

Timothy Graves, Christian L.E. Franzke, Nicholas W. Watkins, Robert B. Gramacy, Elizabeth Tindale
Systematic inference of the long-range dependence and heavy-tail distribution parameters of ARFIMA models

**Article (Accepted version)
(Refereed)**

Original citation: Graves, Timothy, Franzke, Christian L.E., Watkins, Nicholas W., Gramacy, Robert B. and Tindale, Elizabeth (2017) *Systematic inference of the long-range dependence and heavy-tail distribution parameters of ARFIMA models*. [Physica A: Statistical Mechanics and Its Applications](#) . ISSN 0378-4371

DOI: [10.1016/j.physa.2017.01.028](https://doi.org/10.1016/j.physa.2017.01.028)

Reuse of this item is permitted through licensing under the Creative Commons:

© 2016 [Elsevier](#)
CC BY NC-ND

This version available at: <http://eprints.lse.ac.uk/68837/>
Available in LSE Research Online: January 2017

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

Accepted Manuscript

Systematic inference of the long-range dependence and heavy-tail distribution parameters of ARFIMA models

Timothy Graves, Christian L.E. Franzke, Nicholas W. Watkins, Robert B. Gramacy, Elizabeth Tindale



PII: S0378-4371(17)30029-8
DOI: <http://dx.doi.org/10.1016/j.physa.2017.01.028>
Reference: PHYSA 17921

To appear in: *Physica A*

Received date: 16 July 2016
Revised date: 14 December 2016

Please cite this article as: T. Graves, et al., Systematic inference of the long-range dependence and heavy-tail distribution parameters of ARFIMA models, *Physica A* (2017), <http://dx.doi.org/10.1016/j.physa.2017.01.028>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

- Self-Similarity has two contributions: Long-range dependence and heavy-tailed jumps
- Systematic simultaneous estimation of long-range dependence and heavy-tail distribution parameters
- Development of a novel Bayesian method for estimation of these two parameters
- Method flexible to allow choice of heavy-tailed distribution (e.g. t- or α -stable distributed)
- Successful demonstration of effectiveness and accuracy on synthetic data

Systematic Inference of the Long-Range Dependence and Heavy-Tail Distribution Parameters of ARFIMA models

Timothy Graves^a, Christian L. E. Franzke^b, Nicholas W. Watkins^{c,d,e,f},
Robert B. Gramacy^g, Elizabeth Tindale^d

^aArup, London, UK

^bMeteorological Institute and Center for Earth System Research and Sustainability,
University of Hamburg, Germany

^cLondon School of Economics, London, UK

^dUniversity of Warwick, Coventry, UK

^eOpen University, Milton Keynes, UK

^fInstitute of Physics and Astronomy, University of Potsdam, Potsdam-Golm, Germany

^gVirginia Tech, USA

Abstract

Long-Range Dependence (LRD) and heavy-tailed distributions are ubiquitous in natural and socio-economic data. Such data can be self-similar whereby both LRD and heavy-tailed distributions contribute to the self-similarity as measured by the Hurst exponent. Some methods widely used in the physical sciences separately estimate these two parameters, which can lead to estimation bias. Those which do simultaneous estimation are based on frequentist methods such as Whittle's approximate maximum likelihood estimator. Here we present a new and systematic Bayesian framework for the simultaneous inference of the LRD and heavy-tailed distribution parameters of a parametric ARFIMA model with non-Gaussian innovations. As innovations we use the α -stable and t -distributions which have power law tails. Our algorithm also provides parameter uncertainty estimates. We test our algorithm using synthetic data, and also data from the Geostationary Operational Environmental Satellite system (GOES) solar X-ray time series. These tests show that our algorithm is able to accurately and robustly estimate the

*Corresponding author

Email addresses: christian.franzke@uni-hamburg.de (Christian L. E. Franzke)

LRD and heavy-tailed distribution parameters.

Key words: Long-Range Dependence, Heavy-Tails, Bayesian Estimation, ARFIMA

1. Introduction

Long-range dependence (LRD) is an ubiquitous property of many physical, biological and financial systems [1, 31]. Hurst's observation (the "Hurst effect") of the anomalous rate of growth of range in hydrological time series, such as the height of the river Nile, was one of the first natural phenomena for which the need for a non-Brownian statistical description was recognised. Mandelbrot & Van Ness [28] explained the Hurst effect as being due to long range dependence in time, which Mandelbrot & Wallis [29] then dubbed the "Joseph effect". Mandelbrot and his co-authors encapsulated the Joseph effect in their seminal model, fractional Brownian motion (fBm), using its stationary increments, fractional Gaussian noise, to model the Nile time series. Like the more familiar Wiener Brownian motion, fBm has the property of self-similarity under a dilation in time where Δt is replaced by $\lambda\Delta t$:

$$x(\lambda\Delta t) \stackrel{d}{=} \lambda^H x(\Delta t) \quad (1)$$

Throughout our paper we will follow Embrechts & Maejima [11], by defining H as the self-similarity exponent. In fBm H takes values between 0 and 1, with $H = 1/2$ being the Brownian case. To describe the growth of rescaled range (R/S) ("the Joseph effect") due to the persistence seen in the increments of fBm, Mandelbrot & Van Ness [28] used a second exponent J , where

$$R/S \sim \tau^J. \quad (2)$$

The presence of LRD affects the predictability of systems and their long-term behaviour and has thus continued to be controversial.

For reasons that are as much historical as technical [16, 27], the parameters of Gaussian LRD models such as fractional Gaussian noise have typically been inferred either indirectly from the self-similarity exponent H , or directly using J (frequently called the Hurst exponent and also denoted by H), because in such models the self-similarity and Hurst exponents happen to coincide [34]. However, what is frequently still not appreciated is that the self-similarity exponent actually has two contributions: (i) one from

28 LRD (also called the Joseph effect) and (ii) one from non-Gaussian jumps
 29 which are power-law distributed (also called the 'Noah' effect); e.g. α -stable
 30 increments, which have probability density function (pdf) tails decaying as
 31 $f(x) \sim x^{-(1+\alpha)}$ where the index α runs from 0 to 2. As Mandelbrot empha-
 32 sised [e.g 27, p. 157] H can be different from J , and R/S only measures the
 33 latter. We will thus avoid the potentially confusing term 'Hurst exponent'
 34 in this paper and label the contribution of memory to the self-similarity ex-
 35 ponent by J , as Mandelbrot recommended after the ambiguity became clear
 36 to him.

37 A second type of non-Brownian phenomenon had also been recognised
 38 by Mandelbrot [25]. This was the non-Gaussian increments, with "heavy"
 39 power-law tails in the pdf,

$$f(x) \sim x^{-(1+\alpha)} \quad (3)$$

40 seen in financial time series [31] and also in many natural ones. In contrast
 41 to the LRD he called this the "Noah effect". He proposed a second paradig-
 42 matic model, ordinary Levy motion (oLm), for cases when the anomalous
 43 behaviour of the time series originates entirely from this effect, rather than
 44 long temporal memory.

45 Real time series do not necessarily exhibit just one or the other of these
 46 two limiting cases. Mandelbrot & Wallis [30] thus proposed that the effects
 47 modelled by fBm and oLm could be combined in a more general self-similar
 48 additive model, "fractional hyperbolic" [30] motion, a descendent of which
 49 is now referred to as linear fractional stable motion, LFSM, [e.g. 11]. LFSM
 50 has by now been applied to problems as diverse as communications traffic
 51 [24], geophysics [38], magnetospheric physics [47] and solar flares [44]. The
 52 "ambivalent" [4] dual behaviour of such models makes it important to develop
 53 methods which can simultaneously estimate both the Joseph and Noah effects
 54 and their corresponding exponents J and α .

