Bayesian Inference for Mixtures of α-Stable Distributions

Roberto Casarin † ‡

CEREMADE

University Paris IX (Dauphine)

and

Dept. of Economics

University Ca' Foscari, Venice

Abstract

The gaussian model results unsatisfactory and reveals difficulties in fitting data with skewness, heavy tails and multimodality. The use of α-stable distributions allows for modelling skewness and heavy tails but gives rise to inferential problems related to the estimation of the parameters of the distributions. 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 reparameterisation of the mixture, which allows us to make inference on the parameters of the mixture. We use a full Bayesian approach and MCMC simulation techniques for the estimation of the posterior distribution. Some applications of stable mixtures to financial data are provided.

KEYWORDS: mixture model, α-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 *α-stable distributions* or more simply stable distributions have been

†I'm thankful to C.P. Robert, J.M. Marin, P. Müller and M. Billio for the suggestions for 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

introduced as a generalisation of the Gaussian model. Stable distributions 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 [39] and Samorodnitsky and Taqqu [34], 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 estimation techniques.

In order to account for heterogeneity and non-linear dependencies exhibited by the data, stable distributions have been introduced in different kind of stochastic 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 [27]. Stable distributions are also used to model heterogeneity over time. For an introduction to time series models with stable noises, see Qiou and Ravishanker [28] and Mikosch [23].

In finance the use of stable distributions has been motivated also on the basis of empirical evidence from financial markets. The first studies on the hypothesis of stable distributed stock prices can be attributed to Mandelbrot [20], Fama [13], [14] and Fama and Roll [15]. They propose stable distributions and give some statistical instruments for the inference on the characteristic exponent.

For recent works treating the use of stable distributions in finance see for example Bradley and Taqqu [4], Mikosch [23], Mittnik, Rachev and Paolella [24] and Rachev and Mittnik [29] which provides a quite complete analysis of theoretical and empirical aspects of stable distributions.

The problem of multimodality and in general of heterogeneity of data can be better treated using mixtures of distributions. Multimodality in asset return distribution is well documented in the financial literature, also from the earlier studies on the stable distributions. Barnea and Downes [2], following the same estimation approach of Fama and Roll [16], find that for some stocks 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. Beedles and Simkowitz [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. Moreover 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.

Using stable distributions and stable mixtures give rise to inferential problems related to the estimation of the parameters of interest. Different estimation methods for stable distributions have been proposed in the literature. For a full Bayesian approach see Buckle [5], for a maximum likelihood approach see DuMouchel [11] and for MCEM approach with application to time series with symmetric stable innovations see Godsill [19].

The first aim of this work is to propose a stable distribution mixture model in order to capture the 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. We use a full Bayesian approach and MCMC simulation techniques. As suggested in the literature on Gaussians mixtures (see for example Robert [31]), we adopt the data augmentation principle in order to make the inference on the mixture parameters easier. The maximum likelihood approach (see for example McLachlan and Peel [21]) 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 [36], 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 problems 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 presents the Bayesian inference model and the Gibbs sampler for a stable distribution. Section 3 describes the Bayesian model for α-stable mixtures, with particular attention to its missing data structure, and the Gibbs sampler for stable mixture is developed in the case where the number of components is fixed. Some simulation results are provided. In

Section 4 we present an application of the proposed Bayesian stable mixture model and the inference procedure on some financial dataset of general interest. Section 5 concludes.

2 Bayesian Inference for Stable Distributions

In this section we define a stable random variable and briefly describe the Bayesian inference approach for stable distribution.

Stable distributions do not generally have an explicit probability density function and are thus conveniently defined through their characteristic function. The most well known parameterisation is defined in Samorodnitsky and Taqqu [34].

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

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

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 [39]. For a review on all the equivalent definitions of stable distribution and on all their properties see Samorodnitsky, Taqqu [34]. 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

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

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 existence of simulation methods for stable distributions opens the way to Bayesian inference on the parameters of this distribution family. The algorithm we use for simulating a standard stable was first proposed by Chamber, Mallows and Stuck [8] and then discussed

also in Weron [38]. We use this method also to generate synthetic dataset to test the efficiency of our MCMC based Bayesian inference approach.
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 normalizing 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.

2.1 MCMC Methods for Bayesian Models

In Bayesian inference many quantities of interest can be represented in the integral form

$$I(\theta) = \mathbb{E}_{\pi(\theta|\mathbf{x})}\{f(\theta)\} = \int_{\mathcal{X}} f(\theta)\pi(d\theta|\mathbf{x}) \tag{3}$$

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. (3)

$$I_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} f(\theta^{(t)}) \tag{4}$$

which is a consistent estimator of the quantity of interest

$$I_n(\theta) \xrightarrow{a.s.} \mathbb{E}_{\pi(\theta|\mathbf{x})}\{f(\theta)\} \tag{5}$$

In some cases, as in mixture models, it is not possible to simulate directly from the posterior distribution and a further simulation step is needed (see the *completion step* in Section 3).

For an introduction to Markov chains and to Markov Chain Monte Carlo methods see Robert and Casella [33]. Further details on Markov chains can be found for example in Meyn and Tweedie [22] and in Tierney [37].

