Journal of Mathematical Extension
Vol. 14, No. 4, (2020), 123-145
ISSN: 1735-8299
URL: http://www.ijmex.com
Original Research Paper

Multivariate Restricted Skew-Normal Scale Mixture of Birnbaum-Saunders Distribution

H. Samary
Marvdasht Branch, Islamic Azad University

Z. Khodadadi*
Marvdasht Branch, Islamic Azad University

H. Jafarpour
Shiraz Branch, Islamic Azad University

Abstract. In spite of widespread use as well as theoretical properties of the multivariate scale mixture of normal distributions, practical studies show a lack of stability and robustness against asymmetric features such as asymmetry and heavy tails. In this paper, we develop a new multivariate model by assuming the Birnbaum-Saunders distribution for the mixing variable in the scale mixture of the restricted skew-normal distribution. An analytically simple and efficient EM-type algorithm is adopted for iteratively computing maximum likelihood estimate of model parameters. To account standard errors, the observed information matrix is derived analytically by offering an information-based approach. Results obtained from real and simulated datasets are reported to illustrate the practical utility of the proposed methodology.

AMS Subject Classification: 62F10; 62J02
Keywords and Phrases: EM-type algorithm, birnbaum-saunders distribution, multivariate scale mixture distribution, restricted skew-normal distribution

Received: March 2019; Accepted: August 2019
*Corresponding author

1. Introduction

In statistical inference, for mathematical tractability reasons and attractive properties, the normal distribution plays a central role. However, practical studies show that several phenomena are not always in agreement with normality assumption due to exhibiting non-normal features such as asymmetry and heavy tails. Although, a quick way to deal with this drawback is to exploit some transformations on the data, it may cause some other potential deficiencies (see [16]). On the other hand, there are two approaches among researchers to construct more flexible distributions that possess skewness and kurtosis. [4, 5] established the first approach, initially in the univariate case, by modifying the normal density function in a multiplicative manner and imposing shape parameter. The new distribution was called skew-normal (SN) model. Later, the multivariate extension of the SN distribution was proposed by [9]. Even though the SN distribution provides more flexible families than the normal model, it does not accommodate heavy tails data and so that it is not flexible enough to model data sets that simultaneously have skew and heavy-tailed empirical distributions. This leads to proposing some generalizations of the SN distribution that can be found in the work of [7, 8, 13, 14, 17, 41] and the acknowledged articles therein, among others.

An alternative approach for constructing more flexible families than normal model is obtained by normal mean-variance mixture (NMV) representation, introduced by [11]. The NMV models assume not only that the variance is not fixed for all members of the population but also that they have non-constant mean. This is done by introducing randomness into the mean and variance of a normal distribution via a positive mixing variable. Therefore, the NMV class of distributions can control skewness as well as leptokurtosis, simultaneously. Specifically, a random vector $\mathbf{X}$ is distributed by the NMV model if it approaches the following representation:

$$\mathbf{X} \stackrel{d}{=} \mu + W\eta + W^{1/2}\mathbf{Z}, \tag{1}$$

where $\mu, \eta \in \mathbb{R}^p$, $\mathbf{Z} \sim \mathbf{N}_p(\mathbf{0}, \mathbf{\Sigma})$, a p-dimensional multivariate normal (MN) distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$, and W is a non-negative random variable, independent of $\mathbf{Z}$, with cumulative distribution function (cdf) $H(\cdot;\theta)$ parametrized by the vector parameter θ. Details in-depth of the NMV model can be found in [27]. Recently, some new spatial cases of NMV model have been considered. For instance [37] proposed normal mean-variance mixture of Birnbaum-Saunders distribution (NMVBS) by assuming that the mixing random variable is followed by Birnbaum-Saunders (BS) [12] model. [32] showed that the NMVBS takes wider ranges of skewness and kurtosis as compared with the SN distribution and some of its extensions. Moreover, if W in (1) is followed by the Lindley distribution [24], then the normal mean-variance mixture Lindley distribution is obtained [32]. [3] introduced a brilliant class of distributions by assuming generalized inverse Gaussian (GIG) distribution ([21]) for W and replacing normal distribution for $\mathbf{Z}$ with SN model. The new model contains the former and possess some interesting properties.

Owing to the proven proficiency of the two-parameter BS distribution in applied statistics, the main objective of the present work is to introduce a new skew and heavy-tailed distribution by representation (1) when η tends to zeros. The model is constructed by considering W and $\mathbf{Z}$ follow BS and restricted skew-normal (rMSN) distributions, respectively. The new model is called the multivariate scale mixture of restricted skew-normal BS (rMSN-BS) distribution. Some properties of the new model are studied and expectation-maximization (EM) algorithm ([15]) is used to obtain maximum likelihood (ML) estimate of parameters .

To give a short overview of this paper: Section 2 provides a shot review on the previous works. In Section 3, the formulation of new model is described and some of its properties are outlined. The ML estimates of the rMSN-BS distribution via implementing EM algorithm is also presented in Section 4. Some computational aspects for the EM algorithm are summarized in Section 5. Finally, in Sections 6 and 7, we illustrate the performance of the rMSN-BS distribution through analyzing real as well as synthetic datasets.

2. Background

2.1 The rMSN distribution

let $\phi_p(\cdot;\xi,\mathbf{\Sigma})$ be the probability density function (pdf) of $\mathbf{N}_p(\xi,\mathbf{\Sigma})$, and $\Phi(\cdot)$ represents the cdf of the standard normal distribution. A random vector $\mathbf{Z}$ is said to follow a p-variate rMSN distribution ([38]) with location and skewness vectors ξ and λ, respectively, and scale covariance matrix $\mathbf{\Sigma}$, if its pdf is

$$f(\mathbf{z}|\xi,\mathbf{\Sigma},\lambda) = 2\phi_p(\mathbf{z}|\xi,\mathbf{\Omega})\Phi\left(\frac{\lambda^\top\mathbf{\Omega}^{-1}(\mathbf{z}-\xi)}{\sqrt{1-\lambda^\top\mathbf{\Omega}^{-1}\lambda}}\right), \tag{2}$$

where $\mathbf{\Omega} = \mathbf{\Sigma} + \lambda\lambda^\top$. The notation $\mathbf{Z} \sim \mathrm{rMSN}(\xi,\mathbf{\Sigma},\lambda)$ will be used for a random vector $\mathbf{Z}$ with density (2). Following [38], the rMSN distribution can be presented by a convenient stochastic representation

$$\mathbf{Z} = \lambda|X_0| + \mathbf{X}_1, \quad X_0 \perp \mathbf{X}_1, \tag{3}$$

where $X_0 \sim \mathrm{N}(0,1)$, $\mathbf{X}_1 \sim \mathbf{N}_p(\xi,\mathbf{\Sigma})$ and the symbol $\perp$ indicates independence. Obviously, $|X_0|$ follows standard half-normal (HN) distribution, denoted by

$|X_0| \sim HN(0,1)$, and $\mathbf{Z}$ approaches $\mathbf{N}_p(\xi, \mathbf{\Sigma})$ pdf as $\lambda = \mathbf{0}$. The mean and covariance of $\mathbf{Z}$ can be obtained by (3), respectively as

$$E(\mathbf{Z}) = \xi + \sqrt{\frac{2}{\pi}}\lambda \quad and \quad cov(\mathbf{Z}) = \mathbf{\Sigma} + (1 - \frac{2}{\pi})\lambda\lambda^\top. \tag{4}$$

2.2 Birnbaum-saunders distribution

The well-known BS distribution is an asymmetric, non-negative model that has been recently received considerable attention in reliability and lifetime studies. Although the BS distribution was initially pioneered for modeling the fatigue life of structures under cyclic stress, it has been commonly accepted that the BS distribution can be taken as promising alternative to the Weibull, gamma, and log-normal models. See [26, 31, 39], among others, to find some applications of the BS distribution.

A random variable W taking positive real values follows the BS distribution with the shape α and scale β parameters if the pdf of W is given as

$$F(w; \alpha, \beta) = \frac{w + \beta}{\alpha\sqrt{w^3\beta}}\phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{w}{\beta}} - \sqrt{\frac{\beta}{w}}\right)\right], \qquad w > 0,\, \alpha > 0,\, \beta > 0.$$

We denote it by $W \sim BS(\alpha, \beta)$. It can be easily seen that the pdf of W is a mixture of two GIG distributions, i.e,

$$f(w; \alpha, \beta) = \frac{1}{2}f_{GIG}\left(w; \frac{1}{2}, \frac{1}{\beta\alpha^2}, \frac{\beta}{\alpha^2}\right) + \frac{1}{2}f_{GIG}\left(w; \frac{-1}{2}, \frac{1}{\beta\alpha^2}, \frac{\beta}{\alpha^2}\right), \tag{5}$$

where $f_{GIG}(\cdot; \kappa, \chi, \psi)$ denotes the pdf of GIG distribution with parameter set (κ, χ, ψ), denoted by $T \sim \mathrm{GIG}(\kappa, \chi, \psi)$, and pdf

$$f_{GIG}(t; \kappa, \chi, \psi) = (\frac{\psi}{\chi})^{\kappa/2}\frac{t^{\kappa-1}}{2K_\kappa(\sqrt{\psi\chi})}\exp\left\{\frac{-1}{2}\left(t^{-1}\chi + t\psi\right)\right\}, \;\; t > 0,$$

where $K_\kappa(\cdot)$ denotes the modified Bessel function of the third kind with index κ. The parameters of GIG distribution should fulfill in the condition $\chi \geqslant 0, \psi > 0$, if $\kappa > 0$; $\psi \geqslant 0, \chi > 0$, if $\kappa < 0$, and $\chi > 0, \psi > 0$, otherwise. The attractive overviews of BS distribution and its properties can be found in [10, 20, 23].