55 In our paper we use a newer, more flexible time series model: the well-
 56 known Autoregressive Fractional Integrated Moving Average (ARFIMA) [e.g.
 57 [1]] with non-Gaussian increments [e.g. [8]], which also allows an adjustable
 58 high frequency component. In this model J is encapsulated by the standard
 59 LRD parameter d used in statistics, which ranges from $-1/2$ to $1/2$, with
 60 $d = 0$ being the uncorrelated, white noise case. Our algorithm allows for
 61 α -stable but also t -distributed increments and can also easily be extended to
 62 use any distribution characterized by only a shape and scale parameter.

63 Our method is based on the Bayesian ARFIMA inference algorithm of
64 Graves et al. [17] for which we developed a new approximate likelihood for
65 the efficient parameter inference. We will show how nuisance parameters (e.g.
66 short memory effects) can be integrated out in order to focus systematically
67 on the long memory parameter. Here we extend our method to simultane-
68 ously estimate the LRD and the heavy-tailed parameter. As heavy-tailed
69 distributions we use the t -distribution and the α -stable distribution. For
70 computational reasons we have to restrict our inference to the finite mean
71 ($1 < \alpha \leq 2$) case.

72 It was realised as early as 1969 by Mandelbrot & Wallis [30] that
73 non-parametric LRD estimators are not “fooled” by the presence of non-
74 Gaussianity [14], not least because they measure J rather than H . However,
75 it is still advantageous to perform simultaneous parameter estimation in
76 order to minimize estimation bias, and to provide direct estimates of α
77 rather than having to estimate the tail exponent by some other means
78 such as a measurement of the pdf or cdf. This is especially important for
79 ARFIMA type models which contain both Short-Range Dependence (SRD)
80 and LRD characteristics, in contrast to the pure mono-fractal approach
81 originally taken by Mandelbrot with fBm [28].

82 The standard approach [e.g. [1, 46, 7]] in statistics to estimating the
83 parameters of finite variance ARMA models is ultimately derived from a
84 variant of Whittle’s method proposed by Hannan [20]. Successive develop-
85 ments have encompassed Gaussian ARFIMA [13], ARMA with α -stable noise
86 [35], and ARFIMA with such noise [23]. These developments are accessibly
87 summarised by [7] which constructs a new estimator for ARFIMA and impor-
88 tantly can access the negative d range which is not accessible to the original
89 Whittle estimator proposed in [20].

90 Our new inference algorithm is based on the one put forward in [16, 17].
91 The algorithm consists of a systematic Bayesian framework, a new approxi-
92 mate likelihood for ARFIMA processes and an efficient blocked Monte Carlo
93 Markov Chain (MCMC) sampler. Our Bayesian inference algorithm has been
94 designed in a flexible fashion so that, for instance, the innovations can come
95 from a wide class of different distributions such as the α -stable “Levy” class
96 [31], or the t distribution that is also widely employed in finance [3]. Our
97 algorithm can also estimate the SRD parameters, although these can be in-
98 tegrated out if one is only interested in the LRD and heavy-tail parameters.
99 To our knowledge, a t -distribution has not been considered in LRD models
100 so far.

101 The Bayesian approach allows us the direct computation of the proba-
 102 bility of a theory or model parameter [19]. For our purposes, the important
 103 difference between frequentist and Bayesian approaches is that in the for-
 104 mer the parameters ψ of a statistical model are taken as fixed, whereas in
 105 the latter they are uncertain variables. Assume we have data \mathbf{x} which is a
 106 realisation from an unknown distribution. A given hypothesised model dis-
 107 tribution will have a likelihood $L(\mathbf{x}|\psi)$, i.e. the likelihood of getting that
 108 data \mathbf{x} given a set of values of the parameters ψ . The best estimate of these
 109 parameters, e.g. the LRD parameter d , is what we want.

110 Bayes' theorem states that:

$$\pi_{\psi,\mathbf{x}}(\psi|x) \propto p_{\psi}(\psi)L(\mathbf{x}|\psi) \quad (4)$$

111 where p_{ψ} is a *prior* probability density on the parameters ψ , and π_{ψ} is the
 112 desired *posterior* density. The generic 3 stage approach this allows us to use
 113 is i) postulate a p_{ψ} , then ii) multiply the prior by the calculated likelihood
 114 function for that model L and normalise, i.e. apply Bayes' theorem; and so iii)
 115 generate the posterior π_{ψ} . In principle this could be completely analytically
 116 calculable, but in practice one usually has to use computational methods
 117 like Markov Chain Monte Carlo (MCMC) algorithms because it becomes
 118 analytically intractable.

119 Our paper is structured as follows: In section 2 we describe our inference
 120 approach for the memory parameter d which is related to J as $J = d + 0.5$.
 121 Section 3 is the main part of the paper and describes the extensions of the
 122 ARFIMA inference algorithm of Graves [16], Graves et al. [17] to stable
 123 innovations. We summarize in section 4.

124 2. Bayesian inference on Gaussian ARFIMA model for d .

125 We first briefly describe the Bayesian inference of an ARFIMA model
 126 with Gaussian innovations [16, 17].

127 An ARFIMA model has three classes of parameters: those governing the
 128 location (here the mean μ); the innovation distribution (here just the scale σ);
 129 and memory structure (here LRD parameter d , and the AR and MA series, ϕ
 130 and θ respectively, which are of order p and q in an ARFIMA(p, d, q) model).

131 We choose flat priors for μ , $\log \sigma$ and d . Flat priors are non-informative;
 132 i.e. a flat prior in the coin-tossing heads or tails case is $1/2$, while in an
 133 N category case it is $1/N$. As we have no analytic form for the posterior

134 distribution π , we use MCMC sampling. MCMC [9] is a way to simulate
 135 complex, nonstandard multivariate distributions. There are several types of
 136 MCMC, including Metropolis-Hastings [9] which we use extensively.

137 Our approach has several advantages. First, we don't need to assume the
 138 order of the ARFIMA(p, d, q) model, i.e. pre-specify p, q . Rather we use the
 139 reversible jump (RJ) MCMC approach [18]. In this the parameter space of ψ
 140 is extended to include the set of possible models. The Markov chains move
 141 *between* models as well as within them. Reversible-jump MCMC allows the
 142 sampling of the posterior distribution on spaces of varying dimensions using
 143 a transdimensional Markov Chain. Thus, the simulation is possible even if
 144 the number of parameters in the model is not known.

145 Second, our approach allows the reparameterisation of the model to en-
 146 force stationary constraints on ϕ and θ . This reparameterization improves
 147 the computational efficiency of our algorithm [16].

148 Third, our approach allows a fast approximate Gaussian likelihood cal-
 149 culation. The LRD correlation structure, which considerably enlarges the
 150 dimension of the covariance matrix, prevents use of standard likelihood meth-
 151 ods. Previously we [17] proposed a method for the fast evaluation of con-
 152 ditional likelihoods, e.g. $(n \log n)$ in the Gaussian case. Our use of the
 153 Metropolis-Hastings algorithm requires the careful selection of proposal dis-
 154 tributions. In order to propose new parameters we use a truncated Normal
 155 random walk because this sampler has a finite range of $-1/2 < d < 1/2$ for
 156 the LRD parameter.

157 3. Non-Gaussian innovations

158 In the literature on ARFIMA models, Gaussianity of the innovations
 159 is typically assumed. This assumption is made for at least three reasons.
 160 Firstly, Gaussian analysis often turns out to be mathematically convenient
 161 because the form of the multivariate normal likelihood allows many prob-
 162 lems to be solved exactly. Secondly, the normal distribution is a reasonable
 163 model for many “real-life” applications. Thirdly, one often appeals to the
 164 role played by the Gaussian distribution in the central limit theorem (CLT).
 165 By considering the stochastic elements of a problem to be actually composed
 166 of a large number of non-Gaussian small disturbances, then, provided the
 167 variance is finite, the CLT enables us to assume their aggregation is approx-
 168 imately Gaussian.

