Bayesian Inference for Mixtures of Stable Distributions

Roberto Casarin [†] [‡]

CEREMADE

University Paris IX (Dauphine)

and

Dept. of Economics

University Ca' Foscari, Venice

Abstract

In many different fields such as hydrology, telecommunications, physics of condensed matter and finance, the gaussian model results unsatisfactory and reveals difficulties in fitting data with skewness, heavy tails and multimodality. The use of stable distributions allows for modelling skewness and heavy tails but gives rise to inferential problems related to the estimation of the stable distributions' parameters. Some recent works have proposed characteristic function based estimation method and MCMC simulation based estimation techniques like the MCMC-EM method and the Gibbs sampling method in a full Bayesian approach. The aim of this work is to generalise the stable distribution framework by introducing a model that accounts also for multimodality. In particular we introduce a stable mixture model and a suitable reparametrisation of the mixture, which allow us to make inference on the mixture parameters. We use a full Bayesian approach and MCMC simulation techniques for the estimation of the posterior distribution. Finally we propose some applications of stable mixtures to financial data.

KEYWORDS: Mixture model, Stable distributions, Bayesian inference, Gibbs sampling.

1 Introduction

In many different fields such as hydrology, telecommunications, physics and finance, Gaussian models reveal difficulties in fitting data that exhibits a high degree of heterogeneity; thus *stable distributions* have been introduced as a generalisation of the Gaussian model. Stable distributions

---

[†]I'm thankfull to professors C.P. Robert and M. Billio for the suggestions to my Phd. thesis. This work has been presented at the Young Statistician Meeting, Cambridge 14-15 April 2003.

[‡]Adress: CEREMADE, University Paris IX (Dauphine), 16 place du Maréchal de Lattre de Tassigny, 75775 Paris Cédex 16. E-mail: `casarin@ceremade.dauphine.fr`

allow also for infinite variance, skewness and heavy tails. The tails of a stable distribution decay like a power function, allowing extreme events to have higher probability mass than in Gaussian model. For a summary of the properties of the stable distributions see Zoloratev [42] and Samorodnitsky and Taqqu [36], which provide a good theoretical background on heavy-tailed distributions. The practical use of heavy-tailed distributions in many different fields is well documented in the book of Adler, Feldman and Taqqu [1], which also reviews the estimation techniques.

In finance, the first studies on the hypothesis of stable distributed stock prices can be attributed to Mandelbrot [22], Fama [13], [14] and Fama and Roll [15], [14]. They propose stable distributions and give some statistical instruments for the inference on the characteristic exponent. The use of stable distributions has been motivated also on the basis of empirical evidence from financial markets. Brenner [5] uses the notion of stationarity of the time series to explain stability of stock prices. An illuminating work on inference for stable distributions is due to Buckle [6], who makes also an empirical analysis on daily stock prices, using a full Bayesian approach to estimate stable distributions parameters and finding significant evidence of the stable distribution hypothesis.

There are many recent works treating the use of stable distributions in finance. For example see Bradley and Taqqu [4] and Mikosch [26] for an introduction to the use of stable distributions in financial risk modelling. The work of Mittnik, Rachev and Paolella [25] and of Rachev and Mittnik [31] provides a quite complete analysis of the theoretical and empirical aspects of the stable distributions in finance.

Other early studies, performing empirical analysis on stocks prices, suggest to use mixtures of distributions in order to modelling the financial markets heterogeneity. Barnes and Downes [2] use the same estimation techniques of Fama and Roll [16] in order to discuss the results of Teichmoeller [39]. They find that for some stock the property of stability does not hold and that the characteristic exponent varies across the stocks. In order to account for this kind of heterogeneity of the stock prices the authors suggest mixture of stable distributions as an alternative hypothesis. Simkowitz and Beedles [3] perform an empirical analysis focusing on the asymmetry of stock returns. They find that the skewness of the stock returns is frequently positive and dependent on the level of the characteristic exponent. They conclude that securities distributions may be better modelled through mixtures of stable distributions. Finally an extensive empirical analysis due to Fieltz and Rozelle [17] shows that mixtures of Gaussian, or non-Gaussian distributions can better describe stock prices. In particular the authors suggest to use non-Gaussian stable mixtures model with changing scale parameter because it directly accounts for skewness. We can conclude that the problem of multimodality and in general of heterogeneity is well documented in the financial literature, also from the earlier studies on the

stable distributions. Thus an appropriate modelling is needed.

Observe that, in order to account for heterogeneity and non-linear dependencies exhibited by the data, stable distributions have been already introduced in different kind of statistical models. For instance in *survival models*, the heterogeneity within survival times of a population are modelled trough common latent factors, which follow stable distributions, see for example Qiou, Ravishanker and Dey [29]. Stable distributions are also used to model heterogeneity over time. For an introduction to *time series models* with stable noises, see Qiou and Ravishanker [30] and Mikosch [26].

Different estimation methods for stable distributions have been proposed in the literature. For a full Bayesian approach see Buckle [6], for a maximum likelihood approach see DuMouchel [11] and for MCEM approach with application to time series with symmetric stable innovations see Godsill [21].

The first aim of our work is to propose a stable distributions mixture model in order to capture the heterogeneity of data. In particular we want to account for multimodality, which is present, for example, in financial data. The second goal of the work is to provide some inferential tools for stable distributions mixtures. As suggested in the literature on Gaussians mixtures (see for example Robert [34]), we propose a particular reparameterisations of the mixture model in order to make more easy the statistical inference on the mixture parameters. Furthermore we use both a full Bayesian approach and MCMC simulation techniques in order to estimate the parameters. The maximum likelihood approach (see for example McLachlan and Peel [23]) to the mixture model implies numerical difficulties, which rely on the fact that for many parametric density family the likelihood surface has singularities. Furthermore, as pointed out by Stephens [38], the likelihood may have several local maximum and it will be difficult to justify the choice of one of these point estimates. The presence of several local maximum and of singularities implies that the standard asymptotic theory for maximum likelihood estimation and the test theory do not apply in the mixture context. The Bayesian approach avoids these problem as parameters are random variables, with prior and posterior distributions defined on the parameter space. Thus it is no more necessary to choose between several local maximum, because point estimates are obtained by averaging over the parameter space, weighting by the posterior distribution of the parameters or by the simulated posterior distribution.

The structure of the work is as follows. Section 2 defines a stable distribution and the method to simulate from a stable. Section 3 provides an introduction to some basic Markov Chain Monte Carlo(MCMC) methods for mixtures and exhibits the Bayesian model and the Gibbs sampler for a stable distribution. Section 4 describes the Bayesian model for stable mixtures, with particular attention to the missing data structure of the stable mixture model and the Gibbs sampler for stable mixture is developed in the case where the number of components is fixed. Section

5 provides some results of the Bayesian stable mixture model on financial dataset. Section 6 concludes.

2 Simulating from a Stable Distribution

The existence of simulation methods for stable distributions opens the way to Bayesian inference on the parameters of this distribution family. In this section we define a stable random variable and briefly describe the method to simulate from a stable distribution, first proposed by Chamber, Mallows and Stuck [8] and then discussed also in Weron [41]. We use this method in our work, to generate dataset to test the efficiency of the MCMC based Bayesian inference approach. In the following we denote a stable distribution by $S_\alpha(\beta, \delta, \sigma)$. Stable distributions do not generally have an explicit probability density function and are thus conveniently defined through their characteristic function. The most well known parametrisation is defined in Samorodnitsky and Taqqu [36].

Definition 1 *(Stable distribution)*
A random variable X has stable distribution $S_\alpha(\beta, \delta, \sigma)$ if its parameters are in the following ranges: $\alpha \in (0, 2]$, $\beta \in [-1, +1]$, $\delta \in (-\infty, +\infty)$, $\sigma \in (0, +\infty)$ and if its characteristic function can be written as

$$E[exp(i\,\vartheta\,x)] = \begin{cases} exp(-|\sigma\vartheta|^\alpha)(1 - i\,\beta(sign(\vartheta))tan(\pi\alpha/2) + i\delta\vartheta) & if\ \alpha \neq 1; \\ exp(-|\sigma\vartheta|(1 + 2\,i\,\beta ln|\vartheta|sign(\vartheta)/\pi) + i\delta\vartheta) & if\ \alpha = 1. \end{cases} \quad (1)$$

where $\vartheta \in \mathbb{R}$.

The stable distribution is thus completely characterised through the following four parameters: the *characteristic exponent* α, the *skewness parameter* β, the *location parameter* δ and finally the *scale parameter* σ. An equivalent parametrisation is proposed by Zoloratev [42]. For a review on all the equivalent definitions of stable distribution and on all their properties see Samorodnitsky, Taqqu [36]. The distribution $S_\alpha(\beta, 0, 1)$ is usually called *standard stable* and when $\alpha \in (0, 1)$ it is called *positive stable* because the support of the density is the positive half of the real line. In this case the characteristic function reduces to

$$E\left[\exp(i\theta x)\right] = e^{-|\theta|^\alpha} \quad (2)$$

Stable distributions admit explicit representation of the density function only in the following cases: the Gaussian distribution $S_2(0, \sigma, \delta)$, the Cauchy distribution $S_1(0, \sigma, \delta)$ and the Lévy distribution $S_{1/2}(1, \sigma, \delta)$.

The algorithm we used for simulating a standard stable (see Chamber, Mallows and Stuck [8] and Weron [41]) is the following

4

(i) Simulate

$$V \sim U_{[-\frac{\pi}{2}, \frac{\pi}{2}]} \tag{3}$$

$$W \sim \varepsilon xp(1); \tag{4}$$

(ii) If $\alpha \neq 1$ then

$$Z = S_{\alpha,\beta} \frac{sin(\alpha(V + B_{\alpha,\beta}))}{cos(V)^{1/\alpha}} \left(\frac{cos(V - \alpha(V + B_{\alpha,\beta}))}{W} \right)^{\frac{1-\alpha}{\alpha}} \tag{5}$$

$$B_{\alpha,\beta} = \frac{arctan(\beta tan(\frac{\pi\alpha}{2}))}{\alpha}$$

$$S_{\alpha,\beta} = (1 + \beta^2 tan^2(\frac{\pi\alpha}{2}))^{\frac{1}{2\alpha}}$$

If $\alpha = 1$ then

$$Z = \frac{2}{\pi} \left[(\frac{\pi}{2} + \beta V) \tan(V) - \beta \log(\frac{W \cos(V)}{\frac{\pi}{2} + \beta V}) \right] \tag{6}$$

Once a value Z from a standard stable $S_\alpha(\beta, 0, 1)$ has been simulated, in order to obtain a value X from a stable distribution with scale parameter σ and location parameter δ, the following transformation is required

$$X = \begin{cases} Z + \delta & if \ \alpha \neq 1 \\ \sigma Z + \frac{2}{\pi} \beta \sigma \, log(\sigma) + \delta & if \ \alpha = 1 \end{cases}$$

Fig. 1 exhibits simulated data from a stable distribution. We use the parameters estimated by *Buckle* [6] on financial time series.

3 Bayesian Inference for Stable Distributions

In order to make inference on the parameters of a stable distribution in a Bayesian approach it is necessary to specify a *hierarchical model* on the parameters of the distribution. Often, the resulting posterior distribution of the Bayesian model cannot be calculated analytically, thus it is necessary to chose a numerical approximation method. Monte Carlo simulation techniques provide an appealing solution to the problem because, in high dimensional space, they are more efficient than traditional numerical integration methods and furthermore they require the densities involved in the posterior to be known only up to a normalising constant. In the following the basic Markov Chain Monte Carlo (*MCMC*) techniques will be introduced and the Gibbs sampler for a stable distribution will be discussed.
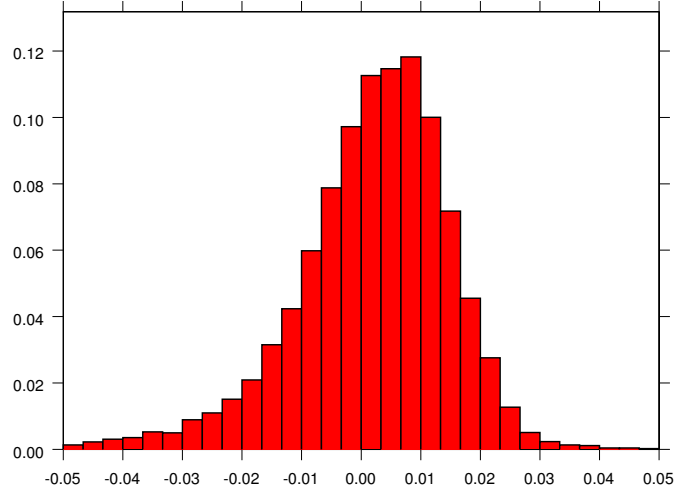
Figure 1: Simulation from stable distribution $S_{1.65}(-0.8, 0.00053, 0.0079)$.