2.2 The Gibbs Sampler for Univariate Stable Distributions

In this paragraph we briefly develop and discuss a Gibbs sampler of the type proposed by Buckle [5] in order to estimate the characteristic exponent, α, of a stable distribution. It is known 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 [5]. The stable density is obtained by integrating with respect to y the following bivariate density function of the pair (x, y)

$$f(x,y|\alpha,\beta,\delta,\sigma) = \frac{\alpha}{|\alpha-1|}\exp\left\{-|\frac{z}{\tau_{\alpha,\beta}(y)}|^{\alpha/(\alpha-1)}\right\}\left|\frac{z}{\tau_{\alpha,\beta}(y)}\right|^{\alpha/(\alpha-1)}\frac{1}{|z|} \quad (6)$$

$$(x,y) \in (-\infty,0)\times(-1/2,l_{\alpha,\beta})\cup(0,\infty)\times(l_{\alpha,\beta},1/2) \quad (7)$$

$$\tau_{\alpha,\beta}(y) = \frac{\sin(\pi\alpha y+\eta_{\alpha,\beta})}{\cos(\pi y)}\left[\frac{\cos(\pi y)}{\cos(\pi(\alpha-1)y)+\eta_{\alpha,\beta}}\right]^{(\alpha-1)/\alpha} \quad (8)$$

$$\eta_{\alpha,\beta} = \beta\, min(\alpha, 2-\alpha)\pi/2 \quad (9)$$

$$l_{\alpha,\beta} = -\eta_{\alpha,\beta}/\pi\alpha \quad (10)$$

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. 1. We use the prior suggested by Buckle [5].

Suppose to observe n realizations $\mathbf{x} = (x_1,\ldots,x_n)$ from a stable distribution $S_\alpha(\beta,\delta,\sigma)$ and simulate a vector of auxiliary variables $\mathbf{y} = (y_1,\ldots,y_n)$, then the *completed likelihood* and the completed posterior distribution are respectively

$$L(\mathbf{x},\mathbf{y}|\theta) = \prod_{i=1}^{n} f(x_i,y_i|\theta) \quad (11)$$

$$\pi(\theta|\mathbf{x},\mathbf{y}) = \frac{L(\mathbf{x},\mathbf{y}|\theta)\pi(\theta)}{\int_\Theta L(\mathbf{x},\mathbf{y}|\theta)\pi(\theta)d\theta} \propto \prod_{i=1}^{n} f(x_i,y_i|\theta)\pi(\theta) \quad (12)$$

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

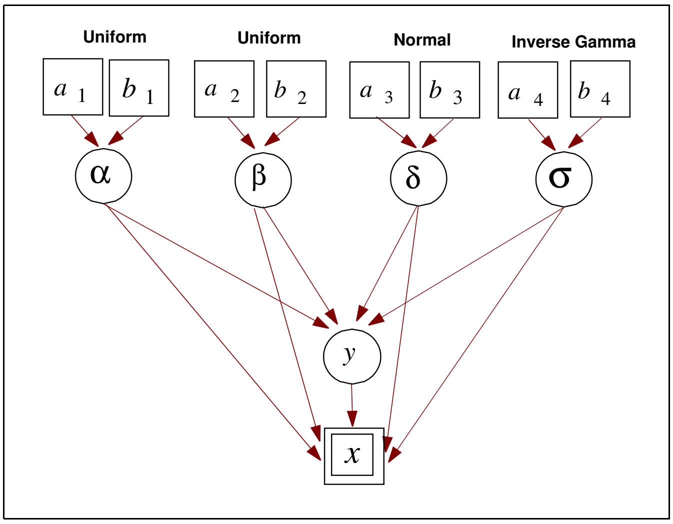

Figure 1: DAG of the Bayesian model for inference on stable distributions. It exhibits the hierarchical structure of priors, and hyperparameters. A single box around a quantity indicates that it is a known constant, a double box indicates the variable is observed and a circle indicates the random variable is not observable. The directed arrows show the dependence structure of the model.

In the following we suppose to observe n values from a stable distribution $S_\alpha(\beta,\delta,\sigma)$. The parameters are estimated by simulating 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*

$$\pi(y_i|\alpha,\beta,\delta,\sigma,z_i) \propto \exp\left\{1-\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}} \tag{13}$$

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

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

$$\pi(\alpha|\beta,\delta,\sigma,\mathbf{x},\mathbf{y}) \propto \frac{\alpha^n}{|\alpha-1|^n}\exp\left\{-\sum_{i=1}^{n}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\prod_{i=1}^{n}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\pi(\alpha) \tag{14}$$

$$\pi(\beta|\alpha,\delta,\sigma,z_i,y_i) \propto \exp\left\{-\sum_{i=1}^{n}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\prod_{i=1}^{n}\left|\frac{1}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\pi(\beta) \tag{15}$$

$$\pi(\delta|\alpha,\beta,\sigma,z_i,y_i) \propto \exp\left\{-\sum_{i=1}^{n}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\right\}\prod_{i=1}^{n}\left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\frac{\alpha}{(\alpha-1)}}\frac{1}{|x_i-\delta|}\pi(\delta) \tag{16}$$

$$\pi(\sigma|\alpha,\beta,\sigma,z_i,y_i) \propto \left(\frac{1}{\sigma^{\alpha/(\alpha-1)}}\right)^n \exp\left\{-\frac{1}{\sigma^{\alpha/(\alpha-1)}}\sum_{i=1}^{n}\left|\frac{(x_i-\delta)}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)}\right\}\pi(\sigma) \tag{17}$$

where $\pi(\alpha)$, $\pi(\beta)$, $\pi(\delta)$, $\pi(\sigma)$ are the prior distributions on the parameters, $\mathbf{y}$ is a vector of auxiliary variables $(y_1,\dots,y_n)$ and $\tau_{\alpha,\beta}$ is a function of y defined in Eq. (8).

In order to simulate from the density function given in Eq. (13) 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 $\tau_{\alpha,\beta}(y^*)=x$.

