p is replaced by q, and $\boldsymbol{\varphi}$ replaced by $\boldsymbol{\vartheta}$. We therefore seek to propose a parameter $\boldsymbol{\xi}^{(p+1,q)} = (\xi_\mu, \xi_\sigma, \xi_d, \boldsymbol{\xi}_{\boldsymbol{\varphi}}^{(p+1)}, \boldsymbol{\xi}_{\boldsymbol{\vartheta}}^{(q)})$, that is somehow based on $\boldsymbol{\psi}^{(p,q)}$. We further simplify by regarding the other three parameters (μ, σ, and d) as having the same interpretation in every model, choosing $\xi_\mu = \mu$, $\xi_\sigma = \sigma$ and $\xi_d = d$. For simplicity we also set $\boldsymbol{\xi}_{\boldsymbol{\vartheta}}^{(q)} = \boldsymbol{\vartheta}^{(q)}$. Now consider the map $\boldsymbol{\varphi}^{(p)} \to \boldsymbol{\xi}_{\boldsymbol{\varphi}}^{(p+1)}$. To specify a bijection we "dimension-match" by adding in a random scalar u. The most obvious map is to specify u so that its support is the interval $(-1, 1)$ and then set: $\boldsymbol{\xi}_{\boldsymbol{\varphi}}^{(p+1)} = \left(\boldsymbol{\varphi}^{(p)}, u\right)$. The corresponding map for decreasing the dimension is $\boldsymbol{\varphi}^{(p+1)} \to \boldsymbol{\xi}_{\boldsymbol{\varphi}}^{(p)}$ is $\boldsymbol{\xi}_{\boldsymbol{\varphi}}^{(p)} = \left(\varphi_1^{(p+1)}, \ldots, \varphi_p^{(p+1)}\right)$. In other words, we either add, or remove the final parameter, whilst keeping all others fixed with the identity map, so the Jacobian is unity. The proposal $q(u|\boldsymbol{\psi}^{(p,q)})$ can be made in many ways – we prefer the simple $\mathcal{U}(-1, 1)$. With these choices the RJ acceptance

ratio is

$$A = \ell_{(p',q')}(x|\xi^{(p',q')}) - \ell_{(p,q)}(x|\psi^{(p,q)}) + \log\left\{\frac{p_{\mathcal{M}}(p',q')}{p_{\mathcal{M}}(p,q)}\frac{U_{(p',q'),(p,q)}}{U_{(p,q),(p',q')}}\right\},$$

which applies to both increasing and decreasing dimensional moves.

Construction of Σ_{ϖ}: much of the efficiency of the above scheme, including within- and between-model moves, depends on the choice of $\Sigma_{\varpi} \equiv \Sigma^{(p,q)}$, the within-model move RW proposal covariance matrix. We first seek an appropriate $\Sigma^{(1,1)}$, as in Sect. 5.1, with a pilot tuning scheme. That matrix is shown on the left below, where we've "blocked it out"

$$\Sigma^{(1,1)} = \begin{pmatrix} \sigma_d^2 & \sigma_{d,\varphi_1} & \sigma_{d,\vartheta_1} \\ & \sigma_{\varphi_1}^2 & \sigma_{\varphi_1,\vartheta_1} \\ \cdots & \cdots & \cdots \\ & & \sigma_{\vartheta_1}^2 \end{pmatrix}, \Sigma^{(p,q)} = \begin{pmatrix} \sigma_d^2 & \Sigma_{d,\boldsymbol{\varphi}^{(p)}} & \Sigma_{d,\boldsymbol{\vartheta}^{(q)}} \\ & \Sigma_{\boldsymbol{\varphi}^{(p)},\boldsymbol{\varphi}^{(p)}} & \Sigma_{\boldsymbol{\varphi}^{(p)},\boldsymbol{\vartheta}^{(q)}} \\ \cdots & \cdots & \cdots \\ & & \Sigma_{\boldsymbol{\vartheta}^{(q)},\boldsymbol{\vartheta}^{(q)}} \end{pmatrix}, \quad (19)$$

(where each block is a scalar) so that we can extend this idea to the (p,q) case in the obvious way – on the right above – where $\Sigma_{\boldsymbol{\varphi}^{(p)},\boldsymbol{\varphi}^{(p)}}$ is a $p \times p$ matrix, $\Sigma_{\boldsymbol{\vartheta}^{(q)},\boldsymbol{\vartheta}^{(q)}}$ is a $q \times q$ matrix, etc. If either (or both) $p, q = 0$ then the relevant blocks are simply omitted. To specify the various sub-matrices, we propose $\varphi_2, \ldots, \varphi_p$ with equal variances, and *independently* of $d, \varphi_1, \vartheta_1$, (and similarly for $\vartheta_2, \ldots, \vartheta_q$). In the context of Eq. (19) the following hold:

$$\Sigma_{d,\boldsymbol{\varphi}^{(p)}} = \begin{pmatrix} \sigma_{d,\varphi_1}\mathbf{0} \end{pmatrix}, \Sigma_{d,\boldsymbol{\vartheta}^{(q)}} = \begin{pmatrix} \sigma_{d,\vartheta_1} & \mathbf{0} \end{pmatrix}, \qquad \Sigma_{\boldsymbol{\varphi}^{(p)},\boldsymbol{\varphi}^{(p)}} = \begin{pmatrix} \sigma_{\varphi_1}^2 & \mathbf{0} \\ \cdots & \cdots \\ \mathbf{0} & \sigma_{\varphi}^2 I_{p-1} \end{pmatrix},$$

$$\Sigma_{\boldsymbol{\vartheta}^{(q)},\boldsymbol{\vartheta}^{(q)}} = \begin{pmatrix} \sigma_{\vartheta_1}^2 & \mathbf{0} \\ \cdots & \cdots \\ \mathbf{0} & \sigma_{\vartheta}^2 I_{q-1} \end{pmatrix}, \qquad \Sigma_{\boldsymbol{\varphi}^{(p)},\boldsymbol{\vartheta}^{(q)}} = \begin{pmatrix} \sigma_{\varphi_1,\vartheta_1} & \mathbf{0} \\ \cdots & \cdots \\ \mathbf{0} & \mathbf{O} \end{pmatrix},$$