3.1   MCMC Methods for Bayesian Models

As evidenced in Chapter **??**, in Bayesian inference many quantities of interest can be represented in the integral form

$$I(\theta) = \mathbb{E}_{\pi(\theta|\mathbf{x})}\{f(\theta)\} = \int_{\mathcal{X}} f(\theta)\pi(d\theta|\mathbf{x}) \tag{7}$$

where $\pi(\theta|\mathbf{x})$ is the posterior distribution of the parameter $\theta \in \mathcal{X}$ given the observed data $\mathbf{x} = (x_1, \ldots, x_k)$. In many cases to find an analytical solution to the integration problem is difficult and a numerical approximation is needed. A way to approximate the integral is to simulate the posterior distribution and to average the simulated values of $f(\theta)$. In particular the MCMC methods consist in the construction of a Markov chain $\left\{\theta^{(t)}\right\}_{t=1}^{n}$ and in the following approximation of the integral given in Eq. (7)

$$I_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} f(\theta^{(t)}) \tag{8}$$

which is a consistent estimator of the quantity of interest

$$I_n(\theta) \xrightarrow{a.s.} \mathbb{E}_{\pi(\theta|\mathbf{x})}\{f(\theta)\} \tag{9}$$

In some cases, as in mixture models, is not possible to simulate directly from the posterior

distribution and a further simulation step (*completion step*) is needed. All MCMC algorithms are based on the construction of a discrete time Markov Chain, through the specification of its transition kernel. Thus the properties of this kind of stochastic process are useful in order to study the convergence of the MCMC simulation algorithms.

We recall that the irreducibility of the chain is a sufficient condition in order to guarantee the convergence of I_n to the quantity of interest given in Eq. (7).

Theorem 1 *(Law of Large Numbers)*
If the Markov chain $\{\Theta\}_{t=0}^{\infty}$ is irreducible and has σ-finite invariant measure π, then $\forall f, g \in L^1(\pi)$, with $\int g(\theta) d\pi(\theta) \neq 0$

$$\lim_{n \to \infty} \frac{I_n(f)}{I_n(g)} = \frac{\int f(\theta) d\pi(\theta)}{\int g(\theta) d\pi(\theta)} \tag{10}$$

where $I_n(h) = \frac{1}{n} \sum_{i=1}^{n} h(\Theta_i)$.

For a brief introduction to Markov chains and to Markov Chain Monte Carlo methods we refer to Chapter **??**. Further details on Markov chains can be found for example in Meyn and Tweedie [24], other theoretical results on convergence are in Tierney [40], finally Robert and Casella [35] provides some techniques for monitoring convergence.

3.2 The Gibbs Sampler

The Gibbs sampler has been introduced in image processing by Geman and Geman [19] (see also Chapter **??** for a general introduction to MCMC methods and to Gibbs sampling) and it is a method of construction of a Markov Chain $\{\Theta^{(t)}\}_{t=0}^{\infty}$ with multivariate stationary distribution $\pi(\theta|(x))$, where $\theta \in \chi$. This simulation method is particularly useful when the posterior density is defined on a high dimension space. If the random vector θ can be written as $\theta = (\theta_1, \ldots, \theta_p)$ and if we can simulate from the full conditional densities

$$(\theta_i | \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p) \sim \pi_i(\theta_i | \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p, \mathbf{x}) \tag{11}$$

then the associated Gibbs sampling algorithm is given by the following transition kernel from $\theta^{(t)}$ to $\theta^{(t+1)}$:

Definition 2 *(Gibbs Sampler)*
Given the state $\Theta^{(t)} = \theta^{(t)}$ at time t, generate the state $\Theta^{(t+1)}$ as follows

   1. $\Theta_1^{(t+1)} \sim \pi(\theta_1 | \theta_2^{(t)}, \ldots, \theta_p^{(t)}, \mathbf{x})$

   2. $\Theta_2^{(t+1)} \sim \pi(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \ldots, \theta_p^{(t)}, \mathbf{x})$

3. ...

4. $\Theta_p^{(t+1)} \sim \pi(\theta_p|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{p-1}^{(t+1)}, \mathbf{x})$

Under some regularity conditions the Markov chain produced by the algorithm converges to the desired stationary distribution (see Robert and Casella [35]).

3.3 The Gibbs Sampler for Univariate Stable Distributions

In this paragraph we give a description of the Gibbs sampler proposed by Buckle [6] in order to estimate the characteristic exponent α of a stable distribution. It is known (see Section 1) how to simulate values from a stable distribution; furthermore it is possible to represent the stable density in integral form, by introducing an auxiliary variable y, as suggested by Buckle [6]. The stable density is obtained by integrating with respect to y the bivariate density function of the pair (x, y)

$$f(x,y|\alpha,\beta,\sigma,\delta) = \frac{\alpha}{|\alpha-1|}\exp\left\{-\left|\frac{z}{\tau_{\alpha,\beta}(y)}\right|^{\alpha/(\alpha-1)}\right\}\left|\frac{z}{\tau_{\alpha,\beta}(y)}\right|^{\alpha/(\alpha-1)}\frac{1}{|z|} \tag{12}$$

$$(x,y) \in (-\infty,0) \times (-1/2, l_{\alpha,\beta}) \cup (0,\infty) \times (l_{\alpha,\beta}, 1/2) \tag{13}$$

$$\tau_{\alpha,\beta}(y) = \frac{\sin(\pi\alpha y + \eta_{\alpha,\beta})}{\cos(\pi y)}\left[\frac{\cos(\pi y)}{\cos(\pi(\alpha-1)y) + \eta_{\alpha,\beta}}\right]^{(\alpha-1)/\alpha} \tag{14}$$

$$\eta_{\alpha,\beta} = \beta\,min(\alpha, 2-\alpha)\pi/2 \tag{15}$$

$$l_{\alpha,\beta} = -\eta_{\alpha,\beta}/\pi\alpha \tag{16}$$

where $z = \frac{x-\delta}{\sigma}$. Previous elements allow to perform simulation based Bayesian inference on the parameters of the stable distribution. The Bayesian model is described through the *Directed Acyclic Graph* (DAG) in Fig. 2. Suppose to observe n realizations $\mathbf{x} = (x_1, \dots, x_n)$ from a stable distribution $S_\alpha(\beta, \sigma, \delta)$ and simulate a vector of auxiliary variables $\mathbf{y} = (y_1, \dots, y_n)$, then the *completed likelihood* and the completed posterior distribution are respectively

$$L(\mathbf{x}, \mathbf{y}|\theta) = \prod_{i=1}^{n} f(x_i, y_i|\theta) \tag{17}$$

$$\pi(\theta|\mathbf{x}, \mathbf{y}) = \frac{L(\mathbf{x}, \mathbf{y}|\theta)\pi(\theta)}{\int_\Theta L(\mathbf{x}, \mathbf{y}|\theta)\pi(\theta)d\theta} \propto \prod_{i=1}^{n} f(x_i, y_i|\theta)\pi(\theta) \tag{18}$$

where $\theta = (\alpha, \beta, \delta, \sigma)$ is the stable parameter vector varying in the parameter space Θ.

In the following we suppose to observe n values from a standard stable distribution $S_\alpha(\beta, 1, 0)$ and we assume the other parameters to be known. Parameters α and β are estimated by simulating
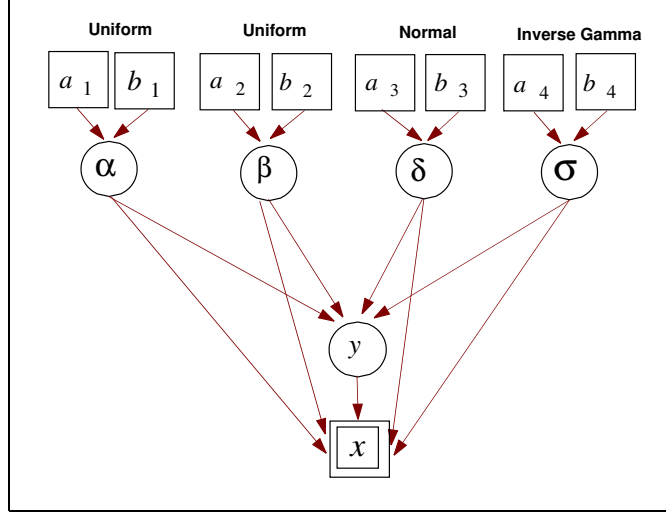
Figure 2: DAG of the Bayesian model for inference on stable distributions. It exhibits the hierarchical structure of priors, and hyperparameters. A single box around a quantity indicates that it is a known constant, a double box indicates the variable is observed and a circle indicates the random variable is not observable. The directed arrows show the dependence structure of the model. We use the the prior suggested by Buckle [6].

from the complete posterior distribution and by averaging simulated values. Simulations from the posterior distribution are obtained by iterating the following steps of the Gibbs sampler

(i) *Update the completing variable*

$$\pi(y_i|\alpha, \beta, \delta, \sigma, z_i) \propto \exp\left\{1 - \left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\right\} \left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}} \tag{19}$$

with $i = 1, \ldots, n$.

(ii) *Simulate from the complete full conditional distributions*

$$\pi(\alpha|\beta, \delta, \sigma, \mathbf{x}, \mathbf{y}) \propto \frac{\alpha^n}{|\alpha-1|^n}\exp\left\{-\sum_{i=1}^n\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\prod_{i=1}^n\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\pi(\alpha) \tag{20}$$

$$\pi(\beta|\alpha, \delta, \sigma, z_i, y_i) \propto \exp\left\{-\sum_{i=1}^n\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\prod_{i=1}^n\left|\frac{1}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\pi(\beta) \tag{21}$$

$$\pi(\delta|\alpha,\beta,\sigma,z_i,y_i) \propto \exp\left\{-\sum_{i=1}^{n}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\prod_{i=1}^{n}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\frac{1}{|x_i-\delta|}\pi(\delta) \qquad (22)$$

$$\pi(\sigma|\alpha,\beta,\sigma,z_i,y_i) \propto \left(\frac{1}{\sigma^{\alpha/(\alpha-1)}}\right)^n\exp\left\{-\frac{1}{\sigma^{\alpha/(\alpha-1)}}\sum_{i=1}^{n}\left|\frac{(x_i-\delta)}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)}\right\}\pi(\sigma) \qquad (23)$$

where $\pi(\alpha)$, $\pi(\beta)$, $\pi(\delta)$, $\pi(\sigma)$ are the prior distributions on the parameters, $\mathbf{y}$ is a vector of auxiliary variables $(y_1,\ldots,y_n)$ and $\tau_{\alpha,\beta}$ is a function of y defined in Eq. (14).

In order to simulate from the density function given in equation (19) we apply the accept reject method (see Devroye [9]), because the density is proportional to a function which has finite support $(-\frac{1}{2},\frac{1}{2})$ and which is bounded with value 1 at the maximum y^*, where y^* is such that $\tau_{\alpha,\beta}(y^*) = x$. To emphasize numerical problems which arise in making inference on stable distributions, we plot in Fig. 4 the density function of y for different values of x. Note that for all values of $\alpha \in (0,1)$, high values of x make the density function spiked around the mode. Thus the basic accept method performs quite poorly. A way to improve the simulation method is to build a histogram with the rejected values and to use it as an envelope in the accept reject algorithm.

Due to the way the parameter α enters in the likelihood, the densities given in Equations (20), (21), (22) and (23) are undulating and rather concentrated, therefore as suggested by Buckle [6] and Ravishanker and Qiou [30] we introduce the following reparametrisations which give a more manageable form of the conditional posteriors of α, β and δ

$$v_i = \tau_{\alpha,\beta}(y_i) \qquad (24)$$

$$\phi_i = \frac{\tau_{\alpha,\beta}}{x_i-\delta} \qquad (25)$$

the resulting posteriors are

$$\pi(\alpha|\beta,\delta,\sigma,\mathbf{x},\mathbf{v}) \propto \frac{\alpha^n}{|\alpha-1|^n}\exp\left\{-\sum_{i=1}^{n}\left|\frac{z_i}{v_i}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\prod_{i=1}^{n}\left|\frac{z_i}{v_i}\right|^{\frac{\alpha}{(\alpha-1)}}\left|\frac{d\tau_{\alpha,\beta}}{dy}\right|^{-1}_{\tau_{\alpha,\beta}(y)=v_i}\pi(\alpha) \qquad (26)$$

$$\pi(\beta|\alpha,\delta,\sigma,z_i,v_i) = \prod_{i=1}^{n}\left|\frac{d\tau_{\alpha,\beta}}{dy}\right|^{-1}_{\tau_{\alpha,\beta}(y)=v_i}\pi(\beta) \qquad (27)$$

$$\pi(\delta|\alpha,\beta,\sigma,z_i,v_i) \propto \prod_{i=1}^{n}\left|\frac{d\tau_{\alpha,\beta}}{dy}\right|^{-1}_{\tau_{\alpha,\beta}(y)=\phi_i(x_i-\delta)}\pi(\delta) \qquad (28)$$

At each step of the reparametrised Gibbs sampler, the Jacobian of the transformation, $\left|\frac{d\tau_{\alpha,\beta}}{dy}\right|^{-1}$, must be evaluated in $y_i = \tau_{\alpha,\beta}^{-1}(v_i)$. Due to the complexity of the function $\tau_{\alpha,\beta}$, its inverse has not

an analytical expression. Therefore, following Buckle [6], the inverse transformation is determined numerically. We use the modified safeguard Newton algorithm proposed in Press et al. [28].