169 Whilst the first and third of these reasons are both reasonably sound and
 170 objective, the second has sometimes been the product of modelers' wishes
 171 rather than observational evidence. In practice many real-life processes sim-
 172 ply cannot be modelled as Gaussian [3, 31, 48, 38]. In particular, overdis-
 173 persion is a common problem, i.e. a Gaussian model cannot account for all
 174 the observed variability. Consequently, if the data suggest leptokurtosis, the
 175 Gaussianity assumption may be inappropriate [12].

176 We define the tail behaviour as follows:

$$\mathbb{P}(X > x) \sim Cx^{-a}, \quad (5)$$

177 for some positive constants a and C . Such a distribution will be referred
 178 to as 'heavy-tailed' if a is between 0 and 2. Clearly for such distributions,
 179 moments only exist up to the a -th one. If X is heavy-tailed with parameter
 180 exponent a then:

$$\begin{aligned} \mathbb{E}|X|^p &< \infty && \text{for any } 0 < p < a, \\ \mathbb{E}|X|^p &= \infty && \text{for any } p \geq a. \end{aligned} \quad (6)$$

181 One example of a heavy-tailed distribution is the family of stable distribu-
 182 tions.

183 3.1. Stable distributions

184 A random variable X is said to have a stable distribution, denoted $X \sim$
 185 $\mathcal{S}_{\alpha,\beta}(\gamma, \delta)$, if there are parameters $0 < \alpha \leq 2$, $-1 < \beta < 1$, γ positive and δ
 186 real, such that its characteristic function has the following form:

$$\log [\varphi_{\mathcal{S}}(\theta)] = \begin{cases} -\gamma^\alpha |t|^\alpha \left(1 - i\beta(\text{sign } t) \tan \frac{\pi\alpha}{2}\right) + i\delta t & \text{if } \alpha \neq 1 \\ -\gamma |t| \left(1 + i\beta \frac{2}{\pi}(\text{sign } t) \log |t|\right) + i\delta t & \text{if } \alpha = 1 \end{cases}. \quad (7)$$

187 The support for the stable distribution is the whole real line, except in
 188 the case where $\alpha < 1$ and $\beta = \pm 1$, in which case it is limited to a semi-
 189 infinite interval of the real line. Note also that if $\alpha = 2$ the parameter β
 190 is irrelevant, in which case it is convention to set $\beta = 0$. There are many
 191 different parametrisations of the stable distribution; the article by [37] pro-
 192 vides an excellent summary. Unfortunately neither the probability density
 193 nor distribution functions have generally applicable analytic forms, with a
 194 few known exceptions [48, 49]: $\mathcal{S}_{2,0}(\gamma, \delta)$ is the $\mathcal{N}(\delta, 2\gamma^2)$, $\mathcal{S}_{1,0}(\gamma, \delta)$ is the
 195 Cauchy distribution, and $\mathcal{S}_{1/2,1}(\gamma, \delta)$ is the Lévy distribution.

196 Because for $\alpha < 1$ the process has no mean, which causes many difficul-
 197 ties in LRD parameter inference, we will assume henceforth that $1 < \alpha \leq 2$.
 198 We denote by parameters δ and γ the ‘location’ and ‘scale’ parameters re-
 199 spectively. Although the density is nearly always non-analytical, the stable
 200 distribution does satisfy a location-scale density [16] of the form:

$$f(x; \delta, \sigma, \boldsymbol{\lambda}) \equiv \frac{1}{\sigma} f\left(\frac{x - \delta}{\sigma}; 0, 1, \boldsymbol{\lambda}\right). \quad (8)$$

201 Consequently one need only be concerned with the ‘standardised’ stable dis-
 202 tributions with $\delta = 0$ and $\gamma = 1$, which will be denoted using the shorthand
 203 $\mathcal{S}_{\alpha, \beta}$ with corresponding density $f_{\mathcal{S}}(\cdot; \alpha, \beta)$.

204 Typically the parameter controlling the tail decay, α , is referred to as
 205 the ‘index of stability’. The parameter β is conventionally called the ‘skew’
 206 parameter since non-zero values induce skewness in the distribution.

207 Throughout the remainder of this section, it will be assumed that a pro-
 208 cess $\{X_t\}$ is both causal and invertible and has Wold expansion:

$$X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}, \quad (9)$$

209 where the coefficients $\{\psi_k\}$ are real and ℓ^2 -convergent, and the innovations
 210 $\{\varepsilon_t\}$ are independent and identically distributed $\mathcal{S}_{\alpha, \beta}(\gamma, 0)$ for some $1 <$
 211 $\alpha < 2$, $-1 \leq \beta \leq 1$ and positive γ . Recall that for such processes, a
 212 stably distributed process has long memory if its Wold expansion decays as
 213 a power-law [16]. Throughout most of this paper, only stably distributed
 214 ARFIMA(0, d , 0) processes will be considered. We note that in the Gaussian
 215 case the condition:

$$-\frac{1}{2} < d < \frac{1}{2}, \quad (10)$$

216 was required to ensure causality and invertibility. In the stable case, the
 217 following stronger condition exists:

$$-\left(1 - \frac{1}{\alpha}\right) < d < 1 - \frac{1}{\alpha}. \quad (11)$$

218 Note that this is consistent with the $\alpha = 2$, Gaussian, case (10). The region
 219 of allowable values of the pair (α, d) is shown in figure 1.

220 *3.1.1. Statistical inference from stable processes*

221 We first discuss how to draw inference in the simplest possible scenario
 222 of independent and identically distributed random variables. Because of the
 223 lack of an analytical density, and consequently likelihood, stable distributions
 224 are notoriously difficult to work with. In many cases, it may only be the
 225 parameter α that is of interest and therefore it may seem reasonable to
 226 estimate this directly from the tail behaviour (5). The naive approach of
 227 calculating the empirical distribution function and using log-regression was
 228 developed by [21] and [10] amongst others. But [32] later showed that this
 229 method is seriously flawed because the tail behaviour is truly *asymptotic*
 230 and for some parametrisations and combinations of parameter values, the
 231 power-law behaviour does not occur until far out into the tail.

232 Bayesian analysis of stable distributions has been limited; [5] considered
 233 the problem of finding the joint posterior distribution of the parameters from
 234 a collection of n independent and identically distributed stable random vari-
 235 ables. Unsurprisingly the posterior is intractable so [5] developed an MCMC
 236 method requiring n auxiliary variables and complicated sampling regimes.

237 *3.1.2. Bayesian inference for stably-distributed ARFIMA processes*

238 The most significant challenge in the stably-distributed ARFIMA pro-
 239 cesses scenario is the efficient computation of the log-likelihood. We need to
 240 be able to compute the logarithm of the density $f_S(x; \alpha, \beta)$ for any $\alpha > 1$,
 241 $-1 < \beta < 1$ and real x . To compute the log-likelihood efficiently, note that
 242 we actually seek to evaluate the same density at n points simultaneously. For
 243 this purpose we use the approach developed by Mittnik et al [36], in which
 244 the stable density is calculated on a regular grid, from which the density at
 245 all the x_k can be interpolated. Their method takes advantage of the fact that
 246 the characteristic function φ of stable processes (7) is the ‘Fourier-dual’ of
 247 the probability density function:

$$f_S(x; \alpha, \beta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \varphi_S(t; \alpha, \beta) dt. \quad (12)$$

248 In particular, [36] showed that this integral can be approximated by a sum
 249 which, using an N -sized FFT, can calculate (to an arbitrary precision) the
 250 values of $f_S(y_k; \alpha, \beta)$ on an N -grid of equally spaced values $\{y_k\}$ where:

$$y_k = h \left(k - 1 - \frac{N}{2} \right), \quad k = 1, \dots, N$$

251 for some N and h . Once this grid has been calculated, the densities
 252 $f_S(x_1; \alpha, \beta), \dots, f_S(x_n; \alpha, \beta)$ can be evaluated by linear interpolation.