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

where the dotted lines indicate further blocking, **0** is a row-vector of zeros, and **O** is a zero matrix. This choice of Σ_{ϖ} is conceptually simple, computationally easy and preserves the positive-definiteness as required (see Graves, 2013).

6   Empirical illustration and comparison

Here we provide empirical illustrations for the methods above: for classical and Bayesian analysis of long memory models, and extensions for short memory. To ensure consistency throughout, the location and scale parameters will always be chosen as $\mu_l = 0$ and $\sigma_l = 1$. Furthermore, unless stated otherwise, the simulated series will be of length $n = 2^{10} = 1024$. This is a reasonable size for many applications; it is equivalent to 85 years' monthly observations. When using the approximate likelihood method we set $P = n$.

6.1   Long memory

Standard MCMC diagnostics were used throughout to ensure, and tune for, good mixing. Because d is the parameter of primary interest, the initial values $d^{[0]}$ will be chosen to systematically cover its parameter space, usually starting five chains at the regularly-spaced points $\{-0.4, -0.2, 0, 0.2, 0.4\}$. Initial values for other parameters are not varied: μ will start at the sample mean $\bar{x}$; σ at the sample SD of the observed series $\boldsymbol{x}$.

6.1.1   Efficacy of approximate likelihood method

Start with the "null case", i.e. how does the algorithm perform when the data are not from a long memory process? One hundred independent ARFIMA$(0, 0, 0)$, or Gaussian white noise, processes are simulated, from which marginal posterior means, SDs, and credibility interval endpoints are extracted. Table 1 shows averages over the runs.

The average estimate for each of the three parameters is less than a quarter of a SD away from the truth. Credibility intervals are nearly symmetric about the estimate and

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

the marginal posteriors are, to a good approximation, locally Gaussian (not shown). Upon, applying a proxy "credible-interval-based hypothesis test" one would conclude in ninety-eight of the cases that $d = 0$ could not be ruled out. A similar analysis for μ and σ shows that hypotheses $\mu = 0$ and $\sigma = 1$ would each have been accepted ninety-

5   six times. These results indicate that the 95 % credibility intervals are approximately correctly sized.

Next, consider the more interesting case of $d_I \neq 0$. We repeat the above experiment except that ten processes are generated with d_I set to each of $\{-0.45, -0.35, \ldots, 0.45\}$, giving 100 series total. Figure 1 shows a graphical analog of results from this ex-

10 periment. The plot axes involve a Bayesian residual estimate of d, $\widehat{d_R}^{(B)}$, defined as $\widehat{d_R}^{(B)} = \widehat{d}^{(B)} - d_I$, where $\widehat{d}^{(B)}$ is the Bayesian estimate of d.

From the figure is clear that the estimator for d is performing well. Plot (a) shows how "tight" the estimates of d are around the input value – recall that the parameter space for d is the whole interval $\left(-\frac{1}{2}, \frac{1}{2}\right)$. Moreover, plot (b) indicates that there is no

15 significant change of posterior bias or variance as d_I is varied.

Next, the corresponding plots for the parameters σ and μ are shown in Fig. 2. We see from plot (a) that the estimate of σ also appears to be unaffected by the input value d_I. The situation is different however in plot (b) for the location parameter μ. Although the bias appears to be roughly zero for all d_I, the posterior variance clearly *is* affected

20 by d_I. To ascertain the precise functional dependence, consider plot (c) which shows, on a semi-log scale, the marginal posterior SD of μ, $\widehat{\sigma_\mu}^{(B)}$, against d_I.

It appears that the marginal posterior SD $\widehat{\sigma_\mu}^{(B)}$ is a function of d_I; specifically: $\widehat{\sigma_\mu}^{(B)} \propto A^{d_I}$, for some A. The constant A could be estimated via least-squares regression. Instead however, inspired by asymptotic results in literature concerning classical

25 estimation of long memory processes (Beran, 1994a) we set $A = n$ and plotted the best fitting such line (shown in plot c). Observe that, although not fitting exactly, the relation $\widehat{\sigma_\mu}^{(B)} \propto n^{d_I}$ holds reasonably well for $d_I \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. Indeed, Beran motivated long memory in this way, and derived asymptotic consistency results for optimum (likelihood-

based) estimators and found indeed that the standard error for μ is proportional to $n^{d-1/2}$ (Theorem 8.2) but the standard errors of all other parameters are proportional to $n^{-1/2}$ (Theorem 5.1).

6.1.2 Effect of varying time series length

5 We now analyze the effect of changing the time series length. For this we conduct a similar experiment but fix $d_I = 0$ and vary n. The posterior statistics of interest are the posterior SDs $\widehat{\sigma_d}^{(B)}$, $\widehat{\sigma_\mu}^{(B)}$ and $\widehat{\sigma_\sigma}^{(B)}$. For each $n \in \{128 = 2^7, 2^8, \ldots, 2^{14} = 16\,384\}$, 10 independent ARFIMA$(0,0,0)$ time series are generated. The resulting posterior SDs are plotted against n (on log-log scale) in Fig. 3.