In order to simulate from the posteriors given in Equations (26), (27) and (28) we use Metropolis-Hastings algorithm (see Chapter **??** for an introduction to Monte Carlo Markov Chain methods)

Given $\theta_i^{(k-1)}$

(***i***) *Generate from the proposal distribution* $\theta_i^* \sim q(\theta_i | \theta_i^{(k-1)})$

(***ii***) *Take:*

$$
\theta_i^{(k)} = \begin{cases} \theta_i^* & \text{with probability} \quad \rho(\theta_i^{(k-1)}, \theta_i^*) \\ \theta_i^{(k-1)} & \text{with probability} \quad 1 - \rho(\theta_i^{(k-1)}, \theta_i^*) \end{cases}
$$

where: $\rho(\theta_i^{(k-1)}, \theta_i^*) = 1 \wedge \left\{ \dfrac{\pi(\theta_i^* | \theta_{-i})}{\pi(\theta_i^{(k-1)} | \theta_{-i})} \dfrac{q(\theta_i^{(k-1)})}{q(\theta_i^*)} \right\}.$

In order to simulate the full conditional posteriors given in Eq. (26) and (27), we use a beta distribution, $\mathcal{Be}(a, b)$ as proposal. The sample generated from the beta distribution is not independent because in order to simulate the k-th value of the M.-H. chain, we pose the mean of the beta distribution to be equal to the $(k-1)$-th value of the chain. In setting a and b, the parameters of the proposal distribution we distinguish the following cases

$$
\begin{cases} a = \dfrac{\alpha_{k-1}^2 (1 - \alpha_{k-1}) - v\, \alpha_{k-1}}{v} \\ b = a \dfrac{1 - \alpha_{k-1}}{\alpha_{k-1}} \end{cases} \qquad \text{when} \quad \alpha \in [0, 1)
$$

and

$$
\begin{cases} a = \dfrac{(\alpha_{k-1} - 1)^2 (2 - \alpha_{k-1}) - v\, (\alpha_{k-1} - 1)}{v} \\ b = a \dfrac{2 - \alpha_{k-1}}{(\alpha_{k-1} - 1)} \end{cases} \qquad \text{when} \quad \alpha \in (1, 2]
$$

where α_{k-1} is the value generated by the Metropolis-Hastings chain at step $(k-1)$ and v is the variance of the proposal distribution. This parameters choice allows us also to avoid numerical problems related to the evaluation of the Metropolis-Hastings acceptance ratio in the presence of fat tailed and quite spiked likelihood functions.

We use a gaussian random walk proposal to simulate the full conditional posterior of the location parameter (Eq. (28)).

In order to complete the description of the hierarchical model and of the associated Gibbs sampler, we consider the following joint prior distribution

$$\mathbb{I}(\alpha)_{(1,2)} \frac{1}{2} \mathbb{I}(\beta)_{[-1,1]} \frac{1}{\sqrt{2\pi b_3}} e^{-\frac{(\delta - a_3)^2}{2b_3}} \mathbb{I}(\delta)_{(-\infty, +\infty)} \frac{b_4^{a_4}}{\Gamma(a_4)} \frac{e^{-\frac{b_4}{\theta}}}{\theta^{a_4+1}} \mathbb{I}(\theta)_{[0,\infty)} \tag{29}$$

where $\theta = \sigma^{\frac{\alpha}{\alpha-1}}$. We use informative priors for the location and scale parameters. For δ we assume a normal distribution. Note that the prior distribution of θ is the *inverse gamma* distribution $IG(a_4, b_4)$, which is a conjugate prior of the distribution given in equation (23). Simulations of the parameter σ can be obtained from the simulated values of θ by a simple transformation. Finally for parameters α and β we assume non informative priors.

We show the efficiency of the MCMC based Bayesian inference, running the Gibbs sampler on simulated dataset. In the following examples we discuss the numerical results and also some computational remarks related to different values of the characteristic exponent.

Example 1 - *(Positive α-Stable Distributions)*

Fig. 5, in *Appendix A*, exhibits the dataset of 1,000 observations generated from the positive stable distribution $S_{0.5}(0.9, 0, 1)$. Figures 6, 7 give the results of 15,000 iterations of the Gibbs sampler relative to the characteristic exponent α, and the asymmetry parameter β, for an increasing number of simulations. Figures 8, 9 exhibit the ergodic averages. Note that convergence has been achieved after 6,000 iterations. In the first experiment we use uniform priors for α and β. We set the variance of the proposal distribution (see Eq. 46) at $v = 0.0001$ for α, at $v = 0.0009$ for β and the starting values of the Gibbs sampler at $\alpha = 0.2$ and $\beta = 0.5$. Figures 10 and 11 exhibit the estimated acceptance rate for the M.-H. steps in the Gibbs sampler, for an increasing number of simulations. The histograms of the posteriors are depicted in figures 12 and 13. Note that through the Gibbs sample it is also possible to obtain the confidence intervals for the estimated parameters and to perform a goodness of fit test.

Example 2 - *(α-Stable Distributions with $\alpha > 1$)*

Note that in the first dataset α is less than 1, therefore the moment of order less than two are infinite. In some applications like in finance, in order to give an interpretation to the result it is preferable to work at least with finite first order moments. Therefore we verify the efficiency of the Gibbs sampler also on a sample generated from a stable distribution with $\alpha \in (1, 2]$. Fig. 14 in *Appendix A* exhibits the dataset of 1,000 values simulated from the stable distribution $S_{1.5}(0.9, 0, 1)$. Gibbs sampler realisations and ergodic averages are represented respectively in figures 15, 16 and in figures 17, 18. Convergence has been achieved after 6,000 iterations. In the experiment we use uniform priors for α and β and set the variance of the proposal distribution

at $v = 0.0005$ for α, at $v = 0.0009$ for β and the starting values of the Gibbs sampler at $\alpha = 1.2$ and $\beta = 0.2$. Figures 19, 20 exhibit the estimated acceptance rate, for an increasing number of simulations. Finally, posterior distributions are depicted in figures 21 and 22.

For each dataset, Table 1 summarizes the estimated parameters, the standard deviations and the estimated acceptance rates of the M.-H. steps of the Gibbs sampler. Results are obtained on a PC with Intel *1063 MHz* processor, using routines implemented in *C/C++*. We validate the MCMC code by checking that without any data the estimated joint posterior distribution correspond to the joint prior.

Table 1: Numerical results - Ergodic Averages over 15,000 Gibbs realisations.

	First Dataset:		$S_{0.5}(0.9, 0, 1)$		
Parameter	True Value	Starting Value	Estimate[*]	Std.Dev.	Acc. Rate
α	0.5	0.2	0.54	0.02	0.11
β	0.9	0.5	0.95	0.07	0.14
	Second Dataset:		$S_{1.5}(0.9, 0, 1)$		
Parameter	True Value	Starting Value	Estimate[**]	Std.Dev.	Acc. Rate
α	1.5	1.2	1.46	0.04	0.49
β	0.9	0.2	0.96	0.04	0.21

[*]*Time (sec): 4259*
[**]*Time (sec): 4128*

4   Bayesian Inference for Mixtures of Stable Distributions

In this section we extend the Bayesian framework, introduced in the previous section, to the mixtures of stable distributions. In many situations data may exhibit simultaneously: heavy tails, skewness, and multimodality. In time series analysis, the multimodality of the empirical distribution can also find a justification in a heterogeneous time evolution of the observed phenomena. For example, the distribution of financial time series like prices or prices volatility may have many modes because the stochastic process evolves over time following different regimes. Stable distributions allow for skewness and heavy tails, but not for multimodality. Thus a way to model these features of the data, is to introduce stable mixtures. Furthermore the use of stable mixtures is appealing also because they have normal mixtures as special case, which is a widely studied topic (see for example Stephens [38], Richardson and Green [32]). Other relevant works on the Bayesian approach to the mixture models estimation are Diebolt and Robert [10], Escobar

and West [12] and Robert [34], [33]. In *Appendix C* some examples of two components stable mixtures are exhibited. We simulate stable mixtures with different parameters setting, in order to understand the influence of each parameter on the shape of the mixture's distribution.

4.1   The Missing Data Model

In the following we define a stable mixture model, while assuming to known the number of mixture components. Under a practical point of view the number of components may be detected by looking at the number of modes in the distribution or by performing a statistical test, see section 5. Let L be the finite number of mixture components and $f(x|\alpha_l, \beta_l, \delta_l, \sigma_l)$ the l-th stable distribution in the mixture, then the mixture model $m(x|\theta, p)$ is

$$m(x|\theta, p) = \sum_{l=1}^{L} p_l f(x|\theta_l) \tag{30}$$

$$\text{with} \quad \sum_{l=1}^{L} p_l = 1, \quad p_l \geq 0, \quad l = 1, \ldots, L$$

where $\theta_l = (\alpha_l, \beta_l, \delta_l, \sigma_l)$, $l = 1, \ldots, L$ are the parameter vector and $\theta = (\theta_1, \ldots, \theta_L)$. In the following we suppose L to be known. In order to perform Bayesian inference two steps of completion are needed. First, we adopt the same completion technique used for stable distributions. The auxiliary variable, y, is introduced in order to obtain an integral representation of the mixture distribution

$$m(x|\theta, p) = \sum_{l=1}^{L} p_l \int_{-1/2}^{1/2} f(x, y|\theta_l) \, dy. \tag{31}$$

The second step of completion is introduced in order to reduce the complexity problem, which arises in simulation based inference for mixtures. The completing variable (or *allocation variable*), $\nu = \{\nu_1, \ldots, \nu_L\}$ is defined as follow

$$\nu_l = \begin{cases} 1 & if \ x \sim f(x, y|\theta_l) \\ 0 & otherwise \end{cases} \quad \text{with } l = 1, \ldots, L \tag{32}$$

and is used to select the mixture component. The allocation variable is not observable and this missing data structure can be estimated by following a simulation based approach. Simulations from the mixture model can be performed in two step: first, simulating the allocation variable; second, simulating a mixture component conditionally on the allocation variable. The resulting demarginalized mixture model is
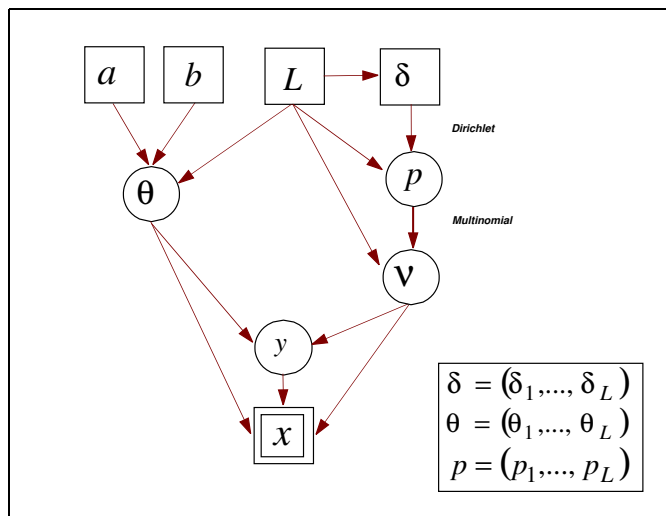
Figure 3: DAG of the Bayesian hierarchical model for inference on stable mixtures. Note that the completing variable ν is not observable. Thus, two levels of completion, y and ν, are needed for a stable mixture model.

$$m(x,\nu|\theta,p) = \prod_{l=1}^{L}\left(p_l \int_{-1/2}^{1/2} f(x,y|\theta_l)\,dy\right)^{\nu_l}, \qquad \sum_{l=1}^{L}\nu_l = 1 \tag{33}$$

This completion strategy is now quite popular in Bayesian inference for mixtures (see Robert [33], Robert and Casella [35], Escobar and West [12] and Diebolt and Robert [10]). For an introduction to Monte Carlo methods in Bayesian inference from data modeled by mixture of distributions see also Neal [27] and for a discussion of the numerical and identifiability problems in mixtures inference see Richardson and Green [32], Stephens [37] and Celeux, Hurn and Robert [7].

4.2 The Bayesian Approach

The Bayesian model for inference on stable mixtures is represented through the DAG in Fig. 3. Before specifying the Bayesian model we introduce two distributions that are quite useful in Bayesian inference form mixtures: the *multinomial distribution* and the *Dirichlet distribution*.

Definition 3 *(Multinomial distribution)*
The random variable $X = (x_1,\ldots,x_L)$ has L dimensional multinomial distribution if its density function is

$$\mathcal{M}(n,p_1,...,p_L) = \begin{pmatrix} n \\ x_1\ldots x_L \end{pmatrix} p_1^{x_1}\cdot\ldots\cdot p_L^{x_L}\mathbb{I}_{\sum x_l = n} \tag{34}$$

where $\sum_{l=1}^{L} p_l = 1$ and $p_l \geq 0$, $l = 1,\ldots,L$.

As suggested in the literature on gaussian mixtures, in the following we assume a multinomial prior distribution for the completing variable ν: $V \sim f_V(\nu) = \mathcal{M}_L(1, p_1, \ldots, p_L)$.