There are several issues to note regarding implementation of this scheme. Firstly, the choice of parameters N and h is important. The number of points in the grid is N , so a larger N generates a larger grid (at mild computational expense since the FFT has complexity $\mathcal{O}(N \log N)$). The spacing between points is h , so a smaller h produces a more detailed grid. If the data are fat-tailed, the maximum of $|y_k|$ would be expected to be large which means that Nh must also be large. This means making N large, at the expense of slowing the FFT, or making h large, at the expense of losing detail in the interpolation. We will therefore use fixed values for N and h ($N = 2^{13}$, $h = 2^{-10}$) and use a series expansion to calculate the remaining outliers by noting that for $1 < \alpha \leq 2$ the density $f_S(x; \alpha, \beta)$ has the following asymptotic expansion for $x \rightarrow \infty$ [2, 33]:

$$f_S(x) = \frac{1}{\pi} \sum_{k=1}^{\infty} c a_k (cx)^{-k\alpha-1}$$

$$\text{where } c = \cos(\beta^*)^{1/\alpha},$$

$$a_k = (-1)^{k-1} \frac{\Gamma(1+k\alpha)}{k!} \sin\left(\frac{k}{2}(\pi\alpha + 2\beta^*)\right),$$

$$\text{and } \tan(\beta^*) = -\beta \tan(\pi\alpha/2).$$

253 Further refinements to this FFT-based approximation of stable densities
 254 were described by [33], which included numerically calculating the integral in
 255 (12) using Simpson's rule, and replacing the linear interpolation with cubic
 256 splines. In practice, we found that there was no noticeable advantage to
 257 using these more costly techniques in the long memory context.

258 To simplify matters, we assume a two-dimensional uniform joint prior
 259 over the allowable region (Fig. 1):

$$p_{d,\alpha}(d, \alpha) \propto \mathbb{1} \left\{ |d| < 1 - \frac{1}{\alpha}, 1 < \alpha < 2 \right\}. \quad (13)$$

260 Note that this prior places zero probability on $\alpha = 2$, i.e. the Gaussian.
 261 Because of their qualitatively different behaviours, we will not try and include
 262 the cases of $\alpha = 2$ and $\alpha < 2$ in the same analysis.

The prior (13) results in the marginal prior for d being no longer uniform:

$$\begin{aligned} p_d(d) &= \int_1^2 p_{d,\alpha}(d, \alpha) d\alpha \\ &= \frac{1}{2 - 2\log 2} \left(2 - \frac{1}{1 - |d|} \right), \quad |d| < \frac{1}{2}. \end{aligned}$$

Similarly, the marginal prior for α :

$$p_\alpha(\alpha) = \frac{1}{1 - \log 2} \left(1 - \frac{1}{\alpha} \right), \quad 1 < \alpha < 2.$$

d and α are updated as follows:

$$\begin{aligned} \xi_d | (d, \alpha) &\sim \mathcal{N}^{(-1 + \frac{1}{\alpha}, 1 - \frac{1}{\alpha})}(d, \sigma_d^2) \\ \xi_\alpha | (d, \alpha) &\sim \mathcal{N}^{(\frac{1}{1-|d|}, 2)}(\alpha, \sigma_\alpha^2), \end{aligned}$$

for some σ_α^2 . These proposals are accepted/rejected according to:

$$\begin{aligned} A_d(d, \xi_d) &= \Delta\ell + \log \left\{ \frac{\Phi[(1 - \frac{1}{\alpha} - d)/\sigma_d] - \Phi[(\frac{1}{\alpha} - 1 - d)/\sigma_d]}{\Phi[(1 - \frac{1}{\alpha} - \xi_d)/\sigma_d] - \Phi[(\frac{1}{\alpha} - 1 - \xi_d)/\sigma_d]} \right\} \\ A_\alpha(\alpha, \xi_\alpha) &= \Delta\ell + \log \left\{ \frac{\Phi[(2 - \alpha)/\sigma_\alpha] - \Phi[(\frac{1}{1-|d|} - \alpha)/\sigma_\alpha]}{\Phi[(2 - \xi_\alpha)/\sigma_\alpha] - \Phi[(\frac{1}{1-|d|} - \xi_\alpha)/\sigma_\alpha]} \right\}. \end{aligned}$$

263 with $\Delta\ell = \ell(\mathbf{x}|\xi_d, \boldsymbol{\psi}_{-d}) - \ell(\mathbf{x}|\boldsymbol{\psi})$

264 An alternative approach would be to propose the pair (d, α) jointly, but
 265 this is unnecessarily complicated in practice. It should be noted that, due to
 266 numerical issues for α near to 1, the lower bound of $\alpha = 1$ used throughout
 267 this procedure is actually replaced by $\alpha = 1.02$ in the computer code. The
 268 code we have written also allows α to be fixed (the Bayesian interpretation
 269 would be a unit mass prior). In this case, any prior can be used for d .
 270 Tuning of the proposal variance σ_α^2 is achieved using the same automatic
 271 tuning procedure outlined in Graves et al. [17]. Initial values of (d, α) are
 272 chosen uniformly randomly over the allowable region. Comparison of our
 273 Figure 1 with Figure 2 of [7] shows that our set of allowable parameter
 274 values is smaller, and in particular they can access part of the infinite mean
 275 range of α for some negative d values, however for both cases they coincide
 276 for positive d .

277 *3.1.3. Application to synthetic data*

278 In this section we evaluate our algorithm. Initially, 60 time series of length
 279 $n = 2^{10}$ were simulated with each value of $\alpha_I \in \{1.75, 1.50, 1.25\}$ being used
 280 20 times. Values of d_I were chosen randomly, conditional on the pair (d_I, α_I)
 281 being in the allowable range.

282 The residuals of d and α are presented in figure 2. From plot (a) we
 283 see that the Bayesian estimate of d appears to be approximately unbiased
 284 and the residual appears to be independent of d_I . It is immediately clear
 285 that accuracy of the Bayesian estimator appears to increase as α decreases.
 286 To confirm this, the average posterior standard deviations for d were 0.015,
 287 0.009, 0.004 for $\alpha_I = 1.75, 1.50, 1.25$ respectively. In summary therefore, our
 288 knowledge about d increases when α is small, i.e. detection of long memory is
 289 easier in the presence of fat tails. Although this may seem initially surprising,
 290 there is a strong intuitive explanation. Long memory is characterised by the
 291 slow decay of the influence of each innovation, or ‘shock’. In the Gaussian
 292 framework, no shock is very much larger than any other so the traceability of
 293 each shock is hard because it gets lost in the ‘noise’. Yet in the fat-tailed case,
 294 extreme shocks are to be expected, and their effects will be easier to observe
 295 through time. As α decreases, such shocks appear more frequently and are
 296 more dramatic, and consequently their decay profile is easier to determine.

297 Naturally we are also interested in the posterior of α . We see from figure
 298 2(b) that the Bayesian estimate of α is essentially unbiased, and the poste-
 299 rior standard deviation of α is roughly independent of d_I and α_I (although
 300 there is some suggestion that the posterior standard deviation is smallest
 301 when $\alpha_I = 1.25$, this is not practically significant). Interestingly, the joint
 302 posterior of (d, α) shows no correlation between the two parameters, indi-
 303 cating that the posteriors are independent. This fact helps to justify not
 304 proceeding with a joint proposal of (d, α) in the Metropolis–Hastings step
 305 in the previous subsection. Furthermore, it inspires the question: “can we
 306 improve our knowledge about d if we *know* the value of α ?” A simple Monte
 307 Carlo comparison shows that this is not the case (see figure 3) leading to
 308 the interesting result that, if we assume no knowledge of α , we sacrifice no
 309 information about d .