10 Observe that all three marginal posterior SDs are proportional to $\frac{1}{\sqrt{n}}$, although the posterior of μ is less 'reliable'. Combining these observations with our earlier deduction that $\sigma_\mu^{(B)} \propto n^{d_I}$, we conclude that for an ARFIMA$0, d_I, 0$ process of length n, the marginal posterior SDs follow those of Beran given previously.

6.1.3 Comparison with common estimators

15 In many practical applications, the long memory parameter is estimated using non/semi-parametric methods. These may be appropriate in many situations, where the exact form of the underlying process is unknown. However when a specific model form is known (or at least assumed) they tend to perform poorly compared with fully parametric alternatives (Franzke et al., 2012). Our aim here is to demonstrate, via
20 a short Monte Carlo study involving ARFIMA$(0, d, 0)$ data, that our Bayesian likelihood-based method significantly outperforms other common methods in that case. We consider the following comparators: (i) rescaled adjusted range, or R/S (Hurst, 1951; Graves, 2013) – we use the $\mathbb{R}$ implementation in the FGN (McLeod et al., 2007) package; (ii) Semi-parametric Geweke–Porter-Hudak (GPH) method (Geweke and Porter-
25 Hudak, 1983) – implemented in $\mathbb{R}$ package fracdiff (Fraley et al., 2012); (iii) de-

trended fluctuation analysis (DFA), originally devised by Peng et al. (1994) – in the ℝ package `PowerSpectrum` (Vyushin et al., 2009). (iv) wavelet-based semi-parametric estimators (Abry et al., 2003) available in ℝ package `fARMA` (Wuertz, 2012).

Each of these four methods will be applied to the same 100 time series with varying d_I as were used earlier experiments above. We extend the idea of a residual, $\widehat{d_R}^{(R)}$, $\widehat{d_R}^{(G)}$, $\widehat{d_R}^{(D)}$, and $\widehat{d_R}^{(W)}$, to accommodate the new comparators, respectively, and plot them against $\widehat{d_R}^{(B)}$ in Fig. 5.

Observe that all four methods have a much larger variance than our Bayesian method, and moreover the R/S is positively biased. Actually, the bias in some cases would seem to depend on d_I: R/S is significantly (i.e. > 0.25) biased for $d_I < -0.3$ but slightly negatively biased for $d > 0.3$ (not shown); DFA is only unbiased for $d_I > 0$; both the GPH and wavelet methods are unbiased for all $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$.

6.2 Extensions for short memory

Known form: we first consider the MCMC algorithm from Sect. 5.1 for sampling under an ARFIMA$(1, d, 0)$ model where the full memory parameter is $\boldsymbol{\omega} = (d, \phi_1)$. Recall that that method involved proposals from a hypercuboid MVN using a pilot-tuned covariance matrix. Also recall that it is a special case of the re-parametrized method from Sect. 5.2.

In general, this method works very well; two example outputs are presented in Fig. 6, under two similar data generating mechanisms.

Plot (a) shows relatively mild correlation ($\rho = 0.21$) compared with (b) which shows strong correlation ($\rho = 0.91$). This differential behavior can be explained heuristically by considering the differing data-generating values. For the process in plot (a) the short memory and long memory components exhibit their effects at opposite ends of the spectrum; see Fig. 7a. The resulting ARFIMA spectrum, with peaks at either end, makes it easy to distinguish between short and long memory effects, and consequently the posteriors of d and ϕ are largely uncorrelated. In contrast, the parameters of the process in plot (b) express their behavior at the same end of the spectrum. With neg-

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

ative d these effects partially cancel each other out, except very near the origin where the negative memory effect dominates; see Fig. 7b. Distinguishing between the effects of ϕ and d is much more difficult in this case, consequently the posteriors are much more dependent.

In cases where there is significant correlation between d and ϕ, it arguably makes little sense to consider only the marginal posterior distribution of d. For example the 95 % credibility interval for d from plots (b) is $(-0.473, -0.247)$, and the corresponding interval for ϕ is $(-0.910, -0.753)$, yet these clearly give a rather pessimistic view of our joint knowledge about d and ϕ – see Fig. 7c. In theory an ellipsoidal credibility set could be constructed, although this is clearly less practical when $\dim \boldsymbol{\omega} > 2$.

Unknown form: the RJ scheme outlined in Sect. 5.2 works well for data simulated with p and q up to 3. The marginal posteriors for d are generally roughly centered around d_I (the data generating value) and the modal posterior model probability is usually the "correct" one. To illustrate, consider again the two example data generating contexts used above.

For both series, kernel density for the marginal posterior for d are plotted in Fig. 8a and b, together with the equivalent density estimated assuming unknown model orders.

Notice how the densities obtained via the RJ method are very close to those obtained assuming $p = 1$ and $q = 0$. The former are slightly more heavy-tailed, reflecting a greater level of uncertainty about d. Interestingly, the corresponding plots for the posteriors of μ and σ do not appear to exhibit this effect (see Fig. 8c and d). The posterior model probabilities are presented in Table 2, showing that the "correct" modes are being picked up consistently.

