JIRSS (2019)
Vol. 18, No. 02, pp 87-113
DOI:10.29252/jirss.18.2.87

Skew–Normal Mean–Variance Mixture of Birnbaum–Saunders Distribution and Its Associated Inference and Application

Mostafa Tamandi [1], Hossein Negarestani [2,3], Ahad Jamalizadeh [4], and Mehdi Amiri[5]

[1] Department of Statistics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran.
[2] Young Researchers Society, Shahid Bahonar University of Kerman, Kerman, Iran.
[3,4] Department of Statistics, Faculty of Mathematics and Computers, Shahid Bahonar University of Kerman, Kerman, Iran.
[5] Department of Statistics, Faculty of Basic Sciences, University of Hormozgan, Bandar abbas, Iran.

Received: 16/02/2018, Revision received: 26/12/2018, Published online: 29/08/2019

Abstract. This paper presents a skew-normal mean-variance mixture based on Birnbaum-Saunders (SNMVBS) distribution and discusses some of its key properties. The SNMVBS distribution can be thought as a flexible extension of the normal mean-variance mixture based on Birnbaum-Saunders (NMVBS) distribution as it possesses one additional shape parameter for providing more flexibility with skewness and kurtosis. Next, we develop a computationally feasible ECM algorithm for the maximum likelihood estimation of the model parameters. Asymptotic standard errors of the ML estimates are obtained through an approximation of the observed information matrix. Finally, the usefulness of the proposed model and its fitting method are illustrated through a Monte-Carlo simulation as well as three real-life datasets.

Keywords. Birnbaum-Saunders, ECM Algorithm, Observed Information Matrix, Ro-

Mostafa Tamandi (tamandi@vru.ac.ir)
Hossein Negarestani (hnegarestani@gmail.com)
Corresponding Author: Ahad Jamalizadeh (A.jamalizadeh@uk.ac.ir)
Mehdi Amiri (m.amiri@hormozgan.ac.ir)

bustness, Scale-Shape Mixtures.

MSC: 60E05; 62F10.

1 Introduction

The normal distribution is a popular model for many practical problems since it has many desirable statistical properties. However, in many applications such as finance, engineering, medicine, and actuarial sciences, the density is usually strongly peaked, skewed, and heavy-tailed than the normal distribution. To model such a data, many authors have proposed distributions that are more flexible than the normal distribution in terms of skewness and tail thickness.

In a well-known attempt, Barndorff–Nielsen (1977) introduced a broad class of normal mean-variance mixture (NMVM) distributions. The idea behind the NMVM distribution is to introduce randomness into the variance and the mean of a normal distribution via a positive mixing variable. Specifically, let Y denote an NMVM random variable. Then, Y can be represented as

$$Y = \xi + W\lambda + W^{1/2}Z, \tag{1.1}$$

where $Z \sim N(0, \sigma^2)$ and $W > 0$ is a scalar-valued random variable that is independent of Z. The parameters ξ and λ are in $\mathbb{R}$ and $\sigma > 0$. Generalized hyperbolic (GH) distribution, introduced by Barndorff–Nielsen (1978), is one of the popular distributions in this class. This class includes several important distributions such as skew Laplace (Arslan , 2010) and Normal inverse Gaussian (NIG) as special cases. The random variable with GH distribution can be represented as an NMVM variable when W in (1.1) is the generalized inverse Gaussian (GIG) random variable. A positive random variable W follows a GIG distribution, denoted by $W \sim GIG(\kappa, \chi, \psi)$, if its probability density function (PDF) is given by

$$f_{GIG}(w; \kappa, \chi, \psi) = (\frac{\psi}{\chi})^{\kappa/2} \frac{w^{\kappa-1}}{2K_\kappa(\sqrt{\psi\chi})} \exp\left\{-\frac{1}{2}(w^{-1}\chi + w\psi)\right\}, \quad w > 0, \tag{1.2}$$

where $K_\kappa(\cdot)$, $\kappa \in \mathbb{R}$, denotes the modified Bessel function of the third kind with property $K_\kappa(\cdot) = K_{-\kappa}(\cdot)$. The parameters χ and ψ are such that $\chi \geq 0, \psi > 0$ if $\kappa > 0$; $\psi \geq 0, \chi > 0$ if $\kappa < 0$, and $\chi > 0, \psi > 0$ if $\kappa = 0$. This unimodal density contains gamma and inverse gamma densities as special cases when $\chi = 0$ and $\psi = 0$, respectively.

Let the random variable Y be represented as in (1.1), where $W \sim GIG(\kappa, \chi, \psi)$ and $Z \sim N(0, \sigma^2)$. Then, the PDF of Y is given by

$$f_{GH}\left(y; \xi, \sigma^2, \lambda, \kappa, \chi, \psi\right) = C \frac{K_{\kappa-1/2}\left(\sqrt{(\frac{(y-\xi)^2}{\sigma^2} + \chi)(\frac{\lambda^2}{\sigma^2} + \psi)}\right)}{\left(\sqrt{(\frac{(y-\xi)^2}{\sigma^2} + \chi)(\frac{\lambda^2}{\sigma^2} + \psi)}\right)^{\frac{1}{2}-\kappa}} e^{\lambda \frac{(y-\xi)}{\sigma^2}}, \ y \in \mathbb{R},$$

where $C = \frac{(\frac{\psi}{\chi})^{\kappa/2}(\frac{\lambda^2}{\sigma^2}+\psi)^{\frac{1}{2}-\kappa}}{\sqrt{2\pi}\sigma K_\kappa\left(\sqrt{\chi\psi}\right)}$. The corresponding cumulative distribution function (CDF) is denoted by $F_{GH}(\cdot; \xi, \sigma^2, \lambda, \kappa, \chi, \psi)$. For further details about GIG and GH distributions, one may refer to Hu (2005).

Pourmousa et al. (2015) presented an NMVBS distribution, which is another NMVM distribution with Birnbaum-Saunders (BS) as a mixing random variable (Birnbaum and Saunders , 1969). A positive random variable W follows a BS distribution with shape parameter α and scale parameter β, denoted by $W \sim BS(\alpha, \beta)$, if its PDF is given by

$$f_{BS}(w; \alpha, \beta) = \frac{w + \beta}{2\alpha\sqrt{\beta w^3}} \phi\left(\frac{1}{\alpha}\left\{\sqrt{\frac{w}{\beta}} - \sqrt{\frac{\beta}{w}}\right\}\right), \ \ w > 0; \alpha > 0, \beta > 0.$$

In a similar vein, Azzalini (1985) proposed the skew-normal (SN) distribution by adding skewness to the normal distribution. Let $Y \sim SN(\xi, \sigma^2, \lambda)$ denote a random variable distributed as the SN distribution with location parameter ξ, scale parameter σ^2, and skewness parameter λ. The PDF of Y is given by

$$f_{SN}(y; \xi, \sigma^2, \lambda) = \frac{2}{\sigma}\phi(u)\Phi(\lambda u), \quad y \in \mathbb{R},$$

where $u = (y - \xi)/\sigma$, and $\phi(\cdot)$ and $\Phi(\cdot)$, respectively, denote the PDF and CDF of standard normal distribution. If $Z \sim SN(0, \sigma^2, \lambda)$ is used in (1.1), then a skew-normal mean-variance mixture distribution is obtained. In a multivariate setup, Arslan (2015) assumed that W is a GIG distribution and studied properties of that model.

This paper aims to introduce a skew-normal mean-variance mixture based on the BS (SNMVBS) distribution. The random variable $Y \in \mathbb{R}$ has an SNMVBS distribution if it has the representation

$$Y = \xi + W\lambda_1 + W^{1/2}Z, \tag{1.3}$$

where $Z \sim SN(0; \sigma^2, \lambda_2)$ and $W \sim BS(\alpha, 1)$ are independent random variables. The parameters ξ, λ_1, and λ_2 are in $\mathbb{R}$ and $\sigma^2 > 0$.

The SNMVBS model extends the NMVBS distribution and serves as an alternative to the skew-t (ST) model introduced by Azzalini and Capitaino (2003). The density of the ST distribution is given by

$$f_{ST}(y; \xi, \sigma^2, \lambda, \nu) = \frac{2}{\sigma} t(u_1; \nu) T\left(\lambda u_1 \sqrt{\frac{\nu + 1}{\nu + u_1^2}}; \nu + 1\right), \tag{1.4}$$

where $u_1 = \frac{y-\xi}{\sigma}$. Also, $t(\cdot; \nu)$, and $T(\cdot; \nu)$ denote the PDF and CDF of Student's t distribution having the degrees of freedom (df) of ν, respectively. We shall adopt the notation $Y \sim ST(\xi, \sigma^2, \lambda, \nu)$ if Y has the PDF given by (1.4).

Although the ST distribution is also a mean-variance mixture of SN distribution, it is well-known that the estimation of the degree of freedom parameter in the ST model poses difficulties. Interestingly, the ML estimates of the parameters in SNMVBS model have been obtained by solving some simple linear equations. But, with an additional scale parameter λ_1, the SNMVBS model possesses more flexibility to provide a better fit than the ST model.