310 There is some mild correlation between the posteriors of α and γ (not
 311 shown). This is not surprising since both parameters affect the ‘variabil-
 312 ity’ of the underlying stable distribution. However the algorithm is able to
 313 successfully disentangle the scaling effects of γ and the shaping effects of

314 α . Finally, an analysis varying the length of times series n , was performed.
 315 From Fig. 4 we see that a $n^{-1/2}$ rule applies for stably distributed processes.
 316 Note also the relative decreases in posterior standard deviation obtained by
 317 decreasing α_I and increasing n . For example, to obtain the same level of
 318 confidence about the parameter d when $\alpha_I = 1.5$ and $n = 2^{10} = 1024$, one
 319 would have to use a time series of length about $2^{14} = 16384$ in the Gaussian
 320 case.

321 3.2. Asymmetric stable distributions

322 We will now briefly consider the asymmetric case, where $\beta \neq 0$. Again,
 323 because of the careful modularisation of the method, adding in this parameter
 324 to the Metropolis–Hastings algorithm presents no difficulty. Any prior on the
 325 support $[-1, 1]$ can be used but for convenience we will use the simplest form:

$$p_\beta(\cdot) \sim \mathcal{U}(-1, 1).$$

326 The proposal distribution is:

$$\xi_\beta | \beta \sim \mathcal{N}^{(-1,1)}(\beta, \sigma_\beta^2),$$

327 for some σ_β^2 . To test the efficacy of the method in this framework, we sim-
 328 ulated twenty processes with $\alpha_I = 1.5$, $\beta_I = 0.5$ and d_I randomly in the
 329 allowable range. The summary statistics are presented in table 1.

330 It is clear that the method can accurately determine the ‘skewness’ pa-
 331 rameter β . Further investigation reveals that the posterior of β is uncorre-
 332 lated with any other parameter (not shown). Also, as $\alpha \rightarrow 2$ the marginal
 333 posterior standard deviation of β becomes increasingly large, and the distri-
 334 bution actually approaches the uniform on $(-1, 1)$ (also not shown). This
 335 is because, as remarked upon earlier, the parameter β becomes increasingly
 336 unidentifiable as α increases, and at the Gaussian limit it is irrelevant.

337 3.3. t -distribution

338 To demonstrate the flexibility of our Bayesian MCMC algorithm, we will
 339 now briefly consider using t -distributed innovations. To our knowledge, there
 340 is no literature concerning long memory models with t -distributed innova-
 341 tions, most likely because of the reasons given at the end of the introduction.
 342 The t -distribution acts as a useful intermediate between the Gaussian and the

343 power law-tailed α stable distributions. To see this, consider its probability
 344 density function:

$$f(x; \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\pi\nu}} \left\{ 1 + \frac{x^2}{\nu} \right\}^{-\frac{\nu+1}{2}}. \quad (14)$$

345 As $x \rightarrow \infty$ the probability density function behaves as $\sim Ax^{-\nu-1}$ for some A
 346 and consequently the tail function behaves as $\mathbb{P}(X > x) \sim Bx^{-\nu}$ for some B .
 347 By comparison with (5) we see that such distributions are power law-tailed.
 348 However, unlike the stable distribution which allowed the tail exponent to
 349 be only $0 < a < 2$, here ν may take any positive value, leading to power law
 350 tail distributions that can have an arbitrary number of finite moments. In
 351 particular, for $\nu > 2$ the t -distribution has finite variance yet is still power
 352 law tailed, in direct contrast to stable distributions (and unlike them being
 353 attracted to the Gaussian under convolution). It is worth remarking that the
 354 limiting distribution of $\nu \rightarrow \infty$ is the standard Gaussian. Furthermore, the
 355 case $\nu = 1$ is the standardised Cauchy distribution. Recall that these two
 356 distributions also correspond to particular values of α -stable distributions
 357 ($\alpha = 1$ and 2 respectively).

358 Turning attention to t -distributed long memory processes, it will be useful
 359 to generalise (14) to obtain a scale-location distribution satisfying (8):

$$f(x; \delta, \gamma, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\gamma\sqrt{\pi\nu}} \left\{ 1 + \frac{1}{\nu} \left(\frac{x - \delta}{\gamma} \right)^2 \right\}^{-\frac{\nu+1}{2}}. \quad (15)$$

360 Such a t -distribution has variance $\gamma^2 \frac{\nu}{\nu-2}$ for $\nu > 2$ (and infinite for $\nu \leq 2$).
 361 Throughout the remainder of this section we will restrict attention to the
 362 ‘intermediate’ t -distributions that have finite variance, $\nu > 2$. As with the
 363 stable distribution, the scale parameter will be notated as γ rather than σ
 364 to avoid implying that it is also a standard deviation.

365 Due to the modularisation of the method outlined [17], it is relatively
 366 trivial to incorporate the t distribution into the Bayesian framework. Calcula-
 367 tion of the log-likelihood is straightforward given the density. The prior for
 368 ν can be chosen to be anything supported on the positive half-line. There is
 369 no standard non-informative prior for ν , so we will use an exponential prior
 370 truncated to the right of $\nu = 2$:

$$p_\nu(\nu) = \lambda e^{-\lambda(\nu-2)}, \quad \nu > 2, \quad (16)$$

371 for some λ to be chosen. Using a prior that is independent of the other model
 372 parameters again allows simplification in the Metropolis–Hastings step. To
 373 propose new values of ν , we will use the same exponentiated random-walk
 374 as for the scale parameter described, although restricted to $\nu > 2$:

$$\log(\xi_\nu - 2) = \log(\nu - 2) + v,$$

375 where $v \sim \mathcal{N}(0, \sigma_\nu^2)$ for some σ_ν^2 . Calculation of the relevant acceptance
 376 probability is trivial. The parameter σ_ν^2 can be automatically ‘tuned’ to
 377 obtain a desired acceptance rate. A suitable initial value for the pair (γ, ν)
 378 are the approximate MLEs which can be found crudely by treating the data as
 379 independent and identically t -distributed and maximising the log-likelihood
 380 numerically.

381 A small Monte Carlo study was conducted, for which we gener-
 382 ated 50 t -distributed ARFIMA(0, d , 0) time series with $\nu_I = 5$ and
 383 $d_I \in \{-0.45, -0.35, \dots, 0.45\}$. Two different priors were used, setting
 384 λ in (16) to be either 0.1 or 0.2. The interesting summary statistics are
 385 presented in table 2.

386 Choosing between the two priors suggested, or indeed any other prior, is
 387 of course up to the modeller. However one fact that may influence this choice
 388 is that, when erroneously applied to Gaussian data, the prior with $\lambda = 0.2$
 389 tends to produce point estimates for ν that are less than 30, whilst $\lambda = 0.1$
 390 leads to most being larger than 30. Since a basic rule-of-thumb is that, for
 391 $\nu > 30$, the t -distribution is practically indistinguishable from the Gaussian,
 392 this might suggest that the prior with $\lambda = 0.1$ might be more useful. It
 393 should be noted however that this analysis would be sensitive to the length
 394 of time series n .

