

# Fully Bayesian Inference for $\alpha$ -Stable Distributions Using a Poisson Series Representation

Tatjana Lemke, Marina Riabiz and Simon J. Godsill

## Abstract

In this paper we develop an approach to Bayesian Monte Carlo inference for skewed  $\alpha$ -stable distributions. Based on a series representation of the stable law in terms of infinite summations of random Poisson process arrival times, our framework leads to a simple representation in terms of conditionally Gaussian distributions for certain latent variables. Inference can therefore be carried out straightforwardly using techniques such as auxiliary variables versions of Markov chain Monte Carlo (MCMC) methods. The Poisson series representation (PSR) is further extended to practical application by introducing an approximation of the series residual terms based on exact moment calculations. Simulations illustrate the proposed framework applied to skewed  $\alpha$ -stable simulated and real-world data, successfully estimating the distribution parameter values and being consistent with other (non-Bayesian) approaches. The methods are highly suitable for incorporation into hierarchical Bayesian models, and in this case the conditionally Gaussian structure of our model will lead to very efficient computations compared to other approaches.

## Index Terms

Asymmetric  $\alpha$ -stable distribution, Lepage series, Poisson series representation, residual approximation, conditionally Gaussian, Markov chain Monte Carlo.

## I. A NOTE OF RELEVANCE TO THIS SPECIAL ISSUE

Professor Bill Fitzgerald was an influential and motivational personality who inspired many in our laboratory and throughout the world to carry out research in Bayesian methods, see e.g. [1] for his much-used text on Bayesian methods for signal processing. The third author would like to warmly acknowledge the inspiration of Bill Fitzgerald for the current paper. Around 1993, while I was carrying out my PhD research in audio signal restoration, Bill introduced me to the beautiful topic of stable law distributions and generalised versions of the central limit theorem. This topic fascinated me and has led to a series of papers on Bayesian methods for inference in the presence of  $\alpha$ -stable distributions, initially focussed on symmetric stable laws, for which very elegant scale mixture of normals representations and associated inference procedures can be devised [2]–[7]. More recently we have studied powerful and general representations of continuous time  $\alpha$ -stable Lévy processes based on series of Poisson random variables [8]–[10], and it is this representation which forms the basis of the current paper.

## II. INTRODUCTION

In a diverse range of fields, including natural science, economics and engineering areas as radar processing, telecommunications and acoustics [11]–[13], real-world processes which exhibit jumps and asymmetric behaviour are present. An underlying Gaussian distribution is rather unsuitable to describe these characteristics. Thus, the extension to the family of stable distributions, which forms a generalisation to the Gaussian law, arising from the generalised version of the central limit theorem (GCLT) as shown in Gnedenko and Kolmogorov, and Feller [14], [15], seems to be the natural way to go. In contrast to the classical CLT the generalised version forgoes the condition of a finite variance and assigns a much less restrictive requirement on tail behaviour. Compared to the Gaussian distribution the general stable distribution depends upon two additional parameters, which represent the asymmetry and heavy tailedness, and is therefore more suited to modelling phenomena showing these empirical features. Other more general classes of distribution can also be considered for heavy-tailed modelling, and we mention just a few which are linked to the  $\alpha$ -stable case. Rosiński [16] formally introduced the class of tempered stable distribution, which combine both alpha-stable and Gaussian properties. Alternative classes of distributions, the so-called modified stable laws and normal modified stable laws provide additional flexibility in modelling the dynamics of financial time series and are discussed by Barndorff-Nielsen and Shephard [17].

Exact simulation of general  $\alpha$ -stable random variates can be carried out using an elegant auxiliary variables framework, see [18]. This same representation was used in an early and fundamental contribution to the Bayesian inference area by Buckle [19], employing Markov chain Monte Carlo (MCMC) methods to infer parameters of the distribution, and extended to time series problems by Quiou and Ravishanker [20]. Despite its elegance, practitioners have typically found these methods hard to tune and apply successfully to real problems. In an alternative approach, Kuruoğlu [21] addressed positive  $\alpha$ -stable probability distributions, providing an analytical approximation based on a decomposition into a product of a Pearson and another positive stable random variable, while inference for autoregressive (AR) processes with possibly asymmetric  $\alpha$ -stable innovations has been presented by Gençağa et al. [22] using a sequential Bayesian approach.

Limiting the models to the symmetric case, Godsill and Kuruoğlu [4], [5] introduced Monte Carlo Expectation-Maximisation (MCEM) and Markov chain Monte Carlo methods, based on the Scale Mixtures of Normals (SMiN) representation of symmetric stable distributions, see also the related methods proposed by Tsonas [23]. An on-line Bayesian filtering and smoothing method, also based on the SMiN, was proposed by Lombardi and Godsill [6]. We will see later that these representations can be viewed as a special case of the approach in this paper for the full asymmetric class of distributions.

In this paper we take a new approach, making use of some remarkable Poisson series representations (PSR) for asymmetric  $\alpha$ -stable distributed random variables in order to provide a conditionally Gaussian framework for inference. By so doing we allow for Bayesian parameter estimation using simple Gibbs sampling-based Markov chain Monte Carlo (MCMC) approaches. We have presented parts of this work in conference publications [8]–[10]. Further details can be found in [24]. This paper however presents for the first time the full technical details and methodology for our proposed sampling schemes. Our motivating examples in the paper are for simple parameter inference in independent and identically distributed (i.i.d.) stable-law data, but we stress that the approach’s real potential is attained when the sampler is embedded in a complex hierarchical model with other parameters to estimate in addition to the stable law distribution parameters, see [9], [24] for some initial examples in this direction.

The original contribution of this paper consists in the use of the approximated PSR and the resulting conditionally Gaussian framework for  $\alpha$ -stable random variables to perform Bayesian MCMC inference for the distribution parameters. The advantages of the proposed approach are twofold. On the one hand our inference technique inherits the benefits of Bayesian methods, enabling the characterization of the whole posterior distribution of the parameters, as opposed to the (approximated) frequentist point estimators present in the literature. In particular in Section VI we make comparisons with the maximum likelihood estimator described by Nolan [25], with the method based on the inversion of the empirical characteristic function, as in [26], and with the estimator based on the quantiles of the distribution introduced by McCulloch [27]. On the other hand the PSR approach results in a Gibbs sampler algorithm with full conditional distributions which are either straightforward to sample from, or easy to target through a Metropolis-Hastings within Gibbs sampler step. This is in contrast to the above cited Bayesian method [19], where proposing the new parameters requires, at each step, either accurate hand-tuning of the proposal distributions, or the implementation of rejection sampling envelopes on the full conditionals. Both our method and Buckle’s method can be embedded in Bayesian hierarchical frameworks for the inference of linear discrete time series models driven by stable innovations. Examples that aim at the simultaneous estimation of the parameters characterizing the stable law and the parameters of the linear autoregressive (AR) model can be found in [9], [24] for the PSR, and in [20] for the Buckle model. However, the very significant advantage of our approach compared with the Buckle method is that we achieve a conditionally Gaussian framework for the sampler. This means that our method may be added in as an extra Gibbs sampling step in a sampler already devised for the Gaussian case. This not only means that existing code can be largely re-used when switching from a Gaussian noise case to an asymmetric stable law, as was done in [9] (based on the earlier code of [28], [29]) but it also means that our methods are amenable to more efficient block-sampling schemes for many parameters simultaneously, Rao-Blackwellised and collapsed Gibbs sampler schemes. A significant consequence of this is that in our sampler the skewness parameter  $\beta$  and scale parameter  $\sigma$  can be indirectly marginalised completely from the formulation, thus further simplifying the sampler and its statistical efficiency. An additional benefit (not detailed in this paper) is a simple representation for stable law Lévy processes, which is not achievable using the Buckle method, see for example sequential state estimation in the continuous time autoregressive model, based on the PSR [8], [24]. The conditionally Gaussian representation allows to use the Kalman filter to estimate the linear part of the state and the Rao-Blackwellised particle filter to estimate the non-linear part. The parameters of the distribution are assumed to be known, but their estimation can in principle also be embedded in the sampler. Finally we remark that the PSR of the  $\alpha$ -stable distribution substantially generalizes the above mentioned works based on the SMiN representation to the asymmetric case.

With respect to our previous work [10], [24], we include in this paper for the first time the estimation of the location parameter, and we also detail additional sampling procedures for data points lying in the tails of the distribution, which are crucial for the practical performance of the sampler.

The paper is organised as follows. In Section III, we introduce  $\alpha$ -stable distributions and state the definition and the PSR of an  $\alpha$ -stable random variable. In Section IV, we present our residual approximation approach. In Section V, we discuss inference for  $\alpha$ -stable distribution parameters via MCMC. In Section VI, we present results of our work and compare them with the mentioned common frequentist approaches. Finally Section VII concludes the paper.

### III. THE $\alpha$ -STABLE DISTRIBUTION

The family of  $\alpha$ -stable distributions forms a rich class of distributions, which allow for asymmetry and heavy tails. In general, closed-form density functions of  $S_\alpha(\sigma, \beta, \mu)$  are not known. Beside the Gaussian distribution for  $\alpha = 2$ , the exceptions are for  $\alpha = 1$  and  $\beta = 0$ , yielding the Cauchy distribution and  $\alpha = 1/2$ ,  $\beta = \pm 1$  for the Lévy distribution. As a generalisation of the classical central limit theorem, the  $\alpha$ -stable distribution follows a generalised central limit theorem, which states that the sum

of a number of random variables with power-law tail distributions decreasing as  $1/|x|^{\alpha+1}$  where  $0 < \alpha < 2$  will converge to an  $\alpha$ -stable distribution as the number of random variables approaches infinity. The  $\alpha$ -stable family of distributions  $S_\alpha(\sigma, \beta, \mu)$  is defined in terms of its characteristic function  $\phi_X(\cdot)$  [15], [30], [31]:

*Theorem 3.1:* A real-valued random variable  $X$  is stable if and only if there exists  $0 < \alpha \leq 2$ ,  $\sigma > 0$ ,  $-1 \leq \beta \leq 1$ ,  $\mu \in \mathbb{R}$  such that for all  $t \in \mathbb{R}$

$$\mathbb{E}[\exp(itX)] = \begin{cases} \exp(-\sigma^\alpha |t|^\alpha [1 - i\beta \text{sign}(t) \tan(\frac{\alpha\pi}{2})] + i\mu t), & \alpha \neq 1 \\ \exp(-\sigma |t| [1 + i\beta \frac{2}{\pi} \text{sign}(t) \ln |t|] + i\mu t), & \alpha = 1, \end{cases} \quad (1)$$

The four parameters and their domains are as follows:

- $\alpha \in (0, 2]$  is referred to as the *characteristic exponent* or *index of stability*, which measures the tail thickness,
- $\beta \in [-1, 1]$  is termed the *skewness parameter*. For  $\beta = 1$  ( $\beta = -1$ ), the family of distributions  $S_\alpha(\sigma, \beta, \mu)$  specialises to the positive (negative) stable family, whereas  $\beta = 0$  indicates the symmetric  $\alpha$ -stable (S $\alpha$ S) distribution,
- $\sigma > 0$  denotes the *scale parameter*,
- $\mu \in \mathbb{R}$  denotes the *location parameter*.

### A. General Scheme Based upon Conditional Gaussians

Our aim in this work is to simplify inference in  $\alpha$ -stable models by making use of powerful auxiliary variables representations of  $\alpha$ -stable random variables and processes. Such formulations are able to represent the  $\alpha$ -stable models exactly, in terms of distributions which may either be evaluated point-wise, or sampled using standard techniques. In particular we will seek conditionally Gaussian representations of these models, where both the mean and precision parameters of the models are considered as random variables. In the symmetric stable case our previous work has demonstrated such a framework, using just scale mixtures of normals, see [4], [5], and also [23]. This is in contrast with the MCMC approach of Buckle [19], in which an exact auxiliary variables approach is proposed, but computations are difficult because no conditionally Gaussian structure arises. Our current work provides a new framework which includes [4], [5] as a special case (as discussed later).

In order to motivate the subsequent developments, suppose we have the following marginal-conditional representation for the marginal stable law,  $X \sim S_\alpha(\sigma, \beta, \mu)$  :

$$(X, \mu_X, \sigma_X) \sim \mathcal{N}(X|\mu_X, \sigma_X^2) p_{\alpha, \beta, \sigma}(\mu_X, \sigma_X),$$

where  $p_{\alpha, \beta, \sigma}(\mu_X, \sigma_X)$  denotes the distribution of the possibly dependent latent variables  $\mu_X, \sigma_X$ , which in turn depends on the distribution parameters  $\alpha, \beta, \sigma$ . This is an auxiliary variables representation of the distribution and would facilitate inference procedures. For example, a conditional distribution for the latent variables  $\mu_X, \sigma_X$  can be formed as follows:

$$p(\mu_X, \sigma_X | X) \propto \mathcal{N}(X|\mu_X, \sigma_X^2) p_{\alpha, \beta, \sigma}(\mu_X, \sigma_X)$$

and this may be sampled in a Gibbs sampler style of implementation. The form of  $p_{\alpha, \beta, \sigma}(\mu_X, \sigma_X)$ , the mixing density, determines the form of the marginal  $p(X)$ , and in subsequent sections we will show how this can be specified exactly, sampled and characterised in the  $\alpha$ -stable case.

### B. Poisson Series Representation

There are a number of possible approximate representations of  $\alpha$ -stable laws, including the series expansions of Bergström [32] and the Gaussian mixtures representations of Kuruoğlu [3]. Here, however, we study a remarkable representation that has not received much attention in the estimation literature for  $\alpha$ -stable models. In the basic result, which is based on a Lepage type series, see for example [33], and which we term the Poisson series representation (PSR), a skewed  $\alpha$ -stable random variable  $Z$  can be represented in terms of a convergent sum based on arrival times of a Poisson process  $\{\Gamma_i\}$ , and some i.i.d. random variables  $\{W_i\}$ . Such series have been used for forward simulation of stable law variables and processes, see e.g. [33], [34], but not to our knowledge in inference frameworks. The basic result is (see [31, Theorem 1.4.5] for a full development and proof):

$$\sum_{i=1}^M \left( \Gamma_i^{-1/\alpha} W_i - k_i^{(\alpha)} \right) \xrightarrow{a.s.} Z \sim S_\alpha(\sigma, \beta, 0), \text{ as } M \rightarrow \infty, \quad (2)$$

which also includes the general skewed case ( $\beta \neq 0$ ). Here the terms are defined as:

$$k_i^{(\alpha)} = \begin{cases} 0 & \text{if } 0 < \alpha < 1, \\ \mathbb{E}[W_1 \int_{|W_1|/i}^{|W_1|/(i-1)} x^{-2} \sin(x) dx] & \text{if } \alpha = 1, \\ \frac{\alpha}{\alpha-1} \left( i^{\frac{\alpha-1}{\alpha}} - (i-1)^{\frac{\alpha-1}{\alpha}} \right) \mathbb{E}[W_1] & \text{if } 1 < \alpha < 2 \end{cases} \quad (3)$$

and

$$\sigma^\alpha = \frac{\mathbb{E}|W_1|^\alpha}{C_\alpha}, \quad \beta = \frac{\mathbb{E}[|W_1|^\alpha \operatorname{sgn} W_1]}{\mathbb{E}|W_1|^\alpha}. \quad (4)$$

The sequences of independent random variables  $\{W_1, W_2, \dots\}$  and  $\{\Gamma_1, \Gamma_2, \dots\}$  are defined as follows. The  $W_i$  are i.i.d. but otherwise essentially arbitrary, subject to having finite moments as follows:

$$\mathbb{E}[|W_1|^\alpha] < \infty \quad \text{if } \alpha \neq 1, \quad (5)$$

$$\mathbb{E}[|W_1 \ln(|W_1|)|] < \infty \quad \text{if } \alpha = 1, \quad (6)$$

and  $\Gamma_i$  are arrival times of a unit rate Poisson process, such that  $(\Gamma_i - \Gamma_{i-1}) \sim \text{Exponential}(1)$ . The constant  $C_\alpha$  is defined as

$$C_\alpha = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha) \cos(\pi\alpha/2)} & \text{if } \alpha \neq 1, \\ 2/\pi & \text{if } \alpha = 1. \end{cases} \quad (7)$$

We refer to Appendix A for a consideration on the relationship between the distribution of  $W_i$  and the possibility of achieving any value of  $\sigma > 0$  and  $\beta \in [-1, 1]$ , for a fixed value of the characteristic exponent  $\alpha$ , through the equations in (4).

Adding the location parameter  $\mu$ , we obtain the PSR for any  $\alpha$ -stable random variable,  $X \sim S_\alpha(\sigma, \beta, \mu)$ :

$$\sum_{i=1}^M \left( \Gamma_i^{-1/\alpha} W_i - k_i^{(\alpha)} \right) + \mu \xrightarrow{a.s.} Z + \mu =: X \sim S_\alpha(\sigma, \beta, \mu), \quad \text{as } M \rightarrow \infty. \quad (8)$$

We further define  $\tilde{k}_i^{(\alpha)} := k_i^{(\alpha)}/\mathbb{E}[W_1]$ . The constant  $k_i^{(\alpha)}$  gives the compensation for the otherwise divergent sum  $\sum_{i=1}^\infty \Gamma_i^{-1/\alpha} W_i$  when  $\alpha > 1$  and the  $W_i$  are asymmetric. Asymptotically, as  $i \rightarrow \infty$ ,  $k_i^{(\alpha)}$  equals to  $\mathbb{E}[\Gamma_i^{-1/\alpha} W_i]$ . For the sake of simplicity, we do not consider the special case  $\alpha = 1$  within this paper, although similar representations do exist for this boundary case [31, Theorem 1.4.5].

*a) Interpretation in terms of Lévy processes:* While this representation looks to have the flavour of a generalised Central Limit Theorem, we should note that in fact the terms  $\Gamma_i^{-1/\alpha} W_i$  are *not* i.i.d., and so something more subtle is involved. A convenient and intuitive interpretation of the result is provided by Samorodnitsky and Taqqu [31] in terms of the increments of an  $\alpha$ -stable Lévy Process. In this interpretation, the terms  $\Gamma_i^{-1/\alpha} W_i$  can be considered as the individual increments of the process over a finite time interval. Since there are almost surely an infinite number of jumps in any finite time interval for the  $\alpha$ -stable Lévy process, the summation is an infinite summation. The jumps are ordered in the series by decreasing expected scale, since the  $W_i$  are i.i.d., while the  $\Gamma_i^{-1/\alpha}$  are strictly decreasing with  $i$ . We notice that the earliest jumps in the series induce the heavy-tailed behaviour of the random variable, since  $\Gamma_1^{-1/\alpha}$  can take very large values with high probability, while later terms in the series correspond to small jumps and less heavy-tailed behaviour. This observation motivates our later Gaussian approximation for the residual terms when the series is truncated at a finite limit. It also motivates use of the  $\alpha$ -stable model in a financial returns application later in the paper, since we interpret the jumps as individual ‘shocks’ to the price of the commodity which are accumulated over time intervals. In the limiting case the number of shocks tends to infinity.