3. Model Formulation

In this section, we start by defining the rMSN-BS distribution and its hierarchical formulation and then introduce some further properties. A random vector $\mathbf{Y}$ is said to follow a p-variate rMSN-BS distribution, denoted by $\mathbf{Y} \sim$ rMSN-BS$(\xi, \boldsymbol{\Sigma}, \lambda, \alpha)$, if it can be generated by

$$\mathbf{Y} = \xi + \sqrt{\tau}\mathbf{Z} = \xi + \sqrt{\tau}(\lambda|X_0| + X_1), \tag{6}$$

where $\mathbf{Z} \sim$ rMSN$(\mathbf{0}, \boldsymbol{\Sigma}, \lambda)$, $\tau \sim$ BS$(\alpha, 1)$ $X_0 \sim N(0,1)$, $\mathbf{X}_1 \sim \mathbf{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ and $\mathbf{Z} \perp \tau$. When λ closes to zero, the rMSN-BS distribution tends to the symmetric class of scale mixture of normal distribution based on the BS distribution, or symmetric NMVBS model. Also, the rMSN-BS distribution tends to rMSN$(\xi, \boldsymbol{\Sigma}, \lambda)$ as α approaches zero.

The following result is an extension of lemma 1 in [6], which is crucial for evaluating some integrations in this paper.

Proposition 3.1. *If $\tau \sim GIG(\kappa, \chi, \psi)$, then for any $a \in \mathbb{R}$*

$$E\left(\Phi(\tau^{-1/2}a)\right) = F_{GH}(a; \kappa, \chi, \psi),$$

where $F_{GH}(\cdot; \kappa, \chi, \psi) = F_{GH_1}(\cdot; 0, 0, 1, \kappa, \chi, \psi)$ in which $F_{GH_p}(\cdot; \mu, \eta, \boldsymbol{\Sigma}, \kappa, \psi, \chi)$ denotes the cdf of p-variate generalized hyperbolic distribution ($GH_p(\mu, \eta, \boldsymbol{\Sigma}, \kappa, \psi, \chi)$) [27].

Proof. Let $V \sim$ N$(0,1)$ be a random variable independent of τ. Then,

$$\begin{aligned} E\left(\Phi(\tau^{-1/2}a)\right) &= E_\tau\left(P(V < \tau^{-1/2}a)|\tau\right) \\ &= E_\tau\left(P(\tau^{1/2}V < a)|\tau\right) = P(T^* < a), \end{aligned}$$

where $T^* \sim$ GH$_1(0, 0, 1, \kappa, \chi, \psi)$. $\square$

Proposition 3.1 is helpful for obtaining the pdf of rMSN-BS distribution. Let $\mathbf{Y} \sim$ rMSN-BS$(\xi, \boldsymbol{\Sigma}, \lambda, \alpha)$. From (6), it can be observed that $\mathbf{Y}|\tau \sim$ rMSN$(\xi, \tau^{-1}\boldsymbol{\Sigma}, \tau^{-\mathbf{1/2}}\lambda)$. Therefore, the density of $\mathbf{Y}$ is

$$\begin{aligned} f(\mathbf{y}; \xi, \boldsymbol{\Sigma}, \lambda, \alpha) =& f_{GH_p}(\mathbf{y}; \xi, \mathbf{0}, \boldsymbol{\Sigma}, \mathbf{0.5}, \alpha^{-\mathbf{2}}, \alpha^{-\mathbf{2}}) \\ & \times F_{GH}\left(A(\mathbf{y}, \boldsymbol{\Theta}); (\mathbf{1} - \mathbf{p})/\mathbf{2}, \mathbf{s} + \alpha^{-\mathbf{2}}, \alpha^{-\mathbf{2}}\right) \\ & + f_{GH_p}(\mathbf{y}; \xi, \mathbf{0}, \boldsymbol{\Sigma}, -\mathbf{0.5}, \alpha^{-\mathbf{2}}, \alpha^{-\mathbf{2}}) \\ & \times F_{GH}\left(A(\mathbf{y}, \boldsymbol{\Theta}); -(\mathbf{1} + \mathbf{p})/\mathbf{2}, \mathbf{s} + \alpha^{-\mathbf{2}}, \alpha^{-\mathbf{2}}\right), \end{aligned} \tag{7}$$

where $\boldsymbol{\Theta} = (\xi, \boldsymbol{\Sigma}, \lambda, \alpha)$, $A(\mathbf{y}, \boldsymbol{\Theta}) = \lambda^\top \boldsymbol{\Omega}^{-\mathbf{1}}(\mathbf{y} - \xi)/\sqrt{\mathbf{1} - \lambda^\top \boldsymbol{\Omega}^{-\mathbf{1}}\lambda}$, $s = (\mathbf{y} - \xi)^\top \boldsymbol{\Omega}^{-\mathbf{1}}(\mathbf{y} - \xi)$ and $f_{GH_p}(\cdot; \mu, \lambda, \boldsymbol{\Sigma}, \kappa, \chi, \psi)$ represents the pdf of GH$_p(\mu, \lambda, \boldsymbol{\Sigma}, \kappa, \psi, \chi)$. The mean and covariance matrix of $\mathbf{Y}$, obtained by (6) and the law of iterative expectations, are

$$
\begin{aligned}
E(\mathbf{Y}) &= \xi + \left(\frac{\mathbf{K_1}(\alpha^{-\mathbf{2}}) + \mathbf{K_0}(\alpha^{-\mathbf{2}})}{\mathbf{K_{-0.5}}(\alpha^{-\mathbf{2}})}\right)\sqrt{\frac{\mathbf{2}}{\pi}}\lambda, \\
cov(\mathbf{Y}) &= (1+0.5\alpha^2)(\mathbf{\Sigma} + (\mathbf{1} - \frac{\mathbf{2}}{\pi})\lambda\lambda^\top) \\
&\quad + \sqrt{\frac{2}{\pi}}\lambda\left((\mathbf{1}+\mathbf{0.5}\alpha^{\mathbf{2}}) - \left(\frac{\mathbf{K_1}(\alpha^{-\mathbf{2}}) + \mathbf{K_0}(\alpha^{-\mathbf{2}})}{\mathbf{K_{-0.5}}(\alpha^{-\mathbf{2}})}\right)^{\mathbf{2}}\right), \qquad \alpha \in \mathbb{R}^+.
\end{aligned}
$$

In the univariate case for $\xi = 0$ and $\sigma^2 = 1$, without lose of generality, the skewness and kurtosis of Y are also obtained as follows:

$$
\gamma_Y = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{1.5}}, \quad \text{and} \quad \kappa_Y = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2} - 3,
$$

where

$$
\begin{aligned}
\mu_1 &= E(Y) = \left(\frac{K_1(\alpha^{-2}) + K_0(\alpha^{-2})}{K_{0.5}(\alpha^{-2})}\right)\sqrt{\frac{1}{2\pi}}\lambda, \\
\mu_2 &= E(Y^2) = \left(1+0.5\alpha^2\right)(1+\lambda^2), \\
\mu_3 &= E(Y^3) = \left(\frac{K_2(\alpha^{-2}) + K_1(\alpha^{-2})}{K_{0.5}(\alpha^{-2})}\right)\sqrt{\frac{1}{2\pi}}\lambda(1+2\lambda^2), \\
\mu_4 &= E(Y^4) = 2(1+\alpha^2+2/3\alpha^4)(1+\lambda^2)^2.
\end{aligned}
$$

To illustrate the tail behavior and asymmetric properties of the rMSN-BS distribution, we display the pdf of a univariate rMSN-BS distribution for $\xi = 0$ and $\sigma^2 = 1$ and various values of α and λ in Figure 1. It is clearly seen that the rMSN-BS distribution can produce very strong skewness and extremely heavier tails than the normal distribution. Also, the contour plots of the skewness and kurtosis of Y in the univariate case are plotted in Figure 2. Observing this figure, the rMSN-BS distribution has negative skewness for both small values of α and negative values of λ. Moreover, it can be observed from Table 1 that the rMSN-BS distribution takes wider ranges of skewness as compared with the skew t and skew-t-normal distributions in [19] based on univarate case.