395 3.4. Comparison with other estimators

396 In [40, 45, 41, 14] various estimator methods such as the Variable Band-
 397 width method [40], wavelets [14], Rescaled range (R/S) [22], Detrended Fluc-
 398 tuation Analysis (DFA) [39], the Whittle estimator [43, 42] and a semi-
 399 parametric power spectral method [15] have been used for estimating the
 400 LRD parameter d in an ARFIMA(1, d , 1) model with α -stable innovations.
 401 Comparison with our results shows that the d parameter is well estimated
 402 with relative small uncertainty bounds compared with the classical estima-
 403 tors [14].

404 4. Application to Solar X ray data

405 The GOES geostationary meteorological satellites have been used to ob-
406 serve X rays from solar flares over several decades starting in the mid 1970s.
407 Burnecki et al. [6], Stanislavsky et al. [44] fitted an ARFIMA model with α -
408 stable innovations to a time series comprising daily aggregates of solar flare
409 events derived from GOES data, and inferred the H , d and α values using
410 the finite impulse response (FIRT), variance of residuals (VaR, or DFA), and
411 McCulloch quantile methods. For this interval [6] found d to be 0.21 and α
412 to be 1.2674

413 Unfortunately the flare series studied by the previous authors was not
414 available at the time of writing from the original public National Oceanic
415 and Atmospheric Administration (NOAA) archive sites. Instead we have
416 obtained the full, 1 minute resolution GOES solar X-ray irradiance (in W/m^2)
417 in the 0.1-0.8 nanometre long wavelength channel for the period that [6]
418 identified, and we show its daily mean values in figure 5.

419 For Solar Cycle 23 the primary and secondary science satellites were
420 GOES 8, 10 and 12, and a correction factor of 0.7 was applied as per the
421 recommendations at the NOAA archive. The data is archived at http://satdat.ngdc.noaa.gov/sem/goes/data/new_avg/
422

423 We note that direct comparison with [6] is thus not possible, as our full X
424 ray series includes the sharp rises due to the onset of the flares, the decay of
425 flares, and the background X ray flux, and so our inferred ARFIMA model
426 should be seen as describing the daily mean of this rather than the daily
427 aggregated flares. For the daily mean data we find posterior mean estimates
428 of α , d to be 1.65 and 0.205.

429 For comparison we have used the maximum likelihood method imple-
430 mented in MATLAB to find a α value of 1.16, and from a power-spectral
431 estimator [15] we find $d = 0.37$. The differences to the Bayesian estimator
432 could be due to our estimator doing a joint estimate and/or they could be
433 due to finite time series length. However, our earlier tests using simulated
434 surrogate data gave very good results. Further work will also be needed to
435 compare in more detail the high resolution X ray time series with the de-
436 rived flare time series studied by [6], but our results suggest that an α -stable
437 ARFIMA model is indeed appropriate and useful for this higher resolution
438 dataset.

439 5. Conclusions

440 Self-similarity is by now well known and well studied, and has found
441 many applications in physics and elsewhere in the sciences of complexity.
442 However, as we discussed, although expressed by a single exponent, H , self-
443 similarity can arise both from long-range dependence and heavy-tailed jumps
444 respectively, thus giving two potential contributions to the exponent. In
445 consequence there is a need to simultaneously estimate both the long-range
446 dependence and heavy-tail distribution parameters, d and α . Although best
447 statistical practice allows joint estimation by some frequentist approaches,
448 the estimation is still sometimes done in the science literature by measuring
449 H and one of d or α . In this paper we presented a novel Bayesian method to
450 directly infer d and α on the hypothesis of an ARFIMA model with heavy
451 tailed innovations. Our method is flexible enough to allow the choice of
452 heavy-tailed distribution (e.g. α - or t -distributed), and we gave a demon-
453 stration of its effectiveness and accuracy on synthetic data and solar X-ray
454 data.

455 6. Acknowledgements

456 CF was supported by the German Research Foundation (DFG) through
457 the cluster of excellence CliSAP (EXC177), the SFB/TRR181 and grant
458 FR3515/3-1. CF and NWW acknowledge travel funding from the Norwegian
459 Research Council KLIMAFORSK project no 229754. NWW acknowledges
460 travel support from the London Mathematical Laboratory, Office of Naval
461 Research grant NICOP-N62909-15-1-N143 at Warwick and Potsdam, and
462 useful discussions with Sandra Chapman. ET is supported by an STFC
463 student ship.

464 **References**

- 465 [1] Beran, J. (1994). *Statistics for Long Memory Processes*. Chapman &
466 Hall, New York.
- 467 [2] Bergström, H. (1952). On some expansions of stable distribution func-
468 tions. *Arkiv för Matematik*, 2(4), 375–378.
- 469 [3] Bouchaud, J.-P. & Potters, M. (2003). *Theory of Financial Risk and*
470 *Derivative Pricing: From Statistical Physics to Risk Management*. Cam-
471 bridge University Press.
- 472 [4] Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of
473 human travel. *Nature*, 439, 462–465.
- 474 [5] Buckle, D. J. (1995). Bayesian inference for stable distributions. *Journal*
475 *of the American Statistical Association*, 90(430), 605–613.
- 476 [6] Burnecki, K., Klafter, J., Magdziarz, M., & Weron, A. (2008). From solar
477 flare time series to fractional dynamics. *Physica A: Statistical Mechanics*
478 *and its Applications*, 387(5), 1077–1087.
- 479 [7] Burnecki, K. & Sikora, G. (2013). Estimation of farima parameters in the
480 case of negative memory and stable noise. *IEEE Transactions on Signal*
481 *Processing*, 61(11), 2825–2835.
- 482 [8] Burnecki, K. & Weron, A. (2014). Algorithms for testing of fractional
483 dynamics: a practical guide to arfima modelling. *Journal of Statistical*
484 *Mechanics: Theory and Experiment*, (pp. P10036).
- 485 [9] Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings
486 algorithm. *The American Statistician*, 49(4), 327–335.
- 487 [10] DuMouchel, W. H. (1983). Estimating the stable index α in order to
488 measure tail thickness: A critique. *The Annals of Statistics*, 11(4), 1019–
489 1031.
- 490 [11] Embrechts, P. & Maejima, M. (2002). *Selfsimilar Processes*. Princeton
491 University Press.
- 492 [12] Feller, W. (1957). *An Introduction to Probability Theory and Its Appli-*
493 *cations*, 3rd ed., volume 1. Wiley, New York.

- 494 [13] Fox, R. & Taqqu, M. S. (1986). Large-sample properties of parameter
495 estimates for strongly dependent stationary gaussian time series. *The*
496 *Annals of Statistics*, (pp. 517–532).
- 497 [14] Franzke, C. L. E., Graves, T., Watkins, N. W., Gramacy, R. B., &
498 Hughes, C. (2012). Robustness of estimators of long-range dependence
499 and self-similarity under non-Gaussianity. *Philosophical Transactions of*
500 *the Royal Society A*, 370(1662), 1250–1267.
- 501 [15] Geweke, J. & Porter-Hudak, S. (1983). The estimation and application
502 of long memory time series models. *Journal of time series analysis*, 4(4),
503 221–238.
- 504 [16] Graves, T. (2013). *A systematic approach to Bayesian inference for long*
505 *memory processes*. PhD thesis, University of Cambridge, UK.
- 506 [17] Graves, T., Franzke, C. L. E., Watkins, N. W., & Gramacy, R. B. (2015).
507 Efficient Bayesian inference for long memory processes. *Nonlin. Proc. Geo-*
508 *phys.*, 22, 679–700.
- 509 [18] Green, P. J. (1995). Reversible jump Markov chain Monte Carlo com-
510 putation and Bayesian model determination. *Biometrika*, 82(4), 711–732.
- 511 [19] Gregory, P. (2005). *Bayesian Data Analysis for the Physical Sciences*.
512 Cambridge University Press.
- 513 [20] Hannan, E. J. (1973). The asymptotic theory of linear time-series mod-
514 els. *J. Appl. Probab.*, (pp. 130–145).
- 515 [21] Hill, B. M. (1975). A simple general approach to inference about the
516 tail of a distribution. *The Annals of Statistics*, 3(5), 1163–1174.
- 517 [22] Hurst, H. E., Black, R. P., & Simaika, Y. M. (1965). *Long-Term Storage:*
518 *An Experimental Study*. Constable, London.
- 519 [23] Kokoszka, P. S. & Taqqu, M. S. (1996). Parameter estimation for infinite
520 variance fractional arima. *The Annals of Statistics*, 24(5), 1880–1913.
- 521 [24] Laskin, N., Lambadatis, I., Harmantzis, F. C., & Devetsikiotis, M.
522 (2002). Fractional levy motion and its application to network traffic mod-
523 eling. *Computer Networks*, 40, 363–275.