Some numerical problems may arise in making inference on stable distributions. In particular for all values of $\alpha \in (0,1)$, high values of x make the full conditional density function of y 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 Eq. (14), (15), (16) and (17) are undulating and rather concentrated, therefore, as suggested by Buckle [5] and Qiou and Ravishanker [28], we introduce the following reparameterisation which gives a more manageable form of the conditional posteriors of α, β and δ

$$v_i = \tau_{\alpha,\beta}(y_i) \tag{18}$$

$$\phi_i = \frac{\tau_{\alpha,\beta}}{x_i-\delta}. \tag{19}$$

The resulting posteriors are

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

$$\pi(\beta|\alpha,\delta,\sigma,z_i,v_i) = \prod_{i=1}^{n}\left|\frac{d\tau_{\alpha,\beta}}{dy}\right|^{-1}_{\tau_{\alpha,\beta}(y)=v_i}\pi(\beta) \tag{21}$$

$$\pi(\delta|\alpha,\beta,\sigma,z_i,v_i) \propto \prod_{i=1}^{n}\left|\frac{d\tau_{\alpha,\beta}}{dy}\right|^{-1}_{\tau_{\alpha,\beta}(y)=\phi_i(x_i-\delta)}\pi(\delta) \tag{22}$$

At each step of the reparameterised Gibbs sampler, the Jacobian of the transformation, $\left|\frac{d\tau_{\alpha,\beta}}{dy}\right|^{-1}$, must be evaluated in $y_i = \tau^{-1}_{\alpha,\beta}(v_i)$. Due to the complexity of the function $\tau_{\alpha,\beta}$, its inverse has not an analytical expression. Therefore, following Buckle [5], the inverse transformation is determined numerically. We use the modified safeguard Newton algorithm proposed in Press et al. [26].
In order to simulate from the posteriors given in Eq. (20), (21) and (22) we use a Metropolis-Hastings algorithm. In general the method generates iteratively the parameter $\theta^{(k)}$ according to the following steps

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

(ii) *Take:*

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

where: $\rho(\theta^{(k-1)},\theta^*) = 1 \wedge \left\{\frac{\pi(\theta^*)}{\pi(\theta^{(k-1)})}\frac{q(\theta^{(k-1)}|\theta^*)}{q(\theta^*|\theta^{(k-1)})}\right\}$.

In order to simulate the full conditional posteriors given in Eq. (20) and (21), we use a beta distribution, $\mathcal{B}e(a,b)$ 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

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

and

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

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. For further details see Appendix A. 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. (22)).

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

$$\mathbb{I}(\alpha)_{(1,2]}\frac{1}{2}\mathbb{I}(\beta)_{[-1,1]}\frac{1}{\sqrt{2\pi b_3}}e^{-\frac{(\delta-a_3)^2}{2b_3}}\mathbb{I}(\delta)_{(-\infty,+\infty)}\frac{b_4^{a_4}}{\Gamma(a_4)}\frac{e^{-\frac{b_4}{\theta}}}{\theta^{a_4+1}}\mathbb{I}(\theta)_{[0,\infty)} \tag{23}$$

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 Eq. (17). 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. The efficiency of the proposed Gibbs sampler has been tested on simulated dataset (see Casarin [6] for further details).

3 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 jointly model these features of the data, is to introduce stable mixtures. The use of stable mixtures is appealing also because they have, as special case, normal mixtures which are a widely studied topic (see for example Stephens [36], Richardson and Green [30] for the discrete normal mixtures). Other relevant works on the Bayesian approach to the mixture models estimation are Diebolt and Robert [10], Escobar and West [12] and Robert [31], [32]. In Appendix B we present some examples of two components stable mixtures with different parameters setting, in order to understand the influence of each parameter on the shape of the mixture's distribution.

3.1 The Missing Data Model

In the following we define a stable mixture model 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.

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

$$m(x|\theta, p) = \sum_{l=1}^{L} p_l f(x|\theta_l) \tag{24}$$

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

where $\theta_l = (\alpha_l, \beta_l, \delta_l, \sigma_l)$, $l = 1, \ldots, L$, are the parameter vectors, $\theta = (\theta_1, \ldots, \theta_L)$ the collection of all the parameters and $p = (p_1, \ldots, p_L)$ the allocation probabilities. 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

$$m(x|\theta, p) = \sum_{l=1}^{L} p_l \int_{-1/2}^{1/2} f(x, y|\theta_l)\, dy. \tag{25}$$

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*) is denoted with $\nu = \{\nu_1, \ldots, \nu_L\}$ and each component $\nu_l \in \{0, 1\}$, with $l = 1, \ldots, L$, allows to select the l-th mixture component, $f(x, y|\theta_l)$.

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 steps: first, simulating the allocation variable; second, simulating a mixture component conditionally on the allocation variable.

The resulting demarginalized mixture model is

$$m(x, \nu|\theta, p) = \prod_{l=1}^{L} \left(p_l \int_{-1/2}^{1/2} f(x, y|\theta_l)\, dy \right)^{\nu_l}, \qquad \sum_{l=1}^{L} \nu_l = 1 \tag{26}$$

This completion strategy is now quite popular in Bayesian inference for mixtures (see Robert [32], Robert and Casella [33], Escobar and West [12] and Diebolt and Robert [10]). For an introduction to Monte Carlo methods in Bayesian inference from data modelled by mixture of distributions see also Neal [25] and for a discussion of the numerical and identifiability problems in mixtures inference see Richardson and Green [30], Stephens [35] and Celeux, Hurn and Robert [7].

3.2 The Bayesian Approach

The Bayesian model for inference on stable mixtures is represented through the DAG in Fig. 2. In the following we specify the prior distributions of the model.

Denote with

$$\mathcal{M}(n, p_1, ..., p_L) = \left(\begin{array}{c} n \\ x_1 \ldots x_L \end{array} \right) p_1^{x_1} \cdot \ldots \cdot p_L^{x_L} \mathbb{I}_{\sum x_l = n} \tag{27}$$

the *multinomial* density of a L dimensional random variable $X = (x_1, \ldots, x_L)$ 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: $V \sim f_V(\nu) = \mathcal{M}_L(1, p_1, \ldots, p_L)$.