*b) Conditionally Gaussian Representation:* The convergence in (2) demonstrates that skewed  $\alpha$ -stable random variables can be simulated using independent and identically distributed random variables  $\{W_i\}_{i=1}^\infty$ , whose distribution is essentially arbitrary, and the arrival times of a unit rate Poisson process  $\{\Gamma_i\}_{i=1}^\infty$ . Note especially that we are free to choose any convenient distribution for the i.i.d. variables  $W_i$ , subject to the finite moment conditions. Making use of the above PSR then permits a Gaussian framework conditional upon  $\{\Gamma_i\}_{i=1}^\infty$  by choosing the i.i.d. random variables  $W_i$  in (2) to be Gaussian,  $W_i \sim \mathcal{N}(\mu_W, \sigma_W^2)$ , which from (4) above implies particular values of the  $\alpha$ -stable distribution parameters,  $\beta = \frac{\mathbb{E}[|W_1|^\alpha \operatorname{sgn} W_1]}{\mathbb{E}|W_1|^\alpha}$  and  $\sigma^\alpha = \frac{\mathbb{E}|W_1|^\alpha}{C_\alpha}$ , and  $\mathbb{E}[W_1] = \mu_W$  in (3). The conditionally Gaussian structure for  $X \sim S_\alpha(\sigma, \beta, \mu)$  can now be obtained directly from (2), and the assumption that  $W_i \sim \mathcal{N}(\mu_W, \sigma_W^2)$ , as:

$$X|\{\Gamma_i\}_{i=1}^\infty \sim \mathcal{N} \left( \mu_W \sum_{i=1}^\infty \left( \Gamma_i^{-1/\alpha} - \tilde{k}_i^{(\alpha)} \right) + \mu, \sigma_W^2 \sum_{i=1}^\infty \Gamma_i^{-2/\alpha} \right) := \mathcal{N}(\mu_X, \sigma_X^2). \quad (9)$$

### C. Relationship with Symmetric Case

We note here that the symmetric case  $\beta = 0$  as studied in [4], [5] can be obtained as a special case of the above form, since with  $\mu_W = 0$  we have:

$$X|\{\Gamma_i\}_{i=1}^\infty \sim \mathcal{N} \left( \mu, \sigma_W^2 \sum_{i=1}^\infty \Gamma_i^{-2/\alpha} \right) := \mathcal{N}(\mu_X = \mu, \sigma_X^2) \quad (10)$$

and we have from (2) that

$$\sum_{i=1}^M \Gamma_i^{-2/\alpha} \xrightarrow{a.s.} \lambda \sim S_{\alpha/2}((C_{\alpha/2})^{-2/\alpha}, 1, 0), \text{ as } M \rightarrow \infty. \quad (11)$$

This latter is the fully skewed mixing distribution used in the scale mixtures of normals case [4], [5]. In that earlier work, the problem was substantially simplified by the full characterisation of the mixing density as a skewed stable random variable. Thus, only one auxiliary variable  $\lambda$  needed to be sampled. In the current (asymmetric) setting, however, the problem is more challenging since we have to sample/characterise a bivariate, dependent, random variable  $\mu_X, \sigma_X^2$ , for which we do not have any convenient form except for the full infinite summations as given in (9).

#### IV. RESIDUAL APPROXIMATION

In order to apply the representation in practice, the infinite series in (9) must be truncated at some finite limit and the residuals need to be approximated. This is the only source of approximation in our models apart from the errors due to standard Monte Carlo inference procedures. We do however, substantially mitigate the approximation error by computing very accurate moment-based approximations to the residual error terms and including these in our inference procedures.

Our approach to approximating the residual series terms is designed to keep the structure of the PSR, where  $\mu_W$  appears as a factor in  $\mu_X$ , and  $\sigma_W^2$  appears as a factor in  $\sigma_X^2$ , see equation (9). We approach the approximation through computing the first two moments of the residual terms of the mean and variance of the conditionally Gaussian framework directly, and then approximating those residuals using a bivariate Gaussian distribution. This residual approximation approach is referred to as the Gaussian approximation of moments approach (GAMA). A bivariate Gaussian distribution for the residuals of the moments of the PSR is an approximation, but preliminary Central Limit Theorem results indicate that the Gaussian approximation is valid as  $M \rightarrow \infty$ , see [24]. We do expect the residuals to be light tailed since the heavy-tailed behaviour is induced by the potentially large first terms of the series appearing in the moments of (9), as the terms  $\Gamma_i^{-1/\alpha}$  are decreasing. Assuming that

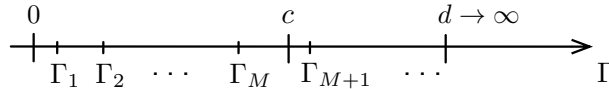


Fig. 1. Setup of the residual approximation approach, showing the increasing sequence of Poisson arrival times  $\{\Gamma_i\}$ , the truncation limit  $c$  and the upper limit  $d \rightarrow \infty$ .

the summation terminates once  $\Gamma_{M+1}$  exceeds some fixed value  $c$ , (see Fig. 1), we consider a fixed interval  $[c, d]$  first. Then, for any interval  $[c, d]$ ,  $\{\Gamma_i\}_{i \geq 1}$  is defined as a unit rate Poisson process on this interval, satisfying the properties

$$|\{\Gamma_i : \Gamma_i \in [c, d]\}| \sim \text{Poisson}(d - c) \text{ for } d > c \quad (12)$$

and given the number of  $\Gamma_i$  in  $[c, d]$ , each  $\Gamma_i$  is uniformly and independently distributed on  $[c, d]$ ,

$$\Gamma_i | |\{\Gamma_i : \Gamma_i \in [c, d]\}| \stackrel{i.i.d.}{\sim} U([c, d]). \quad (13)$$

Note that  $\Gamma_i$  in  $[c, d]$  are now an unordered set of random cardinality. With  $d$  going to infinity we will account for all residual terms in the PSR from  $c$  to  $\infty$ . The consideration of the  $\Gamma$ s in a fixed interval facilitates the computation of moments since the  $\Gamma_i$ s are independent.

We approximate the residual terms  $(R_1, R_2)$  in the summations of the mean and variance of the conditional framework,

$$X | \{\Gamma_i\}_{i=1}^M \stackrel{\text{approx.}}{\sim} \mathcal{N}(\mu_W m + \mu, \sigma_W^2 s), \quad (14)$$

where

$$m := \sum_{i=1}^M \Gamma_i^{-1/\alpha} + R_1, \quad s := \sum_{i=1}^M \Gamma_i^{-2/\alpha} + R_2, \quad (15)$$

by a bivariate Gaussian distribution,  $\mathcal{N}(\mu_R, \Sigma_R)$ , which takes account of the correlation between  $R_1$  and  $R_2$ . Note that the number of summation terms  $M$  is a random variable itself, defined as  $M = |\{i : \Gamma_i < c\}|$ . The residuals  $R_1$  and  $R_2$  are expressed as the limits of

$$R_1^{(d)} := \sum_{i: \Gamma_i \in [c, d]} \Gamma_i^{-1/\alpha} - \sum_{n: \Gamma_n \in [0, d]} \tilde{k}_n^{(\alpha)}, \quad (16)$$

$$R_2^{(d)} := \sum_{i: \Gamma_i \in [c, d]} \Gamma_i^{-2/\alpha} \quad (17)$$

as  $d \rightarrow \infty$ . In a next step the number of terms in the sums is approximated by the expectations  $\mathbb{E}[|\{\Gamma_i : \Gamma_i \in [c, d]\}|] = d - c$  and  $\mathbb{E}[|\{\Gamma_i : \Gamma_i \in [0, d]\}|] = d$ . Now, we can easily compute the mean and variance-covariance matrix of  $(R_1^{(d)}, R_2^{(d)})$ , and finally obtain  $\mu_R^{(\alpha)}$  and  $\Sigma_R^{(\alpha)}$  by letting  $d$  go infinity. The following lemma states the resulting GAMA.

*Lemma 4.1 (Gaussian approximation of moments approach):* The moments for a bivariate Gaussian approximation for the residuals  $(R_1, R_2)$  of a Poisson series representation as given in (9) can be obtained exactly as (see Appendix B):

$$\mu_R^{(\alpha)} = \begin{bmatrix} \frac{\alpha}{1-\alpha} c^{\frac{\alpha-1}{\alpha}} \\ \frac{\alpha}{2-\alpha} c^{\frac{\alpha-2}{\alpha}} \end{bmatrix} \quad (18)$$

and

$$\Sigma_R^{(\alpha)} = \begin{bmatrix} \frac{\alpha}{2-\alpha} c^{\frac{\alpha-2}{\alpha}} & \frac{\alpha}{3-\alpha} c^{\frac{\alpha-3}{\alpha}} \\ \frac{\alpha}{3-\alpha} c^{\frac{\alpha-3}{\alpha}} & \frac{\alpha}{4-\alpha} c^{\frac{\alpha-4}{\alpha}} \end{bmatrix}. \quad (19)$$

Hence, the approximated conditionally Gaussian framework for  $X \sim S_\alpha(\sigma, \beta, \mu)$  can be written as

$$X | \{\Gamma_i\}_{i=1}^M \stackrel{\text{approx.}}{\sim} \mathcal{N}(\mu_X, \sigma_X^2), \quad (20)$$

where

$$\mu_X = \mu_W \left( \sum_{i=1}^M \Gamma_i^{-1/\alpha} + R_1 \right) + \mu \quad (21)$$

$$\sigma_X^2 = \sigma_W^2 \left( \sum_{i=1}^M \Gamma_i^{-2/\alpha} + R_2 \right), \quad (22)$$

with

$$\begin{pmatrix} R_1 \\ R_2 \end{pmatrix} \sim \mathcal{N}(\mu_R^{(\alpha)}, \Sigma_R^{(\alpha)}) \quad (23)$$

and  $\mu_R^{(\alpha)}$  and  $\Sigma_R^{(\alpha)}$  as in (18) and (19).

We remark that we would ideally like to sample the infinite sequence of  $\Gamma$ 's appearing in (9), but this is not possible to achieve without approximation or truncation. The role of  $(R_1, R_2)$  then is to provide a summary approximation of the terms  $(\sum_{i=M+1}^{\infty} \Gamma_i^{-1/\alpha}, \sum_{i=M+1}^{\infty} \Gamma_i^{-2/\alpha})$ , containing the  $\Gamma$ 's missed due to the truncation to  $M$ . This justifies why the Gaussian distribution for the residuals depends on the fixed limit  $c$  and the parameter  $\alpha$ . The implication in the MCMC based parameter estimation is that two approaches are possible when a new  $\alpha$  value is proposed for a move from  $\alpha$  to  $\alpha'$ . On one hand we could consider as target distribution the *ideal* full conditional  $p(\alpha | \mathbf{X}, \{\Gamma_i\}_{i=1}^M, \{\Gamma_i\}_{i=M+1}^{\infty})$ . In terms of the full conditional  $p(\alpha | \mathbf{X}, \{\Gamma_i\}_{i=1}^M, \mathbf{R})$ , which we prove to be available *in practice*, this means that new residuals need to be sampled to evaluate  $p(\alpha | \mathbf{X}, \{\Gamma_i\}_{i=1}^M, \mathbf{R}')$ , the full conditional in the new value  $\alpha'$ . This is because  $(R'_1, R'_2)$  are the residuals corresponding to the unseen terms  $\{\Gamma_i\}_{i=M+1}^{\infty}$ , transformed through the newly drawn  $\alpha'$ . On the other hand, it would be possible to directly consider  $\mathbf{R}$  as the variables entering the sampling scheme. A standard Gibbs sampler could then be adopted, fixing  $\mathbf{R}$  at its current value. However we found in simulations not reported here that the acceptance rate for such a scheme was poor. We thus choose to follow the first alternative. In this, we consider  $\{\Gamma_i\}_{i=M+1}^{\infty}$  to be fixed in the Gibbs sampling step for  $\alpha$ . There remains however statistical uncertainty about the value of  $(R'_1, R'_2) = (\sum_{i=M+1}^{\infty} \Gamma_i^{-1/\alpha'}, \sum_{i=M+1}^{\infty} \Gamma_i^{-2/\alpha'})$  conditioned on the current residuals  $(R_1, R_2)$ , since we do not have access to the individual  $\Gamma_i$  terms from  $(R_1, R_2)$ . Thus, when proposing the new  $\alpha'$ ,  $(R'_1, R'_2)$  must also be drawn from their conditional distribution  $p((R'_1, R'_2) | (R_1, R_2), \alpha, \alpha')$ , which we again characterise exactly through its first two moments, and then approximate the conditional distribution as a bivariate Gaussian with the same moments. The general scheme can be seen as an example of *retrospective sampling* whereby we only simulate the summary stastics of  $\{\Gamma_i\}_{i=M+1}^{\infty}$  as they are required by the sampler. To achieve this conditionally Gaussian framework, we require the Gaussian distribution conditioned on  $\mathbf{R}$ :

$$\begin{pmatrix} R'_1 \\ R'_2 \end{pmatrix} \Big| \begin{pmatrix} R_1 \\ R_2 \end{pmatrix} \sim \mathcal{N}(\mu_R^{\text{cond}}, \Sigma_R^{\text{cond}}).$$

Making the final assumption of joint Gaussianity of  $(\mathbf{R}, \mathbf{R}')$ , the conditional mean and variance can be computed as

$$\mu_R^{\text{cond}} = \mu'_R + \Sigma_{12} \Sigma_{22}^{-1} ([R_1, R_2]^T - \mu_R)$$

and

$$\Sigma_R^{\text{cond}} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

with

$$\Sigma_{12} = \begin{bmatrix} \text{Cov}[R'_1, R_1] & \text{Cov}[R'_1, R_2] \\ \text{Cov}[R'_2, R_1] & \text{Cov}[R'_2, R_2] \end{bmatrix},$$

and  $\mu'_R = \mu_R^{(\alpha')}$  and  $\mu_R = \mu_R^{(\alpha)}$  as in (18),  $\Sigma_{11} = \Sigma_R^{(\alpha')}$ ,  $\Sigma_{22} = \Sigma_R^{(\alpha)}$ , and  $\Sigma_{12} = \Sigma_{21}^T$ . According to (49) in Appendix B, we compute the entries of  $\Sigma_{12}$  with  $d \rightarrow \infty$  as

$$\begin{aligned} \text{Cov}[R'_1, R_1] &\xrightarrow{d \rightarrow \infty} \frac{1}{(1/\alpha' + 1/\alpha) - 1} c^{-(1/\alpha' + 1/\alpha) + 1}, \\ \text{Cov}[R'_1, R_2] &\xrightarrow{d \rightarrow \infty} \frac{1}{(1/\alpha' + 2/\alpha) - 1} c^{-(1/\alpha' + 2/\alpha) + 1}, \\ \text{Cov}[R'_2, R_1] &\xrightarrow{d \rightarrow \infty} \frac{1}{(2/\alpha' + 1/\alpha) - 1} c^{-(2/\alpha' + 1/\alpha) + 1}, \\ \text{Cov}[R'_2, R_2] &\xrightarrow{d \rightarrow \infty} \frac{1}{(2/\alpha' + 2/\alpha) - 1} c^{-(2/\alpha' + 2/\alpha) + 1}. \end{aligned}$$

The GAMA form of approximation is very convenient in that the mean and variance expressions scale directly in terms of the parameters  $\mu_W$  and  $\sigma_W^2$ . As a consequence these two parameters can be marginalised by direct integration in Bayesian inference frameworks. The down-side is that we will have to simulate the residual terms  $R_1$  and  $R_2$  as additional random variables in order to perform inference.

### A. Evaluation of the Residual Approximation

To demonstrate the accuracy of the approximated PSR using our residual approximation GAMA we provide three exemplary distributions with values 0.9, 1.3 and 1.7 for  $\alpha$ . The shift parameter is set to zero and  $(\mu_W, \sigma_W) = (-1, 1)$ , which corresponds to  $(1.7694, -0.8466)$ ,  $(2.0984, -0.8836)$  and  $(2.9570, -0.9098)$  for the distribution parameters  $(\sigma, \beta)$ , respectively. Random variables obtained from the asymmetric stable law, applying the Chambers-Mallows-Stuck (CMS) method [18], serve as a benchmark for our comparison of the representations shown in Fig. 2. The presented residual approximation ('PSR + GAMA') with an average number of summation terms of  $c = 100$  shows a clear improvement to a simple truncation ('truncated PSR,  $c=100$ '), even for a distinctly higher  $c = 500$  ('truncated PSR,  $c=500$ ') and achieves results almost indistinguishable from the benchmark ('CMS benchmark') as can be seen in Fig. 2. Similar improvements were obtained for a wide range of different  $\alpha$ -stable parameter settings. Note the left hand panel of the Figure shows a location shift when simply truncating the PSR, with  $\alpha$  close to, but smaller than 1. This problem increases as  $\alpha \rightarrow 1$  from the left, but is completely eliminated when the moments of the residual are correctly incorporated.

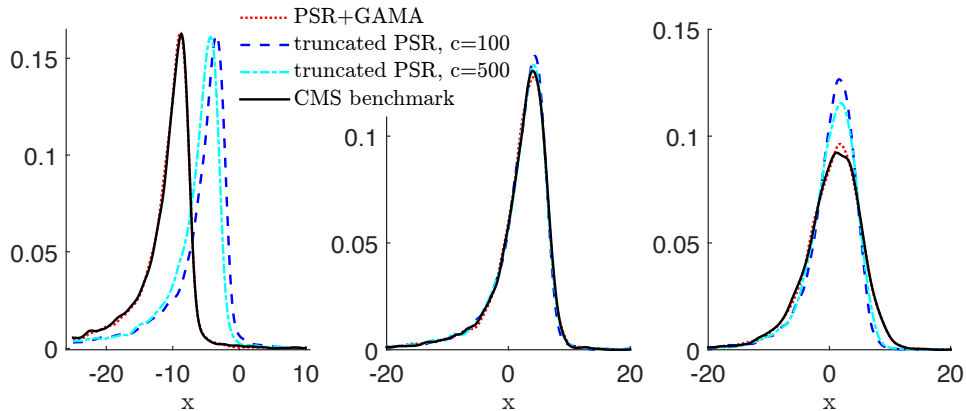


Fig. 2. Comparing the approximated PSR to the real distribution with  $\alpha = 0.9, 1.3$  and  $1.7$  (from left to right), see text for other parameters.