Definition 4 *(Dirichlet distribution)*

The random variable $X = (x_1, \ldots, x_L)$ has L dimensional Dirichlet distribution if its density function is

$$\mathcal{D}(\delta_1, ..., \delta_L) = \frac{\Gamma(\delta_1 + ... + \delta_L)}{\Gamma(\delta_1) \cdot ... \cdot \Gamma(\delta_L)} x_1^{\delta_1 - 1} \cdot ... \cdot x_L^{\delta_L - 1} \mathbb{I}_S(x) \tag{35}$$

where $\delta_l \geq 0, l = 1, \ldots, L$ and $S = \{x = (x_1, \ldots, x_L) \in \mathbb{R}^L | \sum x_l = 1, x_l > 0 \ l = 1, \ldots, L\}$ is the simplex of $\mathbb{R}^L$.

We assume that the parameters of the discrete part of the mixture distribution has the standard conjugate Dirichlet prior: $(p_1, \ldots, p_L) \sim \mathcal{D}_L(\delta_1, \ldots, \delta_L)$, with hyperparameters $\delta_1 = \ldots = \delta_L = \frac{1}{L}$.

Observing n independent values, $\mathbf{x} = (x_1, \ldots, x_n)$, from a stable mixture, the likelihood and the completed likelihood are respectively

$$L(\mathbf{x}, \mathbf{y}|\theta, p) = \prod_{i=1}^{n} \sum_{l=1}^{L} p_l \int_{-1/2}^{1/2} f(x_i, y_i|\theta_l) dy_i \tag{36}$$

$$L(\mathbf{x}, \mathbf{y}, \nu|\theta, p) = \prod_{i=1}^{n} \prod_{l=1}^{L} (p_l f(x_i, y_i|\theta_l))^{\nu_{il}} \tag{37}$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ and $\nu = (\nu_1, \ldots, \nu_n)$ are respectively the auxiliary variable and the allocation variable vectors and $\theta = (\theta_1, \ldots, \theta_L)$ and $p = (p_1, \ldots, p_L)$ are the mixture's parameters vectors. From the completed likelihood and from the priors it follows that the complete posterior distribution of the Bayesian mixture model is:

$$\pi(\theta, p|\mathbf{x}, \mathbf{y}, \nu) \propto \prod_{i=1}^{n} \left\{ \prod_{l=1}^{L} (f(x_i, y_i|\theta_l))^{\nu_{il}} \pi(\nu_i) \right\} \pi(\theta) \pi(p). \tag{38}$$

Bayesian inference on the mixture parameters requires the calculation of the expected value from the posterior distribution. A closed form solution of this integration problem does not exist, thus numerical methods are needed. The introduction of auxiliary variables, that are not observable, simplifies inference for mixtures and also suggests the way to approximate numerically the problem. In fact the auxiliary variables can be replaced by simulated values and the simulated completed likelihood can be used for calculating the posterior distributions. Furthermore in order to approximate numerically the posterior means is necessary to perform simulations from the posterior distributions of the parameters and to average the simulated values.

4.3 The Gibbs Sampler for Mixtures of Stable Distributions

Gibbs sampling allows us to simulate from the posterior distribution avoiding computational difficulties due to the dimension of the parameter vector. Due to the ergodicity of the Markov chain generated by the Gibbs sampler, the choice of the initial values is arbitrary. In particular we choose to simulate them from the prior. The steps of the Gibbs sampler for a mixture model can be grouped in: simulation of the full conditional distributions and augmentation by the completing variables

(i) *Simulate initial values:* $\nu_i^{(0)}, y_i^{(0)}, \quad i = 1, \ldots, n$ *and* $p^{(0)}$ *respectively from*

$$\nu_i^{(0)} \sim \mathcal{M}_L(1, p_1, \ldots, p_L) \tag{39}$$

$$y_i^{(0)} \sim f(y_i|\theta, \nu, x_i) \propto \exp\{1 - \left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)}\} \left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)} \tag{40}$$

$$p^{(0)} \sim \mathcal{D}_L(\delta, \ldots, \delta). \tag{41}$$

(ii) *Simulate from the full conditional posterior distributions*

$$\pi(\theta_l|\theta_{-l}, p, \mathbf{x}, \mathbf{y}, \mathbf{v}) \propto \prod_{i=1}^{n} \{f(x_i, y_i|\theta_l)\, p_l\}^{\nu_{il}}\, \pi(\theta_l) \quad l = 1, \ldots, L \tag{42}$$

$$\pi(p_1, \ldots, p_L|\theta, \mathbf{x}, \mathbf{y}, \mathbf{v}) = \mathcal{D}(\delta + n_1(\nu), \ldots, \delta + n_L(\nu)) \tag{43}$$

(iii) *Update the completing variables*

$$\pi(y_i|\theta, p, \mathbf{x}, \mathbf{y}_{-i}, \mathbf{v}) \propto \exp\left\{1 - \left|\frac{x_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)}\right\} \left|\frac{x_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)} \tag{44}$$

$$\pi(\nu_i|\theta, p, \mathbf{x}, \mathbf{y}, \mathbf{v}_{-i}) = \mathcal{M}_L(1, p_1^*, \ldots, p_L^*) \tag{45}$$

for $i = 1, \ldots, n$, where

$$z = \frac{x - \delta}{\sigma}$$

$$n_l(\nu) = \sum_{i=1}^{n} \nu_{il}, \quad l = 1, \ldots, L$$

$$p_l^* = \frac{p_l f(x_i, y_i|\theta_l)}{\sum_{l=1}^{L} f(x_i, y_i|\theta_l) p_l}, \quad l = 1, \ldots, L.$$

Steps (43) and (45) of the Gibbs sampler are proved in *Appendix D*. Observe that simulations from the conditional posterior distribution of Eq.(43) can be obtained by running the Gibbs sampler given in equations (20)-(23), conditionally to the value of the completing variable ν. To simulate

from the Dirichlet posterior distribution given in Eq. (43), we use the algorithm proposed by Casella and Robert [35], while to draw value from the multinomial posterior distribution of Eq. (45), we use the algorithm proposed by Fishman [18].

In Examples 4.3 and 4.3, we verify the efficiency of the Gibbs sampler on some test samples simulated from stable mixtures. For each mixture's component we assume the joint prior distribution given in equation (29). Furthermore, for the shake of simplicity, we consider $L = 2$. Because of the quite irregular form of the density $f(x_i, y_i|\theta_l)$, during the MCMC based estimation, some computational difficulties were encountered in evaluating the probability p_l^* of each mixture's component. Thus we introduce the following useful reparameterisation and approximation

$$
q_l^* = \begin{cases} \frac{e^{g_1(x_i,y_i)-g_2(x_i,y_i)}}{e^{g_1(x_i,y_i)-g_2(x_i,y_i)}+1} & \text{if} \quad g_1 > g_2 \quad \text{or} \quad g_1 \leq g_2 \\ 1 & \text{if} \quad g_1 \gg g_2 \\ 0 & \text{otherwise} \end{cases}
$$

where $g_l(x_i, y_i) = log(p_l\, f(x_i, y_i|\theta_l))$, with $l = 1, 2$.

Table 2: Numerical results - Ergodic Averages over 15,000 Gibbs realisations.

	Dataset: $0.5S_{1.7}(0.3,1,1) + 0.5S_{1.3}(0.5,30,1)$				
Par.	True Value	Starting Value	Estimate[*]	Std.Dev.	Acc. Rate
α_1	1.7	1.9	1.66	0.09	0.32
α_2	1.3	1.9	1.36	0.07	0.41
β_1	0.3	0.8	0.28	0.09	0.41
β_2	0.5	0.8	0.37	0.10	0.42
p_1	0.5	0.4	0.52	0.02	-
	Dataset: $0.5S_{1.3}(0.3,1,1) + 0.5S_{1.3}(0.8,30,1)$				
Par.	True Value	Starting Value	Estimate[**]	Std.Dev.	Acc. Rate
α	1.3	1.7	1.25	0.08	0.23
β_1	0.3	0.5	0.15	0.03	0.11
β_2	0.8	0.5	0.95	0.05	0.13
p_1	0.5	0.5	0.75	0.09	-
p_2	0.5	0.5	0.25	0.09	-

(*)*Time (sec):9249*

(**)*Time (sec):9525*

18

Example 3 - *(α-Stable Mixture with varying β and α)*

In this example, we apply the Gibbs sampler to a synthetic dataset of 1,000 observations generated from the stable mixture: $0.5S_{1.7}(0.3, 1, 1) + 0.5S_{1.3}(0.5, 30, 1)$. Fig. 29 in *Appendix E* exhibits the dataset. In the M.-H. step of the Gibbs sampler, we set v=0.0001 for β and v=0.005 for α. Result are briefly represented in Table 4.3 and graphically described in Figg. 30-33. Note that the presence in the mixture model, of distribution with different tails behaviour, produces some problems in the convergence of the ergodic averages, due to the label switching of the observations.

Example 4 - *(α-Stable Mixture with constant α and varying β)*

In this experiment we keep α fixed over the mixture components. First we generate 1,000 observations from the following mixture $0.5S_{1.3}(0.3, 1, 1) + 0.5S_{1.3}(0.8, 30, 1)$. Secondly we apply the Gibbs sampler for stable mixture and obtain the results given in Table 4.3. The graphical description of the results is in Figg. 35-38 of *Appendix E* and exhibits a more appreciable mixing of the chain associated to the Gibbs sampler.

To conclude this section, we remark that in developing the Gibbs sampler for α-Stable mixtures and also in previous Monte Carlo experiments the number of components of the mixture is assumed to be known. Thus our research framework can be extended in order to make inference on the number of components. For example, *Reversible Jump MCMC* (RJMCMC) or *Birth and Death MCMC* (BDMCMC) could be applied in this context.

5   Application to Financial Data

Introducing two level of auxiliary variables in the stable mixture model allows us to infer all the parameters of mixture from the data. Gaussian distribution is usually assumed in modelling financial time series, but it performs poorly when data are heavy-tailed and skewed. Moreover the assumption of unimodal distribution becomes too restrictive for some financial time series. In this section, we illustrate how stable mixtures may result particularly useful in modelling different kind of financial variables and present estimates obtained with the MCMC based inferential technique proposed in the previous section.

Example 5 - *(Stock Market)*

In this example we analyse the return rate of the S&P500 composite index from 01/01/1990 to 27/01/2003. The return on the index is defined as: $r_t = (p_t - p_{t-1})/p_{t-1}$. Alternatively, logarithmic returns could be used. The number of observations is 3410. Fig. 39 shows the data histogram and the best normal which is possible to estimate. The QQ-plot in Fig. 40 reveals that data are not normally distributed. We apply the Gibbs sampler for α-Stable mixtures to this dataset. The result is in Tab. 5. Parameter estimates are ergodic averages over the last 10,000 values of the 15,000 Gibbs sampler realisations. Note that index return distribution has tails heavier than Gaussian, because $\hat{\alpha} = 1.674$.

Table 3: Parameter Estimates on S&P 500 Index daily returns.

	Starting Value	Estimate	Std.Dev.	Acc. Rate
α	1.8	1.674	0.005	0.2
β	0.2	0.159	0.004	0.1
σ	0.01	0.070	0.002	-
δ	0.0001	0.000091	0.0001	0.1

Example 6 - *(Bond Market)*

Our second dataset (source: *DataStream*) contains daily price returns on the J.P. Morgan's indexes concerning following countries: France, Germany, Italy, United Kingdom, USA and Japan, between 01/01/1988 and 13/01/2003. Denoting with p_t the price index at time t. The return on the index is defined as: $r_t = (p_t - p_{t-1})/p_{t-1}$. Fig.41 in *Appendix E* exhibits jointly the histogram, the best Gaussian approximation and the density line of the returns distribution. All time series exhibit a certain degree of kurtosis and skewness. Estimation result on the J.P. Morgan Great Britain index is in Tab. 5.

Table 4: Parameter Estimates on JP Morgan - Great Britain index daily returns.

	Starting Value	Estimate	Std.Dev.	Acc. Rate
α	1.5	1.95	0.004	0.2
β	0.02	0.013	0.001	0.2
σ	0.01	0.270	0.003	-
δ	0.005	0.0062	0.02	0.1

Example 7 - *(3 Months Euro-Deposits Interest Rate)*
Our third dataset (source: *DataStream*) contains daily 3-Months Interest Rates on Euro-Deposits concerning following countries: France, Germany, Italy, United Kingdom, USA and Japan, between 01/01/1988 and 13/01/2003. Fig.43 in *Appendix E* exhibits jointly the histogram, the best Gaussian approximation and the density line of the returns distribution. Quite all time series of this dataset exhibit multimodality.
Estimation result on the 3-month Interest Rate for France is in Tab. 5

Table 5: Parameter Estimates on Interest Rates - France. Two components α-stable mixture.