The L dimensional random variable $X = (x_1, \ldots, x_L)$ has *Dirichlet* distribution if

$$\mathcal{D}(\delta_1, ..., \delta_L) = \frac{\Gamma(\delta_1 + ... + \delta_L)}{\Gamma(\delta_1) \cdot \ldots \cdot \Gamma(\delta_L)} x_1^{\delta_1 - 1} \cdot \ldots \cdot x_L^{\delta_L - 1} \mathbb{I}_S(x) \tag{28}$$

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 $\mathbb{R}^L$.

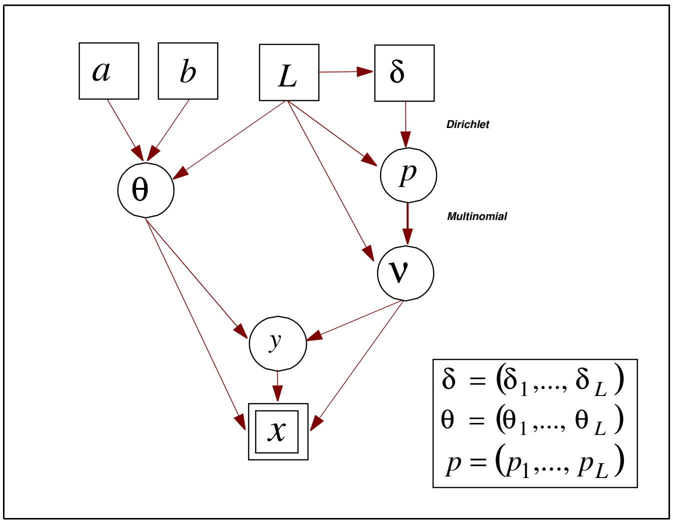

Figure 2: DAG of the Bayesian hierarchical model for inference on stable mixtures. Note that the completing variable ν is not observable. Thus, two levels of completion, y and ν, are needed for a stable mixture model.

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

$$L(\mathbf{x}, \mathbf{y}|\theta, p) = \prod_{i=1}^{n} \sum_{l=1}^{L} p_l \int_{-1/2}^{1/2} f(x_i, y_i|\theta_l) dy_i \tag{29}$$

$$L(\mathbf{x}, \mathbf{y}, \nu|\theta, p) = \prod_{i=1}^{n} \prod_{l=1}^{L} (p_l f(x_i, y_i|\theta_l))^{\nu_{il}} \tag{30}$$

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 completed posterior distribution of the Bayesian mixture model is

$$\pi(\theta, p|\mathbf{x}, \mathbf{y}, \nu) \propto \prod_{i=1}^{n} \left\{ \prod_{l=1}^{L} (f(x_i, y_i|\theta_l))^{\nu_{il}} \pi(\nu_i) \right\} \pi(\theta)\pi(p). \tag{31}$$

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 it is necessary to perform simulations from the posterior distributions of the parameters and to average the simulated values.

3.3 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 into: 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*

$$\nu_i^{(0)} \sim \mathcal{M}_L(1, p_1, \ldots, p_L) \tag{32}$$

$$y_i^{(0)} \sim f(y_i|\theta, \nu, x_i) \propto \exp\{1 - \left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)}\} \left|\frac{z_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)} \tag{33}$$

$$p^{(0)} \sim \mathcal{D}_L(\delta, \ldots, \delta). \tag{34}$$

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

$$\pi(\theta_l|\theta_{-l}, p, \mathbf{x}, \mathbf{y}, \mathbf{v}) \propto \prod_{i=1}^{n} \{f(x_i, y_i|\theta_l)\, p_l\}^{\nu_{il}}\ \pi(\theta_l) \quad l = 1, \ldots, L \tag{35}$$

$$\pi(p_1, \ldots, p_L|\theta, \mathbf{x}, \mathbf{y}, \mathbf{v}) = \mathcal{D}(\delta + n_1(\nu), \ldots, \delta + n_L(\nu)) \tag{36}$$

(iii) *Update the completing variables*

$$\pi(y_i|\theta, p, \mathbf{x}, \mathbf{y}_{-i}, \mathbf{v}) \propto \exp\left\{1 - \left|\frac{x_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)}\right\} \left|\frac{x_i}{\tau_{\alpha,\beta}(y_i)}\right|^{\alpha/(\alpha-1)} \tag{37}$$

$$\pi(\nu_i|\theta, p, \mathbf{x}, \mathbf{y}, \mathbf{v}_{-i}) = \mathcal{M}_L(1, p_1^*, \ldots, p_L^*) \tag{38}$$

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

$$
\begin{aligned}
z &= \frac{x - \delta}{\sigma} \\
n_l(\nu) &= \sum_{i=1}^{n} \nu_{il}\,, \quad l = 1, \dots, L \\
p_l^* &= \frac{p_l f(x_i, y_i|\theta_l)}{\sum_{l=1}^{L} f(x_i, y_i|\theta_l) p_l}\,, \quad l = 1, \dots, L.
\end{aligned}
$$

Steps (36) and (38) of the Gibbs sampler are proved in Appendix C. Observe that simulations from the conditional posterior distribution of Eq. (36) can be obtained by running the Gibbs sampler given in Eq. (14)-(17), conditionally to the value of the completing variable ν.

To simulate from the Dirichlet posterior distribution given in Eq. (36), we use the algorithm proposed by Casella and Robert [33], while to draw value from the multinomial posterior distribution of Eq. (38), we use the algorithm proposed by Fishman [18].

We put on evidence that the key idea for making inference on stable mixtures is to introduce two levels of auxiliary variables (see Eq. 38 and 38). This approach allows us to infer all the parameters of the mixture from the data.

In Examples 1 and 2, we verify the efficiency of the Gibbs sampler on some test samples simulated from stable mixtures. Furthermore, for the sake of simplicity, we consider $L = 2$. For each mixture's component we assume the joint prior distribution given in equation Eq. (23).