## V. INFERENCE FOR $\alpha$ -STABLE DISTRIBUTIONS VIA MCMC

The parameters of interest are  $\sigma$ , the scale parameter,  $\beta$ , the skewness parameter,  $\mu$  the location parameter, and  $\alpha$ , the index of stability. Given the samples  $\mathbf{X}$ , we aim for the posterior of the latent variable set  $\mathbf{\Gamma} := \{\{\Gamma_{i,n}\}_{i=1}^M\}_{n=1}^N$ , the variables  $\mu_W$  and  $\sigma_W$ , as well as the residual approximations  $\mathbf{R} := \{\mathbf{R}_n = (R_{1,n}, R_{2,n})\}_{n=1}^N$ , which all arise from the PSR, and the distribution parameters  $\alpha$  and  $\mu$ .

### A. Marginal and conditional distributions

First, we shall explore the structure of the conditionally Gaussian framework with a view to simplifying inference by deriving closed form conditionally distributions for the parameters  $\sigma_W$ ,  $\mu_W$  and  $\mu$ . In a second step these parameters will be marginalised out to obtain an independent expression for  $\alpha$  and the latent variables  $\mathbf{\Gamma}$  and  $\mathbf{R}$ .

Suppose there are  $N$  i.i.d. samples

$$X_n \sim S_\alpha(\sigma, \beta, \mu) \quad \text{for } n = 1, \dots, N$$

from which we aim to draw conclusions about the distribution parameters, or rather about the parameters  $\sigma_W$ ,  $\mu_W$ ,  $\alpha$  and  $\mu$ , since we make use of the conditionally Gaussian framework for the  $\alpha$ -stable vector  $\mathbf{X} = (X_1, \dots, X_N)^T$ . With standard calculations (see Appendix C) it is possible to show that

$$p(\mathbf{X}|\mathbf{\Gamma}, \sigma_W, \mu_W, \mu, \alpha) = \frac{(2\pi\sigma_W^2)^{-N/2}}{\prod_{n=1}^N \sqrt{s_n}} \exp \left\{ -\frac{1}{2\sigma_W^2} \left( \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} - \tilde{\mathbf{X}}^T \tilde{\mathbf{M}} \boldsymbol{\mu} - \boldsymbol{\mu}^T \tilde{\mathbf{M}}^T \tilde{\mathbf{X}} + \boldsymbol{\mu}^T \tilde{\mathbf{M}}^T \tilde{\mathbf{M}} \boldsymbol{\mu} \right) \right\} \quad (24)$$

with

$$m_n = \sum_{i=1}^{\infty} \left( \Gamma_{i,n}^{-1/\alpha} - \tilde{k}_i^{(\alpha)} \right), \quad s_n = \sum_{i=1}^{\infty} \Gamma_{i,n}^{-2/\alpha}, \quad (25)$$

$$\tilde{\mathbf{X}} = \left( \frac{X_1}{\sqrt{s_1}}, \dots, \frac{X_N}{\sqrt{s_N}} \right), \quad \tilde{\mathbf{M}} = \begin{bmatrix} \frac{m_1}{\sqrt{s_1}} & \frac{1}{\sqrt{s_1}} \\ \vdots & \vdots \\ \frac{m_N}{\sqrt{s_N}} & \frac{1}{\sqrt{s_N}} \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_W \\ \mu \end{bmatrix}. \quad (26)$$

Applying the residual approximation method GAMA, we replace  $m_n, s_n$  in (25) and (26) by their approximated expressions

$$m_n \approx \sum_{i=1}^{M_n} \Gamma_{i,n}^{-1/\alpha} + R_{1,n}, \quad s_n \approx \sum_{i=1}^{M_n} \Gamma_{i,n}^{-2/\alpha} + R_{2,n}.$$

In the following, we aim for a straightforward Gibbs sampler for  $\boldsymbol{\mu}$  and  $\sigma_W^2$ . Their joint conditional distribution can be rewritten as a composition given by

$$p(\boldsymbol{\mu}, \sigma_W^2 | \mathbf{X}, \mathbf{\Gamma}, \mathbf{R}, \alpha) = p(\boldsymbol{\mu} | \mathbf{X}, \mathbf{\Gamma}, \mathbf{R}, \sigma_W, \alpha) p(\sigma_W^2 | \mathbf{X}, \mathbf{\Gamma}, \mathbf{R}, \alpha).$$

Taking uniform priors on  $\mu_W$  and  $\mu$ , we obtain the proportionality

$$p(\boldsymbol{\mu} | \mathbf{X}, \sigma_W, \mathbf{\Gamma}, \mathbf{R}, \alpha) \propto p(\mathbf{X} | \boldsymbol{\mu}, \sigma_W, \mathbf{\Gamma}, \mathbf{R}, \alpha)$$

Now, rearranging (24), it is possible to obtain a bivariate Gaussian with respect to  $\boldsymbol{\mu}$ , as shown in Appendix C

$$p(\boldsymbol{\mu} | \mathbf{X}, \sigma_W, \mathbf{\Gamma}, \mathbf{R}, \alpha) \propto \mathcal{N}(\boldsymbol{\mu} | \mathbf{a}_\mu, \Sigma_\mu), \quad (27)$$

with

$$\mathbf{a}_\mu = (\tilde{\mathbf{M}}^T \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}}^T \tilde{\mathbf{X}}, \quad (28)$$

$$\Sigma_\mu = (\tilde{\mathbf{M}}^T \tilde{\mathbf{M}})^{-1} \quad (29)$$

Next, we derive the conditional distribution for  $\sigma_W^2$ ,  $p(\sigma_W^2 | \mathbf{X}, \mathbf{\Gamma}, \mathbf{R}, \alpha)$ . Through standard calculations (see Appendix C) and taking a uniform prior on  $\sigma_W$  we obtain

$$\begin{aligned} p(\sigma_W^2 | \mathbf{X}, \mathbf{\Gamma}, \mathbf{R}, \alpha) &= p(\boldsymbol{\mu}, \sigma_W^2 | \mathbf{X}, \mathbf{\Gamma}, \mathbf{R}, \alpha) p(\boldsymbol{\mu} | \mathbf{X}, \mathbf{\Gamma}, \mathbf{R}, \sigma_W, \alpha)^{-1} \\ &\propto \mathcal{IG} \left( \sigma_W^2 \left| \frac{N-4}{2}, \frac{B}{2} \right. \right), \end{aligned} \quad (30)$$

with

$$B = \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} - \tilde{\mathbf{X}}^T \tilde{\mathbf{M}} (\tilde{\mathbf{M}}^T \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}}^T \tilde{\mathbf{X}}, \quad (31)$$

and where  $\mathcal{IG}$  denotes an inverse gamma distribution, obtained by identifying with the parameters of the inverted gamma distribution<sup>1</sup>. Explicit conditional distributions are advantageous, since they allow Gibbs sampling, avoiding the risk of low acceptance rates and therefore bad mixing of the MCMC sampler.

<sup>1</sup>**Inverse gamma distribution:** The inverse gamma probability density function is defined over the support  $x > 0$  as

$$f(x) = \frac{t^s}{\Gamma(s)} X^{-(s+1)} \exp \left( -\frac{t}{x} \right)$$

with shape parameter  $s$  and scale parameter  $t$ .



Finally, marginalising  $\boldsymbol{\mu}$  and  $\sigma_W^2$  gives

$$\begin{aligned} p(\mathbf{X}|\boldsymbol{\Gamma}, \mathbf{R}, \alpha) &\propto p(\mathbf{X}|\boldsymbol{\Gamma}, \mathbf{R}, \sigma_W, \boldsymbol{\mu}, \alpha) p^{-1}(\boldsymbol{\mu}|\mathbf{X}, \boldsymbol{\Gamma}, \mathbf{R}, \sigma_W, \alpha) p^{-1}(\sigma_W^2|\mathbf{X}, \boldsymbol{\Gamma}, \mathbf{R}, \alpha) \\ &\propto \frac{\Gamma((N-4)/2) \left| (\tilde{\mathbf{M}}^T \tilde{\mathbf{M}})^{-1} \right|}{(B/2)^{(N-4)/2} (2\pi)^{(N-2)/2} \prod_{n=1}^N \sqrt{s_n}}, \end{aligned} \quad (32)$$

where  $\Gamma(\cdot)$  denotes the gamma function<sup>2</sup>.

Note that the above marginal distribution provides a relation between  $\boldsymbol{\Gamma}, \mathbf{R}, \alpha$  and the given set of data  $\mathbf{X}$ , which can be used for independent samplers from  $\mu_W$  and  $\sigma_W^2$ .

### B. Parameters $\mu_W, \sigma_W, \alpha$ and $\boldsymbol{\mu}$

The  $\alpha$ -stable distribution parameters  $\alpha, \beta$  and  $\sigma$  are reparametrised through the PSR in terms of  $\alpha, \mu_W$  and  $\sigma_W$ , for which we consider possible MCMC sampling methods below.

1) *Gibbs sampler for  $(\mu_W, \mu)$  and  $\sigma_W$* : The parameters  $\boldsymbol{\mu}$  and  $\sigma_W$  can be sampled straight away according to the available joint conditional distribution, which can be written as the product of a bivariate Gaussian and an inverse gamma distribution as derived in (27) and (30):

$$\begin{aligned} p(\boldsymbol{\mu}, \sigma_W^2|\mathbf{X}, \boldsymbol{\Gamma}, \mathbf{R}, \alpha) &= p(\boldsymbol{\mu}|\mathbf{X}, \boldsymbol{\Gamma}, \mathbf{R}, \sigma_W, \alpha) p(\sigma_W^2|\mathbf{X}, \boldsymbol{\Gamma}, \mathbf{R}, \alpha) \\ &= \mathcal{N}(\boldsymbol{\mu}|\mathbf{a}_\mu, \Sigma_\mu) \mathcal{IG}\left(\sigma_W^2 \left| \frac{N-4}{2}, \frac{B}{2} \right.\right), \end{aligned}$$

where  $A, B$  and  $C$  are as in (26) using the approximated  $m_n, s_n$ .

2) *Metropolis-Hastings ('MH') sampler for  $\alpha$* : To sample the  $\alpha$  parameter, we choose the marginalised conditional distribution with a uniform prior on  $\alpha$  to obtain the proportionality

$$p(\alpha|\mathbf{X}, \boldsymbol{\Gamma}, \mathbf{R}) \propto p(\mathbf{X}|\boldsymbol{\Gamma}, \mathbf{R}, \alpha),$$

with  $p(\mathbf{X}|\boldsymbol{\Gamma}, \mathbf{R}, \alpha)$  as derived in (32). Then, the acceptance probability for the 'MH' sampler is computed as

$$\rho(\alpha, \alpha') = \min\left(1, \frac{p(\mathbf{X}|\boldsymbol{\Gamma}, \mathbf{R}', \alpha') q(\alpha|\alpha')}{p(\mathbf{X}|\boldsymbol{\Gamma}, \mathbf{R}, \alpha) q(\alpha'/\alpha)}\right), \quad (33)$$

where  $\alpha$  is proposed from  $q(\alpha'|\alpha) = \mathcal{N}(\alpha, \sigma_\alpha^2)$  with some variance  $\sigma_\alpha^2$ . Observe that, in order to evaluate the full conditional in the new value of the parameter, we sample also new values of the residuals, as explained in Section IV.

### C. Auxiliary variables $\boldsymbol{\Gamma}$ and $\mathbf{R}$

Apart from the parameters of interest, we also need to sample the auxiliary variables  $\boldsymbol{\Gamma}$  and  $\mathbf{R}$  arising from the PSR and the residual approximation.

1) *Metropolis-Hastings sampler for  $\boldsymbol{\Gamma}$  and  $\mathbf{R}$* : One possibility to update the latent variables  $\boldsymbol{\Gamma}$  and  $\mathbf{R}$  is the use of a 'MH' sampling step. Setting the proposals to be the priors  $q(\boldsymbol{\Gamma}'_n|\boldsymbol{\Gamma}_n) = p(\boldsymbol{\Gamma}'_n)$  and  $q(\mathbf{R}'_n|\mathbf{R}_n) = p(\mathbf{R}'_n)$ , the corresponding acceptance probabilities result in

$$\rho(\boldsymbol{\Gamma}_n, \boldsymbol{\Gamma}'_n) = \min\left(1, \frac{p(X_n|\boldsymbol{\Gamma}'_n, \mathbf{R}_n, \boldsymbol{\mu}, \sigma_W^2, \alpha)}{p(X_n|\boldsymbol{\Gamma}_n, \mathbf{R}_n, \boldsymbol{\mu}, \sigma_W^2, \alpha)}\right) \quad (34)$$

$$= \min\left(1, \frac{\mathcal{N}(X_n|\mu'_{X_n}, \sigma'_{X_n})}{\mathcal{N}(X_n|\mu_{X_n}, \sigma_{X_n})}\right) \quad (35)$$

<sup>2</sup>**Gamma function**: The gamma function is defined for complex numbers  $i$  with a positive real part ( $\Re\{i\} > 0$ ) as the integral

$$\Gamma(i) = \int_0^\infty x^{i-1} \exp(-x) dx.$$

combined with the subsequent ‘MH’ step for the residual terms, which are accepted with probability

$$\rho(\mathbf{R}_n, \mathbf{R}'_n) = \min \left( 1, \frac{p(X_n | \Gamma_n, \mathbf{R}'_n, \boldsymbol{\mu}, \sigma_W^2, \alpha)}{p(X_n | \Gamma_n, \mathbf{R}_n, \boldsymbol{\mu}, \sigma_W^2, \alpha)} \right) \quad (36)$$

$$= \min \left( 1, \frac{\mathcal{N}(X_n | \mu'_{X_n}, \sigma'_{X_n})}{\mathcal{N}(X_n | \mu_{X_n}, \sigma_{X_n})} \right). \quad (37)$$

The set of  $\Gamma$ s may be updated jointly with the residuals  $\mathbf{R}$ . Otherwise, the update takes place in two sequential steps.

2) *Rejection sampler for  $\Gamma_n$* : Alternatively to the above presented ‘MH’ samplers we can use rejection sampling to obtain samples for  $\Gamma$  and  $\mathbf{R}$ . Note that  $\Gamma_n$  and  $\mathbf{R}_n$  need to be updated sequentially now to ensure that their full conditionals are bounded. The rejection sampler is expected to be slower than the ‘MH’ sampler, since it proposes samples until one is accepted in each iteration. On the other hand, it provides samples from the exact full conditional while the ‘MH’ sampler might need some period to converge.

As usual, we write the conditionally Gaussian distribution as

$$X_n | \{\Gamma_{i,n}\}_{i=1}^{M_n} \stackrel{\text{approx.}}{\sim} \mathcal{N}(\mu_W m_n + \mu, \sigma_W^2 s_n) = \mathcal{N}(\mu_{X_n}, \sigma_{X_n}^2),$$

where  $m_n$  and  $s_n$  are as in (25). For the set  $\Gamma$  we sample  $\Gamma_n = \{\Gamma_{i,n}\}_{i=1}^{M_n}$  for the  $n$ -th observation from the full conditional distribution using rejection sampling with the envelope function,

$$\begin{aligned} p(\Gamma_n | X_n, \mathbf{R}_n, \boldsymbol{\mu}, \sigma_W, \alpha) &\propto \mathcal{N}(X_n | \mu_{X_n}, \sigma_{X_n}^2) p(\Gamma_n) \\ &< (2\pi\sigma_W^2 R_{2,n})^{-1/2} p(\Gamma_n). \end{aligned} \quad (38)$$

3) *Rejection sampler for  $R_n$* : Residuals are updated using the same scheme as with the set  $\Gamma_n$  with the bounding envelope function

$$p(\mathbf{R}_n | X_n, \{\Gamma_{i,n}\}_{i=1}^{M_n}, \boldsymbol{\mu}, \sigma_W, \alpha) < \left( 2\pi\sigma_W^2 \sum_{i=1}^{M_n} \Gamma_{i,n}^{-2/\alpha} \right)^{-1/2} p(\mathbf{R}_n | \alpha). \quad (39)$$

#### D. Improvements for $\Gamma$ and $\mathbf{R}$

On closer inspection, some of the above introduced samplers might raise an issue concerning their convergence. While the explicit full conditionals for  $\mu_W$  and  $\sigma_W^2$  as well as the marginalised likelihood for  $\alpha$  seem to be rather less problematic, the samplers for the multivariate  $\Gamma_n$  and  $\mathbf{R}_n$  might suffer from bad mixing or high rejection rates. For instance, it is likely that a very small  $\Gamma_{n,1}$  is required to generate a large value for  $m_n$  for an adequate representation of  $X_n$ . Improvements of the acceptance rate, in particular for large observation values, can be achieved through various modifications and/or additional adjustment sampling steps.

A fundamental task for Monte Carlo inference in our framework is the sampling of the latent random variables  $m_n$  and  $s_n$  from their full conditional distribution

$$p(m_n, s_n | X_n, \mu_W, \sigma_W, \mu) \propto \mathcal{N}(X_n | \mu_W m_n + \mu, \sigma_W^2 s_n) p(m_n, s_n).$$

This can be done either indirectly in terms of the latent process  $\Gamma_n$  in combination with the residual approximations  $\mathbf{R}_n$  as presented above, or directly in terms of  $m_n$  and  $s_n$ . Considering  $m_n$  and  $s_n$ , their prior marginals are known, but their joint distribution has not yet been characterised. Thus, one difficulty here is that we need to account for the mutual dependence of  $m_n$  and  $s_n$ . Another issue in case of rejection sampling arises from the unbounded envelope function  $\mathcal{N}(X_n | \mu_W m_n + \mu, \sigma_W^2 s_n)$  as  $s_n \rightarrow 0$ .

Studying the likelihood function, we have for some fixed  $m_n$ ,

$$\mathcal{N}(X_n | \mu_W m_n + \mu, \sigma_W^2 s_n) = \frac{1}{\sqrt{2\pi\sigma_W^2 s_n}} \exp \left( -\frac{1}{2\sigma_W^2 s_n} (X_n - \mu_W m_n - \mu)^2 \right).$$

The supremum with respect to  $s_n$  is achieved when

$$s_n = (X_n - \mu_W m_n - \mu)^2 / (2\sigma_W^2),$$

and the likelihood is finite at this supremum except for  $s_n \rightarrow 0$ , i.e. for  $X_n = \mu_W m_n - \mu$ . Note, however, that  $s_n$  does not hit zero or even come close to it for large observations in practice. In the following, we present an approximation which addresses

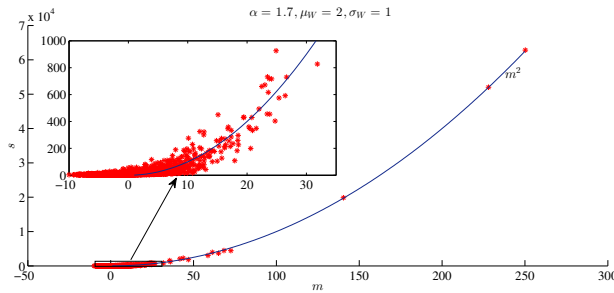


Fig. 3. Scatter plot of  $(m, s)$  samples with  $\alpha = 1.7$ , overlaid on the curve  $s = m^2$ . Series truncated to 1,000 terms with Gaussian approximation of the residual.

the mutual dependence of  $m_n$  and  $s_n$ .