The rest of this paper is organized as follows: Section 2 introduces the SNMVBS distribution and discusses some of its key properties. Section 3 develops an expectation-conditional maximization (ECM) algorithm for the estimation of the parameters of the SNMVBS distribution and also explains a method for obtaining the standard errors of the ML estimates. In Section 4, the performance of this estimation method is evaluated using a Monte Carlo simulation study. Section 5 illustrates the usefulness of the proposed model and its fitting method with three real-data sets. A multivariate version of the proposed model is discussed in Section 6. Finally, some concluding remarks are made in Section 7.

2 SNMVBS Distribution: Properties and Characteristics

Let Y be a random variable following the representation in (1.3). Then, Y is said to follow an SNMVBS distribution, and we denote it by $Y \sim SNMVBS(\xi, \sigma^2, \lambda_1, \lambda_2, \alpha)$. Using (1.3) and the known properties of SN distribution, a hierarchical representation of an SNMVBS random variable can be provided as follows

$$\begin{aligned} Y|W = w &\sim SN(\xi + w\lambda_1, w\sigma^2, \lambda_2), \\ W &\sim BS(\alpha, 1). \end{aligned} \tag{2.1}$$

The following lemma is useful for deriving the density of Y.

Lemma 2.1. *If $W \sim GIG(\kappa, \chi, \psi)$, then:*

(i) $W^{-1} \sim GIG(-\kappa, \chi, \psi)$,

(ii) $E(W^r) = (\frac{\chi}{\psi})^{r/2} R_{(\kappa,r)}(\sqrt{\psi\chi})$,

(iii) $E_W\left(\Phi(aW^{1/2} - bW^{-1/2})\right) = P(Y < a)$,

where $R_{(\kappa,r)}(x) = \frac{K_{\kappa+r}(x)}{K_\kappa(x)}$, $a \in \mathbb{R}$, *and* $b \in \mathbb{R}$. *The random variable Y follows a univariate generalized hyperbolic distribution with* $\xi = 0, \sigma^2 = 1, \lambda = b, \chi$ *and* ψ.

Parts *(i)* and *(ii)* of the above lemma are well-known properties of GIG distribution; see, e.g., Hu (2005) for details. Part *(iii)* has been proved by Arslan (2015).

Theorem 2.1. *Let the random variable Y follow the representation in* (2.1). *Then, the PDF of Y is given by*

$$\begin{aligned} f_{SNMVBS}(y; \xi, \sigma^2, \lambda_1, \lambda_2, \alpha) &= f_{GH}(y; \xi, \sigma^2, \lambda_1, -\frac{1}{2}, \alpha^{-2}, \alpha^{-2})H_1 \\ &+ f_{GH}(y; \xi, \sigma^2, \lambda_1, \frac{1}{2}, \alpha^{-2}, \alpha^{-2})H_2, \end{aligned} \tag{2.2}$$

where H_1 and H_2 refer to $F_{GH}(a; 0, 1, b, \kappa, \delta_1, \delta_2)$ with $\kappa = -1$ and 0, respectively. Further, $\delta_1 = u^2 + \alpha^{-2}$, $\delta_2 = \frac{\lambda_1^2}{\sigma^2} + \alpha^{-2}$, $a = \lambda_2 u$, $b = \frac{\lambda_1\lambda_2}{\sigma}$, *and* $u = (y - \xi)/\sigma$.

Proof. From (2.1), the PDF of Y is given by

$$
\begin{aligned}
f(y) &= \int_0^\infty f_{SN}(y|w) f_{BS}(w) dw \\
&= 2\int_0^\infty \phi(\frac{w^{-1/2}}{\sigma}(y-\xi-w\lambda_1))\Phi(\frac{\lambda_2 w^{-1/2}}{\sigma}(y-\xi-w\lambda_1)) f_{BS}(w) dw \\
&= \int_0^\infty \frac{w^{-2}(1+w)}{(2\pi)\alpha\sigma}\Phi(w^{-1/2}\lambda_2 u - w^{1/2}\lambda_2\lambda_1/\sigma) \\
&\quad \times \exp\left\{-\frac{1}{2}\left(\frac{w^{-1}}{\sigma^2}(y-\xi-w\lambda_1)^2 + \frac{1}{\alpha^2}(\sqrt{w}-\frac{1}{\sqrt{w}})^2\right)\right\} dw \\
&= \frac{\exp\{\lambda_1(y-\xi)/\sigma^2\}}{2(2\pi)\sigma K_{1/2}(\alpha^{-2})} \\
&\quad \times \int_0^\infty w^{-2}(1+w)\Phi(w^{-1/2}a - w^{1/2}b)\exp\left\{-\frac{1}{2}\left(\delta_1 w^{-1} + \delta_2 w\right)\right\} dw \\
&= f_{GH}(y;\xi,\sigma^2,\lambda_1,-1/2,\alpha^{-2},\alpha^{-2}) E_{W_1}\left(\Phi(W_1^{-1/2}a - W_1^{1/2}b)\right) \\
&\quad + f_{GH}(y;\xi,\sigma^2,\lambda_1,1/2,\alpha^{-2},\alpha^{-2}) E_{W_2}\left(\Phi(W_2^{-1/2}a - W_2^{1/2}b)\right).
\end{aligned}
$$

where $W_1 \sim GIG(-1,\delta_1,\delta_2)$ and $W_2 \sim GIG(0,\delta_1,\delta_2)$. The proof is now completed by using Part *(iii)* of Lemma 2.1. □

The notation used in Theorem 2.1 will be used throughout this paper. In the special cases, if $\lambda_1 = 0$, then the PDFin (2.2) reduces to a scale mixture of two SN distributions with the BS model as mixing distribution. In comparison, the ST distribution with the PDF (1.4) is a scale mixture of one SN distribution when the mixing distribution is Gamma with $\nu/2$ as the scale and shape parameters. Also, if $\lambda_2 = 0$, then Y has NMVBS distribution studied by Pourmousa et al. (2015) and the PDF of Y reduces to

$$
\begin{aligned}
f_{NMVBS}(y;\xi,\sigma^2,\lambda_1,\lambda_2,\alpha) &= \frac{1}{2} f_{GH}(y;\xi,\sigma^2,\lambda_1,-\frac{1}{2},\alpha^{-2},\alpha^{-2}) \\
&+ \frac{1}{2} f_{GH}(y;\xi,\sigma^2,\lambda_1,\frac{1}{2},\alpha^{-2},\alpha^{-2}). \qquad (2.3)
\end{aligned}
$$

Figure 1 presents some plots of the density function in (2.2). We have taken $\xi = 0$ and $\sigma^2 = 1$ in all cases. In this figure, we compare the SNMVBS model with NMVBS, SN, and ST models. It is clear that in the cases of SN and ST, the density of SNMVBS has fatter tails when λ_1 is far from zero. Also, with $\lambda_1 < 0$, the SNMVBS model has a heavier left tail as compared to the SN and ST models. Similarly, when $\lambda_1 > 0$, the distribution has a heavier right tail. In comparison with NMVBS distribution, the SNMVBS density has a heavier right tail than the NMVBS when $\lambda_2 > 0$.

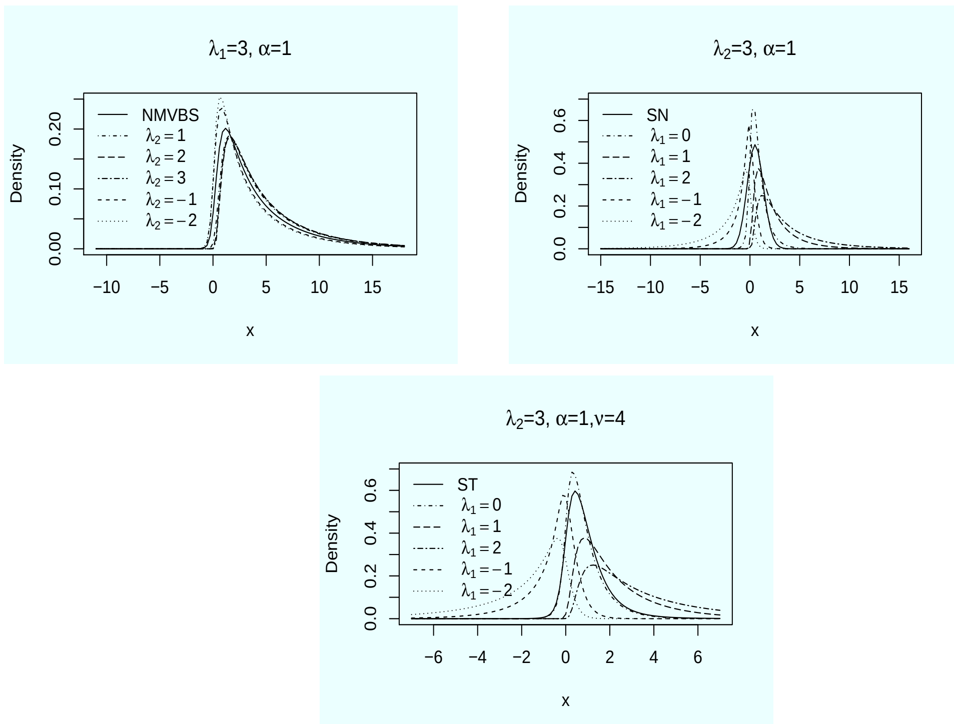


Figure 1: The PDF of the SNMVBS distribution for different parameter settings and a comparison with the NMVBS, skew-normal and skew-*t* distributions.

