Parameter Estimation Using the EM Algorithm for Symmetric Stable Random Variables and Sub-Gaussian Random Vectors

Mahdi Teimouri

Department of Statistics, Faculty of Mathematics and Computer Science,
Amirkabir University of Technology (Tehran Polytechnic), 424 Hafez Ave., Tehran 15914, Iran
Department of Statistics, Faculty of Science, Gonbad Kavous University,
Gonbad Kavous, Iran
teimouri@aut.ac.ir

Saeid Rezakhah*

Department of Statistics, Faculty of Mathematics and Computer Science,
Amirkabir University of Technology (Tehran Polytechnic), 424 Hafez Ave., Tehran 15914, Iran
rezakhah@aut.ac.ir

Adel Mohammadpour

Department of Statistics, Faculty of Mathematics and Computer Science,
Amirkabir University of Technology (Tehran Polytechnic), 424 Hafez Ave., Tehran 15914, Iran
adel@aut.ac.ir

Applying some well-known properties of the class of symmetric α-stable (SαS) distribution, the EM algorithm is extended to estimate the parameters of SαS distributions. Furthermore, we extend this algorithm to the multivariate sub-Gaussian α-stable distributions. Some comparative studies are performed through simulation and for some real data sets to show the performance of the proposed EM algorithm compared with some well-known methods including empirical characteristic function, maximum likelihood, and sample quantile in the univariate and multivariate cases.

Keywords: EM algorithm; Markov Chain Monte Carlo; Symmetric α-stable distribution (SαS), Sub-Gaussian α-stable distribution.

1. Introduction

The theory of stable (or α-stable) distributions has received much interest in different fields as physics, economics, finance, and telecommunications since this class of distributions can

---

*Corresponding author.

account for tails thickness, peak value, and skewness. The empirical density function of many real life data sets found in aforementioned fields has heavy tails so that normal models are clearly inappropriate. For comprehensive accounts of applications of stable distributions, we refer the readers to [13], [19], [20], [26], [28], and [31]. There are several parameterizations for characteristic function of a stable random variable. In the following, Definition 1.1 gives the S_1 parameterization which is more popular.

Definition 1.1. The characteristic function of a stable random variable, say X, has the form

$$\varphi_X(t) = E\exp(itX) = \begin{cases} \exp\left\{-|\sigma t|^\alpha\left[1 - i\beta\,\mathrm{sgn}(t)\tan\left(\frac{\pi\alpha}{2}\right)\right] + i\mu t\right\}, & \alpha \neq 1, \\ \exp\left\{-|\sigma t|\left[1 + i\beta\,\mathrm{sgn}(t)\frac{2}{\pi}\log|t|\right] + i\mu t\right\}, & \alpha = 1, \end{cases} \quad (1.1)$$

where $\mathrm{sgn}(.)$ is the well-known sign function.

A stable distribution is described by four parameters: index of stability $\alpha \in (0,2]$ (also called tail parameter), skewness parameter $\beta \in [-1,1]$, scale parameter $\sigma \in \mathbb{R}^+$, and location parameter $\mu \in \mathbb{R}$. We write $S(\alpha,\beta,\sigma,\mu)$ to denote a stable distribution in this parameterization. If $\beta = 0$, it would be the class of symmetric α-stable (SαS) distributions. We assume throughout that $\beta = 0$, in which case $\varphi_X(t) = \exp\left\{-|\sigma t|^\alpha\right\}$.

In the following, we mention three well-known methods for estimating parameters of this family. The maximum likelihood (ML) estimation for stable distributions approximated theoretically by [9] and evaluated numerically by [21]. Although such approaches lead to some efficient estimates, both involve numerical complexities. The software proposed by [21] called STABLE uses a spline interpolation of stable densities for his numerical method. This package estimates all four parameters of a stable distribution efficiently when $\alpha \geq 0.4$. The empirical characteristic function (CF) method is suggested in [12]. As another approach, sample quantile (SQ) technique proposed in [15] presents consistent estimators for all parameters based on five specific sample quantiles.

Definition 1.2. Let $\boldsymbol{Y}$ be a d-dimensional sub-Gaussian α-stable (sub-Gaussian SαS) random vector. Then, the characteristic function of $\boldsymbol{Y}$ can be written as

$$\varphi_{\boldsymbol{Y}}(\boldsymbol{t}) = E\exp(i\boldsymbol{t}^T\boldsymbol{Y}) = \exp\left\{-\left(\boldsymbol{t}^T\Sigma\boldsymbol{t}\right)^{\alpha/2} + i\boldsymbol{t}^T\boldsymbol{\mu}\right\}, \quad (1.2)$$

where Σ is a $d \times d$ positive definite matrix and $\boldsymbol{\mu} \in \mathbb{R}^d$ is the location parameter.

Computing the maximum likelihood (ML) estimation of a sub-Gaussian SαS distribution directly by maximizing the log-likelihood of sub-Gaussian SαS density, known as full MLE in the literature, has been performed in [24]. Finding the full MLEs is computationally expensive. So, a projecting method is suggested in which a sub-Gaussian SαS random vector is projected into different directions. Then since each component follows a univariate SαS distribution, parameters of each component are estimated using methods such as ML, CF, and SQ which developed for the univariate case, [24]. The obtained estimates in such a way are known as projected estimations and these methods are called the projected ML, CF, and SQ, respectively. In the projection approach, the dispersion matrix of a sub-Gaussian SαS distribution is estimated term-by-term, and so there is no guarantee that the reconstructed matrix to be positive definite. The issue with Σ possibly not being positive definite is addressed in [24]. Also, the projected ML approach still is applicable only for $\alpha \geq 0.4$.

Although the proposed EM algorithm in this work is more computationally expensive than the ML, CF, and SQ methods, the following reasons motivate us to propose it. Contrary to the ML approach, it works satisfactorily for all ranges of α in the univariate and multivariate cases. Also, it gives better performance than the ML, CF, and SQ approaches when α is near one. In the multivariate case, it estimates all entries of dispersion matrix simultaneously so that the estimated dispersion matrix is always positive definite. The proposed EM algorithm can be developed for modelling the mixture of SαS distributions. This advantage of the proposed EM algorithm received much interest in robust mixture modelling since SαS is a heavy tailed distribution. Modelling the mixture of sub-Gaussian SαS distributions suggested in [29] using Gibbs sampling. We can develop the introduced EM algorithm in this work for modelling the mixture of sub-Gaussian SαS distributions that is faster than the Gibbs sampling approach.

This paper is organized as follows. Some preliminaries are given in Section 2. In Section 3, we propose the EM algorithm for estimating the parameters of a univariate SαS distribution. The methodology given in Section 3 is developed for the sub-Gaussian SαS distribution in Section 4. In Section 5, some simulation studies are considered to compare the performance of the presented EM algorithms in the univariate and multivariate cases versus the ML, CF, and SQ approaches. Our method will be applied to some sets of real data in this section. We conclude the paper in Section 6 with a summary of the contributions of this work.

2. Preliminaries

2.1. EM algorithm and extension

Missing or incomplete observations frequently occurs in the statistical literature. The EM algorithm, introduced in [7], is a popular inferential tool for such a situation. The application of the EM algorithm also includes cases with latent variables or models with random parameter provided that they are formulated as a missing value problem. For a brief description of EM methodology, let $\underline{x} = (x_1, \ldots, x_n)$ denote the complete observations follow the density function $f(\underline{x}|\theta)$. We denote $\underline{x} = (y, \underline{z})$, where $\underline{z}$ denotes missing samples and y accounts for the available observations, [16]. The computation and maximization of

$$f(\underline{y}|\theta) = \int_{\underline{z}} f((\underline{y}, \underline{z})|\theta) d\underline{z}, \tag{2.1}$$

are difficult in such a case. So, we propose to use the EM algorithm. Let $L_c(\theta) = f((\underline{y}, \underline{z})|\theta)$ be the complete data likelihood function. The EM algorithm works by maximization of the conditional expectation of the complete log-likelihood function $Q\left(\theta|\theta^{(t)}\right) = E\left(l_c(\theta; \underline{x})|\underline{y}, \theta^{(t)}\right)$ given the available data and a current estimate $\theta^{(t)}$ of the parameter (here, $l_c(\theta; \underline{x})$ indicates on the logarithm of $L_c(\theta)$). Since missing values are unobservable, we replace them by their expectation, given observed values and a current guess (or estimate) for θ. Each iteration of the EM algorithm involves two steps as:

1. Expectation (E) step: Compute $Q\left(\theta|\theta^{(t)}\right)$ at t-th iteration.
2. Maximization (M) step: Maximize $Q\left(\theta|\theta^{(t)}\right)$ with respect to θ to get $\theta^{(t+1)}$.

The E and M steps, in above, are repeated until convergence occurs. When the M step of the EM algorithm becomes analytically intractable, we follow the idea of implying this

step using a sequence of conditional maximization, known as the CM step. The resulting procedure is known as the ECM algorithm, [18]. A faster extension of the EM algorithm is the so-called ECME algorithm introduced in [14]. It should be noted that all the EM, ECM and ECME have the same E step. The ECME algorithm works by maximizing the constrained $Q\left(\theta|\theta^{(t)}\right)$ via some CM steps and maximizing the constrained marginal likelihood function, [1]. Sometimes implementation of the EM algorithm is difficult. In this situation, another extension of this algorithm, called the stochastic EM (SEM) is useful, [2], [3], and [4]. Each SEM works by simulating missing data from conditional density $f(z_i|\theta^{(t)}, y_i)$; for $i = 1, \ldots, n$ and replacing it into the complete likelihood function. Then, we apply the EM algorithm for the pseudo-complete sample where $(Z_1^{(t)}, \ldots, Z_n^{(t)})$ are simulated ones. The updated parameter is then used to simulate from $f(z_i|\theta^{(t+1)}, y_i)$; for $i = 1, \ldots, n$. This process is repeated until convergence occurs for the distribution of the $\{\theta^{(t+1)}\}$. Under some mild regularity conditions, $\{\theta^{(t+1)}\}$ constitutes a Markov chain that converges to a stationary distribution, [11]. The number of iterations is determined via an exploratory data analysis approach such as a graphical display. Another example presented in [11] is to consider a total of 300 iterations of the SEM for estimating the parameter of modulus of rupture of 50 observations of fir woods after fitting a two-parameter Weibull model under Type-I censored scheme (the data were censored at 1.6 which corresponds to a censoring of rate 34%). As pointed out by the authors, the SEM is generally very robust to the starting values. Using a burn-in of M_0 iterations, the sequence $\{\theta^{(t)}\}$ is expected to be close to some stationary point. After a sufficiently large number of iterations, say M, the SEM estimation of θ is given by

$$\hat{\theta} = \frac{1}{M - M_0} \sum_{t=M_0+1}^{M} \theta^{(t)},$$

where M_0 is burn-in size. Summarizing the above statements, each iteration of the SEM algorithm consists of two steps as follows.