Simulations of  $(m_n, s_n)$  pairs indicate that  $s_n$  is quite strongly dependent upon  $m_n$ , and only occupies a fairly small band of values conditional upon  $m_n$ , especially for large values of  $m_n$ , as illustrated in Figure 3. Evidently, to get a very large value of  $m_n$ , the first dominating term in the series  $\Gamma_1^{-1/\alpha}$  needs to be very large as well, hence,

$$\left( \sum_{i=1}^{\infty} \Gamma_{i,n}^{-1/\alpha} \right)^2 \approx \sum_{i=1}^{\infty} \Gamma_{i,n}^{-2/\alpha}.$$

Thus, we have a near-deterministic relationship between  $m_n$  and  $s_n$ . The near-deterministic behaviour is even more pronounced for lower values of  $\alpha$  (more heavy-tailed distributions). Hence, considering the large  $X_n$  case, we expect the conditional posterior to have support only for large values of  $m_n$ . Thus, taking  $s_n \approx m_n^2$ , the likelihood becomes

$$p(X_n | m_n, s_n = m_n^2, \mu_W, \sigma_W) = \frac{1}{\sqrt{2\pi\sigma_W^2 m_n^2}} \exp\left(-\frac{1}{2\sigma_W^2 m_n^2} (X - \mu_W m_n - \mu)^2\right). \quad (40)$$

Now, this can be used to achieve an accurate  $m$  in a first step via a rejection step after maximising over  $m$  and bounding the likelihood (40) in this way. Then, given  $m$  the set of  $\Gamma$ s,  $\Gamma_{i=2:M}$ , and residuals  $(R_1, R_2)$  can be regenerated running a short MCMC chain (for details see Appendix D).

Alternatively, the above described rejection sampling for  $m$  for observations in the tails can be used to propose  $m$ , while  $\Gamma_{i=2:M}$  and  $(R_1, R_2)$  are proposed from their priors. An MCMC chain is then run to target their joint posterior (for details see Appendix E).

While the preceding improvements relate to large observations, we can also improve the acceptance rate of the exact rejection sampler for all  $X_n$  by more tightly bounding the likelihood values based on the idea of the approximation  $s \approx m^2$  for samples in the tails of the  $\alpha$ -stable distribution. This can be done by approximating  $p(m, s)$  to have much more limited support. A simple and effective scheme bounds the support of  $p(m, s)$  between two squares  $(m + k_1)^2$  and  $(m - k_2)^2$ , and a lower limit  $s_{\min}$  such that

$$\begin{aligned} s &< (m + k_1)^2 & \text{for } m > -k_1, \\ s &> (m - k_2)^2 & \text{for } m > k_2, \\ s &> s_{\min}, \end{aligned}$$

as illustrated in Figure 4. Having that particularly convenient constraint region over which to maximise the likelihood at hand, we can maximise the likelihood on each quadratic curve using a slight modification of (51) and maximise on  $s = s_{\min}$  in between the quadratic functions. Then, we choose the maximum of the three as the global maximum of the likelihood over the whole constrained domain. The derivations can be found in Appendix F.

Algorithm 1 outlines the main steps of the MCMC sampler for inference for  $\alpha$ -stable distribution parameters.

## VI. NUMERICAL RESULTS

The Gibbs sampler was run for 10,000 iterations for each of the following examples on real and generated data of length  $N = 500$ . The first half of the MCMC outputs were considered as a burn-in period and neglected for the posterior densities as well as the estimated parameter values. Based on the traceplots, convergence seems to be assured within 5000 iterations for most of the parameter settings tested.

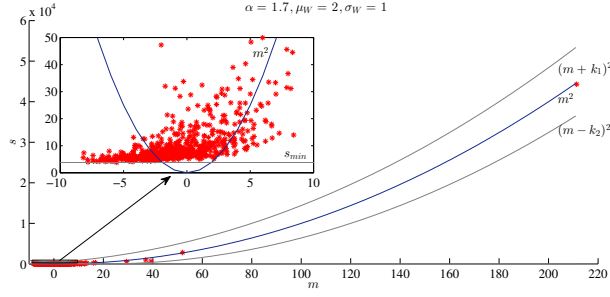


Fig. 4. Scatter plot of  $(m, s)$  samples with  $\alpha = 1.7$ , quadratic bounds and lower bound  $s_{\min}$ .

---

**Algorithm 1**  $\alpha$ -stable distribution parameter estimation - pseudo code -

---

```

1: procedure MCMC SAMPLER
2:   set  $j = 0$  and initialise  $\alpha, \mu, \sigma_W, \Gamma$  and  $\mathbf{R}$ :
3:   assign  $\alpha_0, \mu_0$  and  $\sigma_{W,0}$ 
4:   for  $n=1:N$  do
5:      $\Gamma_{n,0} \sim p(\Gamma)$  ( $\{\Gamma_{i,n} \sim \sum_{j=1}^i e_j, \text{ with } e_j \sim \text{Exp}(1)\}_{i=1}^{M_n}$ , where  $M_n : \Gamma_{M_n,n} < c$  and  $\Gamma_{M_n+1,n} > c$ )
6:      $(R_{1,n,0}, R_{2,n,0}) \sim \mathcal{N}(\mu_{R_0}, \Sigma_{R_0})$  ▷ Lemma 4.1
7:   end for
8:   for  $j=1:\text{MCMC iterations}$  do
9:     sample  $\alpha, \mu, \sigma_W, \Gamma$  and  $\mathbf{R}$  from their full conditional posteriors:
10:    for  $n=1:N$  do
11:       $\Gamma_{n,j} \sim p(\Gamma | X_n, \alpha_{j-1}, \mu_{W,j-1}, \sigma_{W,j-1}, \mathbf{R}_{n,j-1})$  ▷ ‘MH’ or rejection steps
12:       $\mathbf{R}_{n,j} \sim p(\mathbf{R} | X_n, \alpha_{j-1}, \mu_{W,j-1}, \sigma_{W,j-1}, \Gamma_{n,j})$  ▷ as in Section V-C
13:    end for
14:     $\mu_j \sim \mathcal{N}(\cdot, \cdot)$  ▷ Gibbs step (27)
15:     $\sigma_{W,j}^2 \sim \text{IG}(\cdot, \cdot)$  ▷ Gibbs step (30)
16:     $(\alpha_j, \mathbf{R}_j) \sim p(\alpha, \mathbf{R} | X, \Gamma_j, \mathbf{R}_j)$  ▷ ‘MH’ step (33)
17:  end for
18: end procedure

```

---

In particular we run the sampler on a number of datasets simulated using the method of Chambers, Mallows and Stuck [18] with  $(\alpha, \beta, \sigma, \mu) = (0.8, -0.84, 1.71, \mu)$ ,  $(1.6, -0.90, 2.62, \mu)$ ,  $(1.8, -0.92, 3.52, \mu)$ , and with either  $\mu = 0$ , assumed to be known, or  $\mu = 20$ , in which case it is also to be estimated. The values of  $\sigma$  and  $\beta$  are computed in terms of the approximated PSR representation with  $\mu_W = -1$  and  $\sigma_W = 1$ . In examples where  $\mu = 0$  is known, the conditional densities for sampling are simplified slightly from the general case given in expressions (27) and (30), see [10] for details.

We provide illustrative simulations of the basic Metropolis-Hastings schemes, with improvement steps on the  $\Gamma$ s and the residuals added in for large data points, in parameter regimes where the basic algorithm is slow to converge. Observe that, with regards to the sets of parameters tested, we consider large those observations with absolute value greater than 50, but, more generally, quantiles could be used to decide when to execute an enhanced version of the sampler. In the results presented we refer by the abbreviation ‘MH’ to the Metropolis-Hastings step without improvements, as shown in Section V-C, equations (35) and (37). In the case of large observations, the sampling step for the  $\Gamma$ s and  $(R_1, R_2)$  can be enhanced either by regenerating them through an additional short MCMC chain, after sampling a value for  $m$ , as shown in appendix D, termed ‘MH-regeneration’, or by a joint sampling of  $m$ , the  $\Gamma$ s and  $(R_1, R_2)$ , as explained in appendix E, and termed ‘MH-joint’. The alternative sampling scheme presented in Section V-C, equations (38) and (39), involves a rejection step on the  $\Gamma$ s and the residuals. In order to increase the computational speed, we consider directly two improved versions of it: we either re-sample  $m$ , the  $\Gamma$ s and the residuals jointly in the presence of large data points (Appendix E, ‘Rejection-joint’) or constrain the region over which to bound the likelihood for each data point (see Section V-D and Appendix F, ‘Rejection-bounded’).

Realizations of the stable density obtained with our posterior estimates of all four parameters are compared to the point estimates computed with the three frequentist techniques mentioned in Section II (maximum likelihood, quantiles based, empirical characteristic function based estimators). The implementation of these methods is provided in the STABLE Matlab Toolbox, available at [www.RobustAnalysis.com](http://www.RobustAnalysis.com).

Finally as application to a real dataset, we consider daily price returns of Dow Jones Industrial average shares in a 4 year period, and show how the estimated stable density fits the data, compared to a Gaussian model.

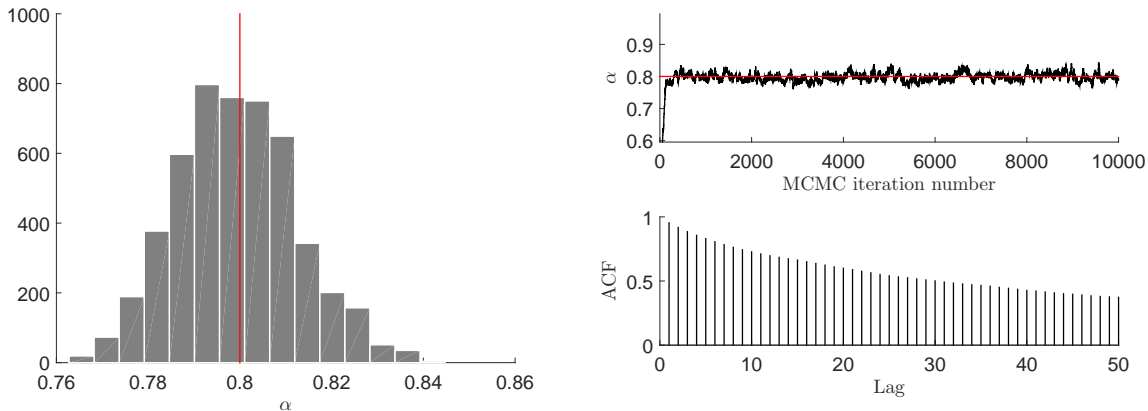


Fig. 5. Inference on the parameter  $\alpha$ , when the true value is  $\alpha = 0.8$ , using the ‘MH’ step for the  $\Gamma$ s and the residuals. Left: Histogram from the MCMC output. The true value is marked by the vertical line. Right, top: MCMC sampled parameter value  $\alpha$ . The true value is marked by the horizontal line. Right, bottom: ACF as a function of lag.

#### A. Example 1: Synthetic data, estimate of $(\alpha, \beta, \sigma)$

We simulate two data sets under the hypothesis that the location  $\mu = 0$  is known, initially with  $\alpha = 0.8$ , corresponding to a very heavy tailed model with occasional very large observations, and then with  $\alpha = 1.8$ , corresponding to less frequent and less extreme observations. We analyse the posterior distribution of the parameters through the traceplots of the chains, the histograms of the marginals, the autocorrelation functions (ACFs)  $\rho(t)$  and the integrated autocorrelation times (IACTs). The latter is defined as  $\tau = 1 + 2 \sum_{t=1}^{\infty} \rho(t)$  and appears as a multiplicative factor in the asymptotic variance of the posterior mean estimator, when this is not computed through i.i.d samples. It corresponds thus to the reduction in the effective number of independent samples due to autocorrelation, with lower IACTs indicating better mixing of the chains, see e.g. [35]. In particular we truncate the series when the sum of two adjacent sample ACFs values is larger than the sum of the previous pair, or when the sum of two adjacent values of the ACF becomes negative, according to the initial monotone sequence estimator (IMS) of the IACT, see [36].

In the case of  $\alpha$  smaller than one, we initialise the sampler to  $(\alpha, \mu_W, \sigma_W) = (0.4, -3, 3)$ , well away from the true values. The Gibbs sampler with the ‘MH’ step shows good performance (chain mixing and computational speed) on a variety of datasets corresponding to the fixed parameters and for several runs over a fixed dataset. We thus present the outputs coming from the simple ‘MH’, as in such parameter regime this basic sampler gives perfectly adequate performance. Figures 5 and 6 show an instance of the MCMC sampled parameter values for  $\alpha$ ,  $\mu_W$  and  $\sigma_W$  on the right-hand side with the ACFs below. On the left-hand side we see the unimodal histograms centred around the true values. With the MCMC samples for the parameters  $\alpha$ ,  $\mu_W$  and  $\sigma_W$  of the PSR we obtain the corresponding paths for the distribution parameters  $\beta$  and  $\sigma$  by reparametrising according to (4). The resulting traceplots and histograms are shown in Figure 7. As with the parameters  $\mu_W$  and  $\sigma_W$ , the samples for  $\beta$  and  $\sigma$  lead to unimodal histograms centred around the true values. For the dataset considered, the sample means averaged over four runs, corresponding each to a different random seed, yield the estimates  $\hat{\alpha} = 0.8$ ,  $\hat{\beta} = -0.84$  and  $\hat{\sigma} = 1.69$ , with average standard deviations 0.01, 0.04 and 0.11, respectively.

We now reduce the weight in the tails of the distribution, focusing on values of  $\alpha$  greater than one. When  $\alpha = 1.8$ , the chains obtained with the basic ‘MH’ scheme may lead to poor mixing for some of the generated datasets, particularly as regards  $\mu_W$ ,  $\sigma_W$  and  $\beta$ . Much longer MCMC runs of the basic sampler would thus be required for such parameter regimes, and hence instead we present the MCMC outputs for simulations based on the four improved samplers detailed above (‘MH-joint’, ‘MH-regeneration’, ‘Rejection-joint’, ‘Rejection-bounded’).

Table I reports the IACTs for the four improved samplers, averaged over four datasets and all significantly improved compared to the basic ‘MH’ approach. Although the IACT values are roughly similar for the four improved methods, we observed that ‘MH-regeneration’ and ‘Rejection-bounded’ have more reliable performance on difficult datasets.

Thus Figures 8, 9 and 10 display the MCMC outputs of an instance of the sampler with the ‘MH-regeneration’ and the ‘Rejection-bounded’ methods, on a dataset ( $D_1$ ). Figure 8 presents the unimodal posterior distribution for  $\alpha$ , centred on the true value. This is a fairly typical result, and indeed gets a closer value to the true  $\alpha$  than the ML method (see Table II). However, it should be observed that the chains are quite correlated over iteration number and certain parameters can get stuck in low probability regions for several iterations, see e.g. the trace plot for  $\beta$  in Figure 10. This means that longer MCMC chains may be needed still for this parameter regime, with  $\alpha$  close to 2.

With regard to the computational cost of the four enhanced versions of the sampler, we have that, denoting with  $T^{CPU}(\cdot)$  the

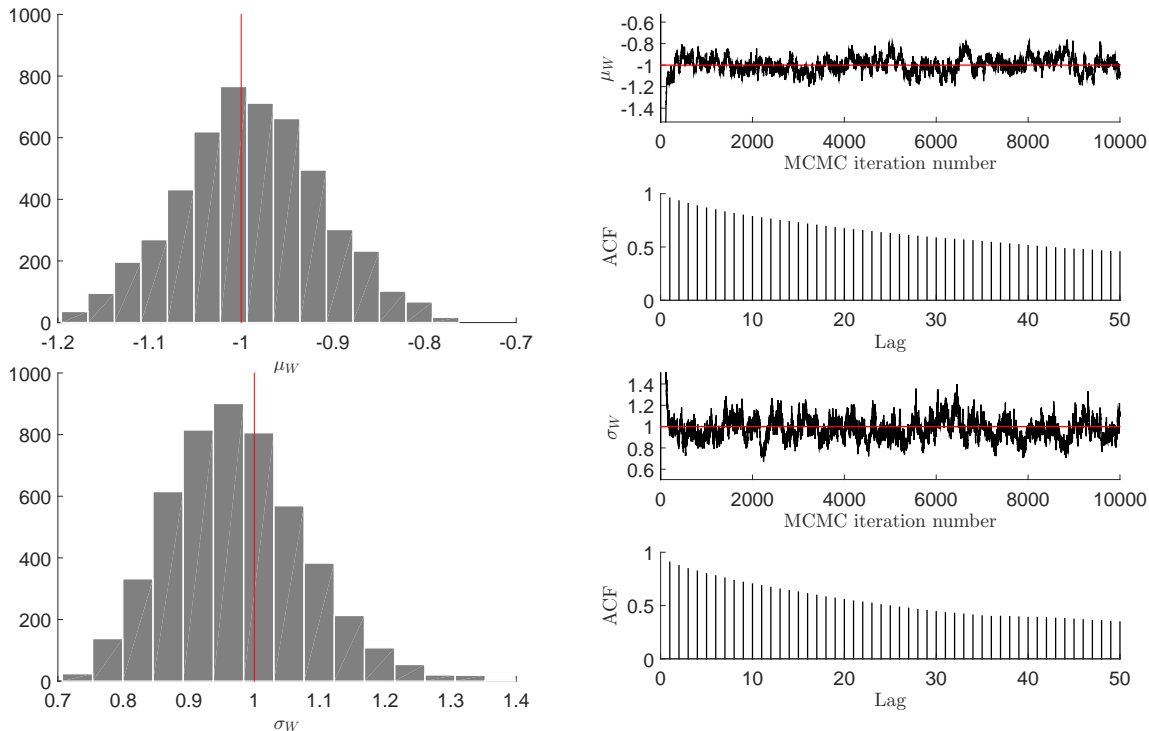


Fig. 6. Inference on the parameters  $\mu_W$  and  $\sigma_W$  when  $\alpha = 0.8$ , using the ‘MH’ step for the  $\Gamma$ s and the residuals. Left: Histograms from the MCMC output. The true values are marked by the vertical lines. Right: MCMC sampled parameter values  $\mu_W$  and  $\sigma_W$  with the ACFs as functions of lag below each. The true values are marked by the horizontal lines.

TABLE I  
IACTS OF THE PARAMETERS  $(\mu_W, \sigma_W, \alpha, \beta, \sigma)$  VARYING THE SAMPLING STEP ON THE  $\Gamma$ s AND THE RESIDUALS, AVERAGED OVER 4 DATASETS, WITH  $\alpha = 1.8$ .