Table 1: The ranges of skewness for different values of α

α	0.05	0.10	1	5	10	100
skewness	(-2.64,2.64)	(-2.62,2.62)	(-1.10,1.10)	(-0.96,0.96)	(-0.93,0.93)	(-0.88,0.88)

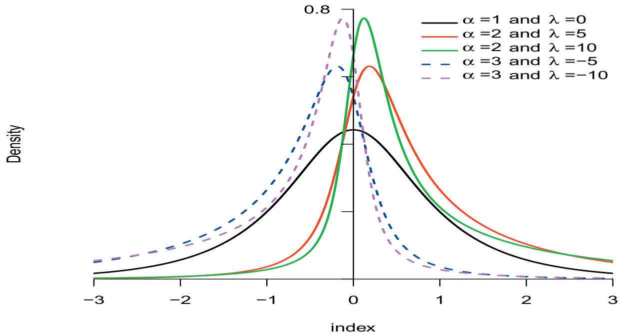

Figure 1. The density plots of rMSN-BS distribution for different values of α and λ

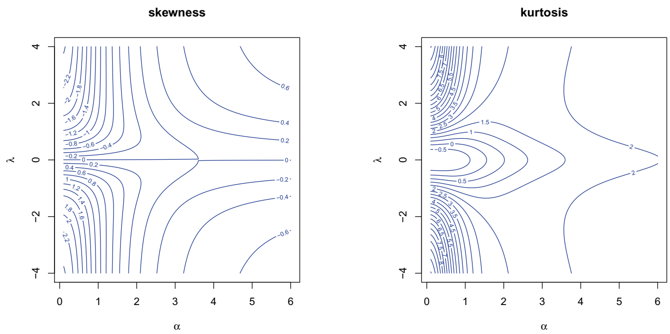

Figure 2. The skewness and kurtosis of the rMSN-BS distribution

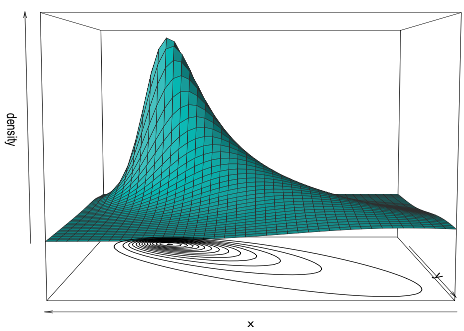

Figure 3. The density and contour plots of the bivariate rMSN-BS distribution

Figure 3, furthermore, displays the density graph of the bivariate rMSN-BS distribution for $\xi = (\mathbf{0}, \mathbf{0})$, $\mathbf{\Sigma} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, $\lambda = (-\mathbf{2}, \mathbf{3})$ and $\alpha = 1.2$.

Theorem 3.2. Let $\mathbf{Y} \sim$ rMSN-BS$(\xi, \mathbf{\Sigma}, \lambda, \alpha)$.

I. For any $\mathbf{A} \in \mathbf{R}^{\mathbf{q}\times\mathbf{d}}$ and $\mathbf{b} \in \mathbf{R}^{\mathbf{q}}$, the q-dimensional random vector $\mathbf{X} = \mathbf{AY} + \mathbf{b}$ is distributed as rMSN-BS$(\mathbf{A}\xi + \mathbf{b}, \mathbf{A\Sigma A}^\top, \mathbf{A}\lambda, \alpha)$.

II. As a consequence of part (I), if we have partition of $\mathbf{Y}$, ξ, λ and $\mathbf{\Sigma}$ as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \end{pmatrix}, \quad \xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \quad \lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma_{11}} & \mathbf{\Sigma_{12}} \\ \mathbf{\Sigma_{21}} & \mathbf{\Sigma_{22}} \end{pmatrix}, \tag{8}$$

then $\mathbf{Y_1} \sim$ rMSN-BS$(\xi_{\mathbf{1}}, \mathbf{\Sigma_{11}}, \lambda_{\mathbf{1}}, \alpha)$ in which $\mathbf{Y_1}$, ξ_1, $\lambda_1 \in \mathbb{R}$ and $\mathbf{\Sigma_{11}} \in \mathbb{R}^{\mathbf{q}\times\mathbf{q}}$.

Proof. The proof is easily completed by using Bayes rule and some algebraic work. □

4. Maximum Likelihood Estimation of the rMSN-BS Distribution

In this section, we demonstrate how to employ EM-type algorithm for ML estimation of the rMSN-BS distribution.

4.1 The model and likelihood

Let $\mathbf{Y} = (\mathbf{Y_1}, \ldots, \mathbf{Y_n})$ be the n independent random variable from rMSN-BS$(\xi, \mathbf{\Sigma}, \lambda, \alpha)$ and denote $\gamma = \sqrt{\tau}|X_0|$. From (3) and (6), the hierarchically representation of $\mathbf{Y}$ is

$$\begin{aligned} \mathbf{Y_j}|\gamma_\mathbf{j}, \tau_\mathbf{j} &\sim N_p(\xi + \lambda\gamma_\mathbf{j}, \tau_\mathbf{j}\mathbf{\Sigma}), \\ \gamma_j|\tau_j &\sim TN(0, \tau_j; (0, \infty)), \\ \tau_j &\sim \text{BS}(\alpha, 1), \end{aligned} \tag{9}$$

where TN$(\mu, \sigma^2; (a, b))$ represents the truncated normal distribution for N(μ, σ^2) lying within the truncated interval (a, b). Hence, the joint pdf of $\mathbf{Y_j}$, γ_j and

τ_j is given by

$$f(\mathbf{y_j}, \gamma_\mathbf{j}, \tau_\mathbf{j}) = \frac{(\tau_\mathbf{j} + \mathbf{1})}{\sqrt{(\mathbf{2}\pi)^{\mathbf{p+1}} \tau_\mathbf{j}^{\mathbf{p-3}}} \alpha |\mathbf{\Sigma}|^{\mathbf{0.5}}}$$
$$\times \exp\left\{-\frac{1}{2\tau_j}\left[(\mathbf{y_j} - \xi - \lambda\gamma_\mathbf{j})^\top \mathbf{\Sigma^{-1}}(\mathbf{y_j} - \xi - \lambda\gamma_\mathbf{j})\right] - \frac{1}{2\alpha^2}(\tau_j + \frac{1}{\tau_j} - 2)\right\}. \tag{10}$$

Integrating out γ_j in (10), we get

$$f(\mathbf{y_j}, \tau_\mathbf{j}) = \frac{2(\tau_j + 1)}{\sqrt{(2\pi)^{p+1} \tau_j^{p-3}} |\Sigma|^{0.5}}$$
$$\times \exp\left\{-\frac{1}{2\tau_j} s_j - \frac{1}{2\alpha^2}(\tau_j + \frac{1}{\tau_j} - 2)\right\} \Phi\left(\tau_j^{-1/2} A(\mathbf{y_j}, \mathbf{\Theta})\right), \tag{11}$$

where $s_j = (\mathbf{y_j} - \xi)^\top \mathbf{\Omega^{-1}}(\mathbf{y_j} - \xi)$. Dividing (10) by (11) gives the conditional distribution of γ_j given $(\mathbf{Y_j} = \mathbf{y_j}, \tau_\mathbf{j})$ as

$$f(\gamma_j \mid \mathbf{y_j}, \tau_\mathbf{j}) = \frac{1}{\Phi\left(\tau_j^{-1/2} A(\mathbf{y_j}, \mathbf{\Theta})\right)}$$
$$\times \exp\left\{\frac{\tau_j^{-1}}{2(1 - \lambda^\top \mathbf{\Omega^{-1}} \lambda)} (\gamma_j - \lambda^\top \mathbf{\Omega^{-1}}(\mathbf{y_j} - \xi))^\mathbf{2}\right\}. \tag{12}$$

It follows from (12) that the conditional distribution of γ_j given $(\mathbf{Y_j} = \mathbf{y_j},\ \tau_\mathbf{j})$ is

$$\gamma_j \mid \mathbf{Y_j} = \mathbf{y_j},\ \tau_\mathbf{j} \sim \mathrm{TN}(\lambda^\top \mathbf{\Omega^{-1}}(\mathbf{y_j} - \xi), \tau_\mathbf{j}(\mathbf{1} - \lambda^\top \mathbf{\Omega^{-1}} \lambda); (\mathbf{0}, \infty)). \tag{13}$$