As a test of the robustness of the method, consider a complicated short memory input combined with a heavy tailed α-stable innovations distribution. Specifically, the time series that will be used is the following ARFIMA($2, d, 1$) process

$$\left(1 - \frac{9}{16}\mathcal{B}^2\right)(1 - \mathcal{B})^{0.25} X_t = \left(1 + \frac{1}{3}\mathcal{B}\right)\varepsilon_t, \qquad \text{where } \varepsilon_t \sim \mathcal{S}_{\alpha=1.75,0}. \tag{20}$$

For more details, see Graves (2013, Sect. 7.1). The marginal posterior densities of d and α are presented in Fig. 9.

Performance looks good despite the complicated structure. The posterior estimate for d is $\widehat{d}^{(B)} = 0.22$, with 95 % CI $(0.04, 0.41)$. Although this interval is admittedly rather wide, it is reasonably clear that long memory is present in the signal. The corresponding interval for α is $(1.71, 1.88)$ with estimate $\widehat{\alpha}^{(B)} = 1.79$. Finally, we see from Table 3 that the algorithm is very rarely in the "wrong" model.

The Nile Data: we conclude with an application of our methods to the famous annual Nile minima data. Because of the fundamental importance of the river to the civilizations it has supported, local rulers kept measurements of the annual maximal and minimal heights obtained by the river at certain points (called gauges). The longest uninterrupted sequence of recordings is from the Roda gauge (near Cairo), between 622 and 1284 AD ($n = 663$).[4] The posterior summary statistics and marginal densities of d and μ for the Nile data are presented in Fig. 10. Posterior model probabilities are presented in Table 4. We see that the model with the highest posterior probability is the ARFIMA$(0, d, 0)$ model with $d \approx 0.4$. This suggests a strong, "pure", long memory feature. Our results compare favorably with other studies (Liseo et al., 2001; Hsu and Breidt, 2003; Ko and Vannucci, 2006a).

7   Conclusions

We have provided a systematic treatment of efficient Bayesian inference for ARFIMA models, the most popular parametric model combining long and short memory effects. Through a mixture of theoretical and empirical work we have demonstrated that the methods can handle the sorts of time series data that are typically confronted with possible long memory in mind.

---

[4]There is evidence (e.g. Ko and Vannucci, 2006b) that the sequence is not actually homogeneous.

Many of the choices made throughout, but in particular those leading to our likelihood approximation stem from a need to accommodate further extension. For example, in future work we intend to extend them to cope with a heavy-tailed innovations distribution. For more evidence of potential in this context, see Graves (2013, Sect. 7). Along similar lines, there is scope for further generalization to incorporate seasonal (long memory) effects. Finally, an advantage of the Bayesian approach is that it provides a natural mechanism for dealing with missing data, via data augmentation. This is particularly relevant for long historical time series which may, for a myriad of reasons, have recording gaps. For example, some of the data recorded at other gauges along the river Nile have missing observations although otherwise span a similarly long time frame. For a demonstration of how this might fit within our framework, see Sect. 5.6 of Graves dissertation.

References

Abry, P., Flandrin, P., Taqqu, M. S., and Veitch, D.: Self-similarity and long-range dependence through the wavelet lens, in: Theory and Applications of Long-Range Dependence, Birkhäuser, Cambridge, MA, USA, 527–556, 2003. 598

Adenstedt, R. K.: On large-sample estimation for the mean of a stationary random sequence, Ann. Stat., 2, 1095–1107, 1974. 580

Barnes, J. A. and Allan, D. W.: A statistical model of flicker noise, Proc. IEEE, 54, 176–178, 1966. 580

Beran, J.: Statistics for Long Memory Processes, Chapman and Hall, New York, 1994a. 574, 596, 597

Beran, J.: On a class of m-estimators for gaussian long-memory models, Biometrika, 81, 755–766, 1994b. 586

Beran, J., Feng, Y., Ghosh, S., and Kulik, R.: Long Memory Processes, Springer, Heidelberg, 2013. 574, 575

Box, G. E. P. and Jenkins, G. M.: Time Series Analysis, Forecasting and Control, Holden-Day, 1970. 578

Brockwell, P. J. and Davis, R. A.: Time Series: Theory and Methods, 2nd Edn., Springer, New York, 1991. 579, 580, 589

5 Ehlers, R. S. and Brooks, S. P.: Bayesian analysis of order uncertainty in ARIMA models, Tech. Rep., Department of Statistics, Federal University of Paraná, Paraná, available at: http://www.est.ufpr.br/rt/ehlb06.pdf (last access: March 2015), 2006. 588, 590

Ehlers, R. S. and Brooks, S. P.: Adaptive proposal construction for reversible jump MCMC, Scand. J. Stat., 35, 677–690, 2008. 576, 588, 590

10 Eğrioğlu, E. and Günay, S.: Bayesian model selection in ARFIMA models, Expert Syst. Appl., 37, 8359–8364, 2010. 577, 588

Fraley, C., Leisch, F., Maechler, M., Reisen, V., and Lemonte, A.: `fracdiff`: Fractionally differenced ARIMA aka ARFIMA(p, d, q) models, R package version 1.4-1,available at: http://CRAN.R-project.org/package=fracdiff (last access: March 2015), 2012. 597

15 Franzke, C. L. E., Graves, T., Watkins, N. W., Gramacy, R. B., and Hughes, C.: Robustness of estimators of long-range dependence and self-similarity under non-Gaussianity, Philos. T. Roy. Soc. A, 370, 1250–1267, 2012. 597

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T.: `mvtnorm`: Multivariate Normal and t Distributions, R package version 0.9-9992, available at: http://CRAN.R-project.org/package=mvtnorm (last access: March 2015), 2012. 589