Lemma 2.2. *Let* $W \sim BS(\alpha, 1)$. *Then:*

(i) The PDF of W can be represented as

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

(ii) $E(W^r) = \frac{1}{2}\left(R_{(\frac{1}{2},r)}(\alpha^{-2}) + R_{(-\frac{1}{2},r)}(\alpha^{-2})\right)$.

In paticular, we have $E(W) = \left(1 + \frac{1}{2}\alpha^2\right)$ *and* $Var(W) = \alpha^2\left(1 + \frac{5}{4}\alpha^2\right)$.

The proof of Lemma 2.2 and some other useful properties of BS distribution can be found in Leiva (2016). Based on Part (i) of Lemma 2.2, the BS distribution is a mixture of two GIG distributions with special parameters. As mentioned in the introduction, Arslan (2015) introduced a mean-variance mixture of the skew-normal distribution with the GIG as mixing distribution. So, the SNMVBS model is an extension of Arslan's model when the proper settings of parameters are derived.

The following theorem gives some orders of moments of the SNMVBS distribution which are useful in studying some properties of this model.

Theorem 2.2. *Let* $Y \sim SNMVBS(\xi, \sigma^2, \lambda_1, \lambda_2, \alpha)$. *Then, the first four moments of* Y *are as follows*

$$\begin{aligned}
\mu_1 &= E(Y) = \xi + \omega_1,\\
\mu_2 &= E(Y^2) = \xi^2 + 2\xi\omega_1 + \omega_2,\\
\mu_3 &= E(Y^3) = \xi^3 + 3\xi^2\omega_1 + 3\xi\omega_2 + \omega_3 + \zeta_1 a_{3/2},\\
\mu_4 &= E(Y^4) = \xi^4 + 4\xi^3\omega_1 + 6\xi^2\omega_2 + 4\xi\left(\omega_3 + \zeta_2 a_{3/2}\right) + \omega_4 + 4\lambda_1\zeta_2 a_{5/2}, \qquad (2.4)
\end{aligned}$$

where

$$\begin{aligned}
\omega_1 &= \lambda_1 a_1 + \sqrt{\frac{2}{\pi}}\delta\sigma a_{1/2},\\
\omega_2 &= \lambda_1^2 a_2 + \sigma^2 a_1 + 2\sqrt{\frac{2}{\pi}}\delta\sigma\lambda_1 a_{3/2},\\
\omega_3 &= \lambda_1^3 a_3 + 3\lambda_1\sigma^2 a_2 + 3\sqrt{\frac{2}{\pi}}\delta\sigma\lambda_1^2 a_{5/2},\\
\omega_4 &= \lambda_1^4 a_4 + 6\sigma^2\lambda_1^2 a_3 + 3\sigma^4 a_2 + 4\sqrt{\frac{2}{\pi}}\delta\sigma\lambda_1^3 a_{7/2},
\end{aligned}$$

and $\zeta_1 = \sqrt{\frac{2}{\pi}}\delta\sigma^3\left(2 + \frac{1}{1+\lambda_2^2}\right)$, $\zeta_2 = \frac{1}{\sqrt{2\pi}}\delta\sigma^3\left(5 + \frac{2}{1+\lambda_2^2}\right)$ *with* $\delta = \frac{\lambda_2}{\sqrt{1+\lambda_2^2}}$, $a_1 = 1 + \frac{\alpha^2}{2}$, *and* $a_2 = 1 + 2\alpha^2 + \frac{3}{2}\alpha^4$. *Moreover,* a_k*'s, for* $k = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \frac{7}{2}$ *can be obtained from Part (ii) of Lemma 2.2 where* $a_k = E(W^k)$.

Proof. The proof is obtained by using iterated expectations on representation (2.1) and Part *(ii)* of Lemma 2.2. □

Thus, using the moments in Theorem 2.2, the variance and the coefficients of skewness (γ_y),and kurtosis (κ_y) of Y can be readily computed.

Figure 2 shows the skewness and kurtosis contours of the SNMVBS, NMVBS, and ST distributions. In all cases, we have taken $\xi = 0$ and $\sigma^2 = 1$. The figures offer helpful information on how γ_y and κ_y change with different values of λ_1, λ_2, α, and ν. It should be noted that the kurtosis of the SNMVBS distribution range over a slightly wider interval than the NMVBS and ST distributions.

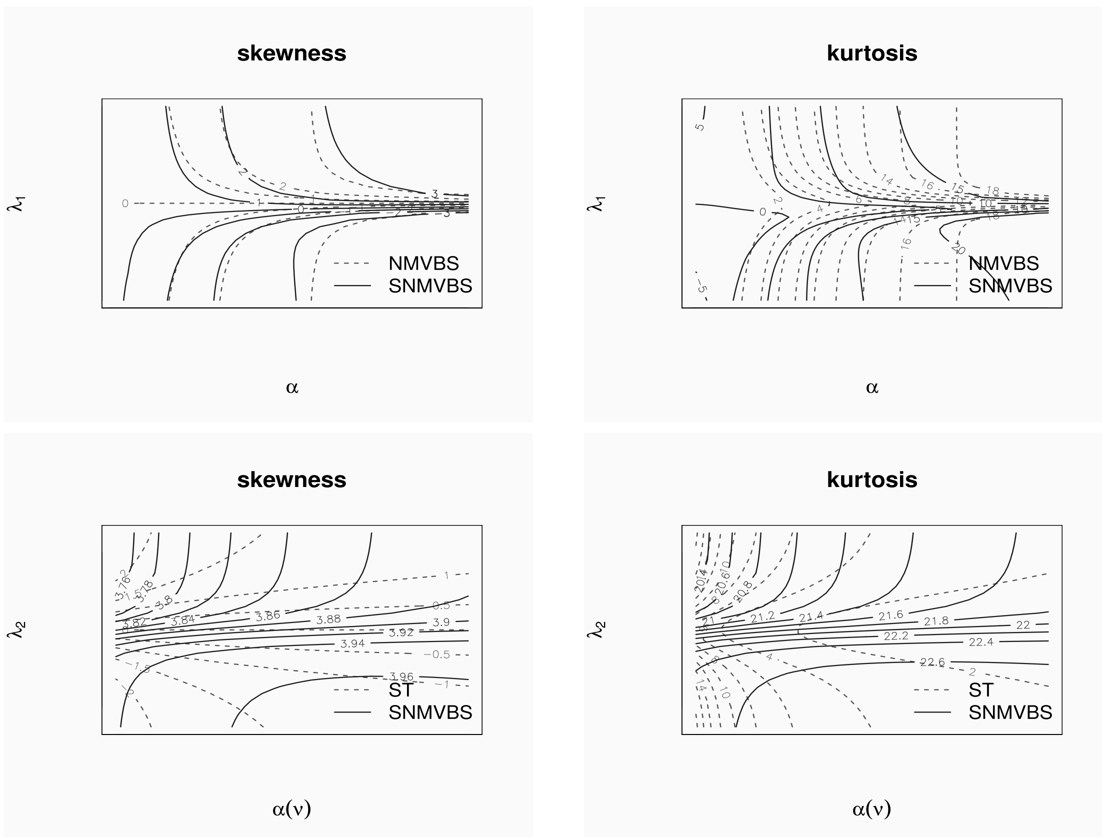


Figure 2: A comparison of skewness and kurtosis contours of the SNMVBS (λ_2 = 1) and NMVBS distributions for different combinations of λ_1 and α (top), and SNMVBS (λ_1 = 2) and ST distributions for different combinations of λ_2 and α (or ν) (bottom).

From the well-known convolution-type of SN and (2.1), we readily obtain

$$Y = \xi + W\lambda_1 + W^{1/2}\sigma\left\{\frac{\lambda_2}{\sqrt{1+\lambda_2^2}}|U_0| + \frac{1}{\sqrt{1+\lambda_2^2}}U_1\right\},$$

where U_0 and U_1 are two independent $N(0,1)$ random variables. Denoting by $\gamma = w^{1/2}\sqrt{1+\lambda_2^2}|U_0|$, then the hierarchical representation of SNMVBS is given by

$$\begin{aligned} Y|(W=w,\gamma) &\sim N\left(\xi + w\lambda_1 + \frac{\sigma\lambda_2\gamma}{1+\lambda_2^2}, w\frac{\sigma^2}{1+\lambda_2^2}\right), \\ \gamma|(W=w) &\sim TN\left(0, w(1+\lambda_2^2)\,;\,(0,\infty)\right), \\ W &\sim BS(\alpha,1), \end{aligned} \tag{2.5}$$

where $TN(\mu,\sigma^2;(a,b))$ represents the truncated normal distribution for $N(\mu,\sigma^2)$ lying within the interval (a,b).