1. Stochastic imputation (S) step: Impute the simulated missing values and constituting the pseudo-complete log-likelihood function at t-th iteration.
2. Maximization (M) step: Find a θ, say $\theta^{(t+1)}$, which maximizes the pseudo-complete log-likelihood function at t-th iteration.

The S and M steps, in above, are repeated until convergence occurs.

2.2. Properties of symmetric and positive stable laws

Suppose that Y_1 and $Y_2 > 0$ are independent stable random variables, $Y_1 \sim S(\alpha_1, 0, \sigma, 0)$ and $Y_2 \sim S(\alpha_2, 1, (\cos(\pi\alpha_2/2))^{1/\alpha_2}, 0)$ for $0 < \alpha_2 < 1$. Then, $Y_1 Y_2^{1/\alpha_1} \sim S(\alpha_1\alpha_2, 0, \sigma, 0)$, known as the product property in the literature. Based on product property, one can write

$$Y \stackrel{d}{=} \sqrt{P} N_1 + \mu, \tag{2.2}$$

where $Y \sim S(\alpha, 0, \sigma, \mu)$, $P \sim S(\alpha/2, 1, (\cos(\pi\alpha/4))^{2/\alpha}, 0)$, and N_1 is a zero-mean Gaussian random variable with variance $2\sigma^2$. So,

$$Y|P = p \sim N_2, \tag{2.3}$$

where N_2 is a Gaussian random variable with mean μ and variance $2p\sigma^2$. As the second useful property, assume that $P \sim S(\alpha/2, 1, (\cos(\pi\alpha/4))^{2/\alpha}, 0)$. Then,

$$\frac{E}{P} \stackrel{d}{=} W(\alpha/2), \qquad (2.4)$$

where E is an exponential random variable with mean one and $W(\nu)$ is a Weibull random variable, independent of E, with density function $f_W(w) = \nu w^{\nu-1}\exp\{-w^\nu\}$; for $w > 0$. Since $0 < \alpha/2 < 1$ and $\beta = 1$, then P sometimes is called the positive stable random variable. For a proof of (2.4), we refer readers to [17] and [30, p. 15].

3. EM Algorithm for SαS

In this section, we introduce two approaches to extend the EM algorithm for SαS distribution. Presenting the first algorithm in Subsection 3.1, we consider a computational method for the first approach in Subsection 3.2. The second approach is presented in Subsection 3.3.

3.1. The first approach

Let $\underline{y} = (y_1, \ldots, y_n)$ be a random sample from a SαS distribution. Refer to (2.3) and consider $\underline{p} = (p_1, \ldots, p_n)$ as the vector of missing observations. The complete data likelihood $L_c(\Theta)$ is factored into the product of the marginal densities of P_i and the conditional densities of Y_i given P_i. So, the complete data log-likelihood can be written as

$$l_c(\Theta) = l_{1c}(\theta) + l_{2c}(\alpha),$$

where $\Theta = (\theta^T, \alpha)^T$, $\theta = (\mu, \sigma)^T$. It follows that

$$l_{1c}(\theta) = C - n\log\sigma - \frac{1}{2}\sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma^2 p_i},$$

and

$$l_{2c}(\alpha) = \sum_{i=1}^n \log f_P(p_i|\alpha),$$

where C is a constant independent of the parameter θ and $f_P(.)$ denotes the density function of the positive stable random variable P. Conditional expectation of the complete data log-likelihood $Q\left(\Theta|\Theta^{(t)}\right) = E\left(l_c(\Theta; \underline{p})|\underline{y}, \Theta^{(t)}\right)$ is given by

$$Q\left(\Theta|\Theta^{(t)}\right) = Q\left(\theta|\Theta^{(t)}\right) + Q\left(\alpha|\Theta^{(t)}\right),$$

$$= C - n\log\sigma - \frac{1}{2}\sum_{i=1}^n \frac{(y_i - \mu)^2 e_i^{(t,s)}}{2\sigma^2} + \sum_{i=1}^n E\left[\log f_P(P_i|\alpha)\Big|y_i, \Theta^{(t)}\right],$$

where $e_i^{(t,s)} = E\left(P^{-1}|y_i, \mu^{(t)}, \sigma^{(s)}, \alpha^{(s)}\right)$, $\Theta^{(t)} = (\theta^{(t)T}, \alpha^{(t)})^T$ and $\theta^{(t)} = (\mu^{(t)}, \sigma^{(t)})^T$. Here, $e_i^{(t,s)}$ is the posterior mean of $1/P$ given y_i and $\Theta^{(t)}$; for $i = 1, \ldots, n$. Thus, the EM algorithm's steps are

- E step: Given current estimation of Θ, $\Theta^{(t)}$, compute $e_i^{(t,t)}$.
- CM step: Update $\theta^{(t)}$ as $\theta^{(t+1)}$ by maximizing $Q\left(\Theta|\Theta^{(t)}\right)$ with respect to θ to obtain

$$\mu^{(t+1)} = \frac{\sum_{i=1}^{n} y_i e_i^{(t,t)}}{\sum_{i=1}^{n} e_i^{(t,t)}}, \tag{3.1}$$

$$\sigma^{(t+1)} = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - \mu^{(t+1)}\right)^2 e_i^{(t+1,t)}}{2n}}. \tag{3.2}$$

- CML step: Update value of $\alpha^{(t)}$ by maximizing the marginal log-likelihood function as

$$\alpha^{(t+1)} = \arg\max_{\alpha} \sum_{i=1}^{n} \log f_Y(y_i|\mu^{(t+1)}, \sigma^{(t+1)}, \alpha),$$

where $f_Y(.)$ is the $S\alpha S$ density function. Under the methodology that we follow here, the CML step itself contains an SEM algorithm (This is because the density function of $S\alpha S$ random variable has no closed-form expression). In the following, two steps of the SEM are described.

- The first step of CML: Let $E_1, E_2, \ldots, E_n$ be independent exponentially distributed random variables with mean one. At $(t+1)$-th iteration, assume that $\mu^{(t+1)}$ and $\sigma^{(t+1)}$ are known. Let $Y_i' = (Y_i - \mu^{(t+1)})/\sigma^{(t+1)}$ and $Y_i'' = Y_i'/\sqrt{E_i}$; for $i = 1, \ldots, n$. Then it follows, from (2.2) and (2.4), that

$$Y_i''|W_i = w_i \sim N\left(0, \frac{2}{w_i^2}\right),$$
$$W_i \sim W(\alpha).$$

Considering $w_1, \ldots, w_n$ as missing observations, the complete data likelihood $L_c(\Theta)$ is factored into the product of the marginal densities of W_i and the conditional densities of Y_i'' given W_i. So, the complete data log-likelihood can be written as

$$l_c(\alpha) = \mathrm{C} + n\log\alpha + (\alpha - 1)\sum_{i=1}^{n} \log w_i - \sum_{i=1}^{n} w_i^{\alpha}. \tag{3.3}$$

- The second step of CML: Simulate the vector $\underline{w} = (w_1, \ldots, w_n)$ from conditional distribution of W_i given y_i''; for $i = 1, \ldots, n$, and $\Theta^{(t)}$ using the Monte Carlo method at the m-th cycle of the S step of the SEM.
- The third step of CML: Substitute $\underline{w}$ in (3.3) and maximize it with respect to α to find $\alpha^{(t+1)}$. Then, repeat the algorithm from E step for M times to obtain $\Theta^{(t)}$ as

$$\hat{\Theta} = \frac{1}{M - M_0} \sum_{t=M_0+1}^{M} \Theta^{(t)},$$

where M_0 is burn-in size. Maximizing (3.3) with respect to α in each cycle is equal to computing the ML estimation of a Weibull distribution with shape parameter α.

3.2. Computational considerations of the first approach

In the E step of the algorithm, we need to calculate the conditional mean $e_i^{(t,t)}$ and $e_i^{(t+1,t)}$ in iteration steps (3.1) and (3.2), respectively. Unfortunately, this quantity cannot be calculated analytically. At $(t+1)$-th iteration of the E step, we have

$$
e_i^{(t,t)} = \frac{\int_0^\infty \frac{u^{-3/2} f_P(u|\alpha^{(t)})}{\sqrt{4\pi}\sigma^{(t)}} \exp\left\{ -\left(\frac{y_i - \mu^{(t)}}{2\sqrt{u}\sigma^{(t)}}\right)^2 \right\} du}{\int_0^\infty \frac{u^{-1/2} f_P(u|\alpha^{(t)})}{\sqrt{4\pi}\sigma^{(t)}} \exp\left\{ -\left(\frac{y_i - \mu^{(t)}}{2\sqrt{u}\sigma^{(t)}}\right)^2 \right\} du},
\tag{3.4}
$$

and after obtaining $\mu^{(t+1)}$ in the CM step, by replacing $\mu^{(t+1)}$ in (3.4) we obtain $e_i^{(t+1,t)}$. Then, we update $\sigma^{(t+1)}$ in the CM step. Direct computation of $e_i^{(t,t)}$ using numerical integration faces with problem when $|y_i|$, for $i = 1, \ldots, n$, is large. To avoid computational issues, we use the series representation for the density function $f_P(.|\alpha^{(t)})$ of $S(\alpha^{(t)}/2, 1, (\cos(\pi\alpha^{(t)}/4))^{2/\alpha^{(t)}}, 0)$ as

$$
f_P(p|\alpha^{(t)}) \simeq \frac{1}{\pi} \sum_{j=1}^k (-1)^{j-1} \frac{\Gamma(\frac{j\alpha^{(t)}}{2} + 1)}{\Gamma(j+1)} \sin\left(\frac{j\pi\alpha^{(t)}}{2}\right) p^{-\frac{j\alpha^{(t)}}{2} - 1},
\tag{3.5}
$$

for large k, [10] and [25]. Substituting the right-hand side of (3.5) into (3.4), after a suitable change of variable and simplifications, we get