Example 1 - *α-stable Mixture with varying α and β*
In this example we apply the proposed 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. 5 in Appendix D 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.
Estimation results are briefly represented in the first panel of Table 1 and graphically described in Fig. 6-9 (Appendix D). Note that the presence, in the mixture model, of distributions with different tails behaviour causes some problems in the convergence of the ergodic averages, due to the label switching of the observations.

Example 2 - *α-stable Mixture with constant α and varying β*

In this experiment we keep α fixed over the mixture components. First the generated dataset consists of 1,000 observations from the following mixture $0.5S_{1.3}(0.3, 1, 1) + 0.5S_{1.3}(0.8, 30, 1)$. The dataset is represented in Fig. 10 (Appendix D). Secondly we apply the proposed Gibbs sampler for stable mixture and obtain the results given in the second panel of Table 1. Raw outputs, ergodic means and acceptance rate of the Gibbs are in Figg. 11-14 of Appendix D. The graphical results exhibit a more appreciable mixing of the chain associated to the Gibbs sampler.

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

	Dataset simulated from: $0.5S_{1.7}(0.3, 1, 1) + 0.5S_{1.3}(0.5, 30, 1)$				
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 simulated from: $0.5S_{1.3}(0.3, 1, 1) + 0.5S_{1.3}(0.8, 30, 1)$				
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*

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.

4 Applications to Financial 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 distributions and stable mixtures may result particularly useful in modelling different kind of financial variables. We present estimates obtained with the MCMC based inferential technique proposed in the previous section.

In the first two examples we test the proposed Gibbs sampler for stable distribution on the S&P500 and on J.P. Morgan's bond index. In the third example we estimate a stable mixture on the 3-month Euro-Deposits interest rate which exhibits multimodality.

Example 3 - In this example we analyse the return rate of the S&P500 composite index from January 01, 1990 to January 27, 2003 (source: *DataStream*). 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. The first graph in Fig. 15 shows the data histogram and the best normal which is possible to estimate. The corresponding QQ-plot, in Fig. 15, reveals that data are not normally distributed. We apply the Gibbs sampler for α-stable distributions to this dataset. The results are presented in Tab. 2. Parameter estimates are ergodic averages over the last 10,000 values of the 15,000 Gibbs sampler realizations. Note that since $\hat{\alpha} = 1.674$ the distribution tails of the index return are heavier than the tails of a Gaussian distribution.

Example 4 - Our second dataset (source: *DataStream*) contains daily price returns on the J.P. Morgan's index concerning Great Britain between January 01, 1988 and January 13, 2003. Denoting with p_t the price index at time t, the return on the index is defined as: $r_t = (p_t - p_{t-1})/p_{t-1}$. The first graph in the second line of Fig.15 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 results on the J.P. Morgan Great Britain index are presented in Tab. 3.

Example 5 - Our third dataset (source: *DataStream*) contains daily 3-Months Interest Rates on Euro-Deposits for France between January 01, 1988 and January 13, 2003. The first graph of the third line in Fig.15 exhibits jointly the histogram, the best Gaussian approximation and the density line of the returns distribution. We note that the time series exhibits multimodality, thus a mixture model is applied. Estimation results on the 3-month Interest Rate for France are presented in Tab. 4.

5 Conclusion

In this contribution we propose an α-stable mixture model and apply a suitable reparameterisation of the mixture in order to perform Bayesian inference. In the literature a statistical inference approach to α-stable mixtures is still missing and in this work we suggest to adopt Bayesian inference due to the flexibility of this approach which allows to simultaneously estimate all the parameters of the model.

The proposed approach to α-stable mixture models estimation is quite general and performed well in our simulation analysis. We discuss in detail some computational issues related to the steps of the Gibbs sampler. Moreover we apply the estimation approach to some financial variables of general interest.

In the proposed example we assume that the number of mixture components is known but the Bayesian approach proposed in this contribution allows to apply goodness of fit tests, Reversible Jump MCMC and Birth and Death MCMC techniques in order to make inference on the number of the components of the mixture.

Directions for further study are the case of symmetric stable mixtures, the noninformative prior on $\alpha \in (0, 2]$ and some computational issues related to the Gibbs sampling from the full conditional: $\pi(y_i|\alpha, \beta, \delta, \sigma, z_i)$.

Table 2: Parameter estimates on S&P 500 index returns. Daily observations over the period January 01, 1990 to January 27, 2003.

	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

Table 3: Parameter estimates on JP Morgan - Great Britain index returns. Daily observations over the period January 01, 1988 to January 13, 2003.

	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

Table 4: Parameter Estimates of a two components α-stable mixture, on the 3-months Euro-Deposit Interest Rates, France. Daily observations over the period January 01, 1988 to January 13, 2003.

	Starting Value	Estimate	Std.Dev.	Acc. Rate
α_1, α_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

Appendix A - Proposal Distributions for the Metropolis-Hastings Algorithm

The shape of the stable distribution suggests to use a beta distribution $\mathcal{B}e(a,b)$ as proposal for the Metropolis-Hastings algorithm

$$\alpha|\alpha_{k-1} \sim \mathcal{B}e(a,b) = \frac{1}{B(a,b)}\alpha^{a-1}(1-\alpha)^{b-1}\mathbb{I}(\alpha)_{(0,1)}. \tag{39}$$

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 M.-H. algorithm.