- 524 [25] Mandelbrot, B. B. (1963). The variation of certain speculative prices.
525 *The Journal of Business*, 36(4), 394–419. Correction: Mandelbrot [26].
- 526 [26] Mandelbrot, B. B. (1972). Correction of an error in Mandelbrot [25].
527 *The Journal of Business*, 45(4), 542–543.
- 528 [27] Mandelbrot, B. B. (2002). *Gaussian Self-Affinity and Fractals: Globality, The Earth, 1/f Noise, and R/S*, volume H of *Selecta*. Springer; New
529 York.
530
- 531 [28] Mandelbrot, B. B. & Van Ness, J. W. (1968). Fractional Brownian
532 motions, fractional noises and applications. *SIAM Review*, 10(4), 422–437.
- 533 [29] Mandelbrot, B. B. & Wallis, J. R. (1968). Noah, Joseph and operational
534 hydrology. *Water Resources Research*, 4(5), 909–918.
- 535 [30] Mandelbrot, B. B. & Wallis, J. R. (1969). Robustness of the rescaled
536 range R/S in the measurement of noncyclic long run statistical dependence.
537 *Water Resources Research*, 5(5), 967–988.
- 538 [31] Mantegna, R. & Stanley, H. E. (2000). *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press.
539
- 540 [32] McCulloch, J. H. (1997). Measuring tail thickness to estimate the stable
541 index α : A critique. *Journal of Business & Economic Statistics*, 15(1), 74–
542 81.
- 543 [33] Menn, C. & Rachev, S. T. (2006). Calibrated FFT-based density approximations for α -stable distributions. *Computational Statistics and Data Analysis*, 50(8), 1891–1904.
544
545
- 546 [34] Mercik, S., Weron, K., Burnecki, K., & Weron, A. (2003). Enigma of
547 self-similarity of fractional Lévy stable motions. *Acta Physica Polonica B*,
548 34(7), 3773–3791.
- 549 [35] Mikosch, T., Gadrich, T., Kluppelberg, C., & Adler, R. J. (1995). Parameter estimation for arma models with infinite variance innovations. *The Annals of Statistics*, (pp. 305–326).
550
551
- 552 [36] Mittnik, S., Doganoglu, T., & Chenyao, D. (1999). Computing the probability density function of the stable Paretian distribution. *Mathematical and Computer Modelling*, 29(10–12), 235–240.
553
554

- 555 [37] Nolan, J. P. (1998). Parameterizations and modes of stable distributions.
556 *Statistics and Probability Letters*, 38(2), 187–195.
- 557 [38] Painter, S. & Paterson, L. (1994). Fractional Lévy motion as a model
558 for spatial variability in sedimentary rock. *Geophysical Research Letters*,
559 21(25), 2857–2860.
- 560 [39] Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E.,
561 & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides.
562 *Physical Review E*, 49(2), 1685–1689.
- 563 [40] Schmittbuhl, J., Vilotte, J.-P., & Roux, S. (1995). Reliability of self-
564 affine measurements. *Physical Review E*, 51(1), 131.
- 565 [41] Sheng, H., Chen, Y., & Qiu, T. (2011). On the robustness of hurst
566 estimators. *IET Signal Processing*, 5(2), 209–225.
- 567 [42] Shimotsu, K. & Phillips, P. C. (2006). Local whittle estimation of
568 fractional integration and some of its variants. *Journal of Econometrics*,
569 130(2), 209–233.
- 570 [43] Shimotsu, K., Phillips, P. C., et al. (2005). Exact local whittle estimation
571 of fractional integration. *The Annals of Statistics*, 33(4), 1890–1933.
- 572 [44] Stanislavsky, A. A., Burnecki, K., Magdziarz, M., Weron, A., & Weron,
573 K. (2005). Farima modeling of solar flare activity from empirical time series
574 of soft x-ray solar emission. *Astrophysical Journal*, 693(2), 1877–1882.
- 575 [45] Stoev, S. & Taqqu, M. S. (2005). Asymptotic self-similarity and wavelet
576 estimation for long-range dependent fractional autoregressive integrated
577 moving average time series with stable innovations. *Journal of Time Series*
578 *Analysis*, 26(2), 211–249.
- 579 [46] Taqqu, M. S. & Teverovsky, V. (1998). On estimating the intensity
580 of long-range dependence in finite and infinite variance time series. *A*
581 *practical guide to heavy tails: statistical techniques and applications*, 177,
582 218.
- 583 [47] Watkins, N. W., Credgington, D., Hnat, B., Chapman, S. C., Freeman,
584 M. P., & Greenhough, J. (2005). Towards synthesis of solar wind and
585 geomagnetic scaling exponents: A fractional Lévy motion model. *Space*
586 *Science Reviews*, 121(1-4), 271–284.

- 587 [48] Waxman, D. & Feng, J. (2005). Implications of long tails in the distri-
588 bution of mutant effects. *Physica D*, 206(3), 265–274.
- 589 [49] Zolotarev, V. M. (1986). *One-dimensional Stable Distributions*, vol-
590 ume 65 of *Translations of Mathematical Monographs*. American Mathe-
591 matical Society.

Table 1: Posterior summary statistics for asymmetric-stable ARFIMA(0, d , 0) process. Average of 20 runs.

	mean	std	95% CI endpoints	
d_R	0.002	0.008	-0.014	0.018
α	1.481	0.044	1.398	1.570
β	0.483	0.084	0.316	0.643

Table 2: Comparison of posterior summary statistics for t_5 -distributed ARFIMA(0, d , 0) process using two different priors. Average of 50 runs.

		mean	std	95% CI endpoints	
$\lambda = 0.1$	d_R	0.003	0.022	-0.039	0.046
$\lambda = 0.2$	d_R	0.003	0.022	-0.039	0.046
$\lambda = 0.1$	γ	1.005	0.039	0.931	1.080
$\lambda = 0.2$	γ	1.002	0.038	0.929	1.077
$\lambda = 0.1$	ν	5.665	1.061	3.857	7.760
$\lambda = 0.2$	ν	5.543	0.987	3.850	7.511

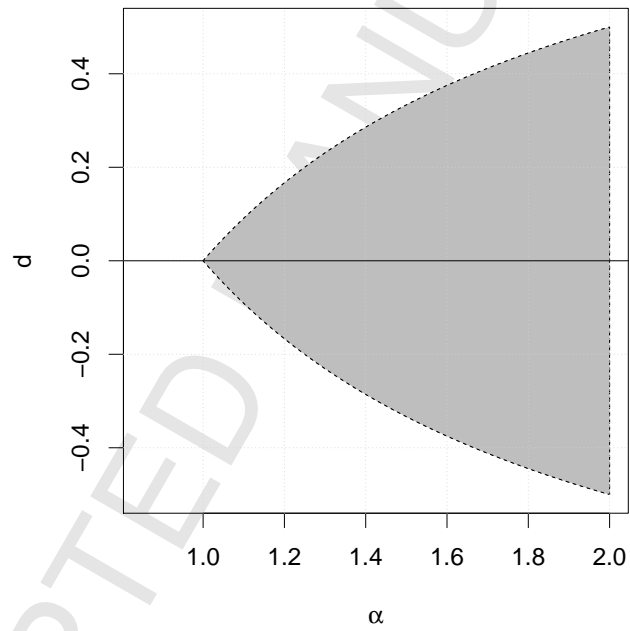


Figure 1: Set of allowable values for (α, d) ; does not include the dotted boundary.