Moreover, dividing (10) by (7) yields

$$f(\tau_j \mid \mathbf{y_j}) = \pi_\mathbf{j} \frac{\mathbf{f_{GIG}}\left(\tau_\mathbf{j}; \frac{\mathbf{1-p}}{\mathbf{2}}, \mathbf{s_j} + \alpha^{-\mathbf{2}}, \alpha^{-\mathbf{2}}\right) \mathbf{\Phi}\left(\tau_\mathbf{j}^{-\mathbf{1/2}} \mathbf{A}(\mathbf{y_j}, \mathbf{\Theta})\right)}{\mathbf{F_{GH}}\left(\mathbf{A}(\mathbf{y_j}, \mathbf{\Theta}); (\mathbf{1} - \mathbf{p})/\mathbf{2}, \mathbf{s_j} + \alpha^{-\mathbf{2}}, \alpha^{-\mathbf{2}}\right)}$$
$$+ (1 - \pi_j) \frac{f_{GIG}\left(\tau_j; -\frac{1+p}{2}, s_j + \alpha^{-2}, \alpha^{-2}\right) \Phi\left(\tau_j^{-1/2} A(\mathbf{y_j}, \mathbf{\Theta})\right)}{F_{GH}\left(A(\mathbf{y_j}, \mathbf{\Theta}); -(\mathbf{1} + \mathbf{p})/\mathbf{2}, \mathbf{s_j} + \alpha^{-\mathbf{2}}, \alpha^{-\mathbf{2}}\right)}, \tag{14}$$

where $s_j = (\mathbf{y_j} - \xi)^\top \mathbf{\Omega^{-1}}(\mathbf{y_j} - \xi)$, and

$$\pi_j = \frac{f_{GH_p}(\mathbf{y_j}; \xi, \mathbf{0}, \mathbf{\Sigma}, \mathbf{0.5}, \alpha^{-\mathbf{2}}, \alpha^{-\mathbf{2}}) \mathbf{F_{GH}}\left(\mathbf{A}(\mathbf{y_j}, \mathbf{\Theta}); (\mathbf{1} - \mathbf{p})/\mathbf{2}, \mathbf{s_j} + \alpha^{-\mathbf{2}}, \alpha^{-\mathbf{2}}\right)}{f(\mathbf{y_j}; \xi, \mathbf{\Sigma}, \lambda, \alpha)}.$$

Proposition 4.1.1. *Let $R_{(\kappa,a)}(c) = K_{\kappa+a}(c)/K_\kappa(c)$. From the conditional density* (14)*, the following statements are obtained:*

(a) The conditional expectation of τ_j given $\mathbf{Y} = \mathbf{y_j}$ is

$$E(\tau_j \mid \mathbf{Y} = \mathbf{y_j}) = \Big(\frac{1}{\alpha^2 s_j + 1}\Big)^{0.5}$$
$$\Big(\pi_j w_{(1,(1-p)/2)} R_{((1-p)/2,1)}\Big(\sqrt{s_j + \alpha^{-2}}/\alpha\Big)$$
$$+ (1-\pi_j) w_{(1,-(1+p)/2)} R_{((p-1)/2,1)}\Big(\sqrt{s_j + \alpha^{-2}}/\alpha\Big)\Big),$$

where

$$w_{(i,k)} = \frac{F_{GH}\left(A(\mathbf{y_j}, \boldsymbol{\Theta}); \mathbf{k}+\mathbf{i}, \mathbf{s_j} + \alpha^{\mathbf{-2}}, \alpha^{\mathbf{-2}}\right)}{F_{GH}\left(A(\mathbf{y_j}, \boldsymbol{\Theta}); \mathbf{k}, \mathbf{s_j} + \alpha^{\mathbf{-2}}, \alpha^{\mathbf{-2}}\right)}.$$

(b) The conditional expectation of τ_j^{-1} given $\mathbf{Y} = \mathbf{y_j}$ is

$$E\Big(\tau_j^{-1} \mid \mathbf{Y} = \mathbf{y_j}\Big) = \left(\alpha^{\mathbf{2}} \mathbf{s_j} + \mathbf{1}\right)^{\mathbf{0.5}}$$
$$\Big(\pi_j w_{(-1,(1-p)/2)} R_{((1-p)/2,-1)}\Big(\sqrt{s_j + \alpha^{-2}}/\alpha\Big)$$
$$+ (1-\pi_j) w_{(-1,-(1+p)/2)} R_{((p-1)/2,-1)}\Big(\sqrt{s_j + \alpha^{-2}}/\alpha\Big)\Big).$$

(c) Specific conditional expectation related to function of τ_j is

$$E\left(\tau_j^{-1/2} \frac{\phi\left(\tau_j^{-1/2} A(\mathbf{y_j}, \boldsymbol{\Theta})\right)}{\Phi\left(\tau_j^{-1/2} A(\mathbf{y_j}, \boldsymbol{\Theta})\right)} \Big| \mathbf{Y} = \mathbf{y_j}\right) =$$
$$\frac{\pi_j \left(\alpha^2 s_j + 1\right)^{p/4}}{\sqrt{2\pi}\sqrt{\alpha}\left(\alpha^2(s_j + A^2(\mathbf{y_j}, \boldsymbol{\Theta})) + \mathbf{1}\right)^{p/4}}$$
$$\times \frac{K_{-\frac{p}{2}}\left(\sqrt{(s_j + A^2(\mathbf{y_j}, \boldsymbol{\Theta}) + \alpha^{\mathbf{-2}})\alpha^{\mathbf{-2}}}\right)}{K_{\frac{1-p}{2}}\left(\sqrt{(s_j + \alpha^{-2})\alpha^{-2}}\right) F_{GH}\left(A(\mathbf{y_j}, \boldsymbol{\Theta}); \frac{\mathbf{1-p}}{\mathbf{2}}, \mathbf{s_j} + \alpha^{\mathbf{-2}}, \alpha^{\mathbf{-2}}\right)}$$
$$+ \frac{(1-\pi_j)\left(\alpha^2 s_j + 1\right)^{(2+p)/4}}{\sqrt{2\pi}\sqrt{\alpha}\left(\alpha^2(s_j + A^2(\mathbf{y_j}, \boldsymbol{\Theta})) + \mathbf{1}\right)^{(2+p)/4}}$$
$$\times \frac{K_{-\frac{2+p}{2}}\left(\sqrt{(s_j + A^2(\mathbf{y_j}, \boldsymbol{\Theta}) + \alpha^{\mathbf{-2}})\alpha^{\mathbf{-2}}}\right)}{K_{-\frac{1+p}{2}}\left(\sqrt{(s_j + \alpha^{-2})\alpha^{-2}}\right) F_{GH}\left(A(\mathbf{y_j}, \boldsymbol{\Theta}); -\frac{\mathbf{1+p}}{\mathbf{2}}, \mathbf{s_j} + \alpha^{\mathbf{-2}}, \alpha^{\mathbf{-2}}\right)}.$$

(d) The conditional expectation of $\tau_j\gamma_j$ *given* $\mathbf{Y}=\mathbf{y_j}$ *is*

$$\begin{aligned}E\Big(\tau_j^{-1}\gamma_j \mid \mathbf{Y}=\mathbf{y_j}\Big) =&\lambda^\top\mathbf{\Omega^{-1}}(\mathbf{y_j}-\xi)\mathbf{E}\Big(\tau_{\mathbf{j}}^{-\mathbf{1}} \mid \mathbf{Y}=\mathbf{y_j}\Big)\\ &+\sqrt{1-\lambda^\top\mathbf{\Omega^{-1}}\lambda}\\ &\times E\left(\tau_j^{-1/2}\frac{\phi\left(\tau_j^{-1/2}A(\mathbf{y_j},\mathbf{\Theta})\right)}{\Phi\left(\tau_j^{-1/2}A(\mathbf{y_j},\mathbf{\Theta})\right)}\Big|\mathbf{Y}=\mathbf{y_j}\right).\end{aligned}$$

(e) The conditional expectation of $\tau_j\gamma_j^2$ *given* $\mathbf{Y}=\mathbf{y_j}$ *is*

$$\begin{aligned}E\Big(\tau_j^{-1}\gamma_j^2 \mid \mathbf{Y}=\mathbf{y_j}\Big) =& (\lambda^\top\mathbf{\Omega^{-1}}(\mathbf{y_j}-\xi))^{\mathbf{2}}\mathbf{E}(\tau_{\mathbf{j}}^{-\mathbf{1}} \mid \mathbf{Y}=\mathbf{y_j})\\ &+(1-\lambda^\top\mathbf{\Omega^{-1}}\lambda)\\ &+(\lambda^\top\mathbf{\Omega^{-1}}(\mathbf{y_j}-\xi))(\mathbf{1}-\lambda^\top\mathbf{\Omega^{-1}}\lambda)\\ &\times E\left(\tau_j^{-1/2}\frac{\phi\left(\tau_j^{-1/2}A(\mathbf{y_j},\mathbf{\Theta})\right)}{\Phi\left(\tau_j^{-1/2}A(\mathbf{y_j},\mathbf{\Theta})\right)}\Big|\mathbf{Y}=\mathbf{y_j}\right).\end{aligned}$$