$$
e_i^{(t,t)} \simeq \frac{\sum_{j=1}^k \frac{(2\sigma^{(t)})^{j\alpha^{(t)}+2}}{|y_i - \mu^{(t)}|^{j\alpha^{(t)}+2}} \frac{(-1)^{j-1}\Gamma\left(\frac{j\alpha^{(t)}}{2}+1\right)\Gamma\left(\frac{j\alpha^{(t)}}{2}+\frac{3}{2}\right)\sin\left(\frac{j\pi\alpha^{(t)}}{2}\right)}{\Gamma(j+1)}}{\sum_{j=1}^k \frac{(2\sigma^{(t)})^{j\alpha^{(t)}}}{|y_i - \mu^{(t)}|^{j\alpha^{(t)}}} \frac{(-1)^{j-1}\Gamma\left(\frac{j\alpha^{(t)}}{2}+1\right)\Gamma\left(\frac{j\alpha^{(t)}}{2}+\frac{1}{2}\right)\sin\left(\frac{j\pi\alpha^{(t)}}{2}\right)}{\Gamma(j+1)}},
\tag{3.6}
$$

where $\Gamma(.)$ denotes the well-known gamma function. It should be noted that series in numerator and denominator of (3.6) converge for

$$
\left| y_i - \mu^{(t)} \right| \in 2\sigma\left(\left[\frac{\Gamma(k\alpha^{(t)}/2 + \alpha^{(t)}/2 + 1)\Gamma(k\alpha^{(t)}/2 + \alpha^{(t)}/2 + 3/2)}{(k+1)\Gamma(k\alpha^{(t)}/2 + 1)\Gamma(k\alpha^{(t)}/2 + 3/2)} \right]^{1/\alpha^{(t)}}, \infty \right),
$$

where k is sufficiently large. This makes the computations to be fast. We found that choosing $k = \left[\min\left(168, 168/\alpha^{(t)}\right) \right]$, where $[x]$ denotes the integer part of x, yields an almost exact evaluation of $f_P(.|\alpha^{(t)})$ in (3.5) and so of $e_i^{(t,t)}$ in (3.6). For those y_i that series in (3.5) does not converge, we use Monte Carlo based approximation of $e_i^{(t,t)}$ given by

$$
e_i^{(t,t)} \simeq \left(\sum_{j=1}^N \frac{P_j^{-3/2}}{\sqrt{4\pi}\sigma^{(t)}} \exp\left\{ -\left(\frac{y_i - \mu^{(t)}}{2\sqrt{P_j}\sigma^{(t)}}\right)^2 \right\} \right) \div \left(\sum_{j=1}^N \frac{P_j^{-1/2}}{\sqrt{4\pi}\sigma^{(t)}} \exp\left\{ -\left(\frac{y_i - \mu^{(t)}}{2\sqrt{P_j}\sigma^{(t)}}\right)^2 \right\} \right),
$$

where $P_j \sim S(\alpha^{(t)}/2, 1, (\cos(\pi\alpha^{(t)}/4))^{2/\alpha^{(t)}}, 0)$; for $j = 1, \ldots, N$, is simulated using the method given in [5]. During our simulation, we set $N = 2000$.

As the CML step itself contains a SEM algorithm, simulation from conditional distribution of W_i given $y_i'' = (y_i - \mu^{(t+1)})/(\sigma^{(t+1)}\sqrt{e_i})$ (e_i, here, denotes an observation from

standard exponential distribution) and $\alpha^{(t)}$ can be performed using Metropolis-Hasting algorithm. We follow our simulation based on Weibull distribution with the shape parameter $\alpha^{(t)}$ as the candidate. The acceptance rate, A_{w_i}, becomes

$$A_{w_i} = \min \left\{ 1, \frac{w_i^{\text{new}} \exp\left\{ -\frac{(y_i'' w_i^{\text{new}})^2}{4} \right\}}{w_i^{(t)} \exp\left\{ -\frac{(y_i'' w_i^{(t)})^2}{4} \right\}} \right\}.$$

The main drawback of the Metropolis-Hasting algorithm is that the new values may have little chance for acceptance in each iteration. In our study, the acceptance rate decays when y_i'' gets large. Due to the heavy tail property of stable distribution, large values are likely to happen. Therefore, we adopt a rejection sampling in m-th cycle of the SEM algorithm of the CML step to generate from the posterior distribution of W_i given y_i'' and $\alpha^{(t+1)}$; for $i = 1, \ldots, n$. To describe this scheme, we note that the density function

$$f_{Y_i''|W_i}(y_i''|w_i) = \frac{w_i}{\sqrt{4\pi}} \exp\left\{ -\frac{(y_i'' w_i)^2}{4} \right\},$$

as a part of the conditional (posterior) density function

$$f_{W_i|Y_i''}(w_i|y_i'', \alpha^{(t+1)}) \propto f_{Y_i''|W_i}(y_i''|w_i) f_{W_i}(w_i) = \frac{\alpha^{(t+1)} w_i^{\alpha^{(t+1)}}}{\sqrt{4\pi}} \exp\left\{ -\frac{(y_i'' w_i)^2}{4} - w_i^{\alpha^{(t+1)}} \right\},$$

is bounded by some constant independent of w_i. More precisely

$$f_{Y_i''|W_i}(y_i''|w_i) \leq \frac{\exp\{-0.5\}}{\sqrt{2\pi}|y_i''|}.$$

Hence, a rejection sampling is employed to generate from the posterior distribution by the following steps.

1. Simulate a sample from a Weibull distribution, say w_i, with shape parameter $\alpha^{(t+1)}$ and scale unity.
2. Generate a sample from a uniform distribution $U\left(0, \exp\{-0.5\}/(\sqrt{2\pi}|y_i''|)\right)$, say u.
3. If $u < \frac{w_i}{\sqrt{4\pi}} \exp\left\{ -\frac{y_i''^2 w_i^2}{4} \right\}$, then accept w_i as a generation from $f_{W_i|Y_i''}(w_i|y'')$; otherwise, go to 1.

While we expect that the rejection sampling works satisfactorily, it faces with the problem when $|y_i''|$ is so large. In this case that the machine sets the quantity $\frac{w_i}{\sqrt{4\pi}} \exp\left\{ -\frac{y_i''^2 w_i^2}{4} \right\}$ to zero and consequently generation from posterior density is so time consuming, we use the mode of the posterior density as a generation from the posterior distribution. It is not hard to check that the mode of the posterior distribution $W_i|Y_i'' = y_i''$ is obtained as a solution of the equation

$$h(w_i) = \alpha^{(t)} - \left(\frac{w_i y_i''}{\sqrt{2}} \right)^2 - \alpha^{(t)} w_i^{\alpha^{(t)}} = 0. \tag{3.7}$$

Our studies reveal that if $|y_i''|$ is so large that machine sets the quantity $\frac{w_i}{\sqrt{4\pi}} \exp\left\{ -\frac{y_i''^2 w_i^2}{4} \right\}$ to zero, then choosing the mode of posterior distribution described by (3.7), as an observation

from this distribution not only improves the accuracy of the simulation, but also accelerates the EM algorithm implementation.

3.3. The second approach

The first approach works efficiently, but it can be improved to reduce computational complexity by relaxing some CM steps for computing $\mu^{(t+1)}$ and $\sigma^{(t+1)}$. This approach significantly reduces the processing times for implementing the EM algorithm. In the CM step of the second approach, we only update μ while both σ and α are updated through the SEM in the CML step. Like the previous approach, the location parameter is updated through (3.1). Let $E_1, E_2, \ldots, E_n$ be independent exponentially distributed random variables with mean one. At $(t+1)$-th iteration of the first step of the CML, assume that $\mu^{(t+1)}$ is known. Let $Y_i' = (\boldsymbol{Y}_i - \boldsymbol{\mu}^{(t+1)})$ and $Y_i'' = Y_i'/\sqrt{E_i}$; for $i = 1, \ldots, n$. It follows that

$$Y_i'' | W_i = w_i \sim N\left(0, \frac{2\sigma^2}{w_i^2}\right),$$

$$W_i \sim W(\alpha).$$

The complete data log-likelihood can be written as

$$l_c(\alpha, \sigma) = \mathrm{C} - n \log \sigma - \frac{1}{2} \sum_{i=1}^{n} \frac{(y_i'' w_i)^2}{2\sigma^2} + n \log \alpha + \alpha \sum_{i=1}^{n} \log w_i - \sum_{i=1}^{n} w_i^{\alpha}. \qquad (3.8)$$

By maximizing (3.8) with respect to σ and α, we find

$$\sigma^{(t+1)} = \sqrt{\frac{\sum_{i=1}^{n} (y_i'' w_i)^2}{2n}}, \qquad (3.9)$$

and $\alpha^{(t+1)}$ is the solution of

$$h(w_i) = \frac{n}{\alpha} + \sum_{i=1}^{n} \log w_i - \sum_{i=1}^{n} w_i^{\alpha} \log w_i = 0, \qquad (3.10)$$

where w_i is simulated from conditional distribution of W_i given y_i'' and $\Theta^{(t)}$; for $i = 1, \ldots, n$, using rejection sampling at the m-th cycle of the SEM step. Once $\sigma^{(t)}$ and $\alpha^{(t)}$ are updated, we implement the E step and repeat the algorithm for a number of M times and obtain $\hat{\Theta}$ as

$$\hat{\Theta} = \frac{1}{M - M_0} \sum_{t=M_0+1}^{M} \Theta^{(t)}.$$

In the first approach, we need to compute $e_i^{(t,t)}$ for updating $\mu^{(t)}$ and $e_i^{(t+1,t)}$ for updating $\sigma^{(t)}$; for $i = 1, \ldots, n$. But in the second approach, the sequence of $e_i^{(t,t)}$ updates $\mu^{(t)}$ and $\sigma^{(t)}$ will be updated in the second step of the CML. This yields a considerable reduction in implementing time of the EM algorithm compared with the first approach.