The conditional distribution of γ, given $Y=y$ and $W=w$, is given by

$$f(\gamma|y,w) = \frac{f(y,\gamma,w)}{f(y,w)} = \frac{f(y|\gamma,w)f(\gamma|w)f(w)}{\int_0^\infty f(y,\gamma,w)d\gamma}.$$

Using representation (2.5) and performing some algebra, we find

$$\gamma|(W=w,Y=y) \sim TN\left(\lambda_2 u_2, w; (0,\infty)\right),$$

where $u_2 = (y-\xi-w\lambda_1)/\sigma$ and

$$\mu_\gamma = \lambda_2 u_2\ ,\ \sigma_\gamma^2 = w.$$

Thus, we have

$$E(\gamma|W=w,Y=y) = \mu_\gamma + \sigma_\gamma\frac{\phi\left(\frac{\mu_\gamma}{\sigma_\gamma}\right)}{\Phi\left(\frac{\mu_\gamma}{\sigma_\gamma}\right)}, \tag{2.6}$$

and by the known properties of conditional expectations, we have

$$\begin{aligned} E(\gamma \mid y) &= E(E(\gamma \mid w,y) \mid y), \\ E(\gamma W^{-1} \mid y) &= E(W^{-1}E(\gamma \mid w,y) \mid y). \end{aligned} \tag{2.7}$$

Theorem 2.3. *Let $Y \sim SNMVBS(\xi, \sigma^2, \lambda_1, \lambda_2, \alpha)$ and $W \sim BS(\alpha, 1)$. Then:*
(i)

$$E(W^n \mid y) = \left(\omega_1(y) H_3 R_{(-1,n)}(\sqrt{\delta_1\delta_2}) + \omega_2(y) H_4 R_{(0,n)}(\sqrt{\delta_1\delta_2})\right)\left(\frac{\delta_1}{\delta_2}\right)^{n/2}, n = \pm 1, \tag{2.8}$$

where

$$\begin{aligned}
\omega_1(y) &= \frac{1}{g(y)} f_{GH}(y; \xi, \sigma^2, \lambda_1, -1/2, \alpha^{-2}, \alpha^{-2}),\\
\omega_2(y) &= \frac{1}{g(y)} f_{GH}(y; \xi, \sigma^2, \lambda_1, 1/2, \alpha^{-2}, \alpha^{-2}),
\end{aligned}$$

and H_3 and H_4 denote the CDF's defined in Theorem 2.1 with $\kappa = n-1$ and n, respectively;
(ii)

$$E\left(W^m q\left(W^{-1/2}\lambda_2 U_2\right) \mid y\right) = \frac{\left(\omega_1^*(y) R_{(\frac{1}{2},m)}(\alpha^{-2}) + \omega_2^*(y) R_{(-\frac{1}{2},m)}(\alpha^{-2})\right)}{\sqrt{2\pi(1+\lambda_2^2)}}, m = \pm 1/2, \tag{2.9}$$

where

$$\begin{aligned}
\omega_1^*(y) &= \frac{1}{g(y)} f_{GH}(y; \xi, \frac{\sigma^2}{1+\lambda_2^2}, \lambda_1, m+1/2, \alpha^{-2}, \alpha^{-2}),\\
\omega_2(y) &= \frac{1}{g(y)} f_{GH}(y; \xi, \frac{\sigma^2}{1+\lambda_2^2}, \lambda_1, m-1/2, \alpha^{-2}, \alpha^{-2}),
\end{aligned}$$

and $q(x) = \frac{\phi(x)}{\Phi(x)}$.

Proof. The conditional density $f(w \mid y)$ is obtained by the Bayes rule, which is a mixture of two GIG PDF's by Part *(i)* of Lemma 2.2. Then, we can get $E(W^n \mid y)$ using Part *(ii)* of Lemma 2.1. The second conditional expectation can be obtained straightforward manner. □

It is difficult to find the parameter estimates for the SNMVBS distribution without resorting to some data augmentation techniques such as the EM algorithm and its variants. The conditional moments given in Theorem 2.3 in this case, become quite useful in the development of an EM-type algorithm discussed in the next section.

3 Parameter Estimation

3.1 ECM Algorithm

The expectation-maximization (EM) algorithm (Dempster *et al.* , 1977) is a versatile tool for the maximum likelihood (ML) estimation in the case of missing data or latent variables. Simplicity in implementation and monotone convergence are two important features of the EM procedure. But, it is not directly applicable to estimate the SNMVBS model since the M-step involves intractable computations. To avoid this complication, we propose to use the ECM algorithm (Meng and Rubin , 1993), which replaces the M-step of EM by a sequence of simpler conditional maximization steps.

Let $y_1, \ldots, y_n$ be n observations on Y, and the corresponding unobserved random values are represented by $\gamma_1, \ldots, \gamma_n$ and $w_1, \ldots, w_n$. Using (2.5), we have the following hierarchical representation:

$$\begin{aligned} Y_i|(w_i, \gamma) &\sim N\left(\xi + w_i\lambda_1 + \frac{\sigma\lambda_2\gamma_i}{1+\lambda_2^2}, w_i\frac{\sigma^2}{1+\lambda_2^2}\right), \\ \gamma_i|w_i &\sim TN\left(0, w_i(1+\lambda_2^2)\ ;\ (0,\infty)\right), \\ W_i &\sim BS(\alpha, 1). \end{aligned} \tag{3.1}$$

Then, under the hierarchical representation in (3.1) and $\nu = \lambda_2/\sigma$, it follows that the complete data log-likelihood function of $\boldsymbol{\theta} = (\xi, \sigma, \lambda_1, \lambda_2, \alpha)$ for the complete data $\mathbf{y}_c = (y_1, \ldots, y_n, \gamma_1, \ldots, \gamma_n, w_1, \ldots, w_n)$ is given by

$$\begin{aligned} \ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = &-\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n} w_i^{-1}\frac{(y-\xi-w_i\lambda_1)^2}{\sigma^2} \\ &-\frac{1}{2}\sum_{i=1}^{n} w_i^{-1}\left(\gamma_i - \nu(y-\xi-w_i\lambda_1)\right)^2 \\ &- n\log\alpha - \frac{1}{2\alpha^2}\sum_{i=1}^{n}(w_i + w_i^{-1} - 2). \end{aligned} \tag{3.2}$$

Now, to perform the ECM algorithm, we start with the E-step, given the current parameter $\hat{\boldsymbol{\theta}}^{(k)} = (\hat{\xi}^{(k)}, \hat{\sigma}^{(k)}, \hat{\lambda}_1^{(k)}, \hat{\lambda}_2^{(k)}, \hat{\alpha}^{(k)})$. Then, we compute the expected value of $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$, denoted by $Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)}) = E(\ell_c(\boldsymbol{\theta} \mid \boldsymbol{y}_c) \mid y_1, \ldots, y_n, \hat{\boldsymbol{\theta}}^{(k)})$, which involves some conditional expectations, including

$$
\begin{aligned}
\hat{q}_{1i}^{(k)} &= E(W_i^{-1} \mid y_i, \hat{\boldsymbol{\theta}}^{(k)}), \quad \hat{q}_{2i}^{(k)} = E(W_i \mid y_i, \hat{\boldsymbol{\theta}}^{(k)}), \\
\hat{q}_{3i}^{(k)} &= E(\gamma_i W_i^{-1} \mid y_i, \hat{\boldsymbol{\theta}}^{(k)}), \quad \hat{q}_{4i}^{(k)} = E(\gamma_i \mid y_i, \hat{\boldsymbol{\theta}}^{(k)}).
\end{aligned} \tag{3.3}
$$

The quantities $\hat{q}_{1i}^{(k)}$ and $\hat{q}_{2i}^{(k)}$ are obtained from (2.3), while $\hat{q}_{3i}^{(k)}$ and $\hat{q}_{4i}^{(k)}$, can be found from (2.7) as

$$
\begin{aligned}
\hat{q}_{3i}^{(k)} &= \hat{q}_{1i}^{(k)} \lambda_2 u_i - \lambda_1 \nu + M_1, \\
\hat{q}_{4i}^{(k)} &= \lambda_2 u_i - \hat{q}_{2i}^{(k)} \lambda_1 \nu + M_2.
\end{aligned}
$$

where M_1 and M_2 refer to $E\left(W^m q\left(W^{-1/2}\lambda_2 U_2\right) \mid y\right)$ with $m = -1/2$ and $1/2$, respectively, which are as defined in (2.3).