Proof. The proof is straightforward. □

4.2 Parameter estimation via ECM algorithm

To compute the ML estimates of unknown parameters, an extension of the EM algorithm, the Expectation Conditional Maximization (ECM) algorithm ([30]) is implemented. The ECM algorithm is a straightforward modification of EM algorithm in which the maximization (M) step is replaced by a sequence of computationally simper conditional maximization (CM) steps. Let $\mathbf{y}=(\mathbf{y_1},\ldots,\mathbf{y_n})$ be a random sample of size n from the rMSN-BS distribution, and $\gamma=(\gamma_\mathbf{1},\ldots,\gamma_\mathbf{n})$ and $\tau=(\tau_\mathbf{1},...,\tau_\mathbf{n})$ represent the hidden variables in the model. The complete data log-likelihood function of $\mathbf{\Theta}=(\xi,\mathbf{\Sigma},\lambda,\alpha)$ given $(\mathbf{y},\gamma,\tau)$, omitting the additive constants, is given by

$$\begin{aligned}\ell_c(\mathbf{\Theta} \mid \mathbf{y},\gamma,\tau) =& -\frac{n}{2}\log\alpha|\Sigma| - \frac{1}{2\alpha^2}\sum_{j=1}^{n}(\tau_j+\frac{1}{\tau_j}-2)\\ &-\frac{1}{2}\sum_{j=1}^{n}\tau_j^{-1}(\mathbf{y_j}-\xi-\lambda\gamma_\mathbf{j})^\top\mathbf{\Sigma^{-1}}(\mathbf{y_j}-\xi-\lambda\gamma_\mathbf{j}).\end{aligned}\qquad(15)$$

The expected value of complete data log-likelihood (15) with respect to the conditional distribution of the missing values (γ, τ) given the observed data $\mathbf{y}$, an evaluated at $\theta = \hat{\theta}^{(\mathbf{k})}$, called Q-function, is

$$
\begin{aligned}
Q(\boldsymbol{\Theta} \mid \hat{\boldsymbol{\Theta}}^{(\mathbf{k})}) = -\frac{\mathbf{n}}{\mathbf{2}} \log \alpha |\boldsymbol{\Sigma}| - \frac{1}{2\alpha^2} \sum_{j=1}^{n} (\hat{\tau}_j^{(k)} + \hat{t}_j^{(k)} - 2) \\
- \frac{1}{2} \sum_{j=1}^{n} \Big\{ \hat{\tau}_j^{(k)} \delta(\mathbf{y_j}, \xi, \boldsymbol{\Sigma}) - \hat{\mathbf{w}}_{\mathbf{1j}}^{(\mathbf{k})} (\mathbf{y_j} - \xi)^\top \boldsymbol{\Sigma}^{-\mathbf{1}} \lambda \\
- \hat{w}_{1j}^{(k)} \lambda^\top \boldsymbol{\Sigma}^{-\mathbf{1}} (\mathbf{y_j} - \xi) + \hat{\mathbf{w}}_{\mathbf{2j}}^{(\mathbf{k})} \lambda^\top \lambda \Big\},
\end{aligned} \tag{16}
$$

where the necessary conditional expectations include $\hat{\tau}_j^{(k)} = E(\tau_j^{-1} \mid \mathbf{y}, \hat{\boldsymbol{\Theta}}^{(\mathbf{k})})$, $\hat{t}_j^{(k)} = E(\tau_j \mid \mathbf{y}, \hat{\boldsymbol{\Theta}}^{(\mathbf{k})})$, $\hat{w}_{1j}^{(k)} = E(\tau_j^{-1} \gamma_j \mid \mathbf{y}, \hat{\boldsymbol{\Theta}}^{(\mathbf{k})})$ and $\hat{w}_{2j}^{(k)} = E(\tau_j^{-1} \gamma_j^2 \mid \mathbf{y}, \hat{\boldsymbol{\Theta}}^{(\mathbf{k})})$ that obtain by Proposition 3.1. In summary, the implementation of the ECM algorithm proceeds as follows:

E-step: Given $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}^{(\mathbf{k})}$, compute $\hat{\tau}_j^{(k)}$, $\hat{t}_j^{(k)}$, $\hat{w}_{1j}^{(k)}$ and $\hat{w}_{2j}^{(k)}$ for $j = 1, \ldots, n$.

CM-step 1: Update $\hat{\alpha}^{(k)}, \hat{\xi}^{(k)}$ and $\hat{\lambda}^{(k)}$, by maximizing (16) over α, ξ and λ, which leads to

$$
\begin{aligned}
\hat{\alpha}^{(k+1)} &= \sqrt{\sum_{j=1}^{n} (\hat{\tau}_j^{(k)} + \hat{t}_j^{(k)} - 2)}, \\
\hat{\xi}^{(k+1)} &= \sum_{j=1}^{n} (\hat{\tau}_j^{(k)} \mathbf{y_j} - \hat{\mathbf{w}}_{\mathbf{1j}}^{(\mathbf{k})} \hat{\lambda}^{(\mathbf{k})}) / \left(\sum_{\mathbf{j=1}}^{\mathbf{n}} \hat{\tau}_{\mathbf{j}}^{(\mathbf{k})} \right), \\
\hat{\lambda}^{(k+1)} &= \sum_{j=1}^{n} \hat{w}_{1j}^{(k)} (\mathbf{y_j} - \hat{\xi}^{(\mathbf{k+1})}) \Big/ \left(\sum_{\mathbf{j=1}}^{\mathbf{n}} \hat{\mathbf{w}}_{\mathbf{2j}}^{(\mathbf{k})} \right).
\end{aligned}
$$

CM-step 2: Update $\hat{\boldsymbol{\Sigma}}^{(k)}$ by

$$
\begin{aligned}
\hat{\boldsymbol{\Sigma}}^{(k+1)} = & \frac{1}{n} \sum_{j=1}^{n} \Bigg\{ \hat{\tau}_j^{(k)} (\mathbf{y_j} - \hat{\xi}^{(\mathbf{k+1})})(\mathbf{y_j} - \hat{\xi}^{(\mathbf{k+1})})^\top + \hat{\mathbf{w}}_{\mathbf{2j}}^{(\mathbf{k})} \hat{\lambda}^{(\mathbf{k+1})} \hat{\lambda}^{(\mathbf{k+1})\top} \\
& - \hat{w}_{1j}^{(k)} \left[\hat{\lambda}^{(k+1)} (\mathbf{y_j} - \hat{\xi}^{(\mathbf{k+1})})^\top + (\mathbf{y_j} - \hat{\xi}^{(\mathbf{k+1})}) \hat{\lambda}^{(\mathbf{k+1})\top} \right] \Bigg\}.
\end{aligned}
$$

5. Computational Strategies Related to Implementation

5.1 Estimation of standard errors

To estimate standard error of the parameter estimation, the information-based method is exploited. Following [29], the Fisher information matrix can be approximated by the information matrix

$$I_o(\hat{\mathbf{\Theta}} \mid \mathbf{y}) = \sum_{\mathbf{j=1}}^{\mathbf{n}} \hat{\mathbf{s}}_{\mathbf{j}} \hat{\mathbf{s}}_{\mathbf{j}}^{\top}, \tag{17}$$

where for $\ell_{cj}(\mathbf{\Theta}; \mathbf{y_{cj}})$, the complete-data log-likelihood (15) computed in the jth individual observation $\mathbf{y_{cj}} = (\mathbf{y_j}, \gamma_{\mathbf{j}}, \tau_{\mathbf{j}})$,

$$\hat{\mathbf{s}}_j = E\left(\frac{\ell_{cj}(\mathbf{\Theta}; \mathbf{y_{cj}})}{\partial \mathbf{\Theta}} \mid \mathbf{y_j}, \hat{\mathbf{\Theta}}\right). \tag{18}$$