When $\alpha \in (0,1)$ the values of the parameters are

$$\left\{ \begin{array}{l} \frac{a}{a+b} = \alpha_{k-1} \\ \frac{ab}{(a+b)^2(1+a+b)} = v \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{l} a = \frac{\alpha_{k-1}^2\,(1-\alpha_{k-1})-v\,\alpha_{k-1}}{v} \\ b = a\frac{1-\alpha_{k-1}}{\alpha_{k-1}} \end{array} \right.$$

where α_{k-1} is the $(k-1)$-th value of the M.-H. chain. 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

$$\alpha_{k-1} \in \left(\frac{3-v}{2} - \frac{\sqrt{v^2-8v+1}}{2}, \frac{3-v}{2} + \frac{\sqrt{v^2-8v+1}}{2} \right). \tag{40}$$

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

$$\alpha|\alpha_{k-1} \sim \mathcal{B}e(a,b) = \frac{1}{B(a,b)}(\alpha-1)^{a-1}(2-\alpha)^{b-1}\mathbb{I}(\alpha)_{(1,2)}. \tag{41}$$

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

$$\left\{ \begin{array}{l} \frac{2a+b}{a+b} = \alpha_{k-1} \\ \frac{ab}{(a+b)^2(1+a+b)} = v \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{l} a = \frac{(\alpha_{k-1}-1)^2\,(2-\alpha_{k-1})-v\,(\alpha_{k-1}-1)}{v} \\ b = a\frac{2-\alpha_{k-1}}{(\alpha_{k-1}-1)} \end{array} \right.$$

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 B - Mixtures of Stable Distributions

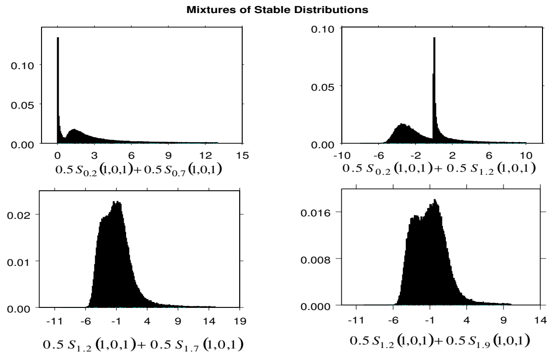

Figure 3: Histograms of $N = 100,000$ values simulated from stable mixtures, with two equally weighted components. In all the examples the location and scale parameters are $\delta_1 = \delta_2 = 0$, $\sigma_1 = \sigma_2 = 1$, while $\beta_1 = \beta_2 = 1$ and α varies across the components.

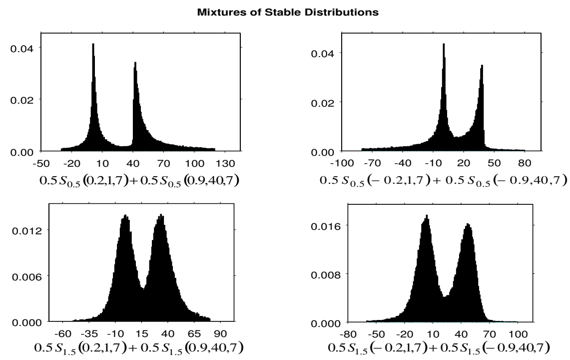

Figure 4: Histograms of $N = 100,000$ values simulated from stable mixtures, with two equally weighted components. In all the examples the location and scale parameters are $\delta_1 = 1$, $\delta_2 = 40$, $\sigma_1 = \sigma_2 = 7$, while $\alpha_1 = \alpha_2 = 0.5\,(\text{or}\,1.5)$ and β varies across the components.

Appendix C - The Gibbs Sampler for a Stable Distributions Mixture

Proof *The posterior distribution of the allocation probabilities* $(p_1, \ldots, p_L)$*, given in Eq. (36), is a Dirichlet and is derived as follows*

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

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

□

Proof *The posterior distribution of the allocation variables given in Eq.(38) is a Multinomial and follows from*

$$
\begin{aligned}
\pi(\nu_1,\ldots,\nu_n|\theta,p,\mathbf{x},\mathbf{y}) &= \frac{L(\mathbf{x},\mathbf{y},\nu|\theta,p)\pi(\theta)\pi(p)}{\int L(\mathbf{x},\mathbf{y},\nu|\theta,p)\pi(\theta)\pi(p)d\nu} \\
&= \frac{\prod_{i=1}^n \left\{\prod_{l=1}^L (f(x_i,y_i|\theta_l))^{\nu_{il}} \prod_{l=1}^L p_l^{\nu_{il}}\right\} \pi(\theta)\pi(p)}{\int \prod_{i=1}^n \prod_{l=1}^L (f(x_i,y_i|\theta_l))^{\nu_{il}} \prod_{l=1}^L p_l^{\nu_{il}} \pi(\theta)\pi(p)d\nu} \\
&= \prod_{i=1}^n \frac{\prod_{l=1}^L \left(f(x_i,y_i|\theta_l)p_l\right)^{\nu_{il}}}{\int \prod_{l=1}^L \left(f(x_i,y_i|\theta_l)p_l\right)^{\nu_{il}} d\nu_i} \\
&= \prod_{i=1}^n \prod_{l=1}^L \left(\frac{f(x_i,y_i|\theta_l)p_l}{\sum_{l=1}^L f(x_i,y_i|\theta_l)p_l}\right)^{\nu_{il}} \\
&= \prod_{i=1}^n \mathcal{M}_L(1,p_1^*,\ldots,p_L^*)
\end{aligned}
$$

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 D - Bayesian Inference for Stable Distributions Mixtures

D.1 Mixtures with varying α and β

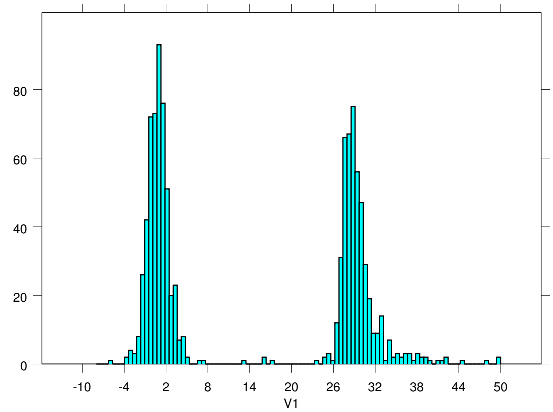

Figure 5: Simulated dataset, 1,000 values from $0.5S_{1.7}(0.3, 1, 1) + 0.5S_{1.3}(0.5, 30, 1)$