	Starting Value	Estimate	Std.Dev.	Acc. Rate
$\alpha_1,\ \alpha_2$	1.5	1.2	0.003	0.15
β_1	0.01	0.02	0.001	0.1
β_2	0.01	0.04	0.001	0.1
σ_1	1.5	0.307	0.002	-
σ_2	1.5	0.873	0.001	-
δ_1	4	3.012	0.02	0.1
δ_2	10	7.301	0.03	0.1

6 Conclusion

In this work we propose a α-Stable mixture model. As result in the literature from many empirical evidences, α-Stable distributions are particularly adapted for modelling financial variables. Moreover some financial empirical studies evidence that mixture models are needed in many cases, due to the presence of multi-modality in the asset returns distribution. We chose Bayesian inference due to the flexibility of the approach, that allows to simultaneously estimate all the parameters of the model. Furthermore we introduce a suitable reparameterisation of the α-Stable mixture in order to perform Bayesian inference. The proposed approach to α-Stable mixture models estimation is quite general and worked well in our simulation analysis, but it needs much more evaluation, with a particular attention to the case of symmetric stable mixtures. Furthermore the Bayesian approach used in this work allows to perform goodness of fit tests and also to use RJMCMC and BDMCMC techniques in order to make inference on the number of components of the mixture.

Appendix A - Bayesian Inference for Stable Distributions
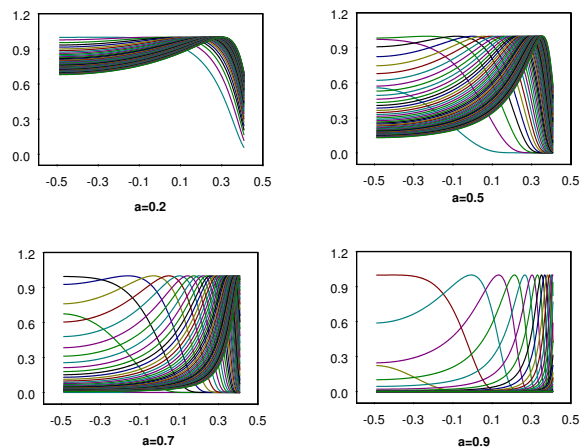
Figure 4: The density function of y for different values of x (0.04,...,0.14) and of α (0.2,...,0.9).
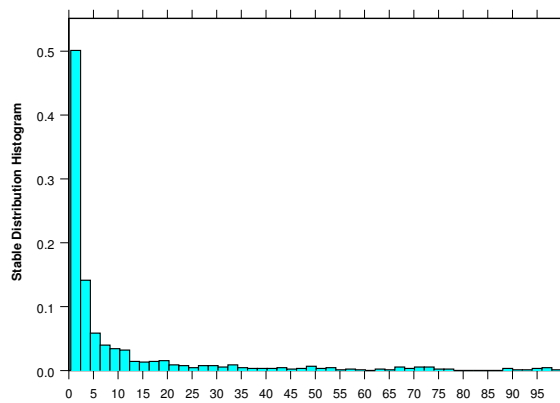
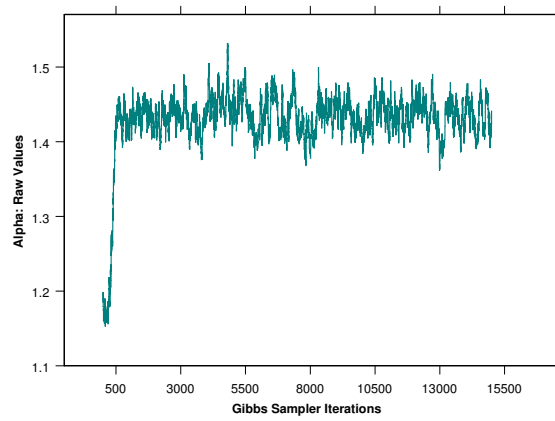Figure 5: Simulated dataset, 1,000 values from $S_{0.5}(0.9, 0, 1)$.
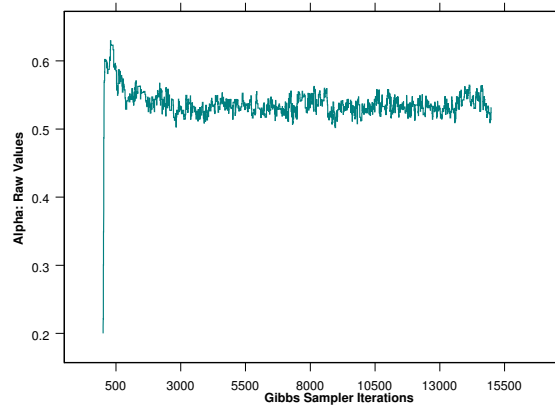
Figure 6: Gibbs sampler realisations, characteristic exponent α.
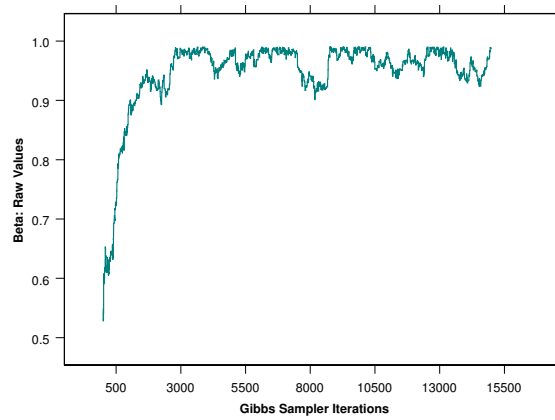
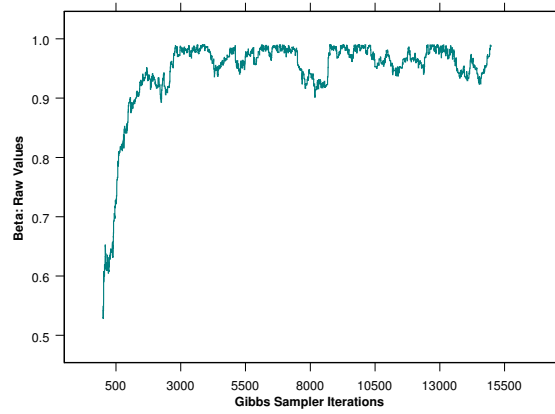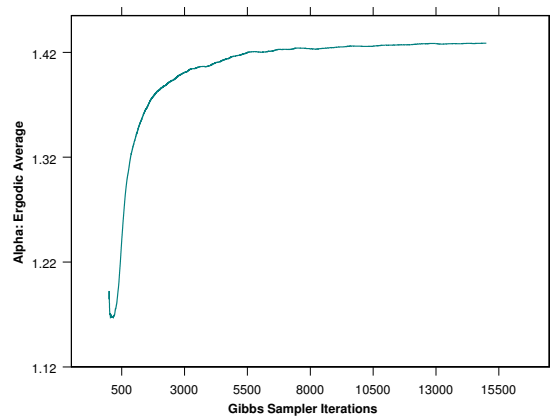Figure 7: Gibbs sampler realisations, asymmetry parameter β.
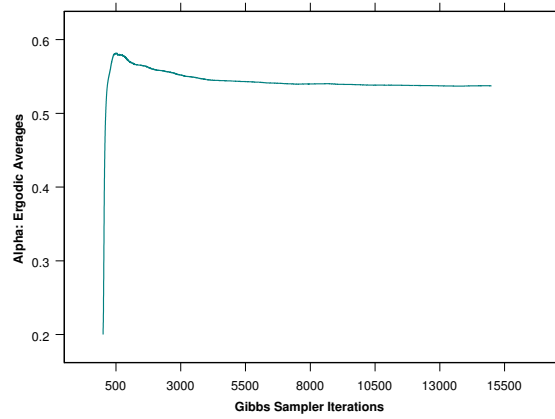
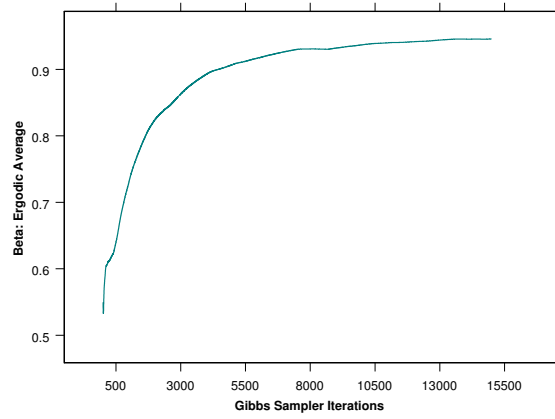Figure 8: Gibbs sampler, ergodic averages for the characteristic exponent α.
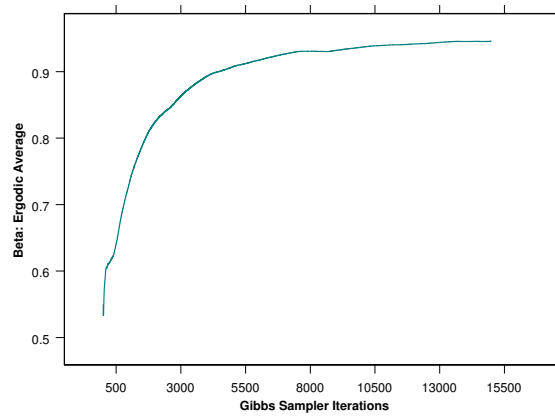
Figure 9: Gibbs sampler, ergodic averages for the asymmetry parameter β.
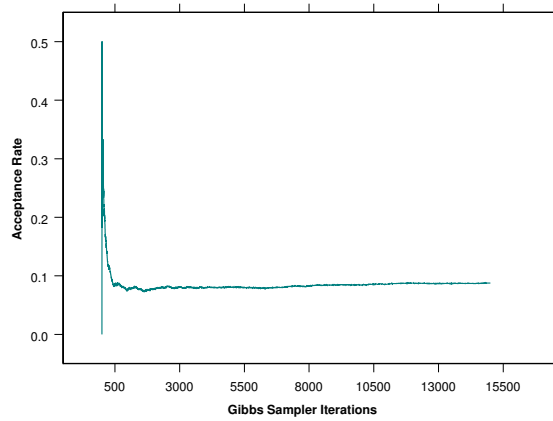
Figure 10: Gibbs sampler, acceptance rate of the M.-H. step, parameter α
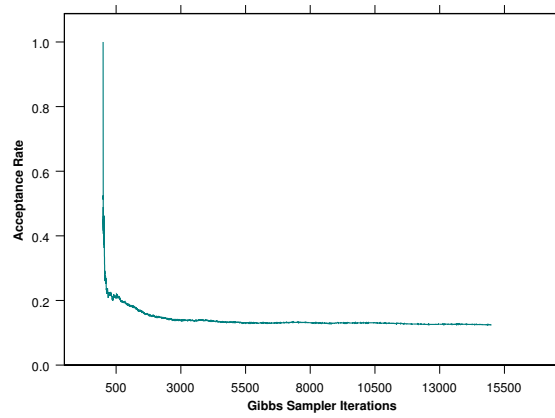
Figure 11: Gibbs sampler, acceptance rate of the M.-H. step, parameter β
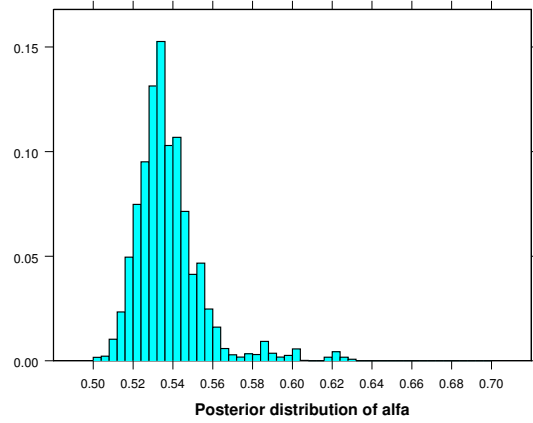
Figure 12: Gibbs sampler, histogram of the posterior $\pi(\alpha|\mathbf{x})$
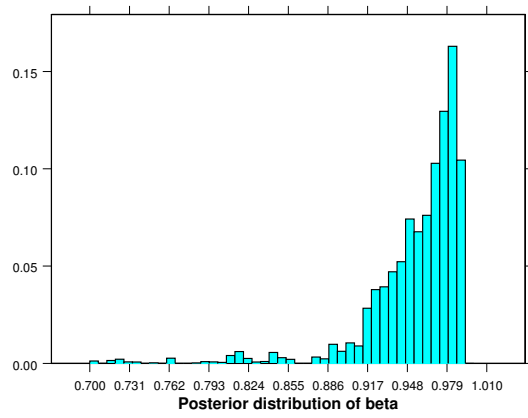
Figure 13: Gibbs sampler, histogram of the posterior $\pi(\beta|\mathbf{x})$

26
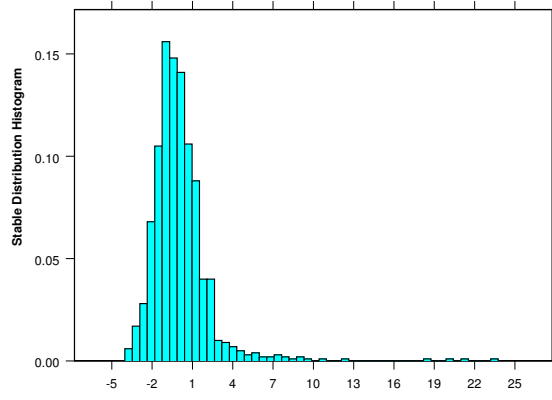
Figure 14: Simulated dataset, 1,000 values from $S_{1.5}(0.9, 0, 1)$.