Thus, the ECM algorithm proceeds as follows:

E-step: Given the current value $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, calculate the Q-function as

$$
\begin{aligned}
Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)}) = &-\frac{n}{2}\log\sigma^2 - n\log\alpha - \frac{1}{2\alpha^2}\sum_{i=1}^{n}(\hat{q}_{1i}^{(k)} + \hat{q}_{2i}^{(k)} - 2) \\
&- \frac{\tau}{2}\sum_{i=1}^{n}\hat{q}_{1i}^{(k)}(y_i - \xi)^2 - \frac{\tau\lambda_1^2}{2}\sum_{i=1}^{n}\hat{q}_{2i}^{(k)} + \nu\sum_{i=1}^{n}\hat{q}_{3i}^{(k)}(y_i - \xi) \\
&- \nu\lambda_1\sum_{i=1}^{n}\hat{q}_{4i}^{(k)} + \tau\lambda_1\sum_{i=1}^{n}(y_i - \xi),
\end{aligned} \tag{3.4}
$$

where $\tau = \frac{1}{\sigma^2} + \nu^2$.

CM-steps: Maximizing (3.4) is done with respect to ξ, σ, λ_1, λ_2 and α, we obtain the following closed-form expressions:

$$\begin{aligned}
\hat{\alpha}^{(k+1)} &= \left(\frac{1}{n}\sum_{i=1}^{n}(\hat{q}_{1i}^{(k)} + \hat{q}_{2i}^{(k)} - 2)\right)^{\frac{1}{2}}, \\
\hat{\xi}^{(k+1)} &= \frac{1}{\sum_{i=1}^{n}\hat{q}_{1i}^{(k)}}\sum_{i=1}^{n}\left(\hat{q}_{1i}^{(k)}y_i - \frac{\hat{\nu}^{(k)}}{\hat{\tau}^{(k)}}\hat{q}_{3i}^{(k)} - \hat{\lambda}_1^{(k)}\right), \\
\hat{\sigma}^{(k+1)} &= \left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{q}_{1i}^{(k)}(y_i - \hat{\xi}^{(k+1)})^2 + \hat{\lambda}_1^{2(k)}\hat{q}_{2i}^{(k)} - 2\lambda_1^{(k)}(y_i - \hat{\xi}^{(k+1)})\right)\right)^{\frac{1}{2}}, \\
\hat{\lambda}_1^{(k+1)} &= \frac{1}{\hat{\tau}^{(k)}\sum_{i=1}^{n}\hat{q}_{2i}^{(k)}}\sum_{i=1}^{n}\left(\hat{\tau}^{(k)}(y_i - \hat{\xi}^{(k+1)}) - \hat{\nu}^{(k)}\hat{q}_{4i}^{(k)}\right), \\
\hat{\nu}^{(k+1)} &= \frac{1}{n\hat{\sigma}^{2(k+1)}}\sum_{i=1}^{n}\left(\hat{q}_{3i}^{(k)}(y_i - \hat{\xi}^{(k+1)}) - \hat{\lambda}_1^{(k+1)}\hat{q}_{4i}^{(k)}\right).
\end{aligned}$$

So, we update $\hat{\lambda}_2^{(k+1)} = \hat{\nu}^{(k+1)}\hat{\sigma}^{(k+1)}$.

The above procedure is repeated until a suitable stopping criterion is satisfied. This stopping rule is specified in the following subsection.

3.1.1 Convergence of the Algorithm

To assess the convergence of the algorithm, there are two effective approaches: *(i)* the difference between two successive log-likelihood values be less than a specified tolerance, and *(ii)* changing all estimates of the parameters by a very small degree. However, in our simulation, we observed that, since the log-likelihood function in each step is related to the CDF of a generalized hyperbolic distribution, the algorithm has a slow convergence. Hence, to speed up the convergence of the algorithm, we used the second approach. In this way, the stopping criterion is related to the relative error of the components of parameter $\boldsymbol{\theta}$ as

$$max_j\left|\frac{\theta_j^{(k)} - \theta_j^{(k+1)}}{\theta_j^{(k+1)}}\right| < \epsilon,$$

where (k) is the iteration index and θ_j is the jth component of $\boldsymbol{\theta}$. We chose $\epsilon = 10^{-6}$ in our implementation of the algorithm.

3.1.2 Initial Values

The EM-type algorithm is likely to get trapped in one of the local maxima of the likelihood function. To deal with this problem, we generated many reasonable initial values and then selected the set of parameters associated with the highest converged log-likelihood value. For the location parameter ξ, the initial value was uniformly generated between the first and third quartiles of the observed sample. The initial scale variance σ^2 was obtained as ds^2, where s^2 is the sample variance and d is a factor uniformly generated between 0.5 and 2. Similarly, the initial value for λ_2 was set to be the sample skewness multiplied by d. As for the initial values of λ_1 and α, they were integers randomly chosen in the interval 1 to 10.

3.2 Provision of Standard Errors

Under some regularlity conditions, the asymptotic covariance matrix of the ML estimates $\hat{\boldsymbol{\theta}}$ can be approximated by the inverse of the observed information matrix. For this, Meilijson (1989) suggested the following empirical information matrix:

$$I_e(\boldsymbol{\theta} \mid \boldsymbol{y}) = \sum_{i=1}^{n} \boldsymbol{s}(y_i \mid \boldsymbol{\theta})\boldsymbol{s}^\top(y_i \mid \boldsymbol{\theta}) - \frac{1}{n}\boldsymbol{S}(\boldsymbol{y} \mid \boldsymbol{\theta})\boldsymbol{S}^\top(\boldsymbol{y} \mid \boldsymbol{\theta}),$$

where $\boldsymbol{S}(\boldsymbol{y} \mid \boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{s}(y_i \mid \boldsymbol{\theta})$ and $\boldsymbol{s}(y_i \mid \boldsymbol{\theta})$ are individual scores that can be determined as

$$\boldsymbol{s}(y_i \mid \boldsymbol{\theta}) = \frac{\partial \log f(y_i \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = E\left\{\frac{\partial \ell_c(\boldsymbol{\theta} \mid y_i, \gamma_i, w_i)}{\partial \boldsymbol{\theta}} \mid y_i, \boldsymbol{\theta}\right\},$$

where $\ell_c(\boldsymbol{\theta} \mid y_i, \gamma_i, w_i)$ is the individual complete-data log-likelihood defined in (3.2) based on a single observation (y_i, γ_i, w_i).

Substituting the ML estimate $\hat{\boldsymbol{\theta}}$, a natural estimator of $I_e(\boldsymbol{\theta} \mid \boldsymbol{y})$ is

$$I_e(\hat{\boldsymbol{\theta}} \mid \boldsymbol{y}) = \sum_{i=1}^{n} \hat{\boldsymbol{s}}_i \hat{\boldsymbol{s}}_i^\top, \tag{3.5}$$

where $\hat{\boldsymbol{s}}_i = (\hat{s}_{i\xi}, \hat{s}_{i\sigma}, \hat{s}_{i\lambda_1}, \hat{s}_{i\lambda_2}, \hat{s}_{i\alpha})$. Explicit expressions for the elements of $\hat{\boldsymbol{s}}_i$ are as follows:

$$
\begin{aligned}
\hat{s}_{i\alpha} &= -\frac{1}{\hat{\alpha}} + \frac{1}{\hat{\alpha}^3}(\hat{q}_{1i} + \hat{q}_{2i} - 2),\\
\hat{s}_{i\xi} &= \hat{\tau}\hat{q}_{1i}(y_i - \hat{\xi}) - \hat{v}\hat{q}_{3i} - \hat{\tau}\hat{\lambda}_1,\\
\hat{s}_{i\lambda_1} &= \hat{\tau}(y_i - \hat{\xi}) - \hat{\tau}\hat{\lambda}_1\hat{q}_{2i} - \hat{v}\hat{q}_{4i},\\
\hat{s}_{i\sigma} &= \frac{1}{\hat{\sigma}^2}\left(\hat{q}_{1i}(y_i - \hat{\xi})^2 + \hat{\lambda}_1^2\hat{q}_{2i} - 2\hat{\lambda}_1(y_i - \hat{\xi})\right) - \frac{1}{\hat{\sigma}},\\
\hat{s}_{i\lambda_2} &= \frac{1}{\hat{\sigma}}\left(\hat{q}_{3i}(y_i - \hat{\xi}) - \hat{v}\hat{q}_{1i}(y_i - \hat{\xi})^2 - \hat{v}\hat{\lambda}_1^2\hat{q}_{2i} - \hat{\lambda}_1\hat{q}_{4i} + 2\hat{v}\hat{\lambda}_1(y_i - \hat{\xi})\right).
\end{aligned}
$$

The standard errors of the ML estimates can then be found by calculating the square roots of the diagonal elements of $I_e^{-1}(\hat{\boldsymbol{\theta}} \mid \boldsymbol{y})$.

4 Simulation Study

To examine the performance of the proposed model and the estimation method, a Monte Carlo simulation experiment is performed in this section. We specifically examine the finite-sample performance of the ML estimates of the SNMVBS parameters. We chose samples of size $n = 250, 500, 1000, 1500$ and 2000 with parameters $\xi = 1, \sigma = 2, \lambda_1 = 1, \lambda_2 = -1$, and $\alpha = 2$ for the simulation study. Each simulated data set was fitted under the true model via the ECM algorithm described in Subsection 3.1. For each sample, this experiment was replicated 300 times.