Let $\sigma = \text{vec}(\mathbf{\Sigma})$ denotes a $(p(p-1)) \times 1$ vector by stacking the column vectors of $\mathbf{\Sigma}$. Using standard matrix differentiations, the individual score vector (18) contains the following elements:

$$\begin{aligned}
\hat{\mathbf{s}}_{j,\xi} &= E\left(\frac{\ell_{cj}(\mathbf{\Theta}; \mathbf{y_{cj}})}{\partial \xi} \mid \mathbf{y_j}, \hat{\mathbf{\Theta}}\right) = \hat{\mathbf{\Sigma}}^{-1}\left\{\hat{\tau}_j(\mathbf{y_j} - \hat{\xi}) - \hat{\lambda}\hat{\mathbf{w}}_{\mathbf{1j}}\right\}, \\
\hat{\mathbf{s}}_{j,\lambda} &= E\left(\frac{\ell_{cj}(\mathbf{\Theta}; \mathbf{y_{cj}})}{\partial \lambda} \mid \mathbf{y_j}, \hat{\mathbf{\Theta}}\right) = \hat{\mathbf{\Sigma}}^{-1}\left\{\hat{w}_{1j}(\mathbf{y_j} - \hat{\xi}) - \hat{\lambda}\hat{\mathbf{w}}_{\mathbf{2j}}\right\}, \\
\hat{\mathbf{s}}_{j,\sigma} &= E\left(\frac{\ell_{cj}(\mathbf{\Theta}; \mathbf{y_{cj}})}{\partial \sigma} \mid \mathbf{y_j}, \hat{\mathbf{\Theta}}\right) = \text{vec}\left(-\frac{1}{2}\left\{\hat{\mathbf{\Sigma}}^{-1} - \hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{\Upsilon}}_j\hat{\mathbf{\Sigma}}^{-1}\right\}\right), \\
\hat{\mathbf{s}}_{j,\alpha} &= E\left(\frac{\ell_{cj}(\mathbf{\Theta}; \mathbf{y_{cj}})}{\partial \alpha} \mid \mathbf{y_j}, \hat{\mathbf{\Theta}}\right) = -\frac{1}{\alpha} + \frac{1}{\alpha^3}(\hat{\tau}_j + \hat{t}_j - 2).
\end{aligned}$$

As a result, the standard errors of the estimator, $\hat{\mathbf{\Theta}}$, can be obtained as the square roots of the diagonal elements of $I_o^{-1}(\hat{\mathbf{\Theta}}|\mathbf{y})$.

5.2 Initial values

To overcome the sensitivity of EM algorithm to the starting values, we consider initial values based on the straightforward way. Bellow, we summarize a convenient way of creating suitable initial values for the implementation of EM-type algorithm to obtain parameter estimates of the rMSN-BS model.

1. Since the new model contains the normal distribution as a special case, the initial component skewness vector and α are chosen as $\hat{\lambda}^{(0)} = \mathbf{0}$, and $\hat{\alpha}^{(0)} = 0.1$. This may facilitates faster convergence when the underlying data has heavy tails.
2. Consequently, the initial values for mixing probabilities, location and scale covariance matrices can be specified as

$$\hat{\xi}^{(0)} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{y_j}, \qquad \hat{\mathbf{\Sigma}}^{\mathbf{(0)}} = \frac{\mathbf{1}}{\mathbf{n}}\sum_{\mathbf{j=1}}^{\mathbf{n}}(\mathbf{y_j} - \hat{\xi}^{\mathbf{(k+1)}})(\mathbf{y_j} - \hat{\xi}^{\mathbf{(k+1)}})^{\top}.$$

5.3 Convergence assessment

The Aitken's acceleration method ([1]) is adopted in this paper as a convergence tool of the EM algorithm. The Aitken's acceleration outperforms the lack of progress criterion and avoids the premature convergence ([28]). Let $a^{(k)} = (\ell^{(k+1)} - \ell^{(k)})/(\ell^{(k)} - \ell^{(k-1)})$ be the Aikten acceleration factor in which $\ell^{(k)}$ denotes the log-likelihood value evaluated at $\hat{\mathbf{\Theta}}^{(k)}$. Then, the asymptotic estimate of the log-likelihood at iteration k is calculated as

$$\ell_{\infty}^{(k+1)} = \ell^{(k)} + \frac{1}{1 - a^{(k)}}(\ell^{(k+1)} - \ell^{(k)}),$$

Now, the EM algorithm can be considered to have converged if $\ell_{\infty}^{(k)} - \ell^{(k)} < \epsilon$. The desired tolerance ϵ considered in this paper is 10^{-5}.

5.4 Model selection

For the sake of compression, two well-known measures based on the penalized log-likelihood are used. We used the Akaike information criterion (AIC; [2]) and Bayesian information criterion (BIC; [40]) which are obtained by $mC(n) - 2\ell_{max}$. Here, ℓ_{max} is the maximized log-likelihood, m is the number of free parameters in the considered model and $C(n) = 2$ for AIC and $C(n) = \log(n)$ for BIC.

6. An Illustration

In this section, the open/closed book (OCB) dataset is analyzed to illustrate the performance of the proposed model. The OCB, originally reported by [25] and subsequently analyzed by [22, 42] among others, includes five proficiency namely mechanics (mc), vectors (ve), algebra (al), analysis (an), and statistics (st) tested on $n = 88$ students. In this example, we focus on a the bivariate sample of two variables (vec, sta) to follow rMS-BS distribution since they exhibit an apparent bimodal asymmetric pattern with some outlying observations. For comparison purposes, bivariate normal (MVN); bivariate-t (MVT), bivariate rMSN, bivariate restricted skew-t (rMST) distributions are also applied to model this dataset.

Table 2 provides a summary of model fitting containing the parameter estimates with associated standard error estimates, and model selection criteria. In light of these two selection criteria, the results show that the rMSN-BS model outperforms MVN, MVT, rMSN-BS and rMST distributions. Figure 4 displays graphical representation of five fitted models. It can be seen that the rMSN-BS model adapts the shape of the scattering pattern more adequately than the other candidate models, showing the superiority of rMSN-BS in the capability of dealing with heterogeneous data.

Table 2: Summary results from fitting various models on the OCB data.

Parameter	MVN		MVT		rMSN		rMST		rMSN-BS	
	ML	se	ML	se	ML	se	ML	se	ML	se
ξ_1	50.590	3.726	50.658	3.122	64.274	3.055	54.051	1.980	64.466	2.423
ξ_2	42.306	3.977	42.086	3.583	37.571	2.941	26.793	2.028	22.374	1.239
σ_{11}	170.878	9.947	163.196	9.053	74.550	8.711	148.066	6.301	143.368	6.431
σ_{12}	97.886	7.056	94.391	6.720	131.150	7.942	117.033	4.320	139.238	5.390
σ_{22}	294.371	11.840	284.445	9.421	282.885	9.862	142.924	7.791	152.540	7.930
λ_1	–	–	–	–	-3.459	3.569	-4.127	1.066	-9.432	1.445
λ_2	–	–	–	–	5.817	1.250	11.560	2.160	18.109	2.384
α	–	–	–	–	–	–	–	–	0.298	0.089
ν	–	–	8.117	2.350	–	–	11.249	2.430	–	–
$\ell(\boldsymbol{\theta})$	-722.769		-720.7649		-712.380		-712.802		-708.858	
AIC	1455.540		1452.530		1438.716		1441.604		1433.717	
BIC	1467.926		1465.394		1456.102		1461.423		1453.536	

7. Simulation Study

This section deals with the performance of our proposed distribution and its computational methodology. In the first simulation, the finite sample properties of the ML estimator of the rMSN-BS parameters are investigated. In each replication, a random sample from the rMSN-BS distribution with presumed parameter values $\xi = (\mathbf{2}, \mathbf{2})$, $\lambda = (\mathbf{2}, \mathbf{3})$, $\alpha = 0.4$ and $\text{Vech}(\boldsymbol{\Sigma}) = (\mathbf{2}, \mathbf{0.4}, \mathbf{1})$ is generated for $n = 30, 50, 100, 200, 500$, and 1000. Figure 5 shows one sample with size 1000. To check performance of the estimates obtained using the EM algorithm, the parameter estimates of the rMSN-BS distribution are computed. Table 3 presents the Bias and mean squared error (MSE) of the ML estimators obtained over 500 trials, respectively, by