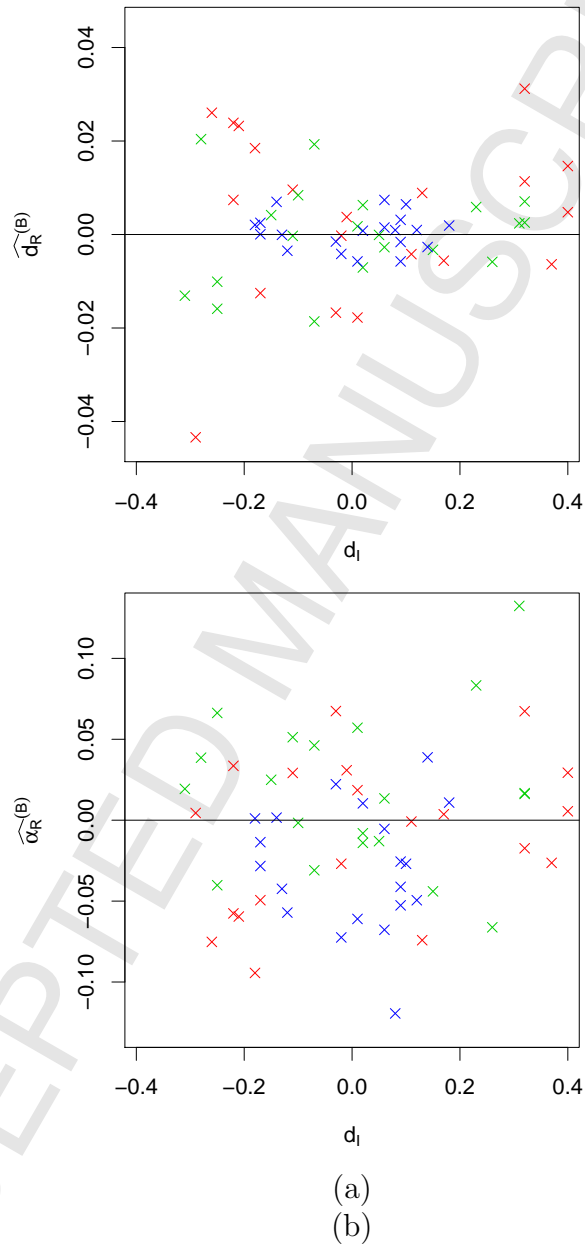


Figure 2: Posterior outputs from α -stable ARFIMA(0, d , 0) series; (a) $\hat{d}_R^{(B)}$ against d_I , (b) $\hat{\alpha}_R^{(B)}$ against d_I . Red is used for $\alpha_I = 1.75$, green for $\alpha_I = 1.50$ and blue for $\alpha_I = 1.25$.

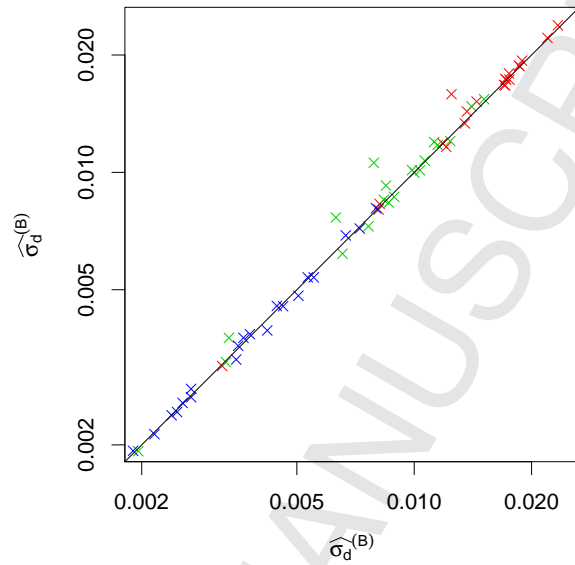


Figure 3: Posterior outputs; log-scale comparison of $\widehat{\sigma}_d^{(B)}$ when α is assumed unknown (y -axis) against unknown (x -axis). Red is used for $\alpha_I = 1.75$, green for $\alpha_I = 1.50$ and blue for $\alpha_I = 1.25$.

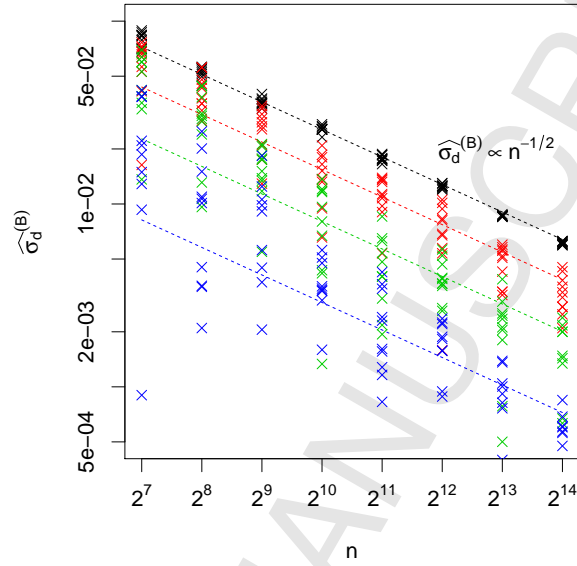


Figure 4: Posterior outputs from ARFIMA(0,0,0) series; (a) $\widehat{\sigma}_d^{(B)}$ against n . Black is used for $\alpha_I = 2$, red for $\alpha_I = 1.75$, green for $\alpha_I = 1.5$ and blue for $\alpha_I = 1.25$. (log-log scale).

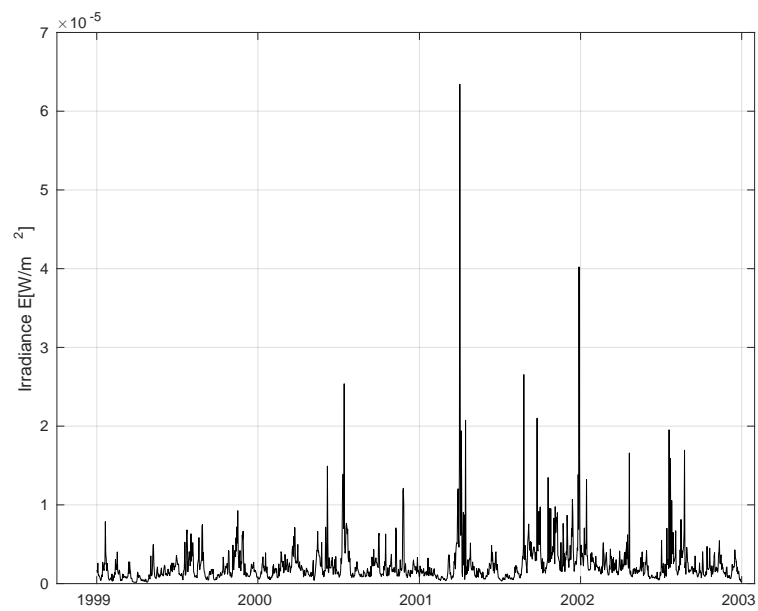


Figure 5: GOES time series.