Figure 15: Gibbs sampler realisations, characteristic exponent α.

27

Figure 16: Gibbs sampler realisations, asymmetry parameter β.

Figure 17: Gibbs sampler, ergodic averages for the characteristic exponent α.

Figure 18: Gibbs sampler, ergodic averages for the asymmetry parameter β.
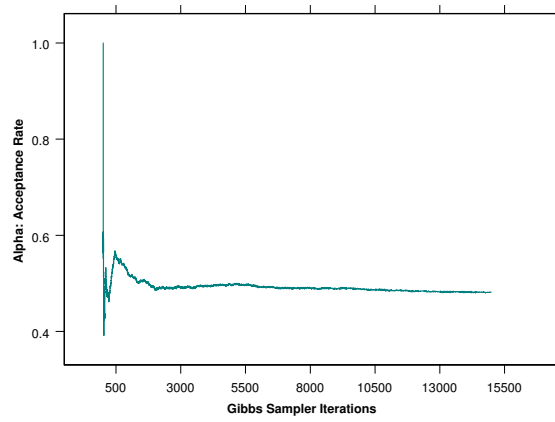
Figure 19: Gibbs sampler, acceptance rate of the M.-H. step, parameter α
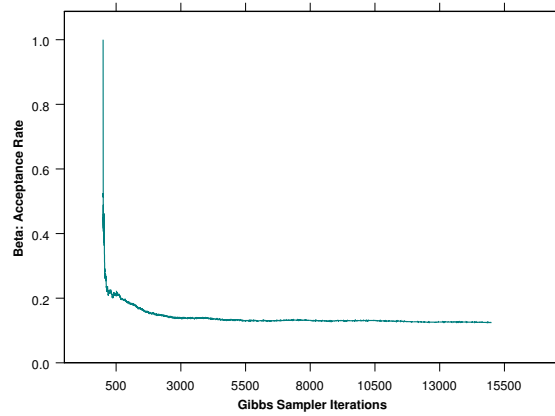
Figure 20: Gibbs sampler, acceptance rate of the M.-H. step, parameter β
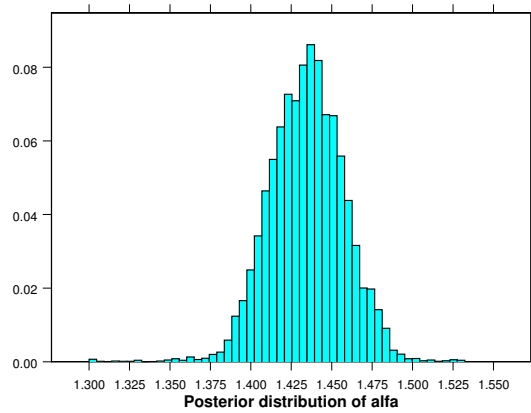
Figure 21: Gibbs sampler, histogram of the posterior $\pi(\alpha|\mathbf{x})$
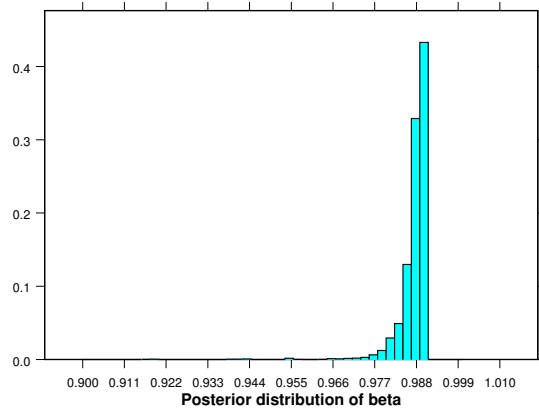
Figure 22: Gibbs sampler, histogram of the posterior $\pi(\beta|\mathbf{x})$

Appendix B - Proposal Distributions for the Metropolis-Hastings Algorithm

The shape of the stable distribution and the presence of skewness suggest us to use a Beta distribution $\mathcal{B}e(a,b)$ as proposal for the Metropolis-Hastings algorithm

$$\alpha|\alpha_{k-1} \sim \mathcal{B}e(a,b) = \frac{1}{B(a,b)}\alpha^{a-1}(1-\alpha)^{b-1}\mathbb{I}(\alpha)_{(0,1)}. \tag{46}$$

We assume that the mean of the distribution is equal to the $(k-1)$-th value of the M.-H. chain and set exogenously the variance equal to v. Through the parameter v it is thus possible to control the acceptance rate of the Metropolis-Hastings algorithm. When $\alpha \in (0,1)$ the value of the parameters of the proposal is

$$\begin{cases} \frac{a}{a+b} = \alpha_{k-1} \\ \frac{ab}{(a+b)^2(1+a+b)} = v \end{cases} \Leftrightarrow \begin{cases} a = \frac{\alpha_{k-1}^2(1-\alpha_{k-1})-v\,\alpha_{k-1}}{v} \\ b = a\frac{1-\alpha_{k-1}}{\alpha_{k-1}} \end{cases}$$

where α_{k-1} is the $(k-1)$-th value of the M.-H. chain and v is the variance. In addition to the previous system of equations, also the positivity constraint on the Beta's parameters: $a > 0$ and $b > 0$ must hold. Thus at each iteration of the M.-H. algorithm the following constraint must be satisfied

$$\alpha_{k-1} \in \left(\frac{3-v}{2} - \frac{\sqrt{v^2-8v+1}}{2}, \frac{3-v}{2} + \frac{\sqrt{v^2-8v+1}}{2}\right). \tag{47}$$

When $\alpha \in (1,2]$ we use a translated Beta distribution

$$\alpha|\alpha_{k-1} \sim \mathcal{B}e(a,b) = \frac{1}{B(a,b)}(\alpha-1)^{a-1}(2-\alpha)^{b-1}\mathbb{I}(\alpha)_{(1,2)}. \tag{48}$$

By imposing the usual constraints on the mean and the variance we obtain the values of the proposal's parameters

$$\begin{cases} \frac{2a+b}{a+b} = \alpha_{k-1} \\ \frac{ab}{(a+b)^2(1+a+b)} = v \end{cases} \Leftrightarrow \begin{cases} a = \frac{(\alpha_{k-1}-1)^2(2-\alpha_{k-1})-v(\alpha_{k-1}-1)}{v} \\ b = a\frac{2-\alpha_{k-1}}{(\alpha_{k-1}-1)} \end{cases}$$

Also in this case the positivity constraints on the Beta's parameters must be considered. We proceed in a similar way for the proposal distribution of the skewness parameter β.

Appendix C - Mixtures of Stable Distributions

C.1 Mixtures with varying α

Observe that in all dataset exhibited in the histograms, N=100,000 values from right skewed ($\beta = 1$) standard stables have been simulated.
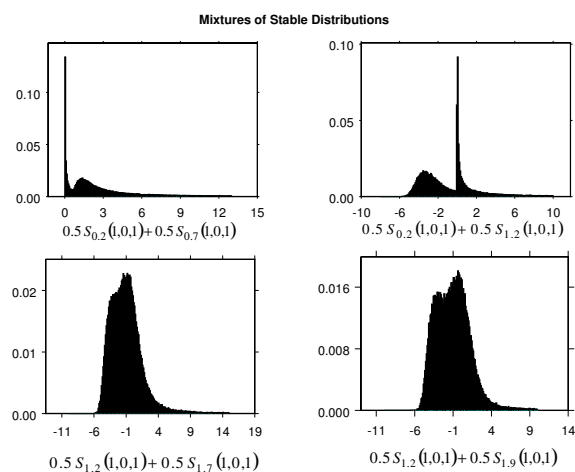

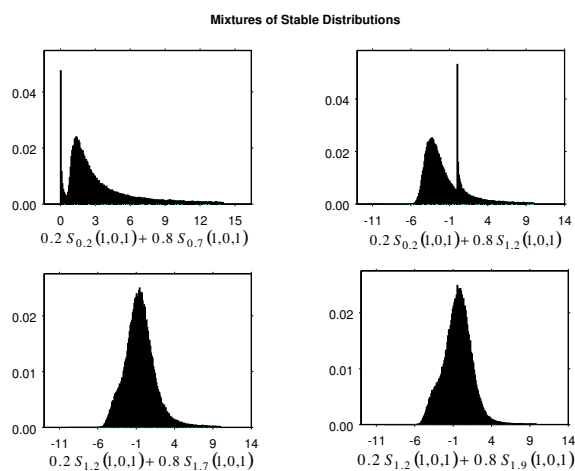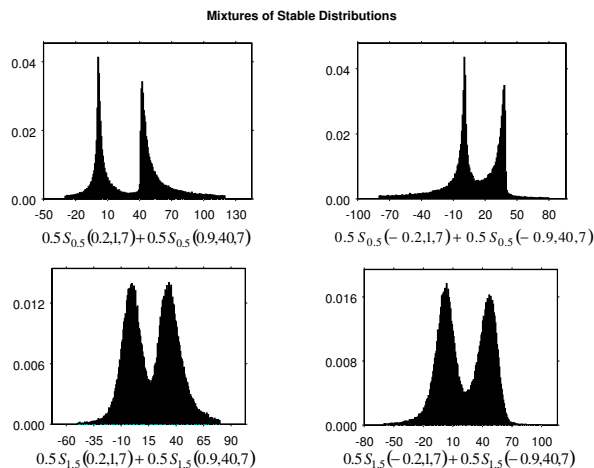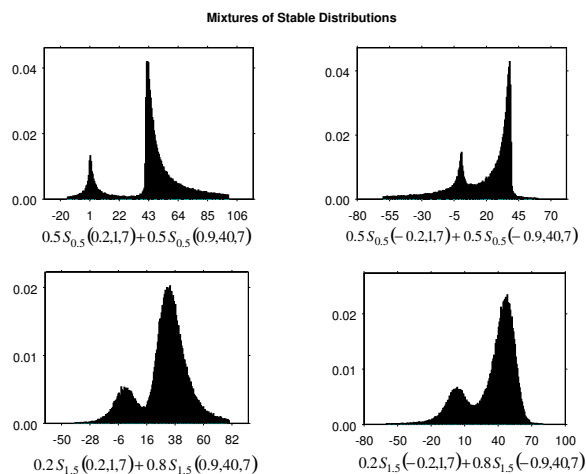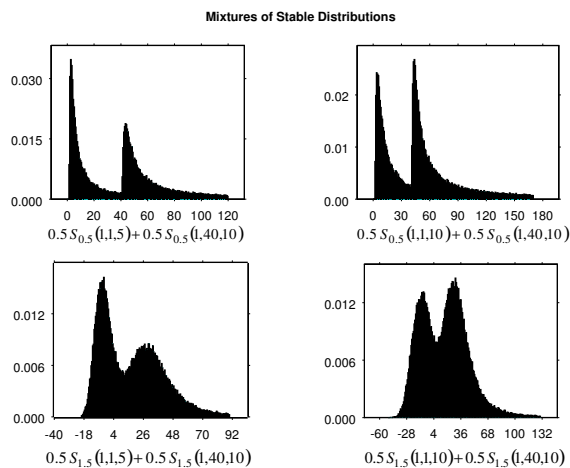Figure 23: Stables mixtures, with equally weighted components.

Figure 24: Stables mixtures, with unequally weighted components.

C.2 Mixtures with varying β

Simulated samples of $N = 100,000$ stable values are exhibited in the following histograms. In all the samples the location and the scale parameters of the mixture are: $\delta_1 = 1$, $\delta_2 = 40$, $\sigma_1 = \sigma_2 = 7$.

Mixtures of Stable Distributions

Figure 25: Stables mixtures, with equally weighted components.

Mixtures of Stable Distributions

Figure 26: Stables mixtures, with unequally weighted components.

C.3 Mixtures with varying σ

Simulated samples of N=100,000 stable values are exhibited in the following histograms. In all the samples the location and the skewness parameters of the mixture are: $\delta_1 = 1$, $\delta_2 = 40$, $\beta_1 = \beta_2 = 1$.

Figure 27: Standard stables mixtures, with equally weighted components.
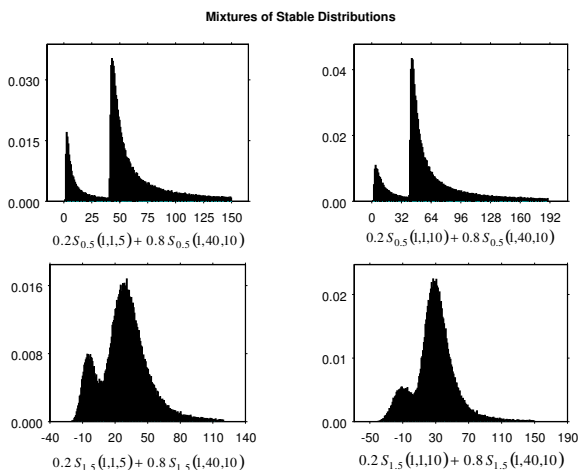
Figure 28: Standard stables mixtures, with unequally weighted components.