4. EM Algorithm for the Multivariate case

Let $\boldsymbol{Y}$ denotes a sub-Gaussian SαS random vector as defined in (1.2). The representation

$$\boldsymbol{Y} = \boldsymbol{\mu} + \sqrt{\Lambda}\boldsymbol{Z}, \qquad (4.1)$$

holds when $\boldsymbol{Z}$ is a zero-mean Gaussian random vector with some $d \times d$ covariance matrix Σ, Λ follows $S(\alpha/2, (\cos(\pi\alpha/4))^{2/\alpha}, 1, 0)$, $\boldsymbol{\mu} \in \mathbb{R}^d$ is the location parameter and $0 < \alpha < 2$. Here, Λ and $\boldsymbol{Z}$ are statistically independent. The stochastic representation (4.1) is the multivariate counterpart of the representation given in (2.2). Hence, methodology introduced in the previous section can be applied to the sub-Gaussian SαS distribution. We have

$$\boldsymbol{Y}|\Lambda = \lambda \sim N(\boldsymbol{\mu}, \lambda\Sigma),$$
$$\Lambda \sim S(\alpha/2, (\cos(\pi\alpha/4))^{2/\alpha}, 1, 0).$$

By the followings, we summarize the EM algorithm for a sub-Gaussian SαS distribution based on the second approach presented in the Subsection 3.3. Let $\underline{\boldsymbol{y}} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ denote a vector of n i.i.d. samples from a sub-Gaussian SαS distribution. The complete data log-likelihood is

$$l_c(\Theta) = l_c(\alpha, \boldsymbol{\mu}, \Sigma) = \mathrm{C} - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}\frac{(\boldsymbol{y}_i - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu})}{\lambda_i} + \sum_{i=1}^{n}\log f_\Lambda(\lambda_i),$$

where $f_\Lambda(.)$ denotes the density function of positive stable random variable Λ. The conditional expectation of the complete data log-likelihood $Q\left(\Theta|\Theta^{(t)}\right) = E\left(l_c(\Theta; \underline{\lambda})|\underline{\boldsymbol{y}}, \Theta^{(t)}\right)$ is given by

$$Q\left(\Theta\Big|\Theta^{(t)}\right) = \mathrm{C} - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu})e_i^{(t,s)} + \sum_{i=1}^{n}E\left[\log f_\Lambda(\lambda_i|\alpha)\Big|\boldsymbol{y}_i, \Theta^{(t)}\right],$$

where $e_i^{(t,s)} = E\left(\Lambda^{-1}|\boldsymbol{y}_i, \boldsymbol{\mu}^{(t)}, \Sigma^{(s)}, \alpha^{(s)}\right)$. So, at the $(t+1)$-th iteration of the CM step we obtain

$$\boldsymbol{\mu}^{(t+1)} = \frac{\sum_{i=1}^{n}\boldsymbol{y}_i e_i^{(t,t)}}{\sum_{i=1}^{n}e_i^{(t,t)}}. \qquad (4.2)$$

Let $E_1, E_2, \ldots, E_n$ be independent exponentially distributed random variables with mean one. At $(t+1)$-th iteration, for known $\boldsymbol{\mu}^{(t+1)}$, define $\boldsymbol{Y}_i' = (\boldsymbol{Y}_i - \boldsymbol{\mu}^{(t+1)})$ and $\boldsymbol{Y}_i'' = \boldsymbol{Y}_i'/\sqrt{E_i}$; for $i = 1, \ldots, n$. We have that

$$\boldsymbol{Y}_i''|W_i = w_i \sim N\left(\boldsymbol{0}, w_i^{-2}\Sigma\right),$$

where W_is are independent and $f_{W_i}(w) = \alpha w^{\alpha-1}\exp\{-w^\alpha\}$; for $i = 1, \ldots, n$. The complete data log-likelihood becomes

$$l_c(\alpha, \Sigma) = \mathrm{C} - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i'')^T\Sigma^{-1}(\boldsymbol{y}_i'')w_i^2 + n\log\alpha + \alpha\sum_{i=1}^{n}\log w_i - \sum_{i=1}^{n}w_i^\alpha, \quad (4.3)$$

where $\boldsymbol{y}_i'' = (\boldsymbol{y}_i' - \boldsymbol{\mu}^{(t+1)})/E_i$. By simulating $\underline{w}$ from conditional distribution of W_i given $\boldsymbol{y}_i''$ and $\alpha^{(t)}$; for $i = 1, \ldots, n$, at the m-th cycle of the SEM, we follow some iterated steps

to find Σ and α in the M step of the SEM algorithm that maximize the pseudo-complete log-likelihood function. Hence, the updated Σ becomes

$$\Sigma^{(t+1)} = \frac{\sum_{i=1}^{n} \boldsymbol{y}_i''(\boldsymbol{y}_i'')^T w_i^2}{n}, \tag{4.4}$$

and $\alpha^{(t+1)}$ is a solution of

$$h(\alpha) = \frac{n}{\alpha} + \sum_{i=1}^{n} \log w_i - \sum_{i=1}^{n} w_i^\alpha \log w_i = 0. \tag{4.5}$$

Updating $\boldsymbol{\mu}^{(t+1)}$, $\Sigma^{(t+1)}$ and $\alpha^{(t+1)}$ are implemented by (4.2), (4.4), and (4.5), respectively for M times.

Corollary 4.1. The updated dispersion matrix $\Sigma^{(t+1)}$ in (4.4) is positive definite.

4.1. Computational considerations

At $(t+1)$-th iteration, updating $\boldsymbol{\mu}^{(t)}$ in (4.2) needs computing $e_i^{(t,t)}$. For this, when $D_i^{(t)} = (\boldsymbol{y}_i - \boldsymbol{\mu}^{(t)})^T \left(\Sigma^{(t)}\right)^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}^{(t)})$ is large, $e_i^{(t,t)}$ is evaluated approximately as

$$e_i^{(t,t)} \simeq \frac{\sum_{j=1}^{k} \frac{(-1)^j}{\Gamma(j+1)} \Gamma\left(\frac{j\alpha^{(t)}+2}{2}\right) \Gamma\left(\frac{j\alpha^{(t)}+d+2}{2}\right) \sin\left(\frac{j\pi\alpha^{(t)}}{2}\right) \left(D_i^{(t)}\right)^{-\frac{j\alpha^{(t)}+d+2}{2}}}{\sum_{j=1}^{k} \frac{(-1)^j}{\Gamma(j+1)} \Gamma\left(\frac{j\alpha^{(t)}+2}{2}\right) \Gamma\left(\frac{j\alpha^{(t)}+d}{2}\right) \sin\left(\frac{j\pi\alpha^{(t)}}{2}\right) \left(D_i^{(t)}\right)^{-\frac{j\alpha^{(t)}+d}{2}}}, \tag{4.6}$$

where

$$D_i^{(t)} \in \left(\left[\frac{\Gamma\left(\frac{k\alpha^{(t)}+\alpha^{(t)}+2}{2}\right) \Gamma\left(\frac{k\alpha^{(t)}+\alpha^{(t)}+2+d}{2}\right)}{(k+1)\Gamma\left(\frac{k\alpha^{(t)}+2}{2}\right) \Gamma\left(\frac{k\alpha^{(t)}+2+d}{2}\right)}\right]^{\frac{2}{\alpha}}, \infty\right), \tag{4.7}$$

and $k = \left[\min\left(168, 168/\alpha^{(t)}\right)\right]$ in (4.6). If $D_i^{(t)}$ does not lie in (4.7), we use the Monte Carlo based approximation of $e_i^{(t,t)}$ as

$$e_i^{(t,t)} \simeq \left(\sum_{j=1}^{N} \Lambda_j^{-d/2-1} \exp\left\{-\frac{D_i^{(t)}}{2\Lambda_j}\right\}\right) \div \left(\sum_{j=1}^{N} \Lambda_j^{-d/2} \exp\left\{-\frac{D_i^{(t)}}{2\Lambda_j}\right\}\right),$$

where $\Lambda_j \sim S(\alpha^{(t)}/2, 1, (\cos(\pi\alpha^{(t)}/4))^{2/\alpha^{(t)}}, 0)$ for large N. As before, we set $N = 2000$ for implementing the EM algorithm. To construct the pseudo-complete data, w_is are generated in (4.3) through a rejection sampling by the following steps.

1. Simulate a sample from a Weibull distribution, say w_i, with shape parameter $\alpha^{(t+1)}$ and scale unity.
2. Define $b = \frac{d^{d/2}((\boldsymbol{y}_i'')^T \Sigma^{-1} \boldsymbol{y}_i'')^{-d/2} \exp\{-d/2\}}{(2\pi)^{d/2}|\Sigma|^{1/2}}$ and generate samples from a uniform distribution $U(0, b)$, say u.
3. If $u < \frac{w_i^d \exp\{-\frac{1}{2}((\boldsymbol{y}_i'')^T \Sigma^{-1} \boldsymbol{y}_i'')w_i^2\}}{(2\pi)^{d/2}|\Sigma|^{1/2}}$, then accept w_i as a generation from $f_{W_i|\boldsymbol{Y}_i''}(w_i|\boldsymbol{y}_i'')$; otherwise, go to 1.

if $(\boldsymbol{y}_i'')^T \Sigma^{-1} \boldsymbol{y}_i''$ is so large, we impute $w_i = \sqrt{d}\left((\boldsymbol{y}_i'')^T \Sigma^{-1} \boldsymbol{y}_i''\right)^{-1/2}$.

5. Simulation study and Real Data Analysis

This section has five parts. In the first part, we analyze the performance of the proposed EM algorithm compared with other estimators in the univariate case through a simulation study. In the second part, we follow the analyses for the performance of the proposed EM algorithm using real data. Simulation studies in the bivariate case are performed in the third subsection. The fourth subsection is devoted to real data analysis for a bivariate case. Simulation studies in three-variate and higher dimensions are given in the fifth subsection. The set of real data used in the univariate and multivariate cases is 9 years of daily returns of 22 major worldwide market indices, which consists of 2535 observations from 1/4/2000 to 9/22/2009. These indices are from four different areas including America (S&P500, NASDAQ, TSX, Merval, Bovespa and IPC), Europe and Middle East (AEX, ATX, FTSE, DAX, CAC40, SMI, MIB and TA100), and East Asia and Oceania (HgSg, Nikkei, StrTim, SSEC, BSE, KLSE, KOSPI and AllOrd). This data set is available at the website of Yahoo finance, [8]. For simplicity, hereafter, we call this data set WMI (Worldwide Market Indices).

5.1. Simulation study in univariate case

We compare the performance of the ML, SQ, CF, and the EM methods for estimating the parameters of SαS distribution through simulation. Comparisons are based on square root of the mean-squared error (RMSE) criterion. In the simulation, the RMSE was computed for sample sizes of $n = 200, 500$, and 1000. The number of replications is $N = 1000$. The parameters were selected as: $\mu = 0$, $\sigma = 0.5, 1, 2$, and $\alpha = 0.4, 0.6, 0.8, 0.9, 1.1, 1.2, 1.4, 1.6, 1.8, 1.9$. The plots of the computed RMSE for all estimators are shown in Figures 1-3. It should be noted that the estimation results for the evaluated ML approach (Nolan's Method [21]) are not reliable for $\alpha < 0.4$ and so are not shown in the Figures 1-3. The smoothed version of the actual values plotted in these figures is obtained using the lowess package (see [6]) with the following settings: the smoothing span is considered to be 2/3, the number of iterations is 3, and the speed of computations is determined by 0.01th of the range of the α values.