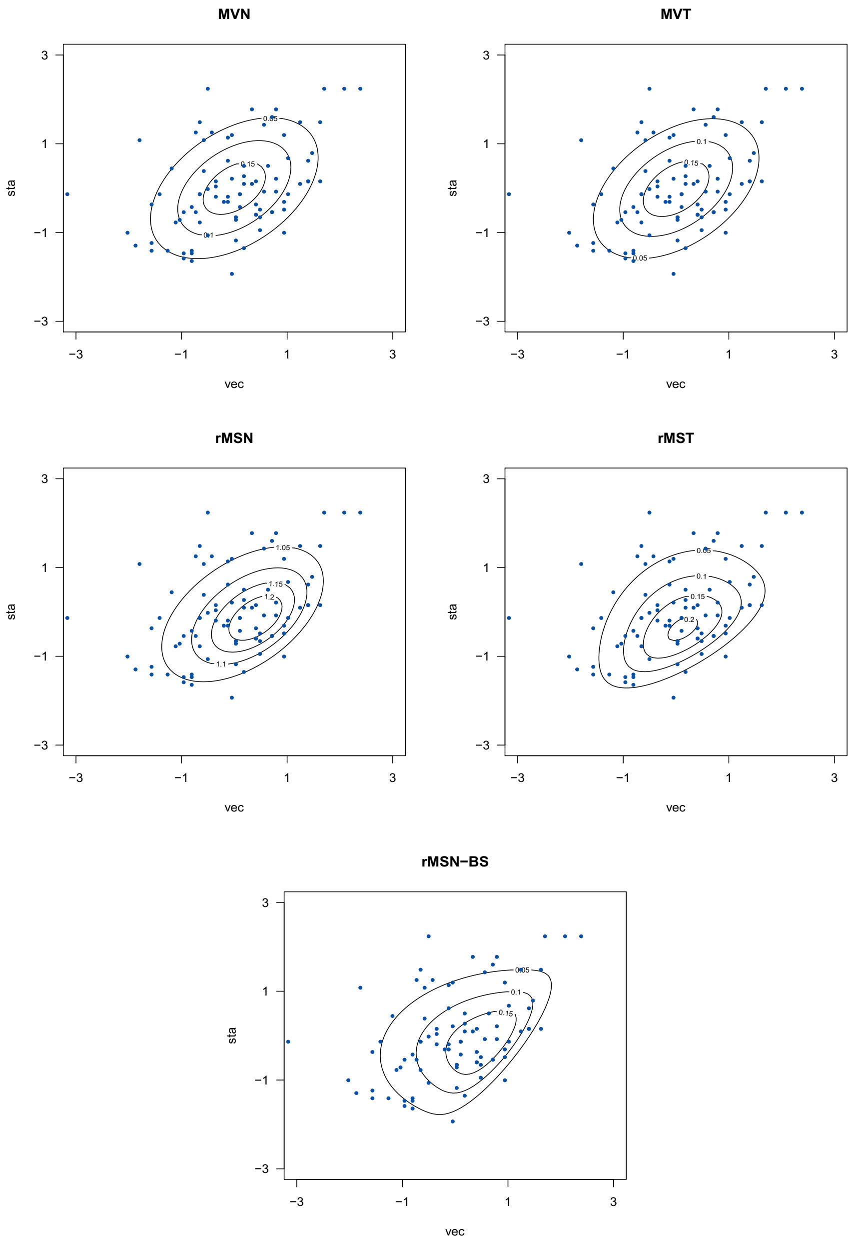

Figure 4. Scatter plot of vec and sta with superimposed contours of various models

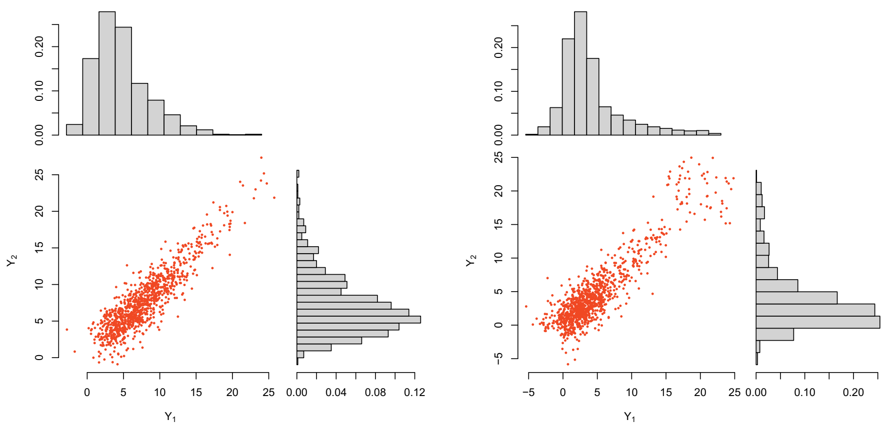

Figure 5. Plot of the generated sample in simulation 1 (left) and in simulation 2 (right)

$$\text{bias}(\theta_i) = \frac{1}{500}\sum_{j=1}^{500}(\hat{\theta}_i^{(j)} - \theta_i), \quad \text{MSE}(\theta_i) = \frac{1}{500}\sum_{j=1}^{500}\left(\hat{\theta}_i^{(j)} - \theta_i\right)^2.$$

The results from this table show that the Bias and MSE tend to approach zero by increasing the sample size. This indicate that the obtained estimates based on the proposed EM algorithm provide good asymptotic properties.

Table 3: Simulation results for assessing the asymptotic properties of parameter estimates

n	measure	ξ_1	ξ_2	λ_1	λ_1	σ_{11}	σ_{12}	σ_{22}	α
30	MSE	1.6729	2.0126	2.1701	2.2914	3.3615	3.2382	3.6019	19650
	bias	0.9611	1.3089	2.1937	2.4720	2.6825	2.1706	2.5630	1.7048
50	MSE	0.9627	1.3461	1.4917	1.5345	2.5711	2.4970	2.7743	1.2018
	bias	0.6709	0.8819	1.6130	1.7043	2.0157	1.4380	1.8901	1.0172
100	MSE	0.7400	1.0310	1.1445	1.2991	2.1595	2.1317	2.3369	0.8499
	bias	0.4861	0.6252	1.2312	1.1954	1.6810	1.2314	1.3990	0.6979
200	MSE	0.3352	0.6335	0.7467	0.6916	1.4622	1.1793	1.3185	0.4737
	bias	0.2295	0.3185	0.7241	0.6893	0.5132	0.6983	0.8694	0.3447
500	MSE	0.0915	0.1172	0.1355	0.1896	0.3301	0.4306	0.3410	0.1419
	bias	0.0101	0.0616	0.0910	0.0862	0.0910	0.0862	0.0921	0.0718
1000	MSE	0.0301	0.0475	0.0353	0.0457	0.0978	0.0980	0.0886	0.0368
	bias	0.0056	0.0038	0.0158	0.0141	0.0234	0.0126	0.0327	0.0242

The second simulation is conducted to check whether the method of approximating standard error of the rMSN-BS parameters, described in Section 5.1, has good asymptotic properties. For each sample size $n = 30, 50, 100, 200, 500$, and 1000, the data is generated 500 times from the rMSN-BS distribution with parameter values indicated in Table 4 (in parentheses). To increase the effect of tail heaviness on the standard errors, we add 5% of n nosies uniformly generated over the interval [10, 25] to the each synthetic sample. Therefore, n becomes 32, 53, 105, 210, 525, and 1050. One sample of size 1000 is plotted in Figure 5. By fitting the rMSN-BS distribution to the generated data and estimating their standard errors, we compute the sample standard errors of parameters (MCSE) and the observed information matrix (A.SE), as measures of verifying consistency of the standard errors estimates, by

$$\mathrm{MCSE}(\theta_i) = \sqrt{\frac{1}{499}\left[\sum_{j=1}^{500} \hat{\theta}_i^{(j)^2} - \frac{1}{500}\left(\sum_{j=1}^{500} \hat{\theta}_i^{(j)}\right)^2\right]},$$

$$\text{and} \quad \mathrm{A.SE}(\theta_i) = \frac{1}{500}\sum_{j=1}^{500} se(\theta_i^{(j)}).$$

The results of the simulation summarized in Table 4 shows that not only the values of MCSE and A.SE are decreased by increasing the sample size, they but also coverage reasonably together.

Table 4: Simulation results for assessing the consistency of standard errors

n	measure	$\xi_1(0)$	$\xi_2(0)$	$\lambda_1(2)$	$\lambda_1(2)$	$\sigma_{11}(2.5)$	$\sigma_{12}(0.6)$	$\sigma_{22}(2)$	$\alpha(0.5)$
30	A.SE	0.9705	1.0120	0.9840	1.2955	2.1190	1.7344	1.8915	0.9610
	MCSE	1.2340	1.3179	1.1029	1.4237	2.5367	2.1052	2.2094	1.1163
50	A.SE	0.8321	0.8609	0.8852	1.1850	1.8750	1.3683	1.5679	0.7930
	MCSE	0.9908	0.9763	0.9593	1.2543	2.2049	1.5917	1.7801	0.9648
100	A.SE	0.6552	0.7856	0.8381	0.9101	1.3565	1.1351	1.2257	0.7123
	MCSE	0.7856	0.8248	0.8919	0.9924	1.6675	1.2098	1.3630	0.8631
200	A.SE	0.4770	0.5854	0.6662	0.7301	0.6197	0.9314	0.8311	0.5795
	MCSE	0.4320	0.5330	0.6273	0.7633	0.7464	0.8546	0.8982	0.6167
500	A.SE	0.2557	0.3138	0.3984	0.4832	0.4498	0.7159	0.6185	0.3502
	MCSE	0.2606	0.3375	0.3858	0.5016	0.4665	0.7358	0.6391	0.3773
1000	A.SE	0.0385	0.0724	0.0937	0.1223	0.0935	0.1109	0.1442	0.0507
	MCSE	0.0348	0.0798	0.0980	0.1298	0.0978	0.1135	0.1492	0.0572

8. Concluding Remarks

This paper has dealt with the proposing new skew distribution by considering the BS model for the mixing variable in the scale mixtures of restricted skew-normal distribution. Calling rMSN-BS model, some statistical as well as mathematical properties of the new model have been investigated. By presenting a convenient hierarchical representation for the rMSN-BS distribution, we have developed a feasible EM-type algorithm for parameter estimation. Finally, we demonstrate our proposed methodology through a real and two simulation datasets. Numerical results show that the proposed method may perform reasonably well for the experimental data.