Appendix D - The Gibbs Sampler for a Stable Distributions Mixture

Proof *(Allocation Probabilities Posterior Distribution)*
The posterior distribution of probabilities $(p_1, \ldots, p_L)$, given in Eq. (43), is a Dirichlet and is derived in the following

$$
\begin{aligned}
\pi(p_1, \ldots, p_L | \theta, \mathbf{x}, \mathbf{y}, \nu) &= \frac{L(\mathbf{x}, \mathbf{y}, \nu | \theta, p)\pi(\theta)\pi(p)}{\int L(\mathbf{x}, \mathbf{y}, \nu | \theta, p)\pi(\theta)\pi(p)dp} = \\
&= \frac{\prod_{i=1}^{n} \prod_{l=1}^{L} \left(p_l f(x_i, y_i | \theta_l)\right)^{\nu_{il}} \pi(\theta)\pi(p)}{\int \prod_{i=1}^{n} \prod_{l=1}^{L} \left(p_l f(x_i, y_i | \theta_l)\right)^{\nu_{il}} \pi(\theta)\pi(p)dp} = \\
&= \frac{\prod_{i=1}^{n} \prod_{l=1}^{L} p_l^{\nu_{il}} \pi(p)}{\int \prod_{i=1}^{n} \prod_{l=1}^{L} p_l^{\nu_{il}} \frac{\Gamma(\delta)\cdots\Gamma(\delta)}{\Gamma(L\,\delta)} p_1^{\delta-1}, \cdots, p_L^{\delta-1} dp_1, \cdots, dp_L} = \\
&= \frac{\prod_{i=1}^{n} \prod_{l=1}^{L} p_l^{\nu_{il}} \pi(p)}{\int \prod_{i=1}^{n} \frac{\Gamma(\delta)\cdots\Gamma(\delta)}{\Gamma(L\,\delta)} \prod_{l=1}^{L} p_l^{\nu_{il}+\delta-1} dp_1 \cdots dp_L} = \\
&= \frac{\prod_{i=1}^{n} \prod_{l=1}^{L} p_l^{\nu_{il}} \pi(p)}{\int \prod_{l=1}^{L} p_l^{\sum_{i=1}^{n} \nu_{il}+\delta-1} \frac{\Gamma(\delta)\cdots\Gamma(\delta)}{\Gamma(L\,\delta)} dp_1, \cdots, dp_L} = \\
&= \frac{\Gamma\left(\delta + \sum_{i=1}^{n}\nu_{i1}\right)\cdots\Gamma\left(\delta + \sum_{i=1}^{n}\nu_{iL}\right)}{\Gamma\left(\delta\,L + \sum_{i=1}^{n}(\nu_{i1}+\ldots+\nu_{iL})\right)} p_1^{\sum_{i=1}^{n}\nu_{i1}+\delta-1} \cdots p_L^{\sum_{i=1}^{n}\nu_{iL}+\delta-1} = \\
&= \mathcal{D}_L\left(\sum_{i=1}^{n}\nu_{i1} + \delta, \ldots, \sum_{i=1}^{n}\nu_{iL} + \delta\right) = \\
&= \mathcal{D}_L(n_1(\nu) + \delta, \ldots, n_L(\nu) + \delta)
\end{aligned}
\tag{49}
$$

where $n_l(\nu) = \sum_{i=1}^{n} \nu_{il}$, with $l = 1, \ldots, L$.

$\square$

Proof *(Allocation Variables Posterior Distribution)*
The posterior distribution of the allocation variables given in Eq.(45) is a Multinomial and follows from

$$\pi(\nu_1, \ldots, \nu_n | \theta, p, \mathbf{x}, \mathbf{y}) = \frac{L(\mathbf{x}, \mathbf{y}, \nu | \theta, p)\pi(\theta)\pi(p)}{\int L(\mathbf{x}, \mathbf{y}, \nu | \theta, p)\pi(\theta)\pi(p)d\nu} =$$

$$= \frac{\prod_{i=1}^{n} \left\{ \prod_{l=1}^{L}(f(x_i, y_i | \theta_l))^{\nu_{il}} \prod_{l=1}^{L} p_l^{\nu_{il}} \right\} \pi(\theta)\pi(p)}{\int \prod_{i=1}^{n} \prod_{l=1}^{L}(f(x_i, y_i | \theta_l))^{\nu_{il}} \prod_{l=1}^{L} p_l^{\nu_{il}} \pi(\theta)\pi(p)d\nu} =$$

$$= \prod_{i=1}^{n} \frac{\prod_{l=1}^{L}(f(x_i, y_i | \theta_l)p_l)^{\nu_{il}}}{\int \prod_{l=1}^{L}(f(x_i, y_i | \theta_l)p_l)^{\nu_{il}} d\nu_i} = \tag{50}$$

$$= \prod_{i=1}^{n} \prod_{l=1}^{L} \left(\frac{f(x_i, y_i | \theta_l)p_l}{\sum_{l=1}^{L} f(x_i, y_i | \theta_l)p_l} \right)^{\nu_{il}} =$$

$$= \prod_{i=1}^{n} \mathcal{M}_L(1, p_1^*, \ldots, p_L^*)$$

where $p_l^* = \frac{f(x_i, y_i | \theta_l)p_l}{\sum_{l=1}^{L} f(x_i, y_i | \theta_l)p_l}$, for $l = 1, \ldots, L$.

$\square$

Appendix E - Bayesian Inference for Stable Distributions Mixtures

E.1 Mixtures with varying α and β
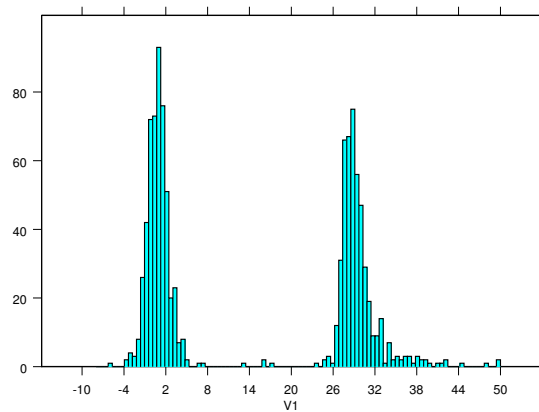
Figure 29: Simulated dataset, 1,000 values from $0.5S_{1.7}(0.3, 1, 1) + 0.5S_{1.3}(0.5, 30, 1)$
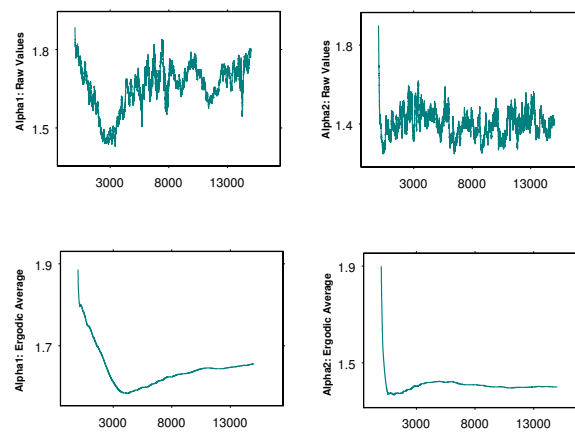
Figure 30: Gibbs sampler realisations and ergodic averages for α_1 and α_2.
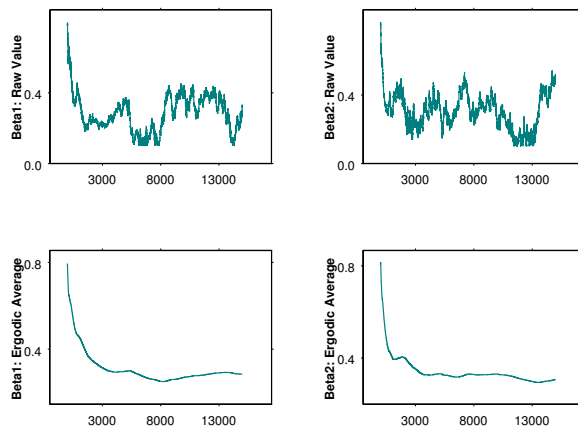
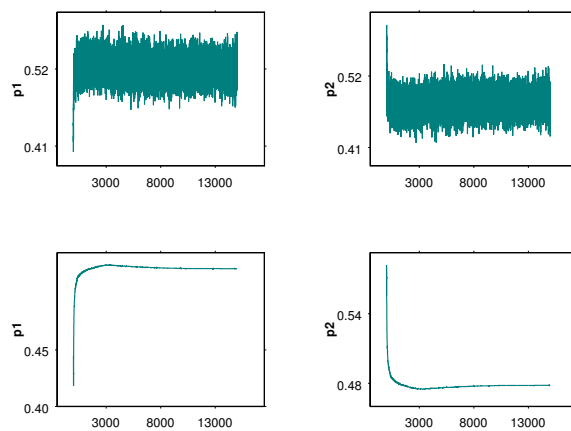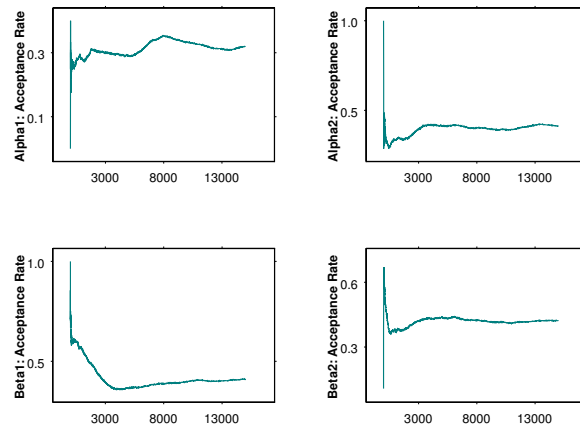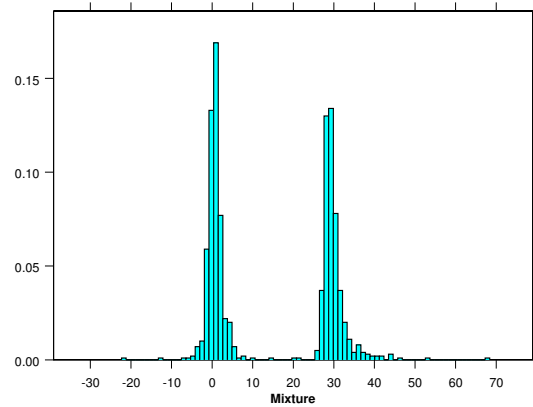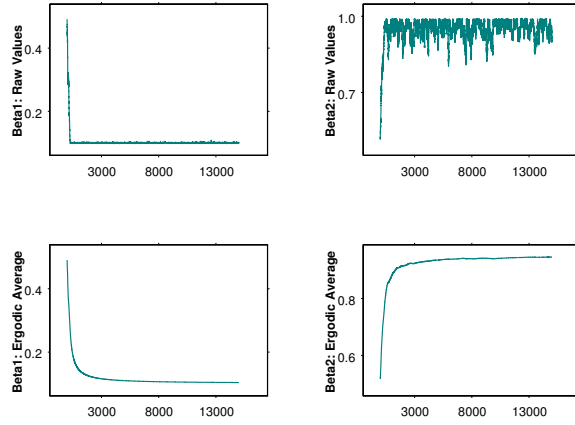Figure 31: Gibbs sampler realisations and ergodic averages for β_1 and β_2.

Figure 32: Gibbs sampler realisations and ergodic averages for p_1 and p_2.

Figure 33: Acceptance rates for α_1, α_2, β_1 and β_2

E.2 Mixtures with fixed α and varying β

Figure 34: Simulated dataset, 1,000 values from $0.7S_{1.3}(0.3, 1, 1) + 0.3S_{1.3}(0.8, 30, 1)$
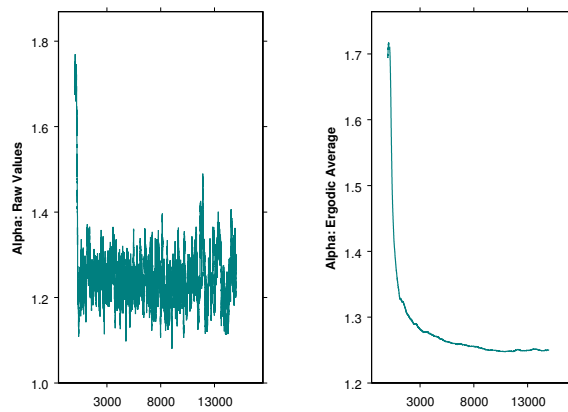
Figure 35: Gibbs sampler realisations and ergodic averages for α

Figure 36: Gibbs sampler realisations and ergodic averages for β_1 and β_2.
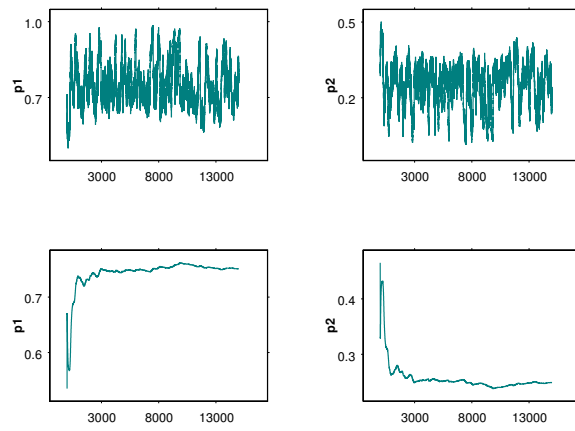
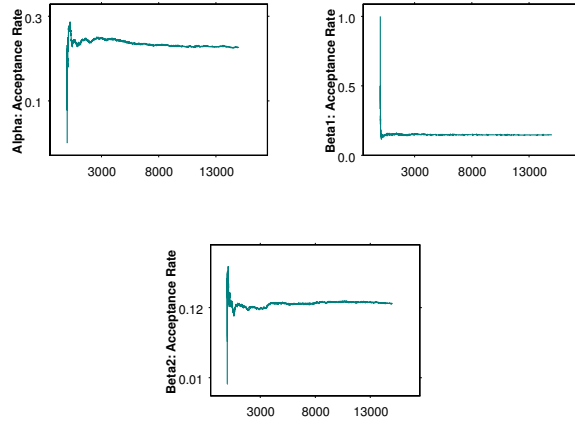Figure 37: Gibbs sampler realisations and ergodic averages for p_1 and p_2.