In each sub-figure of Figures 1-3, the vertical line indicates that the proposed EM algorithm works better than the CF and SQ methods for those α lie in the left side of the line. When the EM algorithm always works better, such a line is not shown. By considering the location parameter as zero, the following results are obtained from simulation and show how RMSE of the estimators varies in terms of σ, α and n:

1. EM-based $\hat{\alpha}$ and $\hat{\sigma}$ have better performance than the SQ-based and CF-based $\hat{\alpha}$ and $\hat{\sigma}$ for $\alpha < 1.1$ in all cases, where shown in Figures 1 and 2, respectively.
2. EM-based $\hat{\mu}$ shows better performance than the SQ and CF for all α, ref. Figure 3.
3. RMSE of the EM-based $\hat{\alpha}$ and $\hat{\sigma}$ are smaller than the SQ-based $\hat{\alpha}$ and $\hat{\sigma}$.
4. We know that the ML method gives the efficient estimation particularly when sample size gets large. But, sometimes the EM-based $\hat{\alpha}$ and $\hat{\sigma}$ outperform the ML counterparts. This occurs, because STABLE gives the approximated ML estimation, not the exact ML estimation.
5. RMSE of all estimators decreases by increasing the sample size n.
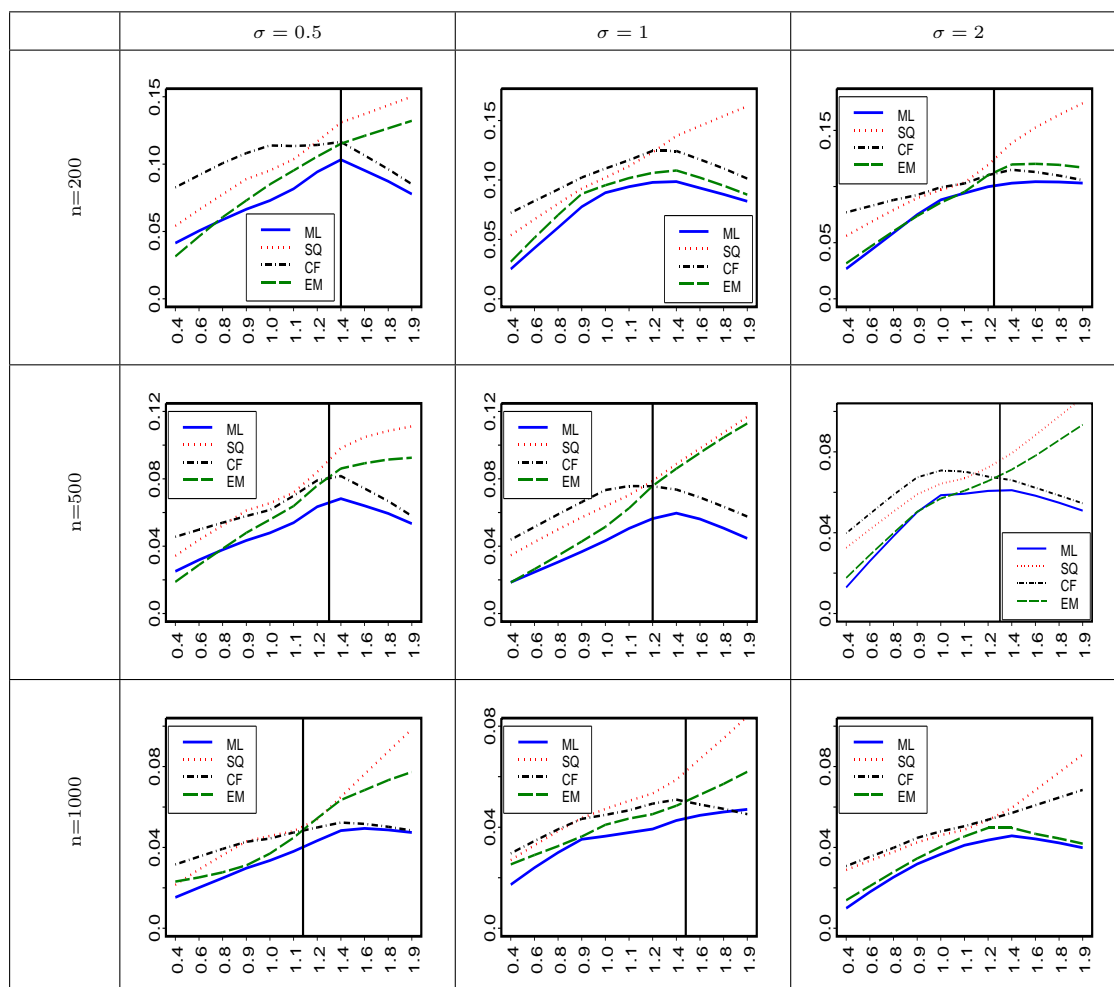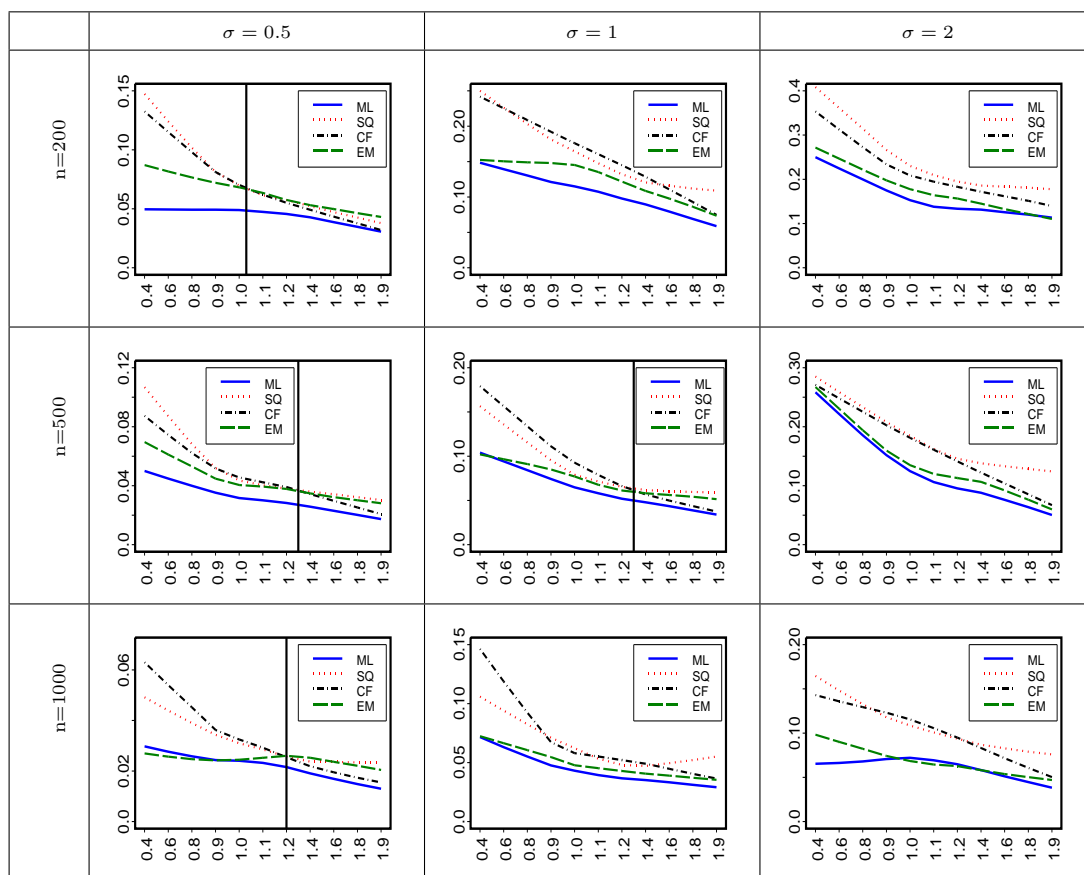
Fig. 1. RMSE of $\hat{\alpha}$ for different values of $\sigma = 0.5, 1, 2$ and for $0.4 \leq \alpha \leq 1.9$. In all sub-figures x-axis is values of α and y-axis is RMSE of $\hat{\alpha}$.

Along with the simulation, we investigate the execution time for different sample sizes and scale parameters in the univariate case. The results are given in Table 1. For this, all runs are performed on a machine with 3.5 GHz Core(TM) i7-2700K Intel(R) processor and 8 GB of RAM. It should be noted that implementing of the ML, CF, and SQ approaches, with settings given in Table 1, takes less than 0.5 second.

Table 1. Average of the execution time in seconds for 50 runs of the EM algorithm when it is applied to SαS distribution. The written program run on a 3.5 GHz Intel processor Core(TM) i7 using a 8 GB RAM. Note that values outside (inside) of parentheses are obtained for $\alpha = 0.5$ ($\alpha = 1.5$). We set $M = 120$, $M_0 = 70$ for implementing the EM algorithm.

Sample size	Scale parameter		
	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 2$
500	30(62)	26(64)	22(62)
1000	70(127)	55(122)	43(124)
5000	334(627)	283(639)	223(621)

Fig. 2. RMSE of $\hat{\sigma}$ for different values of $\sigma = 0.5, 1, 2$ and for $0.4 \leq \alpha \leq 1.9$. In all sub-figures x-axis is values of α and y-axis is RMSE of $\hat{\sigma}$.

5.2. Real data analysis in the univariate case

For real data application, we select the Bovespa (São Paulo Stock Exchange), CAC40 (Bourse de Paris) and DAX (German Stock Index) indices from WMI data set. The empirical distributions of these three sets of data show symmetric pattern around the origin whose tails are heavier than the normal model. We apply the ML, CF, SQ, and EM methods to estimate the parameters of the fitted stable distributions. It should be noted that the sample median is used as an initial value for the location parameter to implement the EM algorithm. Initial values for the tail and scale parameters are obtained by applying the method of the logarithmic moment. This approach, for a symmetric stable distribution, estimates the parameters using the first and the second order moments of the logarithm of a zero-location symmetric stable random variable Y, i.e. $L_1 = E(\log|Y|)$ and $L_2 = E(\log|Y| - L_1)^2$, respectively. Assuming that data, after subtracting from the sample median, follow a zero-location symmetric stable distribution, the parameters α and σ are estimated by equating the sample logarithmic moments to the quantities L_1 and L_2, [20] and [32]. Figure 4 displays cycles of the EM algorithm for $M = 120$ and $M_0 = 70$.