20

Geweke, J. and Porter-Hudak, S.: The estimation and application of long-memory time series models, J. Time Ser. Anal., 4, 221–238, 1983. 597

Granger, C. W. J. and Joyeux, R.: An introduction to long-memory time series models and fractional differencing, J. Time Ser. Anal., 1, 15–29, 1980. 576

25 Graves, T.: A systematic approach to Bayesian inference for long memory processes, PhD thesis, University of Cambridge, Cambridge, UK, 2013. 585, 586, 595, 597, 600, 601

Graves, T., Gramacy, R. B., Watkins, N. W., and Franzke, C.: A brief history of long memory, Tech. rep, The University of Chicago, Chicago, available at: http://arxiv.org/abs/1406.6018, last access: 4 November 2014. 575

30 Green, P. J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, Biometrika, 82, 711–732, 1995. 588

Haslett, J. and Raftery, A. E.: Space-time modelling with long-memory dependence: assessing Ireland's wind power resource, J. Roy. Stat. Soc. C, 38, 1–50, 1989. 582

⏮   ⏭

◀   ▶

Back   Close

Full Screen / Esc

Holan, S., McElroy, T., and Chakraborty, S.: A Bayesian approach to estimating the long memory parameter, Bayesian Anal., 4, 159–190, 2009. 576, 588

Hosking, J. R. M.: Fractional differencing, Biometrika, 68, 165–176, 1981. 576

Hsu, N.-J. and Breidt, F. J.: Bayesian analysis of fractionally integrated ARMA with additive noise, J. Forecasting, 22, 491–514, 2003. 575, 600

Hurst, H. E.: Long-term storage capacity of reservoirs, T. Am. Soc. Civ. Eng., 116, 770–808, 1951. 575, 597

Kay, S. M. and Marple, S. L.: Spectrum analysis – a modern perspective, Proc. IEEE, 69, 1380–1419, 1981. 592

Ko, K. and Vannucci, M.: Bayesian wavelet analysis of autoregressive fractionally integrated moving-average processes, J. Stat. Plan. Infer., 136, 3415–3434, 2006a. 600

Ko, K. and Vannucci, M.: Bayesian wavelet-based methods for the detection of multiple changes of the long memory parameter, IEEE T. Signal Proces., 54, 4461–4470, 2006b. 600

Koop, G., Ley, E., Osiewalski, J., and Steel, M. F.: Bayesian analysis of long memory and persistence using ARFIMA models, J. Econometrics, 76, 149–169, 1997. 575, 584, 587

Liseo, B., Marinucci, D., and Petrella, L.: Bayesian semiparametric inference on long-range dependence, Biometrika, 88, 1089–1104, 2001. 576, 600

Mandelbrot, B. B. and Van Ness, J. W.: Fractional Brownian motions, fractional noises and applications, SIAM Rev., 10, 422–437, 1968. 575

Mandelbrot, B. B. and Wallis, J. R.: Noah, Joseph and operational hydrology, Water Resour. Res., 4, 909–918, 1968. 575

Marriott, J. M., Ravishanker, N., Gelfand, A. E., and Pai, J. S.: Bayesian analysis for ARMA processes: complete sampling based inference under exact likelihoods, in: Bayesian Statistics and Econometrics: Essays in Honor of Arnold Zellner, edited by: Berry, D., Chaloner, K., and Geweke, J., Wiley, New York, 243–256, 1995. 591

McLeod, A. I., Yu, H., and Krougly, Z.: Algorithms for Linear Time Series Analysis: With R Package, J. Stat. Softw., 23, available at: http://www.jstatsoft.org/v23/i05/ (last access: March 2015), 2007. 597

Monahan, J. F.: A note on enforcing stationarity in autoregressive-moving average models, Biometrika, 71, 403–404, 1984. 591, 592

Pai, J. S. and Ravishanker, N.: Bayesian analysis of autoregressive fractionally integrated moving-average processes, J. Time Ser. Anal., 19, 99–112, 1998. 575, 576, 584, 585, 587, 592

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Palma, W.: Long Memory Time Series, Wiley, Hoboken, New Jersey, USA, 2007. 574

Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., and Goldberger, A. L.: Mosaic organization of DNA nucleotides, Phys. Rev. E, 49, 1685–1689, 1994. 598

Robert, C. P.: Simulation of truncated normal variables, Stat. Comput., 5, 121–125, 1995. 586

5 Roberts, G. O. and Rosenthal, J. S.: Optimal scaling for various Metropolis–Hastings algorithms, Stat. Sci., 16, 351–367, 2001. 590

Vermaak, J., Andrieu, C., Doucet, A., and Godsill, S.: Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes, J. Time Ser. Anal., 25, 785–809, 2004. 588, 591

10 Vyushin, D., Mayer, J., and Kushner, P.: PowerSpectrum: Spectral Analysis of Time Series, R package version 0.3, available at: http://www.atmosp.physics.utoronto.ca/people/vyushin/mysoftware.html (last access: March 2015), 2009. 598

Watkins, N. W.: Bunched black (and grouped grey) swans: dissipative and non-dissipative models of correlated extreme fluctuations in complex geosystems, Geophys. Res. Lett., 40, 1–9, doi:10.1002/GRL.50103, 2013. 574

15 Wuertz, D.: fArma: ARMA Time Series Modelling, R package version 2160.77, available at: http://CRAN.R-project.org/package=fArma (last access: March 2015), 2012. 598

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 1. Posterior summary statistics for ARFIMA$(0, 0, 0)$ process. Average of 100 runs.

	mean	std	95 % CI	
d	0.006	0.025	−0.042	0.055
μ	−0.004	0.035	−0.073	0.063
σ	1.002	0.022	0.956	1.041

Table 2. Posterior model probabilities for time series from Figs. 6a and b and 8a and b.