Figure 38: Acceptance rates for α, β_1 and β_2

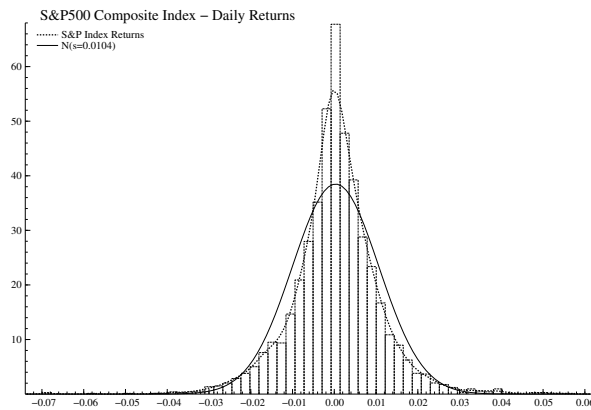# E.3 Stock price indexes
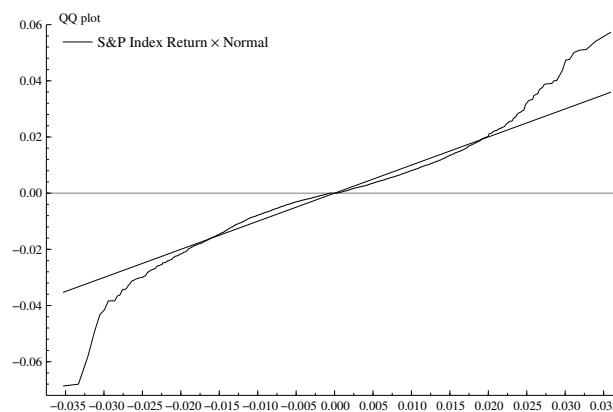

Figure 39: Dataset of daily price returns

Figure 40: QQ-plot of daily price returns distribution against a normal distribution with the same mean and variance
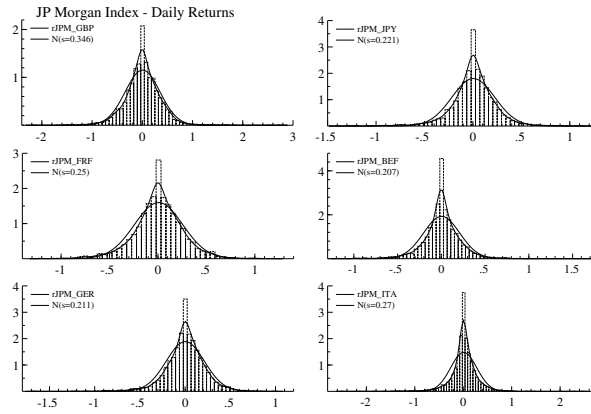
E.4 Bond Indexes

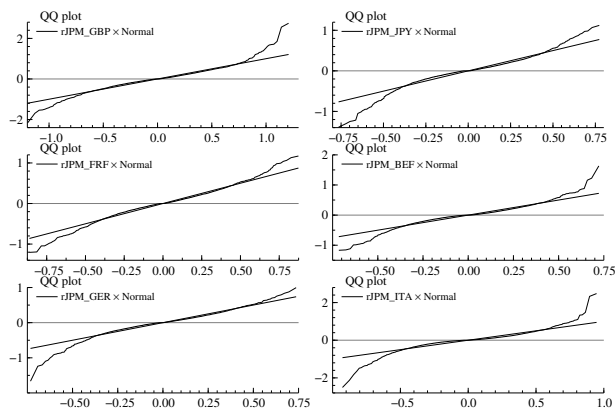Figure 41: Dataset of daily price returns

Figure 42: QQ-plots of daily price returns distribution against a normal distribution with the same mean and variance
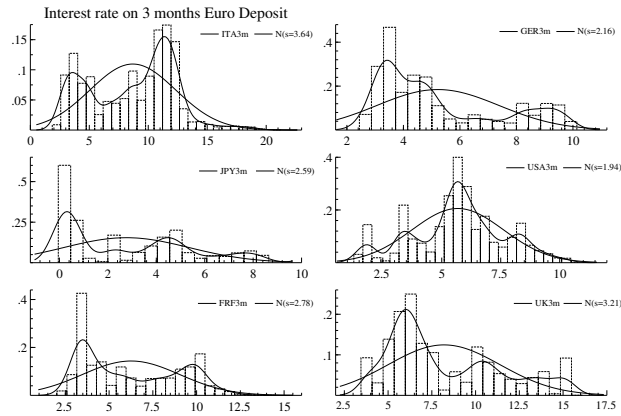
E.5 3-Months Interest Rates

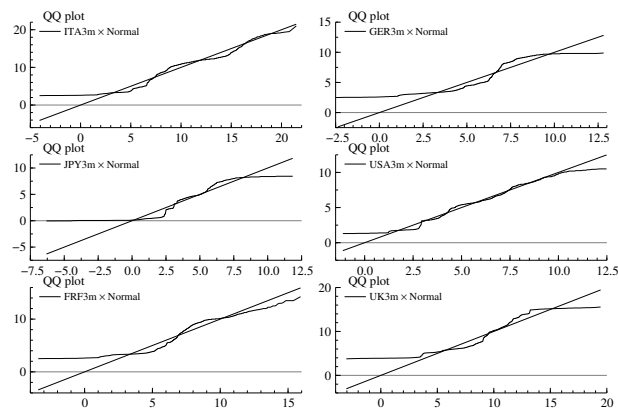Figure 43: Dataset of daily interest rates

Figure 44: QQ-plots of daily interest rates distribution against a normal distribution with the same mean and variance

References

[1] Adler R.J., Feldman R.E., Taqqu M.S. (1998), *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Birkauser Ed., Boston.

[2] Barnea A., Downes D.H. (1973), A Reexamination of the Empirical Distribution of Stock Price Changes, *Journal of the American Statistical Association*, Vol. 68, Issue 342, pp. 348-350.

[3] Beedles W.L., Simkowitz M.A. (1980), Asymmetric Stable Distributed Security Returns, *Journal of the American Statistical Association*, Vol. 75, Issue 370, pp. 306-312.

[4] Bradley B.O., Taqqu M.S. (2001), Financial Risk and Heavy Tails, Preprint, Boston University.

[5] Brennen M. (1974), On the Stability of the Distribution of the Market Component in Stock Price Changes, *Journal of Financial Quantitative Analysis*, Vol. 9, Issue 6, pp. 945-961.

[6] Buckle D.J. (1995), Bayesian inference for stable distribution, *Journal of American Statistical Association*, Vol. 90, pp. 605-613.

[7] Celeux G., Hurn M., Robert C.P. (2000), Computational and inferential difficulties with mixture posterior distributions, *Journal of American Statistical Association*, 95, pp. 957-979.

[8] Chambers J.M., Mallows C.L., Stuck B.W. (1976), A Method for Simulating Stable Random Variables, *Journal of the American Statistical Association*, Vol. 71, Issue 354, pp. 340-344.

[9] Devroye L. (1986), *Nonuniform Random Variate Generation*, Springer Verlag, New York.

[10] Diebolt J., Robert C.P. (1994), Estimation of Finite Mixture Distributions through Bayesian Sampling, *Journal of the Royal Statistical Society*, Series B (Methodological), Vol. 56, Issue 2, pp. 363-375.

[11] DuMouchel W. (1973), On the Asymptotic Normality of the Maximum-Likelihood Estimate when Sampling from a Stable Distribution, *Annals of Statistics*, 3, pp.948-957.

[12] Escobar M.D., West M. (1995), Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association*, 90, pp. 577-588.

[13] Fama E.F. (1965), The Behavior of Stock Market Prices, *Journal of Business*, 34, pp. 420-429.

[14] Fama E.F. (1965), Portfolio Analysis in a Stable Paretian Market, *Management Science*, B 2, pp. 409-419.

[15] Fama E.F., Roll R. (1968), Some Properties of Symmetric Stable Distributions, *Journal of the American Statistical Association*, 68, pp. 817-836.

[16] Fama E.F., Roll R. (1971), Parameter Estimates for Symmetric Stable Distributions, *Journal of the American Statistical Association*, 66, pp. 331-338.

[17] Fielitz B.D., Rozelle J. P. (1983), Stable Distributions and the Mixtures of Distributions Hypotheses for Common Stock Returns, *Journal of the American Statistical Association*, Vol. 78, Issue 1983, pp. 28-36.

[18] Fishman G.S. (1996), *Monte Carlo: concepts, algorithms, and applications*, Springer Verlag, New York.

[19] Geman S., Geman D. (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 6, pp. 721-741.

[20] Gilks W.R., Richardson S. and Spiegelhalter D.J. (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall.

[21] Godsill S. (1999), MCMC and EM-based methods for inference in heavy-tailed processes with α-stable innovations, *Working paper*, Signal Processing Group, Dep. of Engineering, University of Cambridge.

[22] Mandelbrot B., (1963), The Variation of Certain Speculatives Prices, *Journal of Business*,36,394-419.

[23] McLachlan G., Peel D. (2000), *Finite Mixture Models*, John Wiley & Sons.

[24] Meyn S.P., Tweedie R.L. (1993), Markov Chains and Stochastic Stability, Springer-Verlag, New York.

[25] Mittnik S., Rachev T., Paolella M.S. (1998), Stable Paretian modelling in finance: Some empirical and theoretical aspects. In Adler *et. al.*, editor, *A Practical Guide to Heavy Tails*, Birkhauser, Boston, 1998.

[26] Mikosch T. (2000), Modelling Dependence and Tails of Financial Time Series, *Preprint.*

[27] Neal R. M. (1991), Bayesian Mixture Modeling by Monte Carlo Simulation, *Technical Report*, CRG-TR-91-2, Department of Computer Science, University of Toronto.

[28] Press W., Teukolsky S., Vetterling T. and Flannery B.P. (1992), *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge.

[29] Qiou Z., Ravishanker N., and Dey D.K. (1999), Multivariate Survival Analysis with Positive Stable Frailties, *Biometrics*, vol. 55 pp. 81-88.

[30] Qiou Z., Ravishanker N. (1998), Bayesian inference for time series with infinite variance stable innovations. In Adler R., Feldman R. and Taqqu M. editors, *A Practical Guide to Heavy Tails*, pp. 259-282,Birkhauser, Boston, 1998.

[31] Rachev S., Mittnik S. (2000), *Stable Paretian Models in Finance*, John Wiley & Sons.

[32] Richardson S., Green P.J. (1997), On Bayesian analysis of mixtures with an unknown number of components, *Journal of Royal Statistics Society*, Ser. B, 59, pp. 731-792.

[33] Robert C.P. (2001), *The Bayesian Choice, 2nd ed.* Springer Verlag, New York.

[34] Robert C.P. (1996), Mixtures of distributions: Inference and estimation. In *Markov Chain Monte Carlo in Practice* (Eds W.R.Gilks, S.Richardson and D.J.Spiegelhalter). London: Chapman & Hall.

[35] Robert C.P. and Casella G. (1999), *Monte Carlo Statistical Methods*, Springer Verlag, New York.

[36] Samorodnitsky G. and Taqqu M.S. (1994), *Stable Non-Gaussian Random Processes: Stochastic Models with Infinte Variance*,Chambrdige University Press.

[37] Stephens M. (2000), Bayesian Analysis of Mixture Models with an Unknown Number of Components: an alternative to reverstible jump methods, *Annals of Statistics*, 28, pp. 40-74.

[38] Stephens M. (1997), Bayesian Methods for Mixtures of Normal Distributions,*Ph.D. thesis*, University of Oxford.

[39] Teichmoeller J. (1971), A Note on the Distribution of Stock Price Changes, *Journal of the American Statistical Association*, 66, pp. 282-284.

[40] Tierney L. (1994), Markov chains for exploring posterior distributions (with discussion), *Annals of Statistics*, 22, pp. 1701-1786.

[41] Weron R. (1996), On the Chambers-Mallows-Stuck Method for Simulating Skewed Stable Random Variables, *Working Paper*, Technical University of Wroclaw, Poland.

[42] Zoloratev V.M. (1966), On Representation of Stable Laws by Integrals, *Selected Translations in Mathematical Statistics and Probability*, 6, pp. 84-86.