Estimated parameters, along with the Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) test statistics are given in Table 2. More analyses reveal that stable model captures
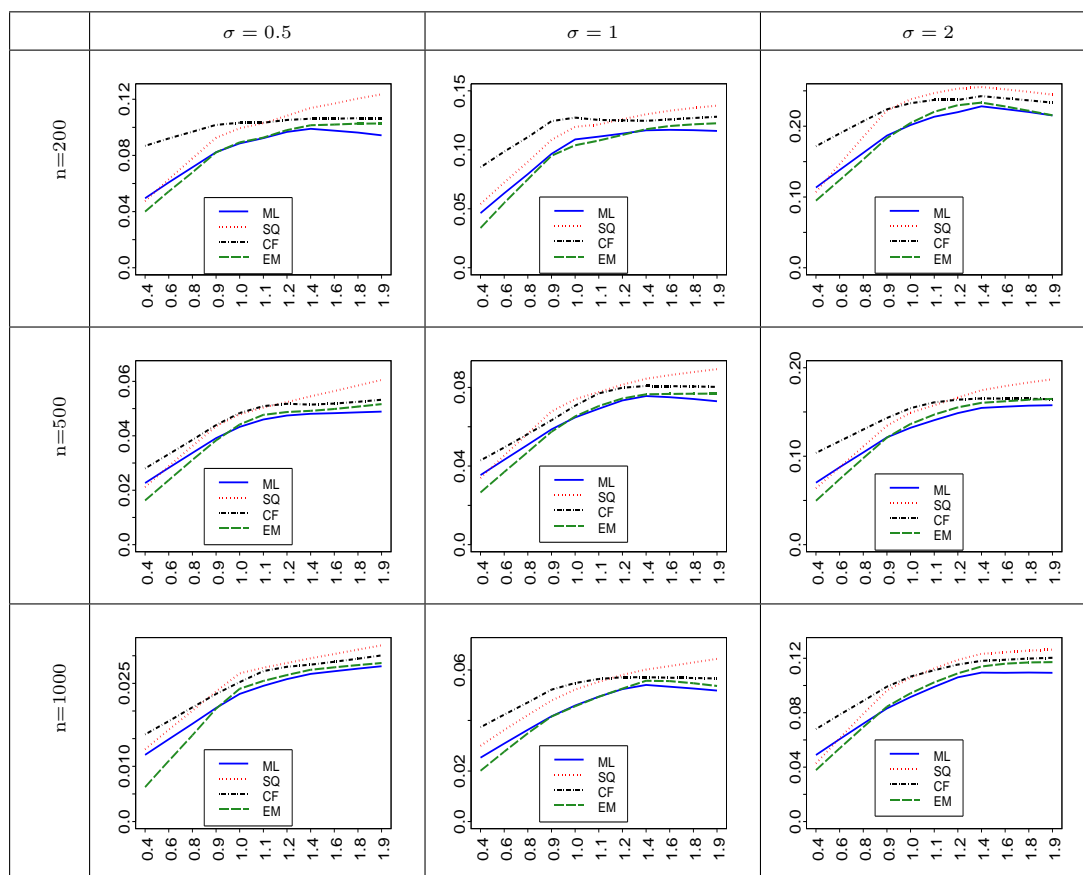
Fig. 3. RMSE of $\hat{\mu}$ for different values of $\sigma = 0.5, 1, 2$ and for $0.4 \leq \alpha \leq 1.9$. In all sub-figures x-axis is values of α and y-axis is RMSE of $\hat{\mu}$.

the general shape of the empirical histogram well and the main difference between fitnesses occurs at the origin for the height of peaks. As it is seen, EM algorithm gives satisfactory results.

5.3. Simulation study in bivariate case

Estimating the parameters of a sub-Gaussian SαS distribution is computationally expensive, [24]. Usually, the parameters of this distribution are estimated using projection method. Hereafter, we write ML, CF, and SQ for the projected ML, CF, and SQ methods, respectively. In the following, we perform a simulation based on a sample of size $n = 500$ to measure the performance of the ML, CF, EM, and SQ approaches in modelling a bivariate sub-Gaussian SαS distribution. For this, we set $\alpha = 1.5, 1.7, 1.9$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^T = (0, 0)^T$, and dispersion matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = 10^{-6} \times \begin{pmatrix} 5.9552 & 4.0783 \\ 4.0783 & 13.9861 \end{pmatrix},$$

as used in [27]. We follow the CF, SQ, and ML approaches by applying the STABLE software [21]. The boxplots, based on 250 runs for all methods, are depicted in Figure 6. It should
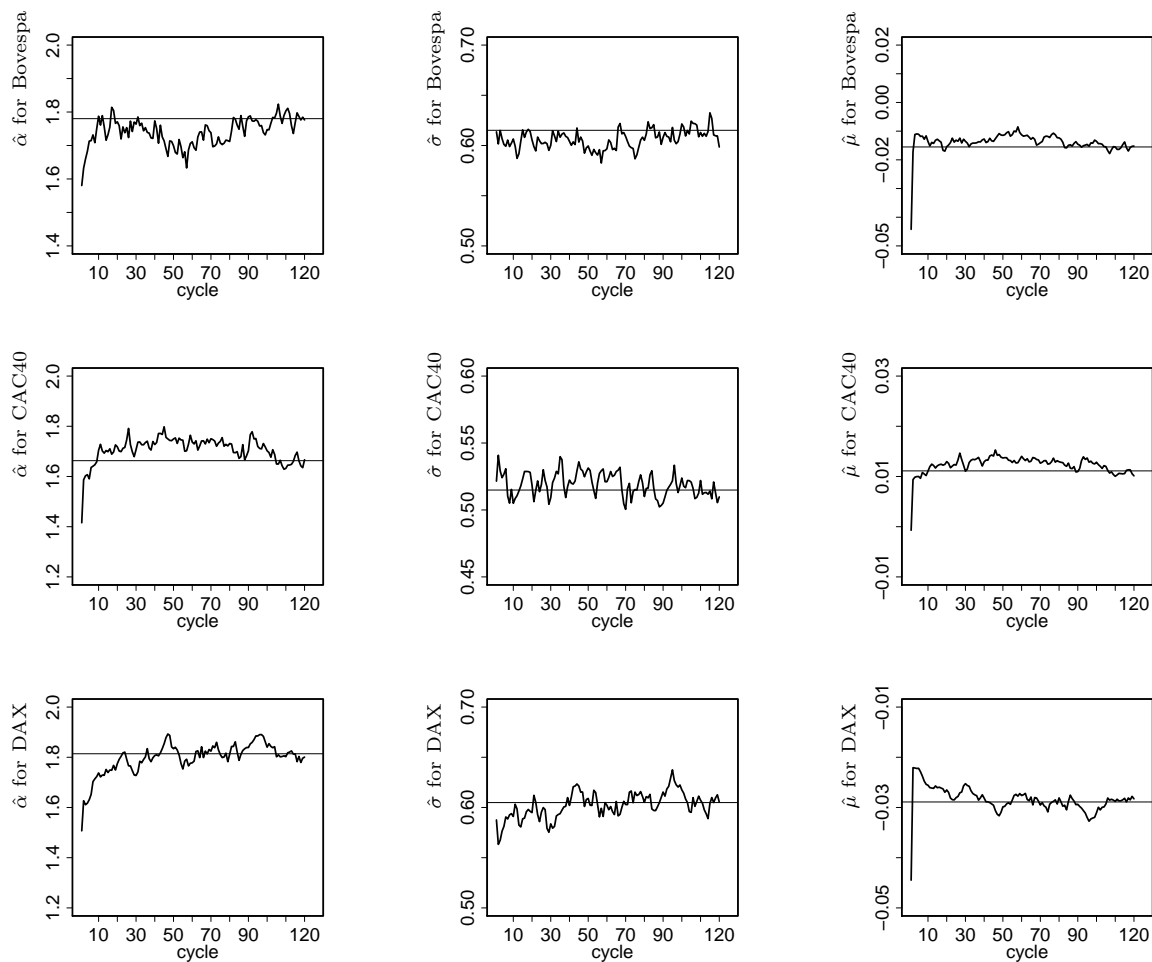
Fig. 4. Estimated values of α, σ, and μ with $M = 120$ cycles each with $n = 2535$ samples for Bovespa, CAC40 and DAX indices selected from WMI data set.

be noted that ML estimations are used here as starting values for EM algorithm. However, our investigations show that the proposed EM algorithm is robust with respect to initial values.

5.4. Real data analysis in bivariate case

Here, we apply ML, SQ, CF, and EM methods to S&P500 and IPC indices selected from WMI data set. The empirical distribution of both S&P500 and IPC indices show a symmetric pattern around the origin. The scatterplot shown in Figure 5 is roughly elliptical contoured. Furthermore, estimated tail parameters after fitting a stable distribution to S&P500 and IPC indices through the ML approach are 1.9003 and 1.9143, respectively, which are assumed to be equal. Thus, the assumption that $X = (S\&P500, IPC)^T$ follows a sub-Gaussian SαS distribution is acceptable, [23]. The results of estimating parameters through ML, CF, EM, and SQ methods are shown in Table 3. The log-likelihood values
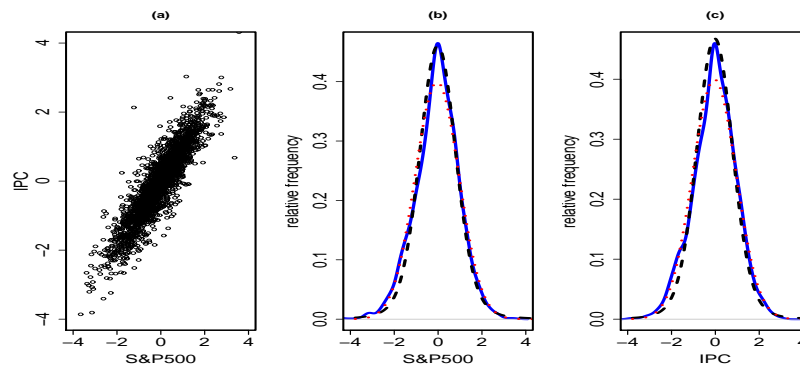
Fig. 5. Scatterplot of the IPC index versus S&P500 index (a) shows that distribution of points on a plane is roughly elliptical. Points in scatterplot are a number of $n = 2535$ daily returns from April 1, 2000 through September 22, 2009. Marginal density plots of the S&P500 (b) and IPC (c) show that the stable distribution (black dashed line) provides a better fit than the normal (red dotted line) distribution. Blue solid line is the fitted empirical density.

Table 2. Estimation results for the return of Bovespa, CAC40, and DAX data through the ML, SQ, CF, and EM approaches applied to the data.