To examine the accuracy of the parameter estimates, we computed the absolute bias and the mean squared error (MSE) as

$$
\mathrm{AB} = \frac{1}{300}\sum_{i=1}^{300} \mid \hat{\theta}_i - \theta_{\text{true}} \mid \quad \text{and} \qquad \mathrm{MSE} = \frac{1}{300}\sum_{i=1}^{300}\left(\hat{\theta}_i - \theta_{\text{true}}\right)^2,
$$

where $\hat{\theta}_i$ denotes the estimate of a specific parameter at the ith replication. Numerical results, displayed in Table 1, confirm the empirical consistency of the ML estimates since the Bias and MSE values shrink to zero when n increases.

Furthermore, we compute the standard deviations (STD) of the ML estimates across 300 simulated samples and compare them with the average values of the approximate standard errors (ASE) obtained by using the method described in Subsection 3.2. As can be seen from the estimation accuracy of standard errors, by an increase in sample size, the value of ASE tends to be closer to the corresponding standard deviations obtained from 300 MC estimates (STD). In small sample sizes, the differences between ASEs and STDs are significantly large, perhaps due to a worse approximation of the values of the Bessel function in R codes. But generally, the information-based method can offer a reasonably satisfactory approximation to the asymptotic covariance matrix of the ML estimates of model parameters, when the associated sample size is sufficiently large.

Table 1: The simulation results of parameter estimates and standard errors for various sample sizes.

Sample size	Measure	ξ	σ	λ_1	λ_2	α
250	AB	0.1430	0.1758	0.2210	0.2765	0.2298
	MSE	0.0333	0.0469	0.0775	0.1530	0.0840
	STD	0.1682	0.2127	0.2785	0.3833	0.2902
	ASE	0.4769	1.1213	1.4980	2.6856	0.8302
500	AB	0.0752	0.1120	0.1170	0.1354	0.1324
	MSE	0.0095	0.0199	0.0229	0.0404	0.0273
	STD	0.0963	0.1401	0.1512	0.2001	0.1641
	ASE	0.2898	0.6192	0.7471	1.3647	0.4661
1000	AB	0.0502	0.0767	0.0774	0.0917	0.0873
	MSE	0.0042	0.0097	0.0108	0.0217	0.0132
	STD	0.0647	0.0976	0.1033	0.1456	0.1143
	ASE	0.1645	0.3230	0.3475	0.5843	0.2541
1500	AB	0.0398	0.0610	0.0578	0.0632	0.0643
	MSE	0.0025	0.0061	0.0058	0.0115	0.0066
	STD	0.0502	0.0777	0.0760	0.1051	0.0811
	ASE	0.1175	0.1419	0.1687	0.2902	0.1623
2000	AB	0.0320	0.0491	0.0451	0.0427	0.0542
	MSE	0.0017	0.0039	0.0040	0.0064	0.0047
	STD	0.0416	0.0621	0.0629	0.0785	0.0687
	ASE	0.0949	0.1058	0.1124	0.1903	0.0981

5 Illustrative Examples

In this section, three examples are presented to illustrate the usefulness of the distribution. In all examples, there are some outlier points and since the SNMVBS distribution has heavier tails than other distributions, the proposed SNMVBS distribution provides a satisfactory fit.

5.1 cDNA Microarray Data Set

In the first example, we consider the cDNA microarray data set of the NCI60 cancer cell lines used earlier by Arslan (2010). This data set contains the measurements of 1400 cancer drugs activity levels on 60 human cancer cell (NCI60) lines. We will use the column "OV: SK-OV-3" in drug activity data set to demonstrate the performance of the proposed distribution. We obtained the maximum likelihood estimates of the SNMVBS model by using the ECM algorithm described in Section 3. We also fitted the SN, ST, NIG, and NMVBS distributions for the sake of comparison. To fit SN and ST, we used the package "*sn*", and for NIG, we used the package "*ghyp*" in R statistical software. Table 2 shows the results assuming $\alpha = \nu$ and $\bar{\alpha}$ for ST and NIG models, respectively. Performance assessments for these models were made on the adequacy of overall fitness in terms of Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC), defined by

$$\text{AIC} = 2m - 2\ell_{max} \quad \text{and} \quad \text{BIC} = m \log n - 2\ell_{max},$$

where m is the number of parameters and $\ell_{\max}$ is the maximized log-likelihood value. Of course, lower values of AIC or BIC indicate a better fit. These criteria have also been shown in Table 2. The Kolmogorov-Smirnov (K-S) test statistic values and the corresponding P-values are also shown in Table 2. As can be seen from the AIC and BIC values and the K-S test statistics, the SNMVBS model provides the best fit for the OV: SK-OV-3 data. Figure 3 depicts the histograms and the fitted densities for this data. The plots demonstrate that the SNMVBS distribution can capture very well the skewness and heavy-tails of the data. Moreover, the P-P plot in Figure 3 shows that the SNMVBS distribution provides a good fit to the data.

5.2 Austrian Bank Interest Rates Data

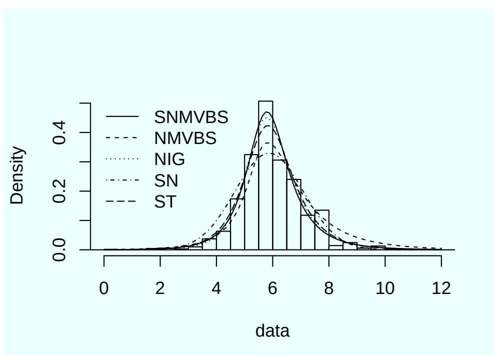
These data consist of 91 monthly interest rates of an Austrian bank. Künsch (1984) fitted an autoregressive time series model of order one, AR(1), to these data. Let $Y(t)$ be the interest rate at month t, and so an AR(1) model is given by

$$Y(t) = \beta_0 + \beta_1 Y(t-1) + \epsilon_t, \quad \beta_0 \in \mathbb{R}, \ \mid \beta_1 \mid \leq 1.$$

where ϵ_t are i.i.d. error variables. The ordinary least squares (OLS) method, $\epsilon_t \sim N(0, \sigma_t^2)$, for fitting an AR(1) model to these data yields $\hat{\beta}_0 = 1.928$ and $\hat{\beta}_1 = 0.792$. Azzalini and Genton (2008) assumed that the error terms ϵ_t are i.i.d. from a skew-t distribution. The maximum likelihood estimates of parameters in this case are found to be $\hat{\beta}_0 = 0.18$ and $\hat{\beta}_1 = 0.98$.

Table 2: The parameter estimates and the estimated standard errors (in parentheses) of the models fitted for OV: SK-OV-3 data.

Parameter	SN	NIG	NMVBS	ST	SNMVBS
ξ	5.95 (0.220)	5.73 (0.016)	5.73 (0.088)	5.47 (0.083)	5.40 (0.046)
σ	1.214 (0.093)	1.18 (0.028)	1.006 (0.045)	0.912 (0.046)	1.79 (1.073)
λ_1	-	0.225 (0.045)	0.168 (0.072)	-	-0.315 (0.390)
λ_2	-0.081 (0.270)	-	-	0.662 (0.145)	1.903 (1.583)
α	-	0.722 (0.016)	0.816 (0.032)	3.526 (0.375)	1.116 (0.237)
$\ell(\hat{\boldsymbol{\theta}})$	-2245.48	-2132.53	-2139.43	-2129.39	-2123.08
AIC	4496.97	4273.07	4286.86	4266.79	4256.16
BIC	4512.69	4294.02	4307.81	4287.75	4282.35
K-S test	0.066	0.032	0.045	0.040	0.027
(P-value)	($< 1e-4$)	(0.093)	(0.008)	(0.038)	(0.25)



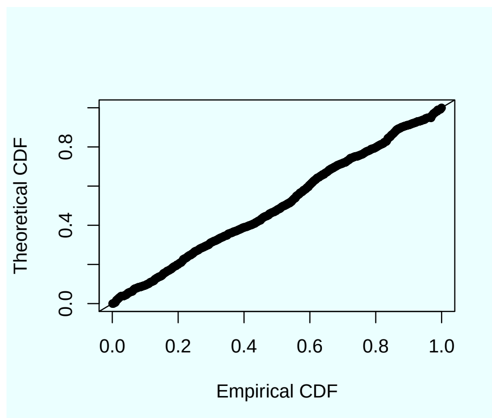


Figure 3: The histogram of data, the fitted distributions (left) and P-P plot for SNMVBS (right) for OV: SK-OV-3 data.