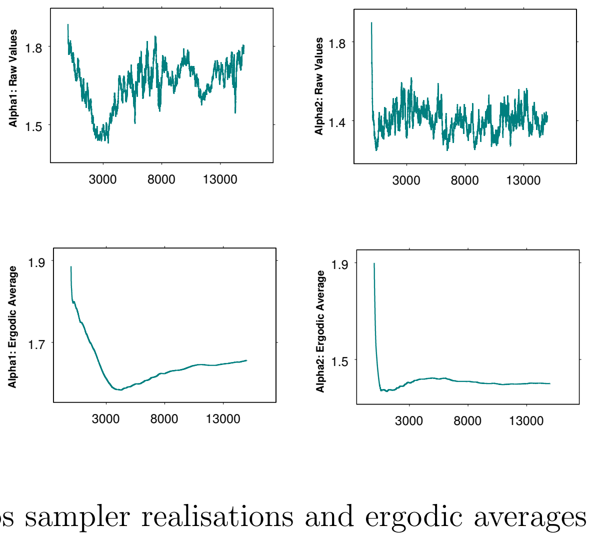

Figure 6: Gibbs sampler realisations and ergodic averages for α_1 and α_2.

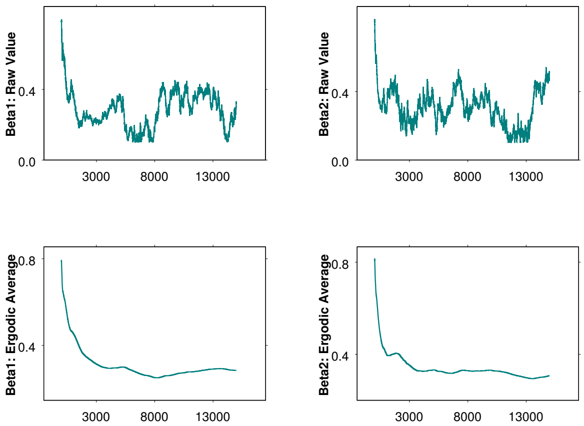

Figure 7: Gibbs sampler realisations and ergodic averages for β_1 and β_2.

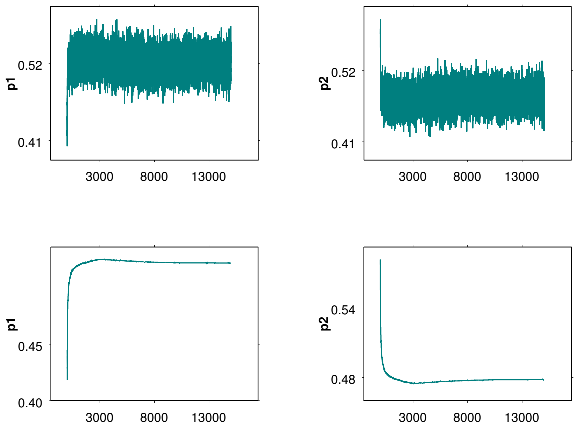

Figure 8: Gibbs sampler realisations and ergodic averages for p_1 and p_2.

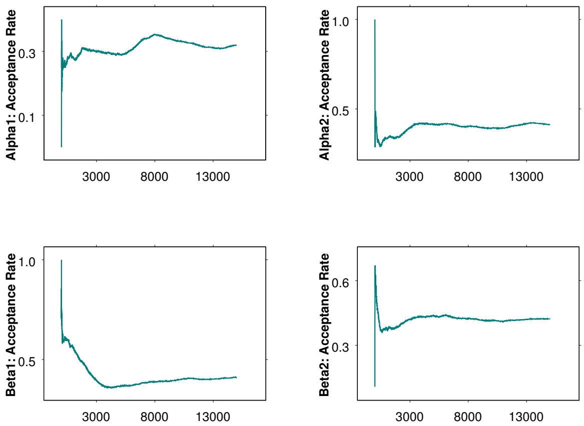

Figure 9: Acceptance rates for α_1, α_2, β_1 and β_2

D.2 Mixtures with fixed α and varying β

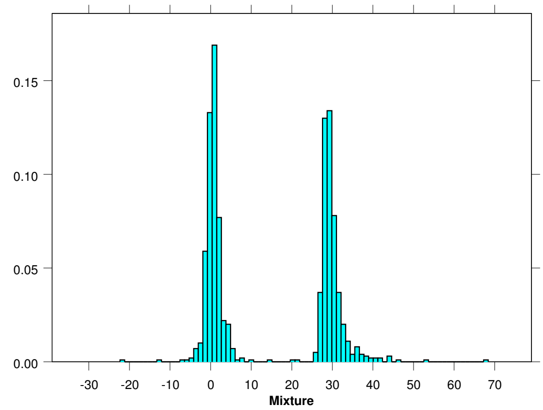

Figure 10: Simulated dataset, 1,000 values from $0.7S_{1.3}(0.3, 1, 1) + 0.3S_{1.3}(0.8, 30, 1)$