(a)	$p\backslash q$	0	1	2	3	4	5	marginal
0		0.000	0.000	0.000	0.000	0.000	0.000	0.000
1		**0.805**	0.101	0.003	0.000	0.000	0.000	0.908
2		0.038	0.043	0.001	0.000	0.000	0.000	0.082
3		0.005	0.004	0.000	0.000	0.000	0.000	0.009
4		0.000	0.001	0.000	0.000	0.000	0.000	0.001
5		0.000	0.000	0.000	0.000	0.000	0.000	0.000
marginal		0.848	0.148	0.004	0.000	0.000	0.000	
(b)	$p\backslash q$	0	1	2	3	4	5	marginal
0		0.000	0.000	0.000	0.000	0.000	0.000	0.000
1		**0.829**	0.125	0.002	0.000	0.000	0.000	0.956
2		0.031	0.013	0.000	0.000	0.000	0.000	0.044
3		0.000	0.000	0.000	0.000	0.000	0.000	0.000
4		0.000	0.000	0.000	0.000	0.000	0.000	0.000
5		0.000	0.000	0.000	0.000	0.000	0.000	0.000
marginal		0.860	0.138	0.002	0.000	0.000	0.000	

Table 3. Posterior model probabilities.

$p\backslash q$	0	1	2	3	4	5	marginal
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	**0.822**	0.098	0.001	0.000	0.000	0.921
3	0.014	0.056	0.004	0.000	0.000	0.000	0.075
4	0.003	0.001	0.000	0.000	0.000	0.000	0.004
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
marginal	0.017	0.880	0.102	0.002	0.000	0.000	

Full Screen / Esc

Table 4. Posterior model probabilities for Nile minima.

$p\backslash q$	0	1	2	3	4	5	marginal
0	**0.638**	0.101	0.010	0.000	0.000	0.000	0.750
1	0.097	0.124	0.011	0.000	0.000	0.000	0.232
2	0.007	0.010	0.000	0.000	0.000	0.000	0.018
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
marginal	0.742	0.236	0.022	0.000	0.000	0.000	
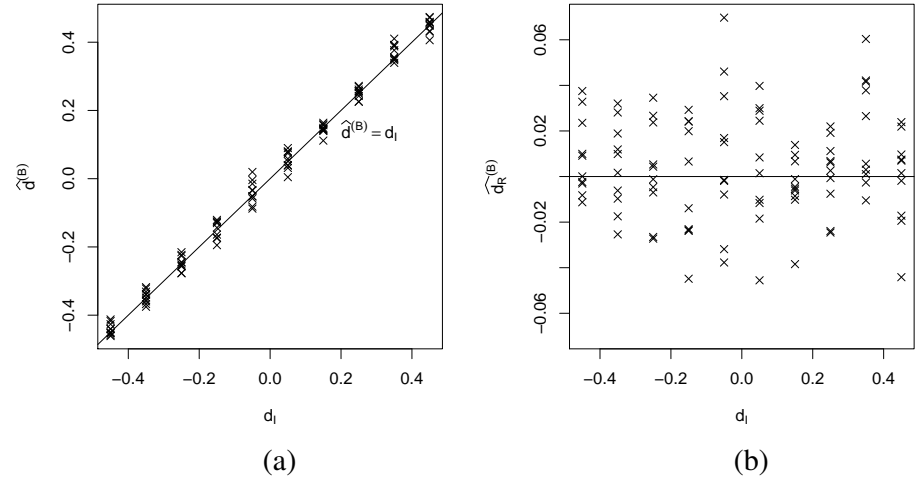
Figure 1. Posterior outputs; **(a)** $\widehat{d}^{(B)}$ against d_I, **(b)** $\widehat{d_R}^{(B)}$ against d_I.
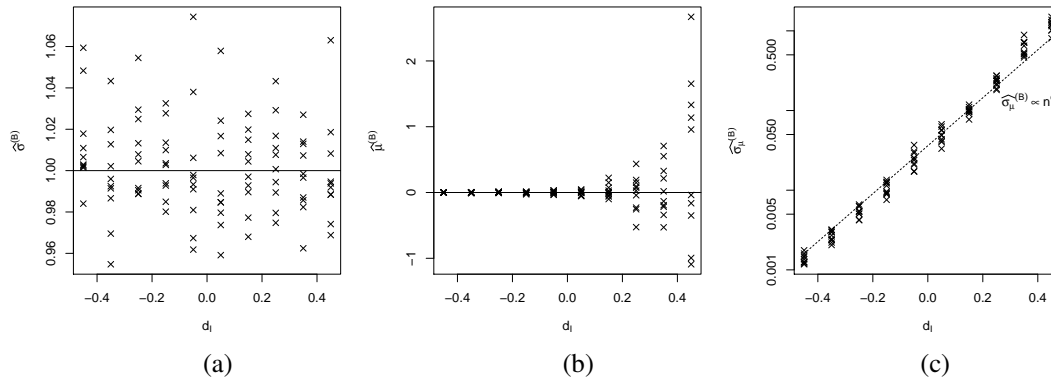
Figure 2. Posterior outputs; **(a)** $\hat{\sigma}^{(B)}$ against d_l, **(b)** $\hat{\mu}^{(B)}$ against d_l, and **(c)** $\widehat{\sigma_\mu}^{(B)}$ against d_l (semi-log scale).