Alternatively, we assume here that $\epsilon_t \sim SNMVBS(0, \sigma^2, \lambda_1, \lambda_2, \alpha)$, or equivalently, $Y_t \mid Y_{t-1} \sim SNMVBS(\beta_0 + \beta_1 Y_{t-1}, \sigma^2, \lambda_1, \lambda_2, \alpha)$. The maximum likelihood estimation procedure described earlier yields $\hat{\beta}_0 = 0.126, \hat{\beta}_1 = 0.985, \hat{\sigma} = 0.147, \hat{\lambda}_1 = 0.0049, \hat{\lambda}_2 = -0.066$ and $\hat{\alpha} = 3.49$. Since the expected value of ϵ_t is not zero, in this case, we must adjust the fitted model by adding $\widehat{E(\epsilon)}$, which is the mean of the fitted error distribution of the model. Hence, we have $\hat{Y}_t = \hat{\beta}_0 + \widehat{E(\epsilon)} + \hat{\beta}_1 \hat{Y}_{t-1}$, where $\hat{\beta}_0 + \widehat{E(\epsilon_t)} = 0.147$. Figure 4 presents the fitted AR models by assuming ϵ_t with Normal, SN, ST, and SNMVBS distributions. We adjust other models in the same way as we did in the SNMVBS model. It is observed that the fitted SNMVBS distribution is quite close to the ST distribution and clearly better than the OLS and SN models. Figure 4 also shows the histogram of the residuals and the fitted SNMVBS density function. Figure 5 shows the P-P plots for all the fitted models for the error component ϵ_t. We observe that the P-P plots of SNMVBS and ST are quite close to the diagonal line, and it is also clear that OLS and SN autoregressive models are inappropriate for these data.

Moreover, to identify the best model studied, we use the summary index of discrepancy ($Q(p)$), proposed by Azzalini and Genton (2008), which is defined as

$$Q(p) = \sum_{t=1}^{n} \mid Y_t - \hat{Y}_t \mid^p, \quad p = 0.5, 1, 2,$$

where $\hat{Y}_t$ is the fitted value of Y_t. Table 3 shows the results for the studied models. The K-S test statistic values and the corresponding P-values for residuals are also shown in this table. This table shows that the SNMVBS model is slightly better than the ST model.

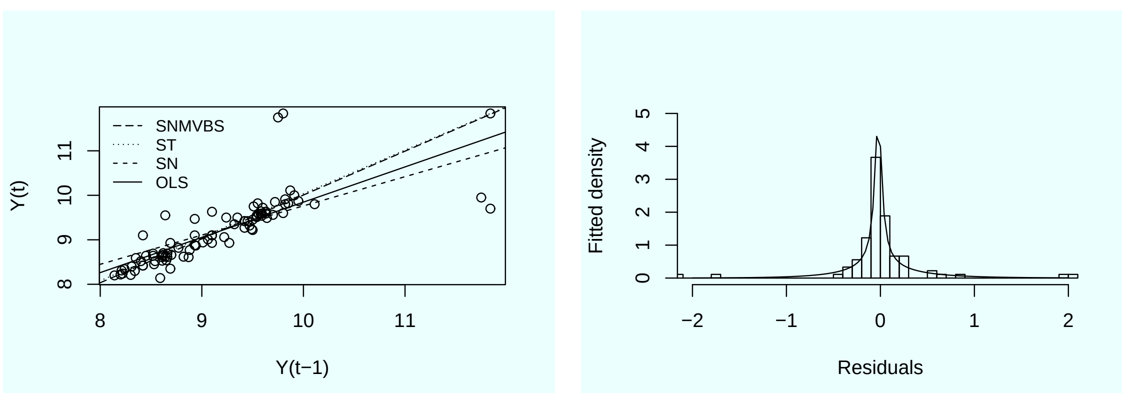


Figure 4: The scatterplot and fitted regression lines (left), the histogram of residuals obtained from the SNMVBS regression model and the fitted SNMVBS distribution (Right) for the Austrian bank data.

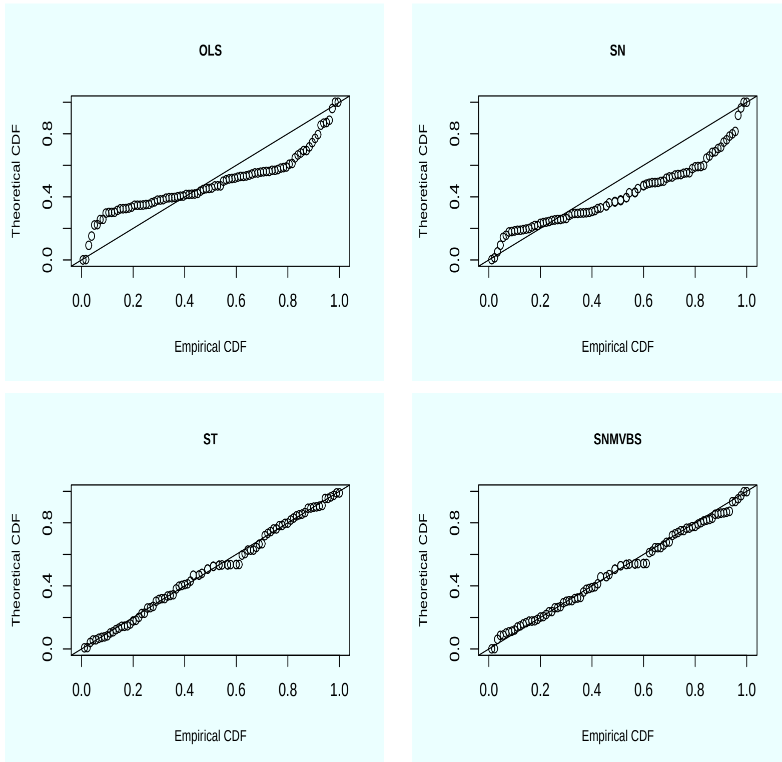


Figure 5: The P-P plots for the four fitted models for the Austrian bank data.

Table 3: The summary index of discrepancy $Q(p)$ for different AR(1) models fitted to the data and the K-S test statistic values and the corresponding p-values (in parentheses) for residuals for the Austrian bank data.

	p=0.5	p=1	p=2	K-S test (p-value)
OLS	37.50	21.48	17.45	0.229 ($< 1e-4$)
SN	42.48	25.39	18.37	0.234 ($< 1e-4$)
ST	34.85	20.18	19.20	0.075 (0.65)
SNMVBS	32.60	19.15	19.20	0.068 (0.76)

5.3 AIS Data: Multiple Regression

Finally, we consider the well-known Australian Institute of Sport (AIS) data set. This data set, which gives somebody indices of Australian athletes, has been analyzed by several authors working on skewed distributions. For example, Arellano-Valle *et al.* (2008) fitted some skewed linear regressions to these data by taking lean body mass (*LBM*) of male athletes as the response variable and height (*Ht*) and weight (*Wt*) as explanatory variables. Here, we consider the variable body fat percentage (*Bfat*), the sum of skin folds (*SSF*) and *Ht* associated with 102 Australian male athletes. Although *Bfat* is a key element of overall fitness, it is difficult to be measured directly. Therefore, one can use other indices to predict it. In this data, we observe that there is a linear relationship between *Bfat* and *SSF*. The variable *Ht* is also added to fit a more accurate model. Thus, we fit the linear regression

$$Bfat = \beta_0 + \beta_1 SSF + \beta_2 Ht + \epsilon,$$

where $\epsilon \sim SNMVBS(0, \sigma^2, \lambda_1, \lambda_2, \alpha)$. The resulting model, after adjustment, is

$$\widehat{Bfat} = 6.61 + 0.16\, SSF - 0.03\, Ht.$$

The estimates of other parameters are $\hat{\sigma} = 0.588$, $\hat{\lambda}_1 = -0.07$, $\hat{\lambda}_2 = 0.725$, and $\hat{\alpha} = 1.88$. For the sake of comparison, we also fitted the OLS, SN and ST regression models.

The resulting $Q(p)$ values for all fitted models are presented in Table 4. This table shows that the SNMVBS model overall provides a better fit than other models. Figure 6 shows the fitted SNMVBS regression plane. Figure 7 shows P-P plots of all the fitted models for the error component ϵ. It is observed that the SNMVBS distribution gives a satisfactory fit among other rival models.

Table 4: The summary index of discrepancy $Q(p)$ between observed and fitted values for the described linear regression models and the K-S test statistic values with the corresponding P-values for residuals for the AIS data .

	p=0.5	p=1	p=2	K-S test (P-value)
OLS	67.07	57.50	62.87	0.107 (0.17)
SN	68.20	58.21	63.30	0.111 (0.13)
ST	67.03	57.32	63.12	0.084 (0.42)
SNMVBS	66.99	57.30	63.70	0.049 (0.95)

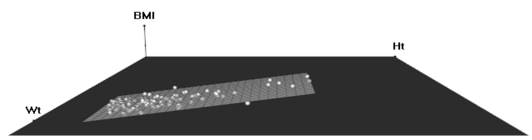


Figure 6: The fitted SNMVBS regression model to AIS data.