Method	IACT( $\mu_W$ )	IACT( $\sigma_W$ )	IACT( $\alpha$ )	IACT( $\beta$ )	IACT( $\sigma$ )
‘MH-joint’	199	367	146	132	10
‘MH-regeneration’	170	317	197	140	9
‘Rejection-joint’	137	227	142	128	6
‘Rejection-bounded’	169	265	124	128	7

CPU time required

$$T^{CPU}(\text{‘MH-joint’}) < T^{CPU}(\text{‘MH-regeneration’}) < T^{CPU}(\text{‘Rejection-joint’}) < T^{CPU}(\text{‘Rejection-bounded’}).$$

The first inequality is due to the extra-computational cost needed for running the short additional Markov chain in the ‘MH-regeneration’ step (200 iterations in the simulations shown). This is, in turn, still lower than the time necessary for the rejection sampler step. The first three methods diverge from the basic ‘MH’ only for large observations, while the last one applies a restricted search of the maximum likelihood for each data point and this justifies its highest computational time. This consideration and the above analyses suggests that the ‘MH-regeneration’ step provides a satisfactory trade off between the IACTs and the computational time when  $\alpha$  approaches 2.

### B. Example 2: Synthetic data, estimate of $(\alpha, \sigma, \beta, \mu)$

Here we present simulation results on data generated from the  $\alpha$ -stable distribution with  $\alpha = 1.6$  and  $\mu = 20$  (dataset  $D_2$ ). We display only the output of the Gibbs sampler with the ‘MH-regeneration’ step, compared to the point estimators available in the STABLE toolbox, but considerations similar to the above can be deduced for the alternative enhanced samplers. Figures 11, 12, 13 and 14 show that the proposed method, initialized to  $(\alpha, \mu_W, \sigma_W, \mu) = (1.2, -3, 3, 30)$ , is capable of targeting the posterior distribution with a good mixing of the chains. For the dataset considered, the posterior of  $\sigma$  is concentrated on values lower than the true one; nevertheless its mean is fairly consistent with the frequentist point estimates, as shown in Table III,

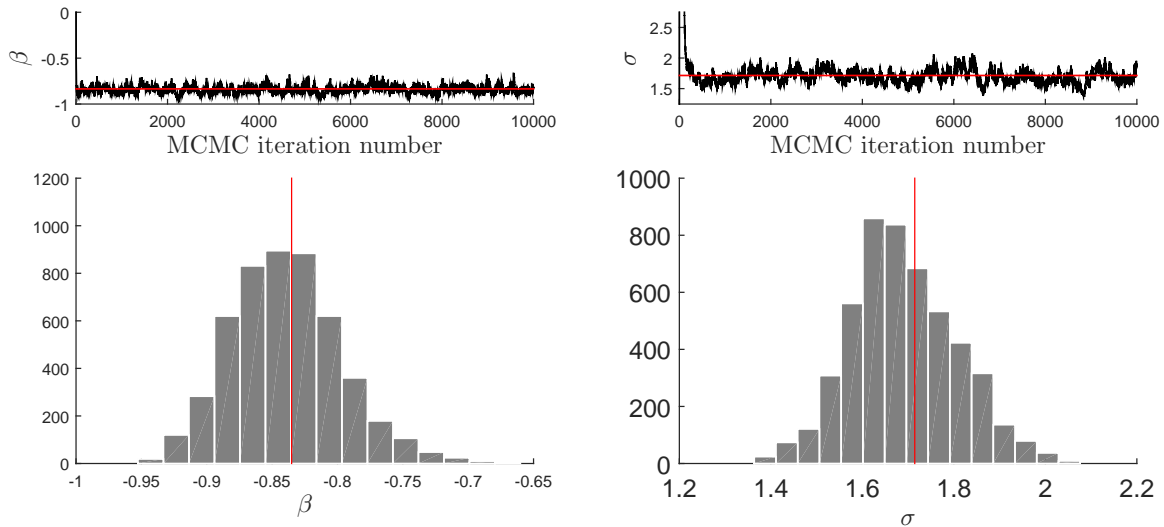


Fig. 7. Inference on the parameters  $\beta$  and  $\sigma$ , when  $\alpha = 0.8$ , using the ‘MH’ step for the  $\Gamma$ s and the residuals. Top: MCMC sampled parameter values  $\beta$  and  $\sigma$ . The true values are marked by the horizontal lines. Bottom: Histograms from the MCMC output. The true values are marked by the vertical lines.

TABLE II

ESTIMATES OF THE DISTRIBUTION PARAMETERS FOR THE DATASET  $D_1$  ( $\alpha = 1.8$ ): TRUE VALUES AND MAXIMUM LIKELIHOOD ESTIMATES; MCMC POSTERIOR MEANS, STANDARD DEVIATIONS, AND IACTS FOR THE FOUR ENHANCED SAMPLERS, AVERAGED OVER 5 RUNS.

Method	$\hat{\alpha}$			$\hat{\beta}$			$\hat{\sigma}$		
True values	1.8			-0.92			3.52		
Maximum likelihood	1.78			-1.00			3.51		
	$\mathbb{E}[\alpha X]$	$std[\alpha X]$	IACT	$\mathbb{E}[\beta X]$	$std[\beta X]$	IACT	$\mathbb{E}[\sigma X]$	$std[\sigma X]$	IACT
‘MH-joint’	1.81	0.04	160	-0.91	0.12	107	3.51	0.14	13
‘MH-regeneration’	1.80	0.04	170	-0.84	0.16	125	3.51	0.15	7
‘Rejection-joint’	1.81	0.04	126	-0.89	0.12	100	3.51	0.14	5
‘Rejection-bounded’	1.81	0.04	115	-0.88	0.13	111	3.51	0.14	5

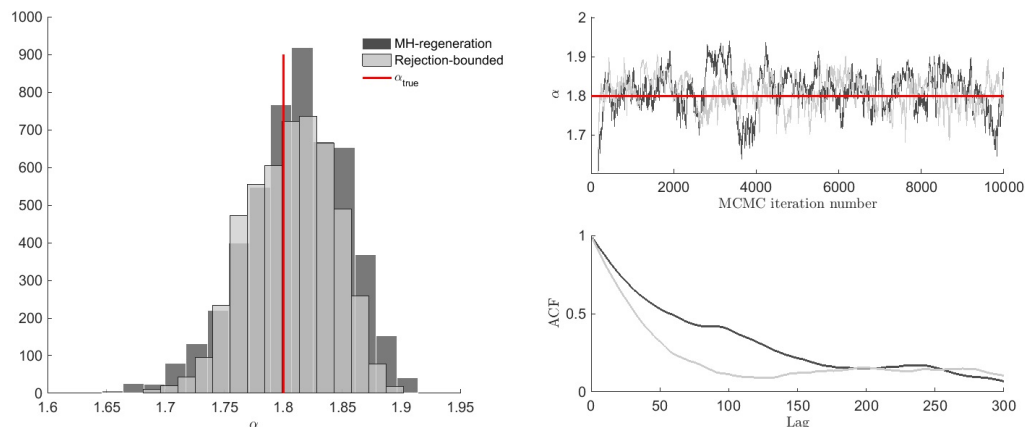


Fig. 8. Inference on the parameter  $\alpha$ , for the dataset  $D_1$  ( $\alpha = 1.8$ ), using the ‘MH-regeneration’ step (dark grey) and the ‘Rejection-bounded’ step (light grey) for the  $\Gamma$ s and the residuals. Left: Kernel density estimates from the MCMC output. The true value is marked by the vertical line. Right, top: MCMC sampled parameter value  $\alpha$ . The true value is marked by the horizontal line. Right, bottom: ACF as a function of lag.

where the MCMC outputs are averaged over 4 runs.

Figure 15 displays the stable density with the true value of the parameters (solid line, plotted thanks to the function `stablepdf` available from the STABLE Toolbox), which overlaps almost perfectly with a kernel smoothing function estimate (K.s.d, dashed line). They are compared to the stable density with the parameters estimated in the run shown in the previous set

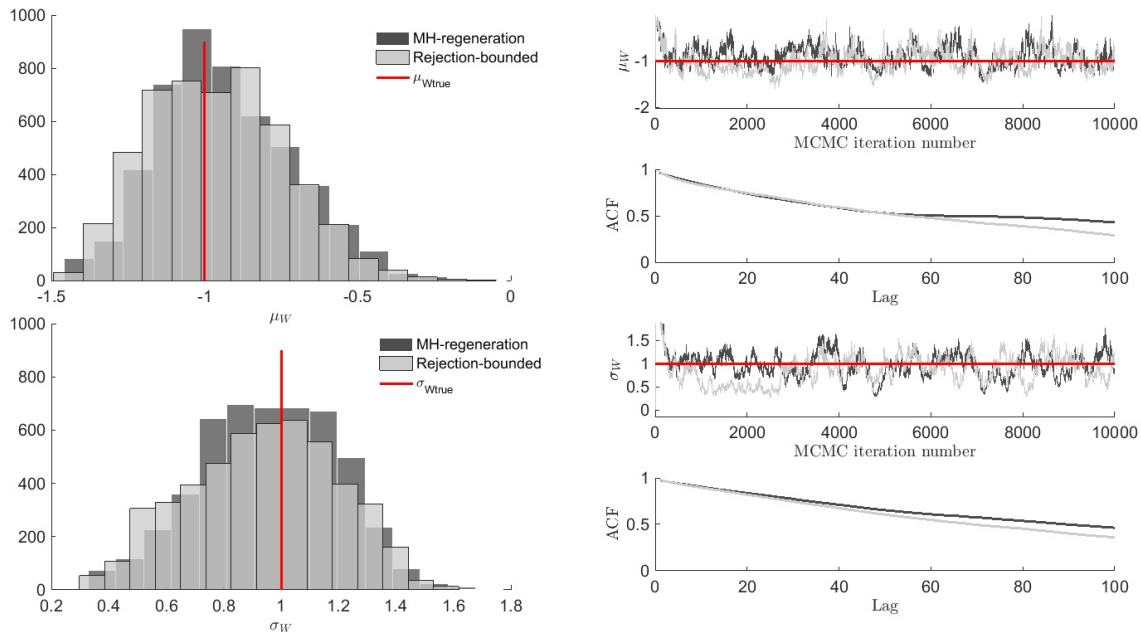


Fig. 9. Inference on the parameters  $\mu_W$  and  $\sigma_W$ , for the dataset  $D_1$  ( $\alpha = 1.8$ ), using the ‘MH-regeneration’ step (dark grey) and the ‘Rejection-bounded’ step (light grey) for the  $\Gamma$ s and the residuals. Left: Histograms from the MCMC output. The true values are marked the vertical lines. Right: MCMC sampled parameter values  $\mu_W$  and  $\sigma_W$  with the ACFs as functions of lag below each. The true values are marked by the horizontal lines.

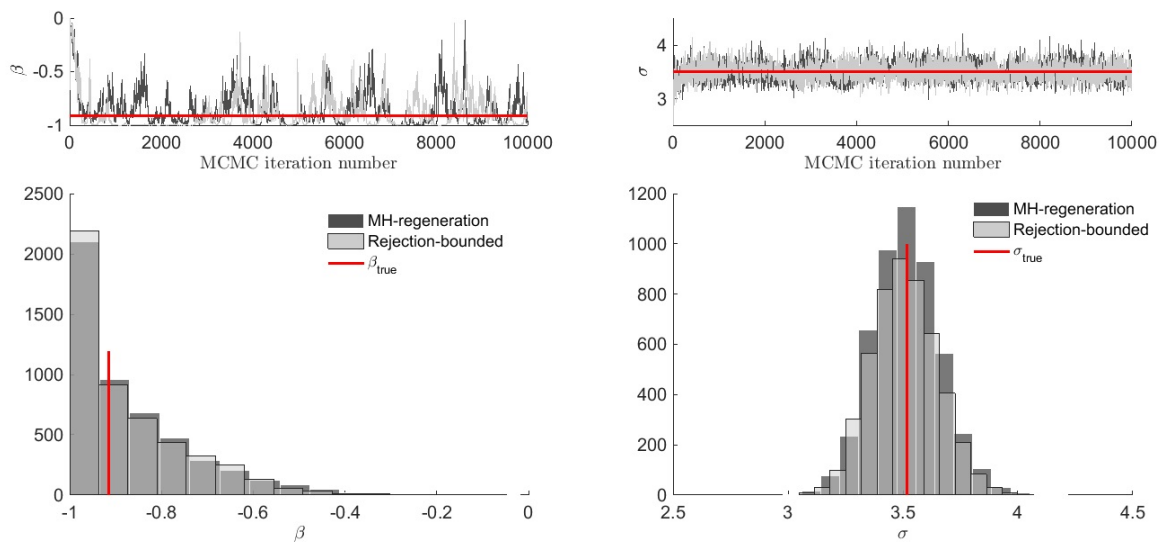


Fig. 10. Inference on the parameters  $\beta$  and  $\sigma$ , for the dataset  $D_1$  ( $\alpha = 1.8$ ), using the ‘MH-regeneration’ step (dark grey) and the ‘Rejection-bounded’ step (light grey) for the  $\Gamma$ s and the residuals. Top: MCMC sampled parameter values  $\beta$  and  $\sigma$ . The true values are marked by the horizontal lines. Bottom: Histograms from the MCMC output. The true values are marked by the vertical lines.

of figures (dashed line). The latter pdf is slightly more peaked to the centre and this could be attributed to the fact that both the estimated  $\hat{\alpha} = 1.58$  and  $\hat{\sigma} = 2.42$  are lower than the true values, assigning more probability mass in the tails. The full Bayesian posterior distribution estimated via the MCMC sampler can be summarised by the confidence interval over the density functions sampled by the MCMC, represented by the grey area in Figure 16 (in log-scale on the ordinate axis). This is formed by considering the mean and standard deviation of 20 sampled stable pdfs (obtained using tabulated density values), corresponding to the parameters sampled by the chains every 250 iterations following burn-in. The interval reassuringly includes the density obtained with the true parameters, as well as the stable densities estimated from the maximum likelihood,



TABLE III  
ESTIMATES OF THE DISTRIBUTION PARAMETERS FOR THE DATASET  $D_2$  ( $\alpha = 1.6$ ): TRUE VALUES, MAXIMUM LIKELIHOOD, QUANTILES METHOD, EMPIRICAL CHARACTERISTIC FUNCTION METHOD ESTIMATES; MCMC POSTERIOR MEANS, STANDARD DEVIATIONS, AND IACTS FOR THE ‘MH-REGENERATION’ SAMPLER, AVERAGED OVER 4 RUNS.

Method	$\hat{\alpha}$			$\hat{\beta}$			$\hat{\sigma}$			$\hat{\mu}$		
True values	1.6			-0.90			2.62			20		
Maximum likelihood	1.58			-0.90			2.40			20.02		
Quantiles method	1.51			-0.78			2.38			19.78		
E.C.F. method	1.58			-0.73			2.40			20.27		
	$\mathbb{E}[\alpha X]$	$std[\alpha X]$	IACT	$\mathbb{E}[\beta X]$	$std[\beta X]$	IACT	$\mathbb{E}[\sigma X]$	$std[\sigma X]$	IACT	$\mathbb{E}[\mu X]$	$std[\mu X]$	IACT
‘MH-regeneration’	1.55	0.06	173	-0.86	0.09	70	2.42	0.10	8	19.93	0.33	32

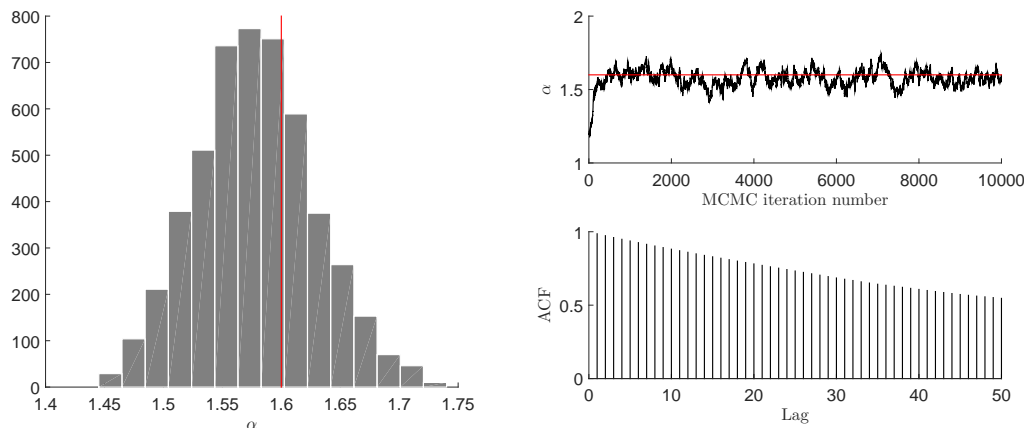


Fig. 11. Inference on the parameter  $\alpha$ , for the dataset  $D_2$  ( $\alpha = 1.6$ ), using the ‘MH-regeneration’ for the  $\Gamma$ s and the residuals. Left: Histogram from the MCMC output. The true value is marked by the vertical line. Right, top: MCMC sampled parameter value  $\alpha$ . The true value is marked by the horizontal line. Right, bottom: ACF as a function of lag.

the quantile and the empirical characteristic function methods: all of the estimates are a plausible explanation of the data. Despite the greater computational time required for the MCMC, the advantage and main motivation for using the Bayesian approach is that it provides a useful measure of the uncertainty of the estimate, through the whole posterior distribution of the parameters. Moreover, inference of the distribution parameters based on the PSR can be embedded into more complex hierarchical model structures (aimed at discrete and continuous time stochastic processes inference) in a way that is not at all straightforward with the classical methods.

### C. Example 3: Dow Jones Industrial Average data

For our real data example we consider daily price returns of Dow Jones Industrial average (DJIA) shares between 14/05/2010 and 14/05/2014, yielding 1005 values. The returns  $R_t$  are defined as the percentage change in price  $R_t = (p_t - p_{t-1})/p_{t-1}$ , with the share price on day  $t$  being denoted by  $p_t$ .

We initialize the sampler to  $(\alpha, \mu_W, \sigma_W, \mu) = (1.4, 0, 0.1, 0)$ . Figures 17 and 18 show the MCMC sampled parameter values for  $\alpha$  and  $\mu$  on the right-hand side with the ACFs below. In particular the step used on the  $\Gamma$ s and the residuals is the basic ‘MH’ without improvements for large observations, which produces good mixing of the chains and unimodal histograms for the posterior distributions, as displayed on the left-hand side of the figures. Good results are provided also for the reconstructed parameters  $\beta$  and  $\sigma$ , as presented in Figure 19. An average over 4 runs of the same algorithm provides the posterior mean estimates  $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \hat{\mu}) = (1.59, -0.09, 5 \times 10^{-3}, 5 \times 10^{-4})$ .

A comparison of the  $\alpha$ -stable distribution with the estimated parameters with respect to a fitted Gaussian pdf shows that the first one is more suitable to capture the presence of heavy tails in the given data (see Fig. 20).