Figure 3. Posterior outputs from ARFIMA$(0,0,0)$ series; **(a)** $\widehat{\sigma}_d^{(B)}$ against n, **(b)** $\widehat{\sigma}_\mu^{(B)}$ against n, **(c)** $\widehat{\sigma}_\sigma^{(B)}$ against n (log-log scale).

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

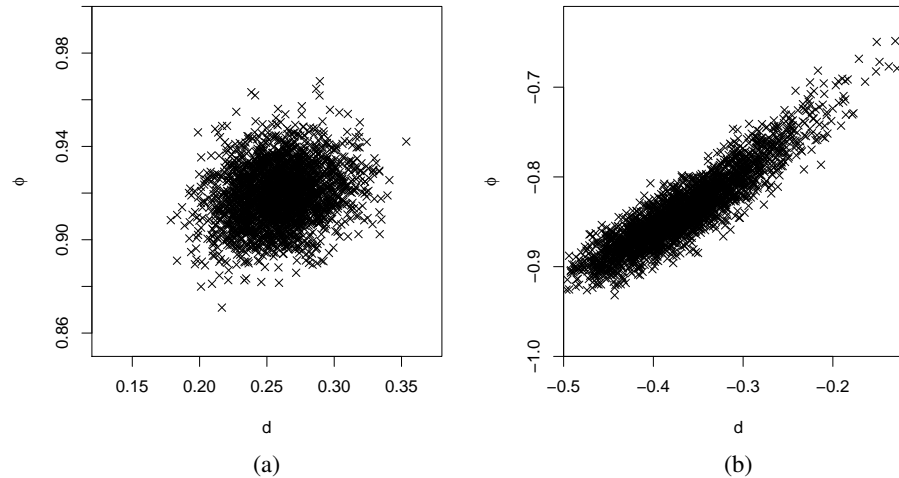Full Screen / Esc

Printer-friendly Version

Interactive Discussion

n	mean difference
128	0.057
256	0.029
512	0.015
1024	0.007

(a)　　　　　　　　(b)
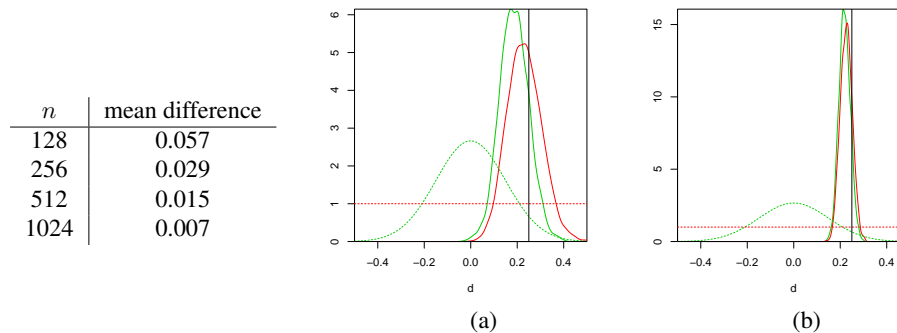
Figure 4. *Table:* Mean difference of estimates $\hat{d}^{(B)}$ under alternative prior assumption. *Plots:* Comparison of posteriors (solid lines) obtained under different priors (dotted lines). Time series used: ARFIMA$(0, 0.25, 0)$; **(a)** $n = 2^7 = 128$, **(b)** $n = 2^{10} = 1024$.
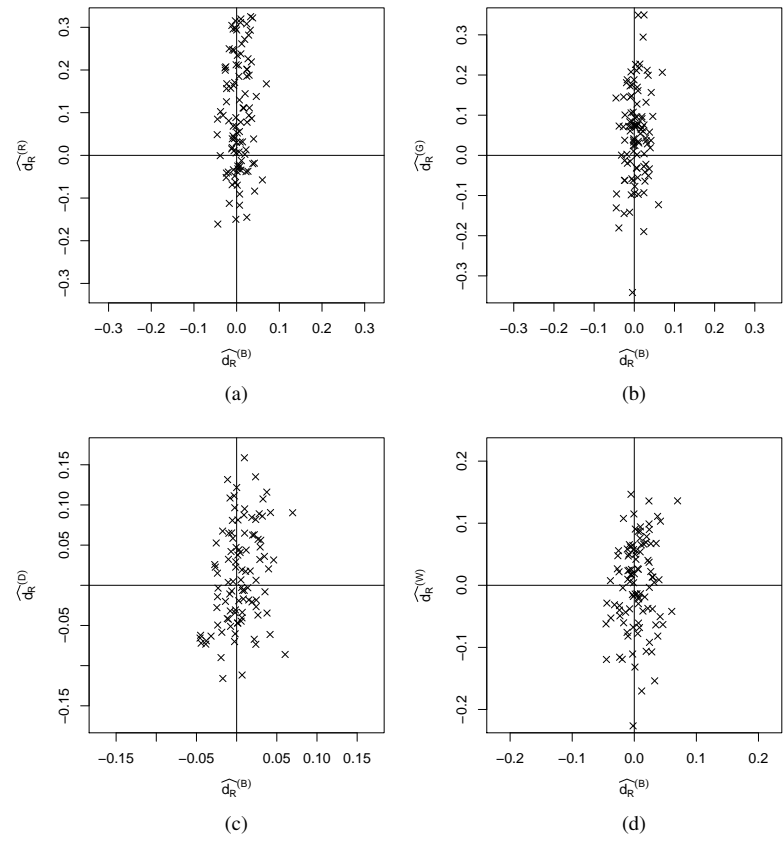
Figure 5. Comparison of Bayesian estimator with common classical estimators; **(a)** R/S, **(b)** GPH, **(c)** DFA, **(d)** wavelet.