	Bovespa				CAC40			
Parameters	ML	CF	EM	SQ	ML	CF	EM	SQ
α	1.8616	1.8976	1.78017	1.7544	1.7074	1.7898	1.6629	1.6076
β	-0.5908	-0.828	—	-0.3021	0.0578	0.3089	—	0.0974
σ	0.6407	0.6431	0.6151	0.6171	0.5461	0.5561	0.5148	0.5287
μ	-0.0572	-0.0541	-0.0155	-0.0805	0.0222	0.0487	0.0111	0.0193
K-S	0.0314	0.0302	0.0305	0.0388	0.0237	0.0251	0.0192	0.0191
A-D	1.7552	1.9720	2.4888	2.5601	1.2075	2.1783	1.6430	1.0461
	DAX							
Parameters	ML	CF	EM	SQ				
α	1.8541	1.8856	1.8140	1.6642				
β	-0.2978	-0.508	—	-0.1368				
σ	0.6401	0.6378	0.6048	0.6029				
μ	-0.0507	-0.0550	-0.0289	-0.0493				
K-S	0.0268	0.0251	0.0220	0.0171				
A-D	1.9657	2.1389	2.1374	1.1655				

indicate that the EM algorithm gives better performance than the CF and SQ approaches.

5.5. Analysis in higher dimensions

In this subsection, firstly we perform a simulation to compare the performance of the ML, CF, EM, and SQ methods where a sample of size $n = 1500$ is generated from three-variate sub-Gaussian SαS distribution. For this purpose, the considered settings are: $\alpha = 1.86$,
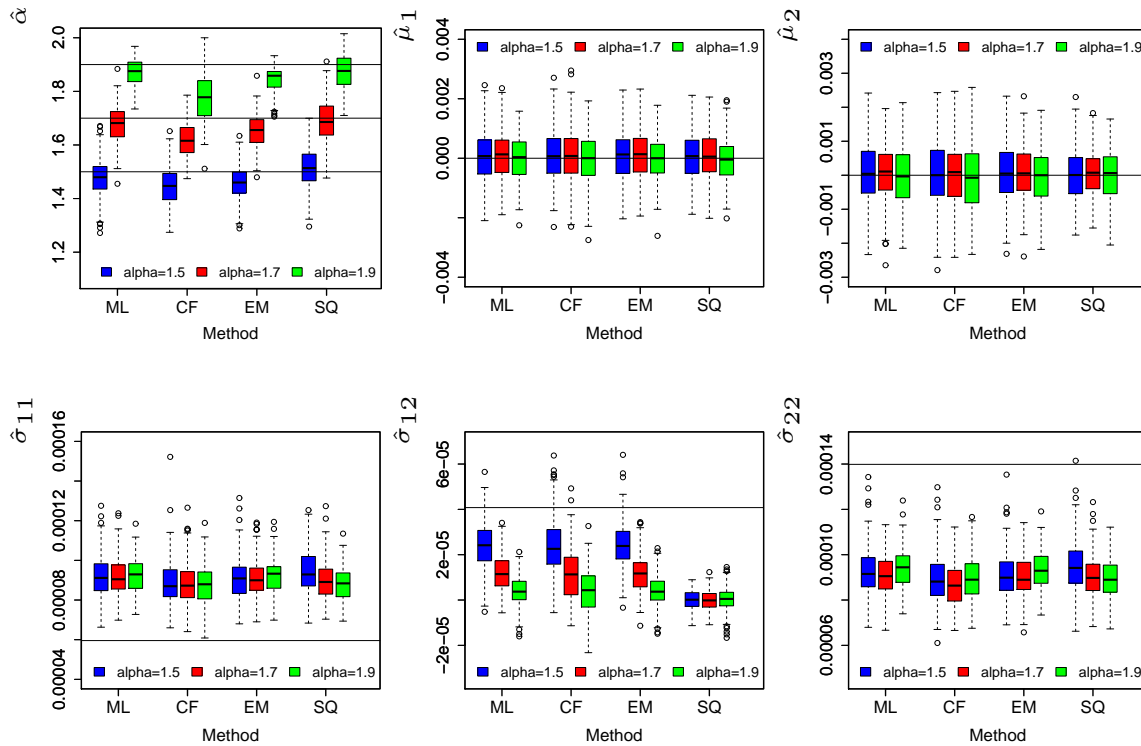
Fig. 6. Boxplot of the ML, CF, EM, and SQ estimations for α, $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$, and entries of Σ when $n = 500$ vectors generated from a bivariate sub-Gaussian SαS with $\alpha = 1.5, 1.7, 1.9$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^T = (0, 0)^T$, $\sigma_{11} = 0.000059552$, $\sigma_{12} = 0.000040783$, and $\sigma_{22} = 0.000139861$. In each sub-figure, horizontal line denotes the real value of parameter. Each boxplot is constructed based on $N = 250$ runs. The used color scheme under each method for boxplots is: blue, red, and green for $\alpha=1.5$, 1.7, and 1.9, respectively.

Table 3. Estimation results for the return of (S&P500, IPC) when the ML, CF, EM, and SQ approaches are applied to the data.

| Method | Parameters | | | Log-likelihood |
	α	Σ	$\boldsymbol{\mu}$	
ML	1.9073	$\begin{pmatrix} 0.45396 & 0.41991 \\ 0.41991 & 0.45626 \end{pmatrix}$	$\begin{pmatrix} 0.00528 \\ 0.00697 \end{pmatrix}$	-4967.706
CF	1.9530	$\begin{pmatrix} 0.45668 & 0.41957 \\ 0.41957 & 0.46008 \end{pmatrix}$	$\begin{pmatrix} 0.02505 \\ 0.02437 \end{pmatrix}$	-4978.067
EM	1.7116	$\begin{pmatrix} 0.37830 & 0.34573 \\ 0.34573 & 0.37589 \end{pmatrix}$	$\begin{pmatrix} -0.02494 \\ -0.02215 \end{pmatrix}$	-4975.581
SQ	1.8179	$\begin{pmatrix} 0.42751 & 0.35074 \\ 0.35074 & 0.43147 \end{pmatrix}$	$\begin{pmatrix} 0.01559 \\ 0.02313 \end{pmatrix}$	-5176.572

$\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^T = (0, 0, 0)^T$, and

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} = 10^{-5} \times \begin{pmatrix} 6.293 & 3.289 & 3.643 \\ 3.289 & 9.133 & 3.921 \\ 3.643 & 3.921 & 7.871 \end{pmatrix},$$
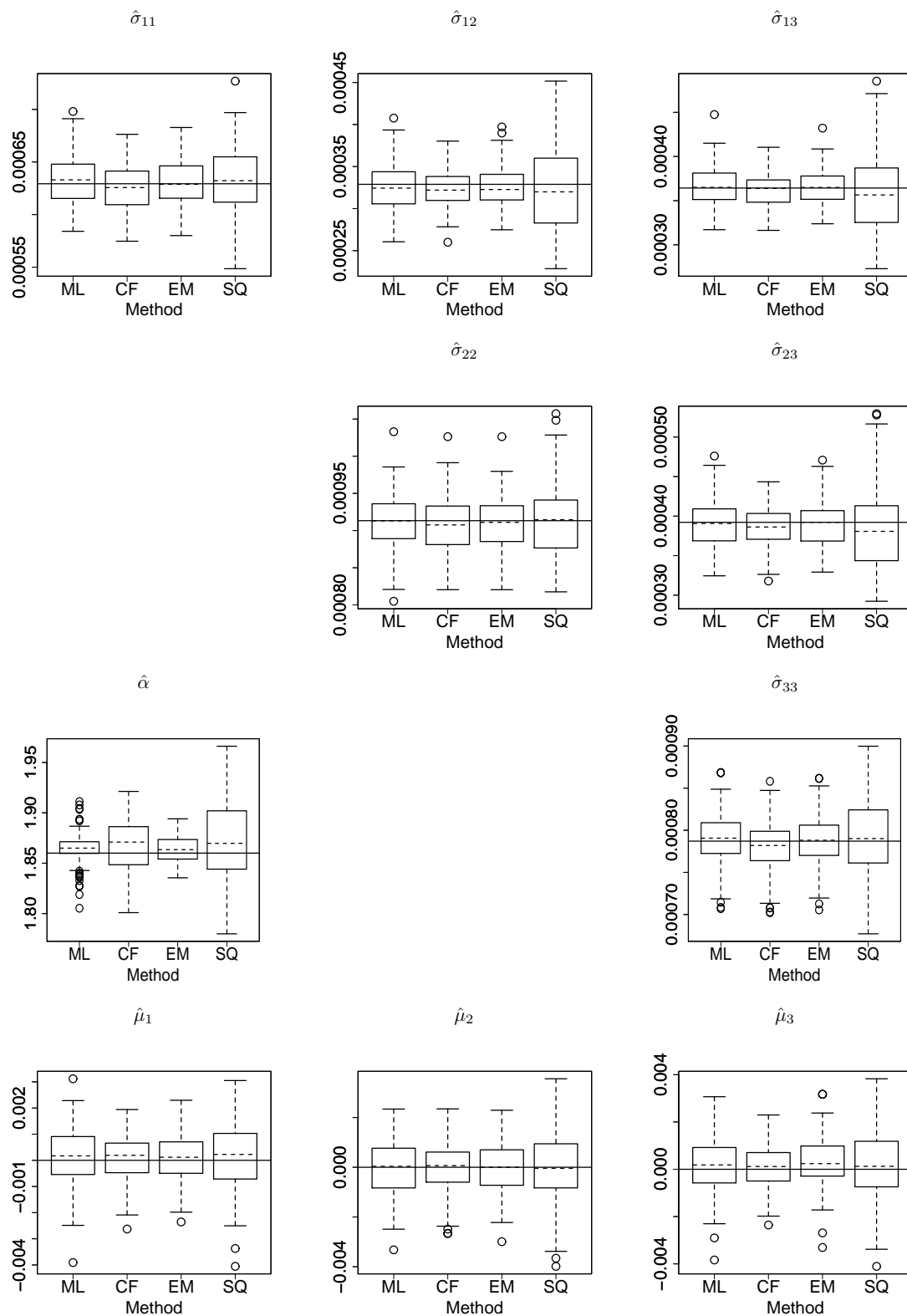
Fig. 7. Boxplot of the ML, CF, EM, and SQ estimations for α, $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^T$, and entries of Σ when $n = 1500$ vectors generated from a three-variate sub-Gaussian SαS with $\alpha = 1.86$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^T = (0, 0, 0)^T$, $\sigma_{11} = 0.0006293$, $\sigma_{12} = 0.0003289$, $\sigma_{13} = 0.0003643$, $\sigma_{22} = 0.0009133$, $\sigma_{23} = 0.0003921$, and $\sigma_{33} = 0.0007871$. In each sub-figure, horizontal line denotes the real value of parameter. Each boxplot is constructed based on $N = 250$ runs.
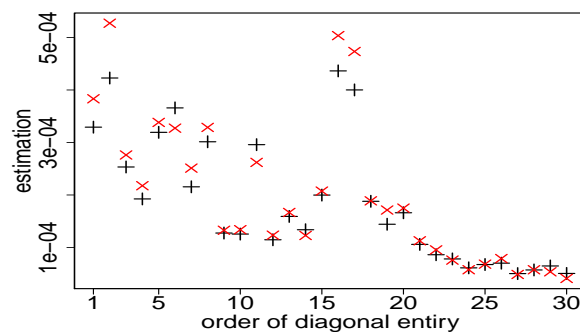
Fig. 8. Estimated values for diagonal entries of dispersion matrix when the EM and ML methods are applied to 30 indices of Dow Jones data. The signs $\times$ and $+$ correspond to the EM and ML methods.

which were used by [24]. Secondly, we fit a thirty-dimensional sub-Gaussian SαS distribution to the 30 indices of Dow Jones data described in [24].