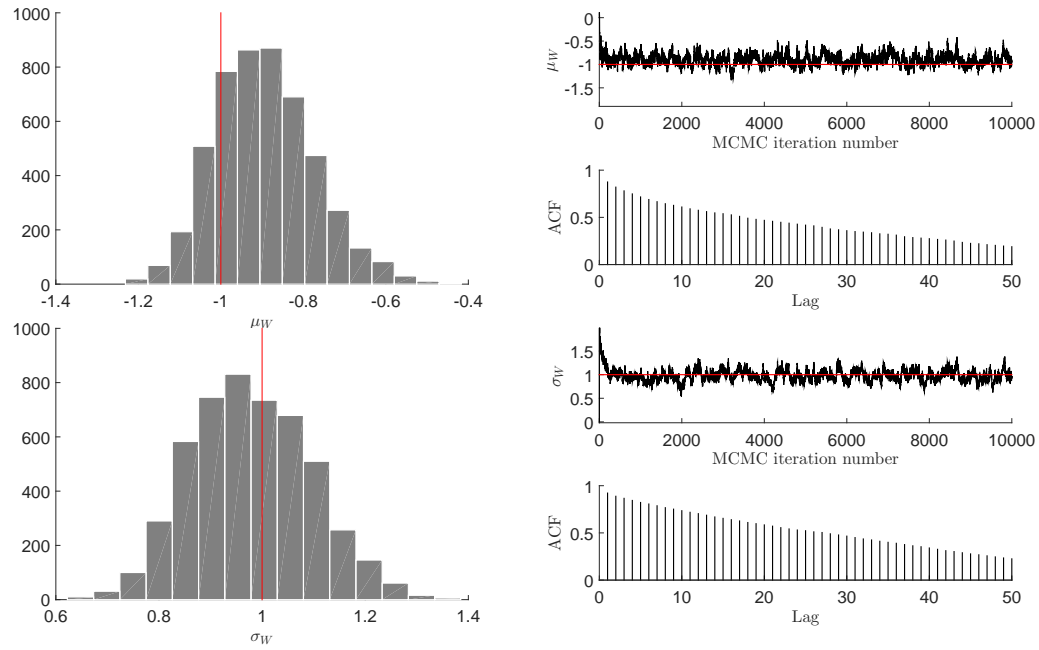


Fig. 12. Inference on the parameters  $\mu_W$  and  $\sigma_W$ , for the dataset  $D_2$  ( $\alpha = 1.6$ ), using the ‘MH-regeneration’ for the  $\Gamma$ s and the residuals. Left: Histograms from the MCMC output. The true values are marked the vertical lines. Right: MCMC sampled parameter values  $\mu_W$  and  $\sigma_W$  with the ACFs as functions of lag below each. The true values are marked by the horizontal lines.

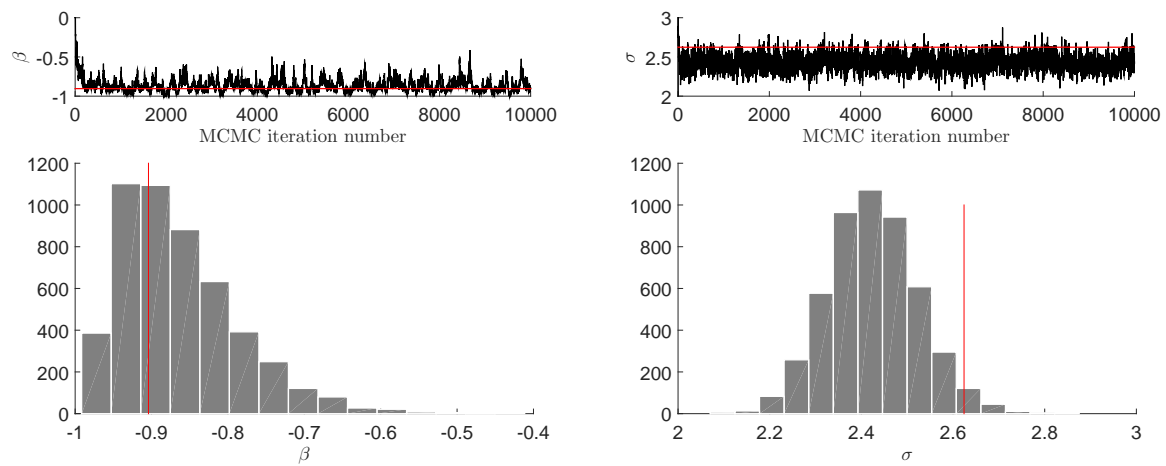


Fig. 13. Inference on the parameters  $\beta$  and  $\sigma$ , for the dataset  $D_2$  ( $\alpha = 1.6$ ), using the ‘MH-regeneration’ for the  $\Gamma$ s and the residuals. Top: MCMC sampled parameter values  $\beta$  and  $\sigma$ . The true values are marked by the horizontal lines. Bottom: Histograms from the MCMC output. The true values are marked by the vertical lines.

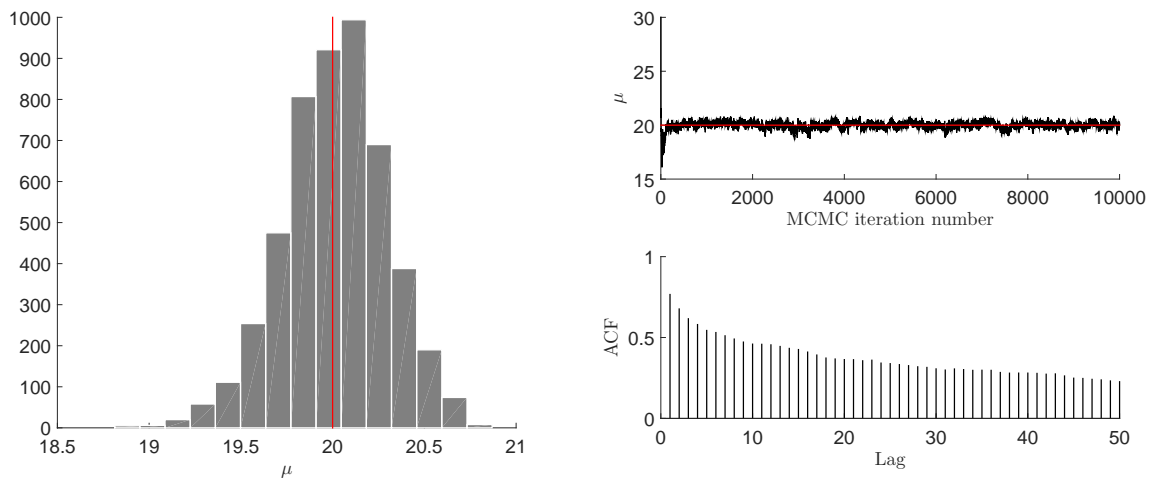


Fig. 14. Inference on the parameter  $\mu$ , for the dataset  $D_2$  ( $\alpha = 1.6$ ), using the ‘MH-regeneration’ for the  $\Gamma$ s and the residuals. Left: Histogram from the MCMC output. The true value is marked by the vertical line. Right, top: MCMC sampled parameter value  $\mu$ . The true value is marked by the horizontal line. Right, bottom: ACF as a function of lag.

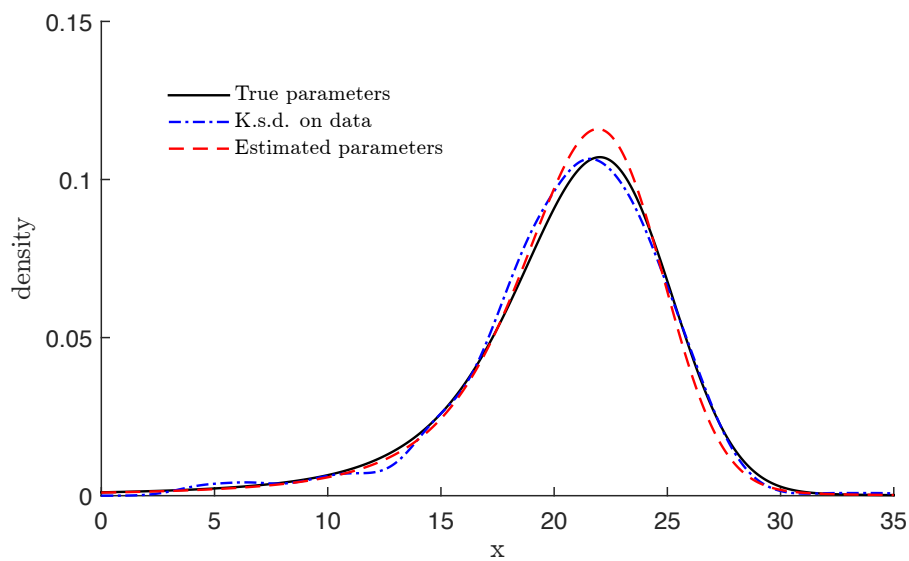


Fig. 15. Stable density corresponding to the true value of the parameters (solid line), for the dataset  $D_2$  ( $\alpha = 1.6$ ), kernel density estimate on the dataset (dashed-dotted line), overlapped with the stable pdf corresponding to the MCMC posterior means of one run (dashed line).

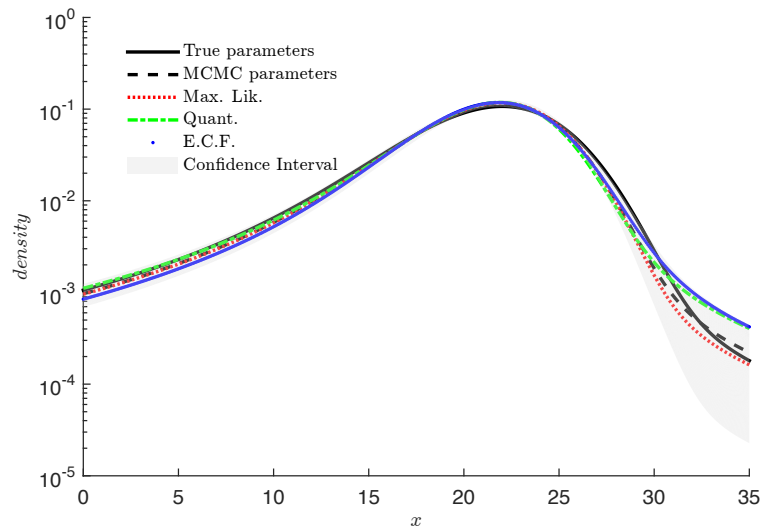


Fig. 16. Confidence interval (mean  $\pm$  2 standard deviations) on the stable pdf corresponding to the dataset  $D_2$  ( $\alpha = 1.6$ ), obtained with 20 densities curves (grey region); stable density corresponding to the true parameters (black solid line), to the MCMC posterior means estimates (black dashed line), to the maximum likelihood estimator (red line), to the quantile estimator (green line) and to the empirical characteristic function method (blue line). For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

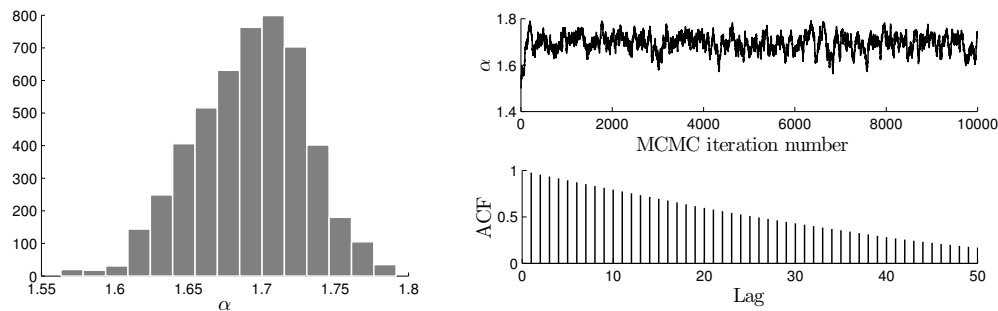


Fig. 17. Inference on the parameter  $\alpha$  for DJIA shares data, using the ‘MH’ step for the  $\Gamma$ 's and the residuals. Left: Histogram from the MCMC output. Right, top: MCMC sampled parameter value  $\alpha$ . Right, bottom: ACF as a function of lag.

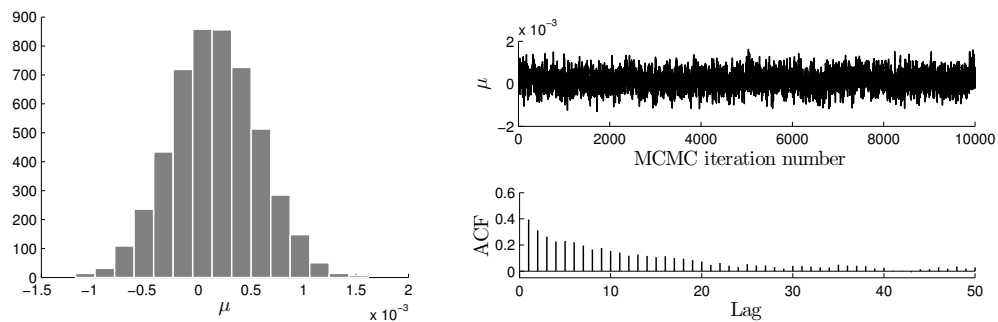


Fig. 18. Inference on the parameter  $\mu$  for DJIA shares data, using the ‘MH’ step for the  $\Gamma$ 's and the residuals. Left: Histogram from the MCMC output. Right, top: MCMC sampled parameter value  $\mu$ . Right, bottom: ACF as a function of lag.

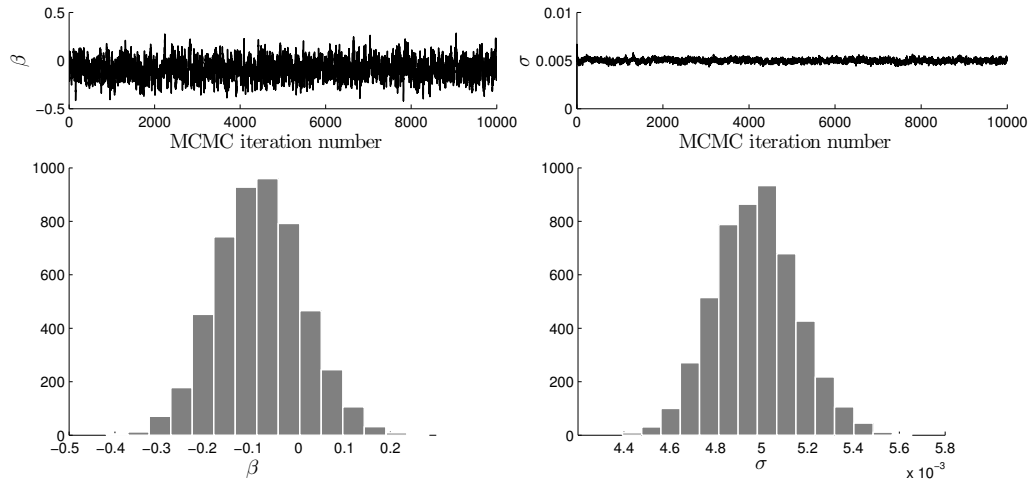


Fig. 19. Inference on the parameters  $\beta$  and  $\sigma$  for DJIA shares data, using the 'MH' step for the  $\Gamma$ 's and the residuals. Top: MCMC sampled parameter values  $\beta$  and  $\sigma$ . Bottom: Histograms from the MCMC output.

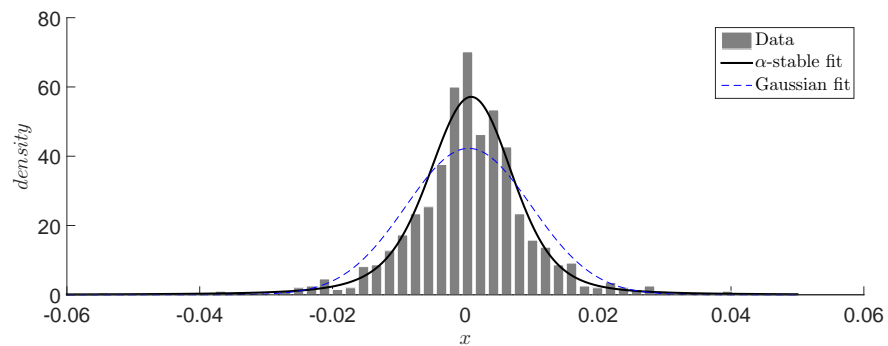


Fig. 20. DJIA shares data: comparison of the  $\alpha$ -stable density with the estimated parameters (averaged over 4 runs) and a Gaussian fit.

## VII. CONCLUSION

We have achieved satisfactory results for parameter estimation for asymmetric  $\alpha$ -stable distributions applying Bayesian inference methods, which are based on our conditionally Gaussian framework resulting from an approximated Poisson series representation. To this end, we introduced a novel residual approximation method of the series residual terms, which exactly characterises the mean and variance of the approximation. Additional improvement steps for updating the latent variables given by the PSR ensure that observations far in the tails can be represented accurately enough.

We would like to stress that this paper exposes only the basic parameter estimation ideas for the Poisson Series Representation. In this we find that it is competitive in accuracy and sometimes better in performance than other competitive parameter estimation procedures. However, it is quite computationally intensive, certainly compared to some of the approximate frequentist procedures available. Where our approach is likely to win out more significantly is in more complex hierarchical models in which the  $\alpha$ -stable noise model drives a time series or stochastic process model. Then we will really see the advantages of having a simple conditionally Gaussian inference framework, since then conditional Kalman filter-smoothers and other standard Gaussian modelling tools can be routinely applied. In fact, it would in practice be straightforward to convert any Bayesian Monte Carlo algorithm currently running on a Gaussian model into the corresponding  $\alpha$ -stable model via the addition of some extra sampling steps for the auxiliary random variables  $\mu_X$  and  $\sigma_X$ . Initial examples of such models have already been given in our conference papers [8]–[10], for both continuous time and discrete time models, and using batch MCMC approaches and sequential Particle Filters, but we do plan to elaborate these ideas significantly in future work.