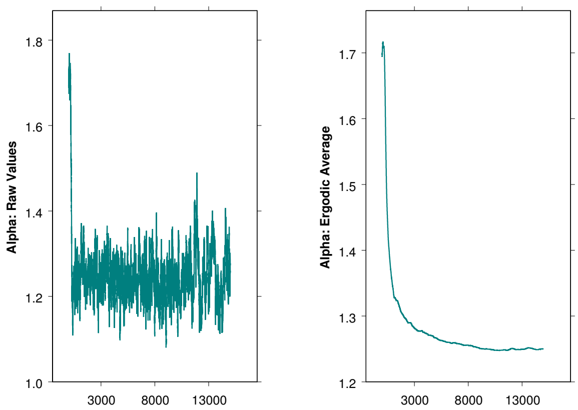

Figure 11: Gibbs sampler realisations and ergodic averages for α

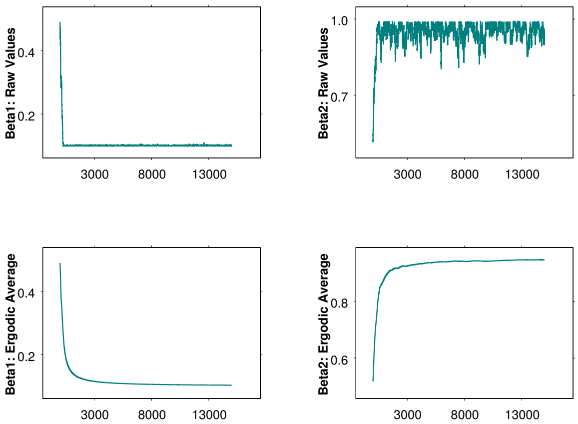

Figure 12: Gibbs sampler realisations and ergodic averages for β_1 and β_2.

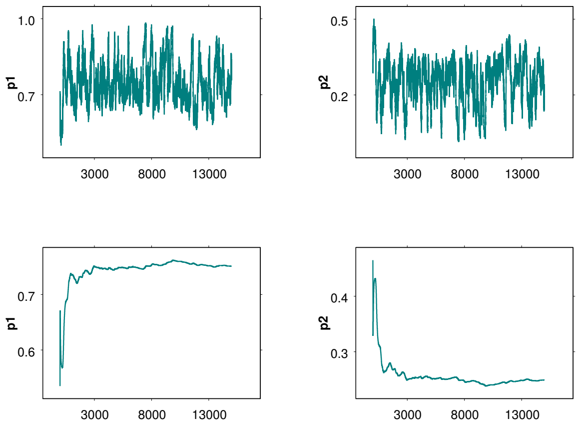

Figure 13: Gibbs sampler realisations and ergodic averages for p_1 and p_2.

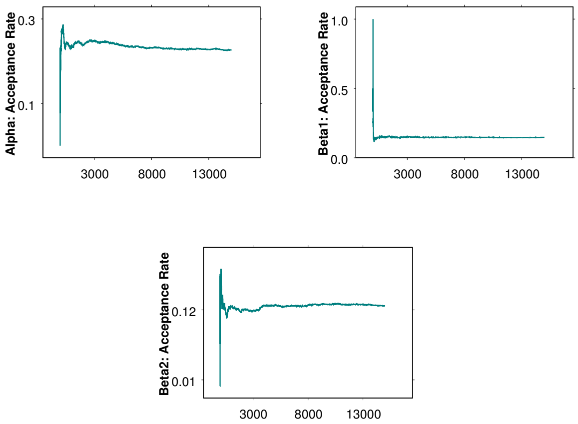

Figure 14: Acceptance rates for α, β_1 and β_2

Appendix E - Financial data

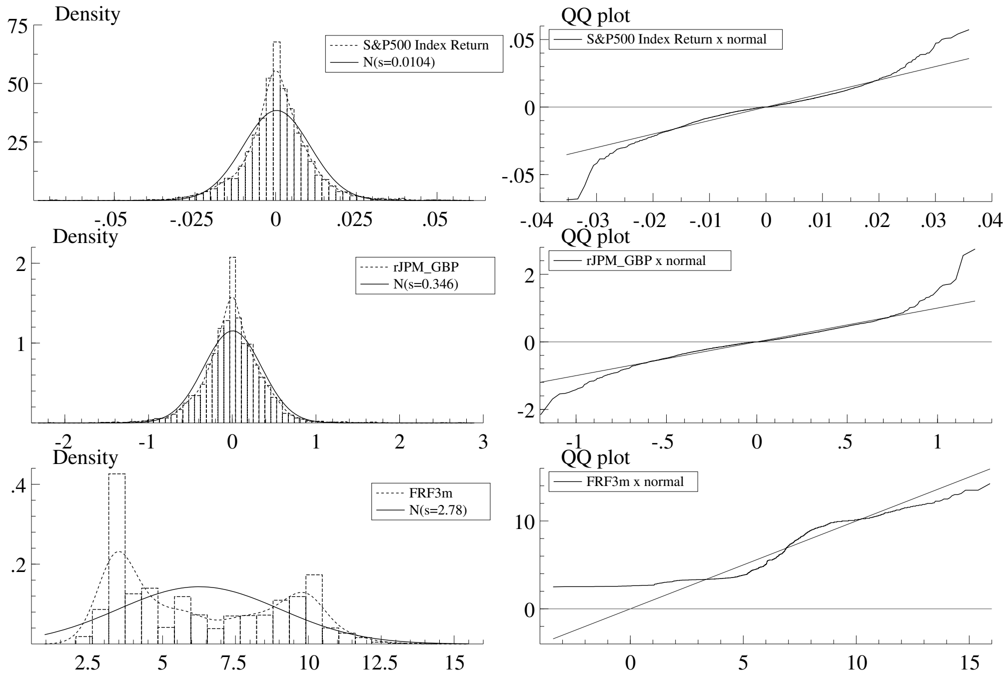

Figure 15: Histograms with kernel density and best normal estimates (left column) and QQ-plots (right column) of daily returns on S&P500 and JP Morgan index for Great Britain bond market and daily interest rates on 3-month Euro-deposits for France.

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

[6] Casarin, R., (2004a), Bayesian Inference for Mixture of Stable Distributions, *Cahier du CEREMADE* N. 0428, University Paris Dauphine.

[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] 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.

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

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

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

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

[24] 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.

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

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

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

[28] 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.

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

[30] 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.

[31] 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.

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

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

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

[35] 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.

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

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

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

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