Bayesian Inference

T. Graves et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

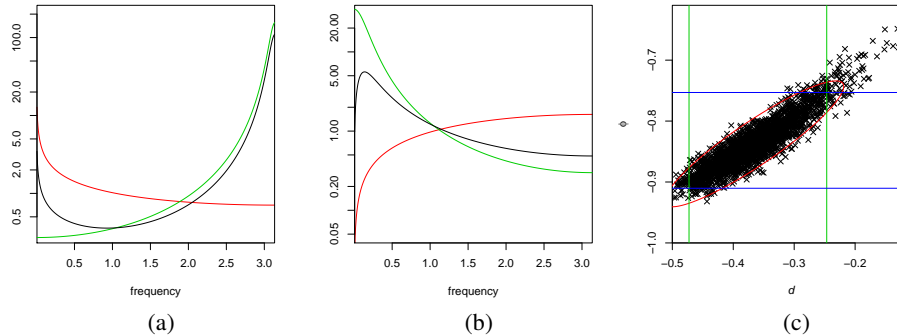Full Screen / Esc

Printer-friendly Version

Interactive Discussion

(a)

(b)

Figure 6. Posterior samples of (d, ϕ); input time series **(a)** $(1 + 0.92\mathcal{B})(1 - \mathcal{B})^{0.25} X_t = \varepsilon_t$, **(b)** $(1 - 0.83\mathcal{B})(1 - \mathcal{B})^{-0.35} X_t = \varepsilon_t$.

Figure 7. Spectra for processes in Fig. 6. Green line is relevant ARMA(1,0) process, red line is relevant ARFIMA(0, d, 0) process, black line is ARFIMA(1, d, 0) process; **(a)** $(1 + 0.92\mathcal{B})(1 - \mathcal{B})^{0.25} X_t = \varepsilon_t$; **(b)** $(1 - 0.83\mathcal{B})(1 - \mathcal{B})^{-0.35} X_t = \varepsilon_t$. **(c)** shows posterior samples of (d, ϕ) from series considered in **(b)** with credibility sets: red is 95 % credibility set for (d, ϕ), green is 95 % credibility interval for d, blue is 95 % credibility interval for ϕ.
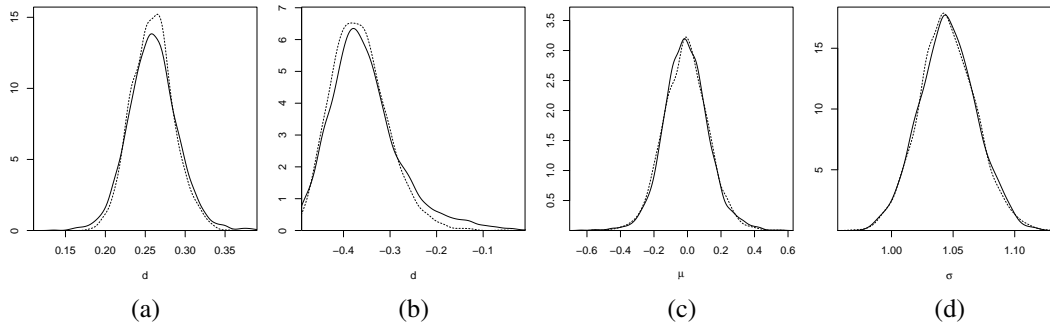
Figure 8. Marginal posterior density of d from series in Fig. 6, **(a, b)** respectively. Solid line is density obtained using reversible-jump algorithm. Dotted line is density obtained using fixed $p = 1$ and $q = 0$. **(c)** and **(d)** show the posterior densities for μ and σ, respectively, corresponding to the series in Fig. 6a; those for Fig. 6b look similar.
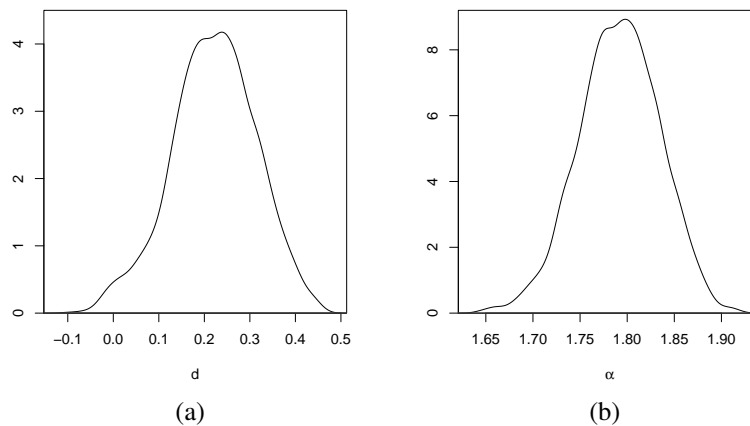
Figure 9. Marginal posterior densities **(a)** d, **(b)** α.

	mean	95% CI	
d	0.039	0.336	0.482
μ	62	1037	1284
σ	1.91	66.46	73.97

(a)

(b)

Figure 10. *Table:* summary posterior statistics for Nile minima. *Plots:* marginal posterior densities for Nile minima; **(a)** d, **(b)** μ.