### APPENDIX A TRANSFORMATIONS FOR $\sigma$ AND $\beta$

Here we show that the mappings expressed in (4) allow to obtain any value of  $\sigma > 0$  and  $\beta \in [-1, 1]$ , given  $\alpha$  and the distribution of  $\{W_i\}_{i=1}^{\infty}$ . Assume that the latter is continuous, with density  $f_W(w)$ ; we can introduce the following density function

$$g(w) = \frac{|w|^\alpha f_W(w)}{\int_{\mathbb{R}} |w|^\alpha f_W(w) dw} = \frac{\tilde{g}(w)}{I},$$

where  $\tilde{g}(w)$  and  $I$  denote the unnormalized density and the normalizing constant, respectively. Then the first transformation in (4) can be rewritten as

$$\sigma = \frac{\int_{\mathbb{R}} \tilde{g}(w) dw}{C_\alpha} = \frac{I}{C_\alpha}.$$

Given that  $C_\alpha > 0$ , it results that  $\sigma > 0$ . Moreover, for any fixed value of  $\alpha$  (and consequently  $C_\alpha$ ), it is possible to achieve any  $\sigma > 0$ , by determining the parameters of the distribution of  $W_i$  that give the necessary value of  $I$  (the reader can think for example to a scale parameter, which in the Gaussian case corresponds to  $\sigma_W$ ). As regard the second transformation expressed in (4), we can write

$$\begin{aligned} \beta &= \frac{-\int_{-\infty}^0 \tilde{g}(w) dw + \int_0^{\infty} \tilde{g}(w) dw}{I} \\ &= -\int_{-\infty}^0 g(w) dw + \int_0^{\infty} g(w) dw \\ &= -(1 - I^+) + I^+ \\ &= 2I^+ - 1, \end{aligned}$$

where  $I^+$  is the probability mass assigned by  $g(w)$  to  $\mathbb{R}^+$ . Then any  $\beta \in [-1, 1]$  can be obtained by setting the parameters of the distribution of  $W_i$  to obtain the necessary  $I^+ \in [0, 1]$  (the reader can think for example to a location parameter, corresponding to  $\mu_W$  in the Gaussian case).

### APPENDIX B GAUSSIAN APPROXIMATION OF MOMENTS

The remaining summation terms are studied by reverting to the Poisson process representation of the  $\Gamma_m$ s on a finite interval  $[c, d]$ . Specifically, since  $\{\Gamma_m\}$  is a unit rate Poisson process, the number of the  $\Gamma$ s in the interval follows a Poisson distribution,

$$|\{\Gamma_m : \Gamma_m \in [c, d]\}| \sim \text{Poisson}(d - c) \quad \text{for } d > c, \quad (41)$$

and each  $\Gamma_m$  is uniformly and independent distributed on  $[c, d]$ ,

$$\Gamma_m | |\{\Gamma_m : \Gamma_m \in [c, d]\}| \sim U([c, d]). \quad (42)$$

Taking the limit as  $d \rightarrow \infty$  accounts for all residual terms. Taking (16) and (17) with  $d \rightarrow \infty$  we compute the limits for  $\mu_R^{(\alpha)}$  and  $\Sigma_R^{(\alpha)}$  as

$$\begin{aligned} \mu_R^{(\alpha)} &= \begin{bmatrix} \mathbb{E}[R_1] \\ \mathbb{E}[R_2] \end{bmatrix} = \begin{bmatrix} \lim_{d \rightarrow \infty} \frac{\alpha}{\alpha-1} \left( d^{\frac{\alpha-1}{\alpha}} - c^{\frac{\alpha-1}{\alpha}} \right) & \text{if } \alpha < 1, \\ \lim_{d \rightarrow \infty} \frac{\alpha}{\alpha-1} \left( d^{\frac{\alpha-1}{\alpha}} - c^{\frac{\alpha-1}{\alpha}} \right) - d^{\frac{\alpha-1}{\alpha}} & \text{if } \alpha > 1 \\ \lim_{d \rightarrow \infty} \frac{\alpha}{\alpha-2} \left( d^{\frac{\alpha-2}{\alpha}} - c^{\frac{\alpha-2}{\alpha}} \right) & \end{bmatrix} \\ &= \begin{bmatrix} \lim_{d \rightarrow \infty} \frac{\alpha}{\alpha-1} \left( d^{\frac{\alpha-1}{\alpha}} - c^{\frac{\alpha-1}{\alpha}} \right) & \text{if } \alpha < 1, \\ \frac{\alpha}{\alpha-1} \left( -c^{\frac{\alpha-1}{\alpha}} \right) & \text{if } \alpha > 1 \\ \lim_{d \rightarrow \infty} \frac{\alpha}{\alpha-2} \left( d^{\frac{\alpha-2}{\alpha}} - c^{\frac{\alpha-2}{\alpha}} \right) & \end{bmatrix} = \begin{bmatrix} \frac{\alpha}{1-\alpha} c^{\frac{\alpha-1}{\alpha}} \\ \frac{\alpha}{2-\alpha} c^{\frac{\alpha-2}{\alpha}} \end{bmatrix}. \end{aligned} \quad (43)$$

In order to obtain the variance-covariance matrix

$$\Sigma_R^{(\alpha)} = \begin{bmatrix} \text{Var}[R_1] & \text{Cov}[R_1, R_2] \\ \text{Cov}[R_2, R_1] & \text{Var}[R_2] \end{bmatrix} \quad (44)$$

on the limit  $d \rightarrow \infty$ , we work out each of the components as

$$\text{Var}[R_1] = \lim_{d \rightarrow \infty} \frac{\alpha}{\alpha-2} \left( d^{\frac{\alpha-2}{\alpha}} - c^{\frac{\alpha-2}{\alpha}} \right) - \frac{1}{d-c} \frac{\alpha^2}{(\alpha-1)^2} \left( d^{\frac{\alpha-1}{\alpha}} - c^{\frac{\alpha-1}{\alpha}} \right)^2 \quad (45)$$

$$= \frac{\alpha}{2-\alpha} c^{\frac{\alpha-2}{\alpha}}, \quad (46)$$

$$\text{Var}[R_2] = \lim_{d \rightarrow \infty} \frac{\alpha}{\alpha-4} \left( d^{\frac{\alpha-4}{\alpha}} - c^{\frac{\alpha-4}{\alpha}} \right) - \frac{1}{d-c} \frac{\alpha^2}{(\alpha-2)^2} \left( d^{\frac{\alpha-2}{\alpha}} - c^{\frac{\alpha-2}{\alpha}} \right)^2 \quad (47)$$

$$= \frac{\alpha}{4-\alpha} c^{\frac{\alpha-4}{\alpha}}, \quad (48)$$

and

$$\begin{aligned} \text{Cov}[R_1, R_2] &= \text{Cov}[R_2, R_1] = \lim_{d \rightarrow \infty} \sum_{m: \Gamma_m \in [c, d]} \sum_{n: \Gamma_n \in [c, d]} \text{Cov} \left( \Gamma_m^{1/\alpha}, \Gamma_n^{-2/\alpha} \right) \\ &= \lim_{d \rightarrow \infty} \sum_{m: \Gamma_m \in [c, d]} \frac{\alpha}{\alpha-3} \left( d^{\frac{\alpha-3}{\alpha}} - c^{\frac{\alpha-3}{\alpha}} \right) - \frac{1}{d-c} \frac{\alpha^2}{(\alpha-1)(\alpha-2)} \left( d^{\frac{\alpha-1}{\alpha}} - c^{\frac{\alpha-1}{\alpha}} \right) \left( d^{\frac{\alpha-2}{\alpha}} - c^{\frac{\alpha-2}{\alpha}} \right) \\ &= \frac{\alpha}{3-\alpha} c^{\frac{\alpha-2}{\alpha}}. \end{aligned} \quad (49)$$

## APPENDIX C

### MARGINAL AND CONDITIONAL DISTRIBUTIONS

Expression (24) is based on the conditionally Gaussian representation of stable random variables, and can be derived in the following way

$$\begin{aligned} p(\mathbf{X} | \mathbf{\Gamma}, \sigma_W, \mu_W, \mu, \alpha) &= \prod_{n=1}^N \mathcal{N} \left( X_n \mid \mu_W \sum_{i=1}^{\infty} \left( \Gamma_{i,n}^{-1/\alpha} - \tilde{k}_i^{(\alpha)} \right) + \mu, \sigma_W^2 \sum_{i=1}^{\infty} \Gamma_{i,n}^{-2/\alpha} \right) \\ &= \prod_{n=1}^N \mathcal{N} \left( X_n \mid \mu_W m_n + \mu, \sigma_W^2 s_n \right) \\ &= \frac{(2\pi\sigma_W^2)^{-N/2}}{\prod_{n=1}^N \sqrt{s_n}} \exp \left\{ -\frac{1}{2\sigma_W^2} \sum_{n=1}^N (X_n - \mu_W m_n - \mu)^2 / s_n \right\} \\ &= \frac{(2\pi\sigma_W^2)^{-N/2}}{\prod_{n=1}^N \sqrt{s_n}} \exp \left\{ -\frac{1}{2\sigma_W^2} \sum_{n=1}^N \left( \frac{X_n}{\sqrt{s_n}} - \begin{bmatrix} m_n/\sqrt{s_n} \\ 1/s_n \end{bmatrix}^T \begin{bmatrix} \mu_W \\ \mu \end{bmatrix} \right)^2 \right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{(2\pi\sigma_W^2)^{-N/2}}{\prod_{n=1}^N \sqrt{s_n}} \exp \left\{ -\frac{1}{2\sigma_W^2} \left| \left( \tilde{\mathbf{X}} - \tilde{\mathbf{M}}\boldsymbol{\mu} \right) \right|^2 \right\} \\
&= \frac{(2\pi\sigma_W^2)^{-N/2}}{\prod_{n=1}^N \sqrt{s_n}} \exp \left\{ -\frac{1}{2\sigma_W^2} \left( \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} - \tilde{\mathbf{X}}^T \tilde{\mathbf{M}}\boldsymbol{\mu} - \boldsymbol{\mu}^T \tilde{\mathbf{M}}^T \tilde{\mathbf{X}} + \boldsymbol{\mu}^T \tilde{\mathbf{M}}^T \tilde{\mathbf{M}}\boldsymbol{\mu} \right) \right\},
\end{aligned}$$

with  $m_n, s_n$  as in (25),  $\tilde{\mathbf{X}}, \tilde{\mathbf{M}}, \boldsymbol{\mu}$  as in (26).

The conditional distribution for  $\boldsymbol{\mu}$  in equation (27) is obtained by rearranging (24) with respect to  $\boldsymbol{\mu}$ , aiming for an expression proportional to a bivariate Gaussian distribution,

$$p(\mathbf{X}|\boldsymbol{\Gamma}, \sigma_W, \boldsymbol{\mu}, \alpha) \propto \mathcal{N}(\boldsymbol{\mu}|\cdot, \cdot). \quad (50)$$

To this end, we write

$$\begin{aligned}
p(\mathbf{X}|\boldsymbol{\Gamma}, \mathbf{R}, \sigma_W, \boldsymbol{\mu}, \alpha) &= \frac{1}{(2\pi\sigma_W^2)^{N/2} \prod_{n=1}^N \sqrt{s_n}} \exp \left\{ -\frac{1}{2\sigma_W^2} \left( (\boldsymbol{\mu} - \mathbf{a}_\boldsymbol{\mu})^T \Sigma_\boldsymbol{\mu}^{-1} (\boldsymbol{\mu} - \mathbf{a}_\boldsymbol{\mu}) + B \right) \right\} \\
&= \frac{|\Sigma_\boldsymbol{\mu}|^{1/2}}{(2\pi\sigma_W^2)^{(N-2)/2} \prod_{n=1}^N \sqrt{s_n}} \mathcal{N}(\boldsymbol{\mu}|\mathbf{a}_\boldsymbol{\mu}, \Sigma_\boldsymbol{\mu}) \exp \left\{ -\frac{B}{2\sigma_W^2} \right\},
\end{aligned}$$

where  $B$  is defined as in (31).

Finally the conditional distribution for  $\sigma_W$  in equation (30) is computed taking a uniform prior on  $\sigma_W$ , considered independent on  $\boldsymbol{\mu}$ , which gives the proportionality  $p(\boldsymbol{\mu}, \sigma_W^2|\mathbf{X}, \boldsymbol{\Gamma}, \mathbf{R}, \alpha) \propto p(\mathbf{X}|\boldsymbol{\Gamma}, \mathbf{R}, \sigma_W, \boldsymbol{\mu}, \alpha)$

$$\begin{aligned}
p(\sigma_W^2|\mathbf{X}, \boldsymbol{\Gamma}, \mathbf{R}, \alpha) &= p(\boldsymbol{\mu}, \sigma_W^2|\mathbf{X}, \boldsymbol{\Gamma}, \mathbf{R}, \alpha) p(\boldsymbol{\mu}|\mathbf{X}, \boldsymbol{\Gamma}, \mathbf{R}, \sigma_W, \alpha)^{-1} \\
&\propto p(\mathbf{X}|\boldsymbol{\Gamma}, \mathbf{R}, \sigma_W, \boldsymbol{\mu}, \alpha) \mathcal{N}(\boldsymbol{\mu}|\mathbf{a}_\boldsymbol{\mu}, \Sigma_\boldsymbol{\mu})^{-1} \\
&= \frac{|\Sigma_\boldsymbol{\mu}|^{1/2}}{(2\pi\sigma_W^2)^{(N-2)/2} \prod_{n=1}^N \sqrt{s_n}} \mathcal{N}(\boldsymbol{\mu}|\mathbf{a}_\boldsymbol{\mu}, \Sigma_\boldsymbol{\mu})^{-1} \\
&\exp \left\{ -\frac{1}{2\sigma_W^2} \left( \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} - \tilde{\mathbf{X}}^T \tilde{\mathbf{M}} (\tilde{\mathbf{M}}^T \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}}^T \tilde{\mathbf{X}} \right) \right\} \\
&\propto \frac{B^{\frac{N-4}{2}}}{\Gamma(\frac{N-4}{2})} (\sigma_W^2)^{(N-2)/2} \exp \left\{ -\frac{B}{2\sigma_W^2} \right\} \\
&= \mathcal{IG} \left( \sigma_W^2 \left| \frac{N-4}{2}, \frac{B}{2} \right. \right).
\end{aligned}$$

#### APPENDIX D

##### SAMPLING $m_n$ FOR OBSERVATIONS IN THE TAILS OF THE DISTRIBUTION

The supremum of (40) is the solution to

$$\frac{\partial p(X_n|m_n, s_n = m_n^2, \mu_W, \sigma_W)}{\partial m_n} \propto \sigma_W^2 m_n^2 + X_n \mu_W m_n - X_n^2 = 0$$

computed as

$$\hat{m}_n = \frac{-X_n(\mu_W - \sqrt{\mu_W^2 + 4\sigma_W^2})}{2\sigma_W^2}. \quad (51)$$

We can determine the approximate support of the likelihood as a function of  $m_n$ , say  $m_{n,\min}$  and  $m_{n,\max}$ . Then, one possibility would be to generate tail samples from  $p(m)$  over the support of the likelihood using the Pareto approximation

$$p(m) \propto m^{-(\alpha+1)} \quad (52)$$

for large  $m_n$ . The cumulative distribution function over the interval  $[m_{n,\min}, m_{n,\max}]$  is given by

$$F(m_n) = \frac{m_{n,\min}^{-\alpha} - m_n^{-\alpha}}{m_{n,\min}^{-\alpha} - m_{n,\max}^{-\alpha}}.$$



Hence, to sample the approximation over  $[m_{n,\min}, m_{n,\max}]$  we generate a uniform random variable  $u \in [0, 1]$  and obtain  $m_n$  by the inverse transformation method<sup>3</sup> as

$$m_n^{-\alpha} = (1 - u)m_{n,\min}^{-\alpha} + um_{n,\max}^{-\alpha}.$$

To choose  $m_{m,\min}$  and  $m_{n,\max}$  we search for the value of  $m_n$  for which the posterior

$$p(m_n|X_n, s_n = m_n^2, \mu_W, \sigma_W) \propto p(X_n|m_n, s_n = m_n^2, \mu_W, \sigma_W)p(m_n)$$

peaks. The maximisation problem requires the solution to

$$(\alpha + 2)\sigma_W^2 m_n^2 + X_n \mu_W m_n - X_n^2 = 0,$$

which is given by

$$\tilde{m}_n = \frac{-X_n(\mu_W - \sqrt{\mu_W^2 + 4(\alpha + 2)\sigma_W^2})}{2(\alpha + 2)\sigma_W^2}.$$

Scaling  $\tilde{m}_n$  up and down gives us  $m_{n,\max}$  and  $m_{n,\min}$ . Finally, the rejection step can be performed with the bounding envelope function

$$\frac{1}{\sqrt{2\pi\sigma_W^2 \hat{m}_n^2}} \exp\left(-\frac{1}{2\sigma_W^2 \hat{m}_n^2}(X_n - \mu_W \hat{m}_n)^2\right) p(m_n).$$

Some suitable procedure needs to be used to decide whether to use the full rejection sampler, which is expected to be slow for certain observations, or the fast approximate tail rejection sampler as introduced above. Note, however, that the presented approximate tail rejection sampler just allows for the sampling of  $\mu_W$  and  $\sigma_W^2$  in a subsequent step, but not of the distribution parameter  $\alpha$ . To be able to apply the ‘MH’ sampler for  $\alpha$  for all  $n$  data points  $X_n$ , the set  $\Gamma_n$  and  $\mathbf{R}_n$  given  $m_n$  and the current  $\alpha$  need to be obtained first.

#### Regenerating the set of $\Gamma$ s and $(\mathbf{R}_1, \mathbf{R}_2)$ given $m$ with an MCMC chain

Given some accepted  $m_n$  from the method described in the previous section, a short MCMC chain is started to regenerate the set of  $\Gamma$ s,  $\{\Gamma_{i,n}\}_{i=2}^{M_n}$ , the residual approximations,  $(R_{1,n}, R_{2,n})$ , and  $s_n$ . To shorten the notation, the observation index  $n$  shall be omitted throughout the rest of this section. We then have

$$\begin{aligned} p(m|\Gamma_{2:M}, \mathbf{R}, \alpha) &= p\left(\Gamma_1^{-1/\alpha} \middle| \Gamma_{2:M}, \mathbf{R}, \alpha\right) = \left|\frac{\partial \Gamma_1^{1/\alpha}}{\partial \Gamma_1}\right|^{-1} p(\Gamma_1|\Gamma_{2:M}) \\ &= \mathcal{I}_{\Gamma_1 > 0} \left|-\frac{1}{\alpha} \Gamma_1^{-(1/\alpha+1)}\right|^{-1} \frac{1}{\Gamma_2} = \mathcal{I}_{\Gamma_1 > 0} \frac{\alpha}{\Gamma_2} \Gamma_1^{(1/\alpha+1)}. \end{aligned}$$

Note that  $\Gamma_1$  is not sampled in this scheme, but proposed through the relation  $m = \sum_{i=1}^M \Gamma_i^{-1/\alpha} + R_1$ . The obtained  $\Gamma_1$  is then used to propose  $s$ . Thus, since  $s$  is not fixed we also need to consider dependence upon the data  $X$ , leading to the full likelihood

$$p(X, m|\Gamma_{2:M}, \mathbf{R}, \alpha) = p(X|m, s(m, \Gamma_{2:M}, \mathbf{R}, \alpha))p(m|\Gamma_{2:M}, \mathbf{R}, \alpha),$$

where the first term  $p(X|m, s(m, \Gamma_{2:M}, \mathbf{R}, \alpha))$  is the conditional Gaussian  $\mathcal{N}(\mu_W m, \sigma_W^2 s)$  as before. An independent ‘MH’ procedure can now be run, proposing from the priors  $p(\Gamma_{2:M}|\mathbf{R}, \alpha) = p(\Gamma_{2:M})$  and  $p(\mathbf{R}, \alpha) = p(\mathbf{R})$ , and targeting the conditional distribution

$$\begin{aligned} p(\Gamma_{2:M}, \mathbf{R}|X, m, \alpha) &\propto p(X, m|\Gamma_{2:M}, \mathbf{R}, \alpha)p(\Gamma_{2:M}, \mathbf{R}|\alpha) \\ &= \mathcal{I}_{\Gamma_1 > 0} \frac{1}{\sqrt{2\pi\sigma_W^2 s}} \exp\left(-\frac{(X - \mu_W m)^2}{2\sigma_W^2 s}\right) \frac{\alpha}{\Gamma_2} \Gamma_1^{(\frac{1}{\alpha}+1)} p(\Gamma_{2:M})p(\mathbf{R}). \end{aligned}$$

<sup>3</sup>**Inverse transformation method:** If  $Y$  is uniformly distributed on  $[0, 1]$  and if  $X$  has a cumulative distribution function  $F_X$ , then the cumulative distribution function of the random variable  $F_X^{-1}(Y)$  is  $F_X$ .