For three-variate case, results of the simulation are displayed in Figure 7 where each boxplot is based on 250 runs. If we define the length of each box as a performance criterion, it follows that the EM algorithm works better than the SQ method in all cases. The performance of the EM and CF approaches is almost the same except in the case of estimating α where the EM algorithm performs much better than the CF method. In the case of case of thirty-dimensional, implementing the EM algorithm takes 303 seconds when $M = 200$ and $M_0 = 150$. To avoid unnecessary details, we only focus on estimating the location parameter and diagonal entries of the dispersion matrix using the EM and ML methods. The result for estimating these parameters are given in Figures 8-9, respectively. As it is seen, the EM algorithm gives results close to the ML approach for estimating diagonal entries of the dispersion matrix. The results of simulation show that the computational cost increases as dimensions increase but differences are not significant. For example, the execution time for implementing the EM and ML methods are given in Table 4. For this, we generated $n = 1000$ random vectors of a sub-Gaussian SαS distribution.

Table 4. Average of the execution time in seconds for 50 runs of the EM and ML methods where they are applied to a d-dimensional sub-Gaussian SαS distribution. The written program run on a 3.5 GHz Intel processor Core(TM) i7 using a 8 GB RAM. During simulation, we set $M = 200$, $M_0 = 150$, and Σ is a positive definite matrix whose eigenvalues are randomly generated from a uniform distribution on $(0, 2)$. The value inside of the parentheses corresponds to the ML method.

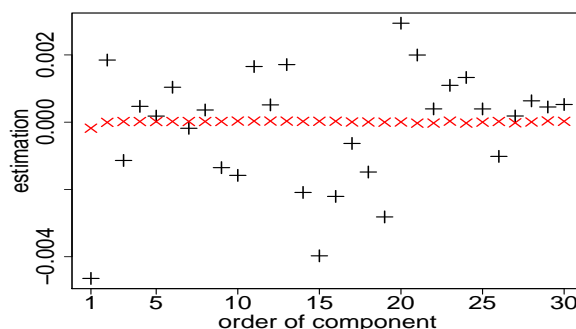| d	α=1.5	α=1
5	201(1.2)	185(0.6)
10	222(3.3)	200(3.5)
20	230(7.6)	223(7.8)
50	275(42)	258(40)
100	456(193)	389(173)

Fig. 9. Estimated values for location parameter when the EM and ML methods are applied to 30 indices of Dow Jones data. The signs × and + correspond to the EM and ML methods.

6. Concluding remarks

We propose an EM algorithm for estimating the parameters of SαS and sub-Gaussian SαS distributions. For SαS case, the performance of the proposed EM algorithm is compared with known approaches such as ML, CF, and SQ through a comprehensive simulation study. The comparisons are based on square root of the mean-squared error (RMSE). It follows that the proposed EM algorithm is a good competitor for the ML, CF, and SQ methods. Furthermore, in the univariate case, its performance is proved using three sets of real data. For the sub-Gaussian SαS case, the performance of the proposed EM algorithm is investigated via simulation and real data modelling. To avoid unnecessary details, we consider only two-, three-, and thirty-dimensional cases. The simulation reveals that the proposed EM algorithm provides satisfactory performance. Particularly, in two-dimensional case, it gives better performance than the CF and SQ methods in modelling a set of real data. The proposed EM algorithm has some advantages over other methods studied in this work. Among them we mention: 1- it works for all ranges of parameters. 2- In the multivariate case, it can estimate all entries of dispersion matrix simultaneously so that the estimated dispersion matrix is always positive definite. 3- it motivates further researches for many other distributions where dividing by an auxiliary random variable in the CML step, simplifies the M step of the EM algorithm. This idea can be applied for exponential power distribution. 4- it outperforms the CF and SQ methods in estimating the location parameter when α is near to one. 5- it can be developed for modelling the mixture of SαS distributions. This advantage of the proposed EM is a privilege as the evaluated ML, SQ and CF are not applicable to the mixture models. As possible future works, we are developing the proposed EM algorithm to estimate parameters of a stable, mixture of SαS, and mixture of sub-Gaussian SαS distributions.

Acknowledgments

written in R package for implementing EM algorithm is freely available from corresponding author on request.

References

[1] R. M. Basso, V. H. Lachos, C. R. B. Cabral and P. Ghosh, Robust mixture modeling based on scale mixtures of skew-normal distributions, Computational Statistics and Data Analysis, 54 (2010) 2926–2941.

[2] M. Broniatowski, G. Celeux and J. Diebolt, Reconnaissance de Mélanges de densités par un algorithme d'apprentissage probabiliste, Data Analysis and Informatics, 3 (1983) 359–374.

[3] G. Celeux and J. Diebolt, The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for mixture problem, Computational Statistics Quarterly, 2 (1) (1985) 73–82.

[4] G. Celeux and J. Diebolt, Une version de type recuit simulé de l'algorithme EM, Notes aux Comptes Rendus de l'Académie des Sciences, 310 (1990) 119–124.

[5] J. M. Chambers, C. L. Mallows and B. W. Stuck, A method for simulating stable random variables, Journal of the American Statistical Association, 71 (354) (1976) 340–344.

[6] W. S. Cleveland, LOWESS: A program for smoothing scatterplots by robust locally weighted regression, Journal of the American Statistical Association, 35 (1981) 54.

[7] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society, Series B, 39 (1977) 1–38.

[8] Y. Dominicy and D. Veredas, The Method of Simulated Quantiles, Journal of Econometrics, 172 (2013) 235–247.

[9] W. H. DuMouchel, Stable distributions in statistical inference, (Ph. D. Thesis, Univ. Ann Arbor, Univ. Microflms, Ann Arbor, Mich, 1971).

[10] W. Feller, An Introduction to Probability Theory and its Applications, (Volume 2, second edition, John Wiley, New York, 1971).

[11] E. H. S. Ip, A stochastic EM estimator for handling missing data, Unpublished Ph. D. Thesis, Department of Statistics, Stanford University, 1994.

[12] S. M. Kogon and D. B. Williams, Characteristic function based estimation of stable parameters, In: Adler, R. Feldman, R. and Taqqu, M. (Eds.), A Practical Guide to Heavy Tailed Data, (1988), 311–338, Birkhäuser, Boston.

[13] F. Lamantia, S. Ortobelli and S. T. Rachev, An empirical comparison among VaR models and time rules with elliptical and stable distributed returns, Investment Management and Financial Innovations, 3 (2006) 8–29.

[14] C. Liu and D. B. Rubin, The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence, Biometrika, 81 (1994) 633-648.

[15] J. H. McCulloch, Simple consistent estimators of stable distribution parameters, Communications in Statistics-Simulation and Computation, 5 (1986) 1109–1136.

[16] G. J. McLachlan and T. Krishnan, The EM Algorithm and Extensions, (second edition, John Wiley, New York, 2008).

[17] S. G. Meintanis, Test for exponentiality against Weibull and gamma decreasing hazard rate alternatives, Kybernetika, 43(3) (2007) 307–314.

[18] X. L. Meng and D. B. Rubin, Maximum likelihood estimation via the ECM algorithm: A general framework, Biometrika, 80 (1993) 267–278.

[19] S. Mittnik and M. S. Paolella, Prediction of financial downside risk with heavy tailed conditional distributions, (In Rachev, S. T. (Ed.), Handbook of Heavy Tailed Distributions in Finance, Elsevier Science, Amsterdam, 2003).

[20] C. L. Nikias and M. Shao, Signal Processing with α-Stable Distributions and Applications, (John Wiley, New York, 1995).

[21] J. P. Nolan, Maximum likelihood estimation of stable parameters, In Barndorff-Nielsen, O. E., Mikosch, T., and Resnick, I. (Eds.), Lévy Processes: Theory and Applications (2001) 379–400, Birkhöuser, Boston.

[22] J. P. Nolan, Modeling financial distributions with stable distributions, Volume 1 of Handbooks in Finance, (Chapter 3, pp. 105-130, Elsevier, Amsterdam, 2013).

[23] J. P. Nolan, Multivariate stable densities and distribution functions: general and elliptical case, (Deutsche Bundesbank's 2005 Annual Fall Conference).

[24] J. P. Nolan, Multivariate elliptically contoured stable distributions: theory and estimation, Computational Statistics, 28 (2013) 2067–2089.

[25] J. P. Nolan, Stable Distributions-Models for Heavy Tailed Data, (Birkhäuser, Boston, Unfinished manuscript, Chapter 1 online at academic2.american.edu/∼ jpnolan, 2016).

[26] S. Ortobelli, S. T. Rachev and F. J. Fabozzi, Risk management and dynamic portfolio selection with stable Paretian distributions, Journal of Empirical Finance, 17(2) (2010) 195–211.

[27] S. T. Rachev, Handbook of Heavy Tailed Distributions in Finance, (Elsevier, Amsterdam, 2003).

[28] S. T. Rachev and S. Mittnik, Stable Paretian Models in Finance, (John Wiley, New York, 2000).

[29] D. Salas-Gonzalez, E.E. Kuruoglu and D. P. Ruiz, Modelling with mixture of symmetric stable distributions using Gibbs sampling, Signal Processing, 90 (2010) 774–783.

[30] G. Samorodnitsky and M. S. Taqqu, Stable Non-Gaussian Random Processes: Stochastic Models and Infinite Variance, (Chapman and Hall, London, 1994).

[31] V. V. Uchaikin and V. M. Zolotarev, Chance and Stability: Stable Distributions and Their Applications, (Utrecht, VSP Press, 1999)

[32] V. M. Zolotarev, One-Dimensional Stable Distributions. American Mathematical Society, (Providence, R. I., 1986)