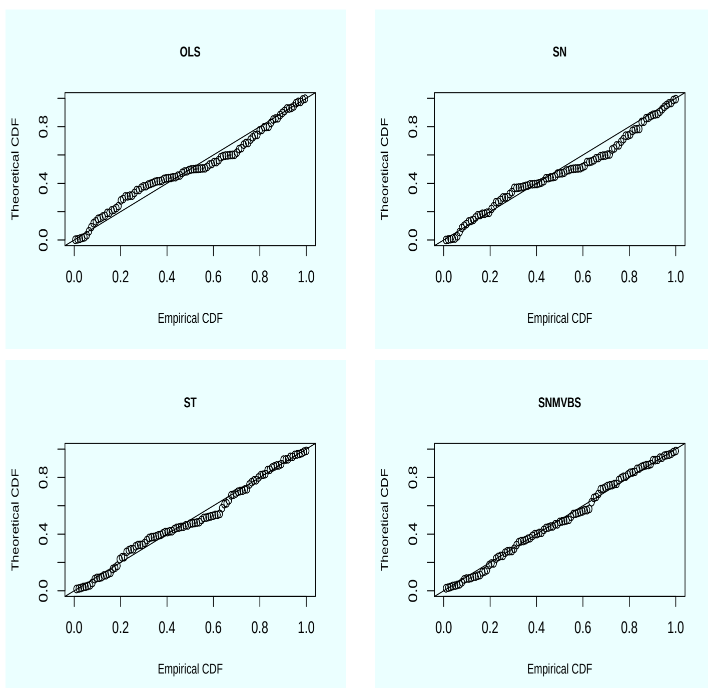


Figure 7: The P-P plots of four fitted models for the AIS data.

6 The Multivariate SNMVBS Distribution

In this section, we briefly discuss a multivariate version of the univariate SNMVBS distribution discussed in the preceding sections. As in the univariate case, we shall use the multivariate SN distribution to define the multivariate SNMVBS (MSNMVBS) distribution. The multivariate version of SN (MSN) distribution was introduced by Azzalini and Dalla Valle (1996) and Arellano–Valle *et al.* (2005), which extends the multivariate normal model by allowing a shape parameter to account for skewness.

A random vector $\mathbf{Y}$ is said to follow a p-variate MSN with location vector $\boldsymbol{\xi}$, scale covariance matrix $\boldsymbol{\Sigma}$ and skewness parameter vector $\boldsymbol{\lambda} \in \mathbb{R}^p$, denoted by $\mathbf{Y} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ if it has a density

$$g(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = 2\phi_p(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Sigma})\Phi\left(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{\xi})\right), \tag{6.1}$$

where $\phi_p(\cdot; \boldsymbol{\xi}, \boldsymbol{\Sigma})$ is the PDF of $N_p(\boldsymbol{\xi}, \boldsymbol{\Sigma})$ and $\Phi(\cdot)$ stands for the CDF of univariate standard normal distribution.

So, extending the representation (1.3), the random variable $\mathbf{Y} \in \mathbb{R}^p$ is said to have a multivariate SNMVBS (MSNMVBS) distribution if

$$\mathbf{Y} = \boldsymbol{\xi} + W\boldsymbol{\lambda}_1 + W^{1/2}\mathbf{Z}, \tag{6.2}$$

where $\mathbf{Z} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}_2)$, $W \sim BS(\alpha, 1)$, and W and $\mathbf{Z}$ are independent. The parameter vectors $\boldsymbol{\xi}, \boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are in $\mathbb{R}^p$, and $\boldsymbol{\Sigma}$ is a positive definite matrix. We denote the random variable $\mathbf{Y}$ in (6.2) by $\mathbf{Y} \sim MSNMVBS_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \alpha)$. Using (6.2) the stochastic representation of an MSNMVBS random variable can be obtained as follows

$$\mathbf{Y}|W = w \quad \sim \quad \mathbf{SN}_p(\boldsymbol{\xi} + w\boldsymbol{\lambda}_1, w\boldsymbol{\Sigma}, \boldsymbol{\lambda}_2), \tag{6.3}$$

Theorem 6.1. *Let the random vector* $\mathbf{Y}$ *follow the representation* (6.3)*, then the PDF of* $\mathbf{Y}$ *is given by*

$$\begin{aligned} g(\boldsymbol{y}) =& f_{GH_p}(\boldsymbol{x}; \boldsymbol{\xi}, \ \boldsymbol{\Sigma}, \boldsymbol{\lambda}_1, -1/2, \alpha^{-2}, \alpha^{-2}) H_{\{-\frac{p+1}{2}\}}(a) \\ &+ f_{GH_p}(\boldsymbol{x}; \boldsymbol{\xi}, \ \boldsymbol{\Sigma}, \boldsymbol{\lambda}_1, 1/2, \alpha^{-2}, \alpha^{-2}) H_{\{-\frac{p-1}{2}\}}(a), \ \ \mathbf{y} \in \mathbb{R}^p, \end{aligned} \tag{6.4}$$

where $H_{\{\kappa\}}(x) = F_{GH_1}(x;\ 0, 1, b, \kappa, \delta_1, \delta_2)$ *and* $f_{GH_p}(\cdot)$ *and* $F_{GH_p}(\cdot)$ *denote the PDF and CDF of a multivariate GH distribution, respectively. In addition* $\delta_1 = (\boldsymbol{y} - \boldsymbol{\xi})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\xi}) + \alpha^{-2}$, $\delta_2 = \boldsymbol{\lambda}_1^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\lambda}_1 + \alpha^{-2}$, $a = \boldsymbol{\lambda}_2^\top \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{\xi})$ *and* $b = \boldsymbol{\lambda}_2^\top \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\lambda}_1$.

The mean vector and covariance matrix of $\mathbf{Y} \sim MSNMVBS_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \alpha)$ are obtained by iterated expectations on representation (6.3) and the part *(ii)* of Lemma 2.2, as

$$\begin{aligned} E(\mathbf{Y}) &= \boldsymbol{\xi} + \boldsymbol{\lambda}_1 E(W) + \sqrt{\frac{2}{\pi}} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta} E(W^{1/2}), \\ Cov(\mathbf{Y}) &= \boldsymbol{\lambda}_1 \boldsymbol{\lambda}_1^\top Var(W) + \boldsymbol{\Sigma} E(W) - \frac{2}{\pi} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta} \boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{1/2} E^2(W^{1/2}) \\ &+ 2\sqrt{\frac{2}{\pi}} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta} \boldsymbol{\lambda}_1^\top cov(W, W^{1/2}). \end{aligned}$$

Finally, we must note that we do not discuss here more details on MSNMVBS distribution, including marginal and conditional distributions, estimation and investigating real data examples. They will be of great interest for further researches.

7 Conclusion

We have introduced a skew-normal mean-variance mixture distribution (SNMVBS) using BS distribution as a mixing distribution. We have discussed some key properties of this distribution and have developed a feasible EM-type scheme for the maximum likelihood estimation of model parameters. We have demonstrated the performance of the proposed EM algorithm with some simulations. Moreover, A brief discussion on a multivariate version of the SNMVBS distribution is also provided. The SNMVBS distribution has many key desirable properties and provides a flexible class of distributions for modeling skewed and heavy-tailed data. As compared to the skew-t model, all parameters of the SNMVBS distribution have been obtained by solving some simple linear equations in the proposed EM algorithm. There are still several of possible directions for future research. For instance, the robust mixture models of SNMVBS distributions and the estimation for the multivariate version will be of great interest.

Acknowledgements

We gratefully acknowledge the incisive comments and suggestions of the two anonymous referees, which led to this greatly improved version of the article.

References

Arellano-Valle, R. B., Castro, L. M., Genton, M. G., and Gómez H. W. (2008), Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis. *Bayesian Analysis*. **3**, 513–540.

Arellano–Valle, R. B., and Genton, M. G. (2005), On fundamental skew distributions. *Journal of Multivariate Analysis*. **96**, 93–116.

Arslan, O. (2010), An alternative multivariate skew Laplace distribution: properties and estimation. *Statistical Papers*. **51**, 865–887.

Arslan, O. (2015), Variance–mean mixture of the multivariate skew normal distribution. *Statistical Papers*. **56**, 353–378.

Azzalini, A. (1985), A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*. **12**, 171–178.

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

Azzalini, A. and Dalla Valle, A. (1996), The multivariate skew–normal distribution. *Biometrika*. **83**, 715–726.

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

Barndorff–Nielsen, O. (1977), Exponentially decreasing distributions for logarithm of particle size. *Proceedings of the Royal Society of London*, Series A. **353**, 401–419.

Barndorff–Nielsen, O. (1978), Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics*. **9**, 43–46.

Birnbaum, Z. W. and Saunders, S. C. (1969), A new family of life distributions. *Journal of Applied Probability*. **6**, 319–327.

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

Hu, W. (2005), *Calibration of multivariate generalized hyperbolic distributions using the EM algorithm, with applications in risk management, portfolio optimization and portfolio credit risk*. Ph.D. dissertation, Florida State University.

Künsch, H. (1984), Infinitesimal robustness for autoregressive processes. *Annals of Statistics*. **12**, 843–863.

Leiva, V. (2016), *The Birnbaum–Saunders Distribution*. London: Elsevier.

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

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

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