APPENDIX E  
TARGETING  $m$ , THE SET OF  $\Gamma$ S AND  $(\mathbf{R}_1, \mathbf{R}_2)$  JOINTLY

To speed up the sampler it is helpful to avoid the additional MCMC chain to regenerate the set of  $\Gamma$ s and the residual terms. Thus, we present an alternative sampler with a joint move for  $m, \Gamma$  and  $\mathbf{R}$  in the case of large observations. Again, we make sure that  $m$  is proposed from the correct region using a Pareto prior as in (52). Suppose that  $m, \Gamma$  and  $\mathbf{R}$  are proposed as follows,

$$\begin{aligned} q(m|X, \mu_W, \sigma_W, \alpha) &\propto \\ m^{-(\alpha+1)} \frac{1}{\sqrt{2\pi\sigma_W^2 m^2}} \exp\left(-\frac{(X - \mu_W m)^2}{2\sigma_W^2 m^2}\right), \\ q(\Gamma_{2:M}|m, X, \mu_W, \sigma_W, \alpha) &= p(\Gamma_{2:M}), \\ q(\mathbf{R}|\Gamma_{2:M}, m, X, \mu_W, \sigma_W, \alpha) &= p(\mathbf{R}). \end{aligned}$$

Samples from  $q(m|X, \mu_W, \sigma_W, \alpha)$  are obtained via rejection sampling with the bounding envelope function

$$m^{-(\alpha+1)} \frac{1}{\sqrt{2\pi\sigma_W^2 \hat{m}^2}} \exp\left(-\frac{(X - \mu_W \hat{m})^2}{2\sigma_W^2 \hat{m}^2}\right),$$

where  $\hat{m}$  is as in (51). The importance weight  $\frac{p(x)}{q(x' \rightarrow x)}$  for such a joint move is computed as

$$\begin{aligned} \rho &= \frac{p(m, \Gamma_{2:M}, \mathbf{R}|X, \mu_W, \sigma_W, \alpha)}{q(m|X, \mu_W, \sigma_W, \alpha)q(\Gamma_{2:M}|m, X, \mu_W, \sigma_W, \alpha)q(\mathbf{R}|\Gamma_{2:M}, m, X, \mu_W, \sigma_W, \alpha)} \\ &= \frac{p(X|m, \Gamma_{2:M}, \mathbf{R}, \mu_W, \sigma_W, \alpha)p(m, \Gamma_{2:M}, \mathbf{R}|\mu_W, \sigma_W, \alpha)}{p(X|\mu_W, \sigma_W, \alpha)} \\ &\quad \cdot \frac{1}{p(m|X, \mu_W, \sigma_W, \alpha)p(\Gamma_{2:M})p(\mathbf{R})} \\ &= \frac{p(X|m, \Gamma_{2:M}, \mathbf{R}, \mu_W, \sigma_W, \alpha)p(m|\Gamma_{2:M}, \mathbf{R}, \mu_W, \sigma_W, \alpha)p(\Gamma_{2:M}, \mathbf{R}|\mu_W, \sigma_W, \alpha)}{(X|\mu_W, \sigma_W, \alpha)} \\ &\quad \cdot \frac{p(X|\mu_W, \sigma_W, \alpha)}{p(m|\mu_W, \sigma_W, \alpha)p(X|m, \mu_W, \sigma_W, \alpha)p(\Gamma_{2:M})p(\mathbf{R})} \\ &\propto \mathcal{I}_{\Gamma_1 > 0} \frac{\alpha}{\Gamma_2} \Gamma_1^{(1/\alpha+1)} m^{(\alpha+2)} s^{-1/2} \exp\left(-\frac{m^2 - s}{2\sigma_W^2 m^2 s} (X - \mu_W m)^2\right). \end{aligned}$$

Alternatively, we could neglect the information given by the observation when choosing the proposal for  $m$  and set

$$q(m|X, \mu_W, \sigma_W, \alpha) \propto m^{-(\alpha+1)}.$$

Thus, the acceptance probability becomes

$$\rho((m, \Gamma_{2:M}, \mathbf{R}), (m', \Gamma'_{2:M}, \mathbf{R}')) = \min\left(1, \frac{\rho'}{\rho}\right),$$

where, in the first case

$$\begin{aligned} \frac{\rho'}{\rho} &= \left(\frac{\Gamma'_1}{\Gamma_1}\right)^{(1/\alpha+1)} \left(\frac{m'}{m}\right)^{(\alpha+2)} \left(\frac{s'}{s}\right)^{-1/2} \\ &\quad \cdot \exp\left(-\frac{m'^2 - s'}{2\sigma_W^2 m'^2 s'} (X - \mu_W m')^2 + \frac{m^2 - s}{2\sigma_W^2 m^2 s} (X - \mu_W m)^2\right), \end{aligned}$$

while in the second case

$$\begin{aligned} \frac{\rho'}{\rho} &= \left(\frac{\Gamma'_1}{\Gamma_1}\right)^{(1+\frac{1}{\alpha})} \left(\frac{m'}{m}\right)^{(\alpha+1)} \left(\frac{s'}{s}\right)^{-1/2} \\ &\quad \cdot \exp\left(-\frac{1}{2\sigma_W^2 s'} (X - \mu_W m')^2 + \frac{1}{2\sigma_W^2 s} (X - \mu_W m)^2\right). \end{aligned}$$

APPENDIX F  
BOUNDED THE LIKELIHOOD

We compute the supremums  $\hat{m}_{k_1}$  and  $\hat{m}_{k_2}$  as solutions to

$$\frac{\partial p(X|m, s = (m + k_1)^2, \mu_W, \sigma_W)}{\partial m} \propto \sigma_W^2 m^2 + (X + \mu_W k_1) \mu_W m - (X + \mu_W k_1)^2 = 0$$

and

$$\frac{\partial p(X|m, s = (m - k_2)^2, \mu_W, \sigma_W)}{\partial m} \propto \sigma_W^2 m^2 + (X - \mu_W k_2) \mu_W m - (X - \mu_W k_2)^2 = 0,$$

obtaining

$$\hat{m}_{k_1} = \frac{-(X + \mu_W k_1)(\mu_W - \sqrt{\mu_W^2 + 4\sigma_W^2})}{2\sigma_W^2},$$

$$\hat{m}_{k_2} = \frac{-(X - \mu_W k_2)(\mu_W - \sqrt{\mu_W^2 + 4\sigma_W^2})}{2\sigma_W^2}.$$

In case the unconstrained  $\hat{m}_{k_1/k_2}$  values sit outside the constrained region, we take the  $m$  value at the quadratic boundary intersection with  $s = s_{\min}$ , i.e.  $m_{k_1} = \sqrt{s_{\min}} - k_1$  and  $m_{k_2} = \sqrt{s_{\min}} + k_2$ . The maximum along the line  $s = s_{\min}$  is obtained when  $\exp(-(X - \mu_W m)^2 / (2\sigma_W^2 s_{\min}))$  reaches its maximum value of one at  $m = X/\mu_W$  if this value of  $m$  sits on the boundary, otherwise we take the end-points  $m_{k_1} = \sqrt{s_{\min}} - k_1$  and  $m_{k_2} = \sqrt{s_{\min}} + k_2$  as before. Suitable values of  $k_1$ ,  $k_2$  and  $s_{\min}$  can be determined off-line for a range of  $\alpha$  values based on a very large set of simulated  $(m, s)$  pairs. For example,  $k_1 = k_2 = 20$ ,  $s_{\min} = 1.3$ , works well for  $\alpha = 1.5$ .

Moreover, bounding the likelihood allows for joint rejection sampling for  $\Gamma$  and  $\mathbf{R}$ , which might speed up the sampler.

ACKNOWLEDGMENT

The authors would like to thank J. P. Nolan, for generously granting a version of the STABLE Matlab Toolbox, used for the comparative analyses. We would also like to warmly acknowledge Omiros Papaspiliopoulos, Neil Shephard, Ralf Korn and Chris Rogers for illuminating discussions on this topic which aided our development, Simon Godsill acknowledges partial funding for the work from the EPSRC BTaRoT project EP/K020153/1, and Tatjana Lemke acknowledges PhD funding from Fraunhofer ITWM, Kaiserslautern.

REFERENCES

- [1] J. Ruanaidh and W. Fitzgerald, *Numerical Bayesian methods applied to signal processing*, ser. Statistics and Computing. Springer New York, 1996.
- [2] E. Kuruoğlu, C. Molina, S. Godsill, and W. Fitzgerald, "A new analytic representation for the symmetric  $\alpha$ -stable probability density function," in *Proceedings of the 5th World Meeting of the International Society for Bayesian Analysis (ISBA)*. ASA: American Statistical Association, 1997, pp. 229–233.
- [3] E. Kuruoğlu, C. Molina, and W. Fitzgerald, "Approximation of  $\alpha$ -stable probability densities using finite Gaussian mixtures," in *Eusipco: European signal processing conference*, 1998, pp. 989–992.
- [4] S. Godsill and E. Kuruoğlu, "Bayesian inference for time series with heavy-tailed symmetric  $\alpha$ -stable noise processes," in *Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, Washington DC, USA, June 1999.
- [5] S. Godsill, "MCMC and EM-based methods for inference in heavy-tailed processes with  $\alpha$ -stable innovations," in *In Proceedings of the IEEE Signal Processing Workshop on Higher-order Statistics*, Caesarea, Israel, June 1999, pp. 228–232.
- [6] M. Lombardi and S. Godsill, "On-line Bayesian estimation of AR signals in symmetric  $\alpha$ -stable noise," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 775–779, 2006.
- [7] G. Yang and S. Godsill, "Bayesian inference for continuous-time ARMA models driven by jump diffusions," in *Statistical Signal Processing, 2007. SSP '07. IEEE/SP 14th Workshop on*, Aug 2007, pp. 99–103.
- [8] T. Lemke and S. Godsill, "Enhanced Poisson sum representation for alpha-stable processes," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May 2011, pp. 4100–4103.
- [9] T. Lemke and S. J. Godsill, "Linear Gaussian computations for near-exact Bayesian Monte Carlo inference in skewed alpha-stable time series models," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, March 2012, pp. 3737–3740.
- [10] T. Lemke and S. J. Godsill, "A Poisson series approach to Bayesian Monte Carlo inference for skewed alpha-stable distributions," in *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014.
- [11] W. Schoutens, Ed., *Lévy Processes in Finance*. Wiley, 2003.
- [12] J. Nolan, "Modeling financial data with stable distributions," in *Handbook of Heavy Tailed Distributions in Finance*, S. Rachev, Ed. Elsevier-North Holland, Amsterdam, 2005.
- [13] S. Godsill and P. Rayner, *Digital Audio Restoration*. Springer, Berlin, 1998.
- [14] B. Gnedenko and A. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Cambridge, MA, 1968, second edition.
- [15] W. Feller, *An Introduction to Probability Theory and its Applications*. New York: Wiley, 1966, second edition.
- [16] J. Rosiński, "Tempering stable processes," *Stochastic Processes and Their Applications*, vol. 117, no. 6, pp. 677–707, 2007.
- [17] O. Barndorff-Nielsen and N. Shephard, "Normal modified stable processes," *Economics Papers*, vol. 72, 2001.
- [18] J. Chambers, C. Mallows, and B. Stuck, "A method for simulating stable random variables," *Journal of American Statistical Association*, vol. 71, pp. 340–344, 1976.
- [19] D. J. Buckle, "Bayesian inference for stable distributions," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 605–613, 1995.

- [20] Z. Quiou and N. Ravishanker, "Bayesian inference for time series with stable innovations," *Journal of Time Series Analysis*, vol. 19, no. 2, pp. 235–250, 1998.
- [21] E. Kuruoğlu, "Analytical representation for positive  $\alpha$ -stable densities," in *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 6, Hong Kong, 2003, pp. 729–732.
- [22] D. Gençağa, E. E. Kuruoğlu, A. Ertüzün, and S. Yildirim, "Estimation of time-varying autoregressive symmetric alpha stable processes by particle filters," *Signal Processing*, vol. 88, no. 10, pp. 2564–2572, 2008.
- [23] E. Tsionas, "Monte Carlo inference in econometric models with symmetric stable disturbances," *Journal of Econometrics*, vol. 88, pp. 365–401, 1999.
- [24] T. Lemke, *Poisson series approaches to Bayesian Monte Carlo inference for skewed  $\alpha$ -stable distributions and stochastic processes*. PhD Thesis, Dr. Hut, ISBN 978-3-8439-1454-3, 2014.
- [25] J. P. Nolan, "Maximum likelihood estimation of stable parameters," in *Lévy processes: Theory and applications*. Birkhäuser Boston, 2001, pp. 379–400.
- [26] I. A. Koutrouvelis, "Regression-type estimation of the parameters of stable laws," *Journal of the American Statistical Association*, vol. 75, no. 372, pp. pp. 918–928, 1980.
- [27] J. H. McCulloch, "Simple consistent estimators of stable distribution parameters," *Communications in Statistics - Simulation and Computation*, vol. 15, no. 4, pp. 1109–1136, 1986.
- [28] P. Troughton and S. Godsill, "A reversible jump sampler for autoregressive time series," in *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, Seattle, WA, USA, 1998, pp. 2257–2260.
- [29] S. Godsill, "On the relationship between MCMC model uncertainty methods," *Journal of Computational and Graphical Statistics*, vol. 10, no. 2, pp. 230–248, 2001.
- [30] D. Applebaum, *Lévy Processes and Stochastic Calculus*. Cambridge University Press, 2004.
- [31] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, 1994.
- [32] H. Bergström, "On some expansions of stable distribution functions," *Arkiv för Matematik*, vol. 2, no. 4, pp. 375–378, 1952.
- [33] A. Janicki and A. Weron, *Simulation and Chaotic Behavior of  $\alpha$ -stable Stochastic Processes*. CRC Press, 1993, vol. 178.
- [34] N. Azzaoui and L. Clavier, "Statistical channel model based on  $\alpha$ -stable random processes and application to the 60 ghz ultra wide band channel," *Communications, IEEE Transactions on*, vol. 58, no. 5, pp. 1457–1467, 2010.
- [35] M. B. Thompson, "A comparison of methods for computing autocorrelation time," *ArXiv e-prints*, Oct. 2010.
- [36] C. J. Geyer, "Practical Markov Chain Monte Carlo," *Statist. Sci.*, vol. 7, no. 4, pp. 473–483, 11 1992.



**Tatjana Lemke** received the Dipl.-Math. oec. degree in Business Mathematics majoring in Financial Mathematics, and the Ph.D. degree (Dr. rer. nat.) in Financial Mathematics from the Kaiserslautern University of Technology, Faculty of Mathematics (Germany), in 2010 and 2015, respectively. During her Ph.D. she visited the Signal Processing and Communications Laboratory in the Engineering Department at the University of Cambridge (UK) for a one and a half year research stay, supervised by Prof. Simon J. Godsill. The focus of her research is on the Bayesian Inference for skewed  $\alpha$ -stable distributions and stochastic processes. She is currently working in risk controlling at the Federal Republic of Germany - Finance Agency.



**Marina Riabiz** received the Bachelor's and Masters degree in Mathematical Engineering from Politecnico di Milano (Italy) in 2010 and 2013, respectively. She is currently working towards the Ph.D degree in Information Engineering, at the University of Cambridge (UK), in the Signal Processing and Communications Laboratory. She is also affiliated as a student at Clare Hall College, Cambridge. Her current research interests include Bayesian (parametric and non-parametric) inference methods for heavy tailed statistical distributions that model extreme events.



**Simon J. Godsill** is Professor of Statistical Signal Processing in the Engineering Department at Cambridge University. He is also a Professorial Fellow and tutor at Corpus Christi College, Cambridge. He coordinates an active research group in Signal Inference and its Applications within the Signal Processing and Communications Laboratory at Cambridge, specializing in Bayesian computational methodology, multiple object tracking, audio and music processing, and financial time series modelling. A particular methodological theme over recent years has been the development of novel techniques for optimal Bayesian filtering and smoothing, using Sequential Monte Carlo or Particle Filtering methods. Prof. Godsill has published extensively in journals, books and international conference proceedings, and has given a number of high profile invited and plenary addresses at conferences such as the Valencia conference on Bayesian Statistics (twice), the IEEE Statistical Signal Processing Workshop and the Conference on Bayesian Inference for Stochastic Processes (BISP) and the IEEE Workshop on Machine Learning in Signal Processing (2013). He co-authored a seminal Springer text Digital Audio Restoration with Prof. Peter Rayner in 1998. He was technical chair of the successful IEEE NSSPW workshop in 2006 on sequential and nonlinear filtering methods, and has been on the conference panel for numerous other conferences/workshops. Prof.

Godsill has served as Associate Editor for IEEE Tr. Signal Processing and the journal Bayesian Analysis. He was Theme Leader in Tracking and Reasoning over Time for the UKs Data and Information Fusion Defence Technology Centre (DIF-DTC) and Principal Investigator on many grants funded by the EU, EPSRC, QinetiQ, General Dynamics, MOD, Microsoft UK, Citibank, Mastercard, Google, DSO Singapore, Huawei and Jaguar Landrover. In 2009-10 he was co-organiser of an 18 month research program in Sequential Monte Carlo Methods at the SAMSI Institute in North Carolina and in 2014 he co-organised a research programme at the Isaac Newton Institute on Sequential Monte Carlo methods. In 2018 he will be General Chair of the FUSION Conference in Cambridge. Two of his journal papers have recently received Best Paper awards from the IEEE and IET. He continues to be a Director of CEDAR Audio Ltd. (which has received numerous accolades over the years, including a technical Oscar), and for which he was a founding staff member in 1988. The company has commercialised many of the ideas from Professor Godsills research over the years.