The utility of our current approach can be extended to introduce new distribution based on Lindley model ([33]). In addition, the rMSN-BS distribution can be used to construct a new finite mixture model for analyzing multimodal and fat tails datasets [34, 32, 36]. It may also be interesting to propose a model for clustering right-skewed positive data via fuzzy classification maximum likelihood algorithm [18].

Acknowledgements
If you'd like to thank anyone, place your comments here.

References

[1] A. C. Aitken, On bernoulli's numerical solution of algebraic equations. *Proceedings of the Royal Society of Edinburgh*, 46 (1927), 289–305.

[2] H. Akaike, Information theory and an extension of the maximum likelihood principle. *In: Springer Series in Statistics*, Springer New York, (1998), 199–213.

[3] O. Arslan, Variance-mean mixture of the multivariate skew normal distribution. *Statistical Papers*, 56 (2) (2014), 353–378.

[4] A. Azzalini, A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, (1985), 171–178.

[5] A. Azzalini, Further results on a class of distributions which includes the normal ones, *Statistica*, (1986), 46.

[6] A. Azzalini and A. Capitanio, Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61 (3) (1999), 579–602.

[7] A. Azzalini and A. Capitanio, Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65 (2) (2003), 367–389.

[8] A. Azzalini and M. G. Genton, Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review*, 76 (1) (2008), 106–129.

[9] A. Azzalini and A. D. Valle, The multivariate skew-normal distribution. *Biometrika*, 83 (4) (1996), 715–726.

[10] N. Balakrishnan and D. Kundu, Birnbaum-Saunders distribution: A review of models, analysis, and applications. *Applied Stochastic Models in Business and Industry*, (2018).

[11] O. Barndorff-Nielsen, Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of statistics*, (1978), 151–157.

[12] Z. W. Birnbaum and S. C. Saunders, A new family of life distributions. *Journal of applied probability*, 6 (2) (1969), 319–327.

[13] C. R. B. Cabral, V. H. Lachos, and M. O. Prates, Multivariate mixture modeling using skew-normal independent distributions. *Computational Statistics & Data Analysis*, 56 (1) (2012), 126–142.

[14] C. da Silva Ferreira, H. Bolfarine, and V. H. Lachos, Skew scale mixtures of normal distributions: properties and estimation. *Statistical Methodology*, 8 (2) (2011), 154–171.

[15] 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 (1) (1977), 1–38.

[16] A. M. Garay, V. H. Lachos, H. Bolfarine, and C. R. B. Cabral, Linear censored regression models with scale mixtures of normal distributions. *Statistical Papers*, 58 (1) (2015), 247–278.

[17] A. Gupta, Multivariate skew t-distribution. *Statistics: A Journal of Theoretical and Applied Statistics*, 37 (4) (2003), 359–363.

[18] F. Hashemi, M. Naderi, and M. Mashinchi, Clustering right-skewed data stream via Birnbaum-Saunders mixture models: A flexible approach based on fuzzy clustering algorithm. *Applied Soft Computing*, (2019), 105—539.

[19] H. J. Ho, S. Pyne, and T.I Lin, Maximum likelihood inference for mixtures of skew Student-t-normal distributions through practical EM-type algorithms. *Statistics and Computing*, 22 (2012), 287–299.

[20] A. Jamalizadeh, F. Hashemi, and M. Naderi, Discussion of “Birnbaum-Saunders distribution: A review of models, analysis, and applications”. *Applied Stochastic Models in Business and Industry*, (2018).

[21] B. Jörgensen, Statistical properties of the generalized inverse Gaussian distribution. *Springer Science & Business Media*, (1982).

[22] H. M. Kim, M. Maadooliat, R. B. Arellano-Valle, and M. G. Genton, Skewed factor models using selection mechanisms. *Journal of Multivariate Analysis*, 145 (2016), 162–177.

[23] A. J. Lemonte and G. M. Cordeiro, Discussion of "Birnbaum-Saunders distribution: A review of models, analysis, and applications". *Applied Stochastic Models in Business and Industry*, (2018).

[24] D. V. Lindley, Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, (1958), 102–107.

[25] K. V. Mardia, J. T. Kent, and J. M. Bibby, Multivariate Analysis, *Academic Press*, INC., 1979.

[26] C. Marchant, V. Leiva, M. F. Cavieres, and A. Sanhueza, Air Contaminant Statistical Distributions with Application to PM10 in Santiago, *Chile. Springer New York*, New York, NY, (2013), 1–31.

[27] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative risk management: Concepts, techniques and tools.* 3 (2005), Princeton university press Princeton.

[28] P. D. McNicholas, T. B. Murphy, A. F. McDaid, and D. Frost, Serial and parallel implementations of model-based clustering via parsimonious Gaussian mixture models. *Computational Statistics and Data Analysis*, 54 (2010), 711–723.

[29] I. Meilijson, A fast improvement to the EM algorithm to its own terms. *Journal of the Royal Statistical Society. Series B. Statistical Methodology)*, 51 (1989), 127–138.

[30] X. L. Meng and D. B. Rubin, Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80 (2) (1993), 267–278.

[31] K. Mohammadi, O. Alavi, and J. G. McGowan, Use of Birnbaum-Saunders distribution for estimating wind speed and wind power probability distributions: A review. *Energy Conversion and Management*, 143 (2017), 109–122.

[32] M. Naderi, A. Arabpour, and A. Jamalizadeh, Mixture of normal mean-variance of Lindley distributions. *Journal of Statistical Research of Iran*, 13 (2) (2017a), 197–214.

[33] M. Naderi, A. Arabpour, and A. Jamalizadeh, Multivariate normal mean-variance mixture distribution based on lindley distribution. *Communications in Statistics-Simulation and Computation*, 47 (4) (2017b), 1179–1192.

[34] M. Naderi, A. Arabpour, and A. Jamalizadeh, On the finite mixture modelling via normal mean-variance Birnbaum-Saunders dstribution. *Journal of the Iranian Statistical Society*, 16 (1) (2017c), 33–52.

[35] M. Naderi, A. Arabpour, T. I Lin, and A. Jamalizadeh, Nonlinear regression models based on the normal mean–variance mixture of Birnbaum–Saunders distribution, *Journal of the Korean Statistical Society*, 46 (3) (2017), 476–485.

[36] M. Naderi, W. L. Hung, T. I Lin, and A. Jamalizadeh, A novel mixture model using the multivariate normal mean-variance mixture of Birnbaum-Saunders distributions and its application to extrasolar planets, *Journal of Multivariate Analysis*, 171 (2019), 126–138.

[37] R. Pourmousa, A. Jamalizadeh, and M. Rezapour, Multivariate normal mean-variance mixture distribution based on Birnbaum-Saunders distribution, *Journal of Statistical Computation and Simulation*, 85 (13) (2015), 2736–2749.

[38] S. Pyne, X. Hu, K. Wang, E. Rossin, T. Lin, L. M. Maier, C. Baecher-Allan, G. J. McLachlan, P. Tamayo, D. A. Hafler, P. L. D. Jager, and J. P. Mesirov, Automated high-dimensional flow cytometric data analysis. *Proceedings of the National Academy of Sciences*, 106 (21) (2009), 8519–8524.

[39] H. Saulo, V. Leiva, F. A Ziegelmann, and C. Marchant, A nonparametric method for estimating asymmetric densities based on skewed Birnbaum–Saunders distributions applied to environmental data, *Stochastic Environmental Research and Risk Assessment*, 27 (6) (2013), 1479–1491.

[40] G. Schwarz, Estimating the dimension of a model, *Ann. Statist*, 6 (1978), 461–464.

[41] T. Wang, B. Li, and A. K. Gupta, Distribution of quadratic forms under skew normal settings, *Journal of Multivariate Analysis*, 100 (3) (2009), 533–545.

[42] M. Watanabe and K. Yamaguchi, The EM Algorithm and Related Statistical Models, *in: Statistics: A Series of Textbooks and Monographs*, CRC Press, (2003).

Hossein Samary
Ph.D Student of Statistics
Department of Statistics
Marvdasht Branch, Islamic Azad University
Marvdasht, Iran
E-mail: Hossein.samary@yahoo.com

Zahra Khodadadi
Assistant Professor of Statistics
Department of Statistics
Marvdasht Branch, Islamic Azad University
Marvdasht, Iran
E-mail: Zkhodadadi@miau.ac.ir

Hedieh Jafarpour
Assistant Professor of Statistics
Department of Statistics
Shiraz Islamic Azad University
Shiraz, Iran
Email: Hedieh.jafarpour@gmail.com