ATLANTIS PRESS

Journal of Statistical Theory and Applications
Vol. **18(3)**, September 2019, *pp.* 244–258
DOI: https://doi.org/10.2991/jsta.d.190617.001; ISSN 1538-7887
https://www.atlantis-press.com/journals/jsta

A Multivariate Skew-Normal Mean-Variance Mixture Distribution and Its Application to Environmental Data with Outlying Observations

M. Tamandi[1,*], N. Balakrishnan[2], A. Jamalizadeh[3], M. Amiri[4]

[1] *Department of Statistics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran*
[2] *Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada*
[3] *Department of Statistics, Faculty of Mathematics and Computers, Shahid Bahonar University of Kerman, Kerman, Iran*
[4] *Department of Statistics, Faculty of Basic Sciences, University of Hormozgan, Bandar Abbas, Iran*

ARTICLE INFO

Article History

Received 21 Feb 2019
Accepted 14 June 2019

Keywords

Multivariate distribution
Birnbaum–Saunders
ECM algorithm
Outliers
Mean-variance mixtures

ABSTRACT

The presence of outliers, skewness, kurtosis, and dependency are well-known challenges while fitting distributions to many data sets. Developing multivariate distributions that can properly accomodate all these aspects has been the aim of several researchers. In this regard, we introduce here a new multivariate skew-normal mean-variance mixture based on Birnbaum–Saunders distribution. The resulting model is a good alternative to some skewed distributions, especially the skew-t model. The proposed model is quite flexible in terms of tail behavior and skewness, and also displays good performance in the presence of outliers. For the determination of maximum likelihood estimates, a computationally efficient Expectation-Conditional-Maximization (ECM) algorithm is developed. The performance of the proposed estimation methodology is illustrated through Monte Carlo simulation studies as well as with some real life examples.

1. INTRODUCTION

Although normal distribution has a central role in statistical analysis, the presence of outliers or atypical observations becomes problematic in some applications. This problem often occurs in some fields such as environmental and financial sciences. For example, in a random sample of some variables from water quality or air pollution in a specific area, it is quite common to encounter some outliers in the data. Similarly, outlying observations may be seen in stock returns analysis. For further details on this issue, one may refer to [1] and [2]. When data are noisy, fitting a suitable distribution does become a challenge. Adding a tail or skewness parameter to a normal distribution is a recognized method to accommodate the presence of possible outliers. Two well-known formulations that follow this approach are by Barndorff-Nielsen [3] and Azzalini [4].

Normal mean-variance mixture (NMVM) distributions were introduced by Barndorff-Nielsen [3]. Specifically, if $\boldsymbol{Y}$ be a $p\times 1$ random vector from NMVM distribution, then $\boldsymbol{Y}$ can be represented as

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

where $\mathbf{Z} \sim N_p(\mathbf{0},\boldsymbol{\Sigma})$ and $W > 0$ is a scalar-valued random variable that is independent of $\mathbf{Z}$. The parameter vectors of $\boldsymbol{\xi}$ and $\boldsymbol{\lambda}$ are in $\mathbb{R}^p$ and $\boldsymbol{\Sigma}$ is a positive definite matrix. Generalized hyperbolic (GH) distribution, introduced by Barndorff-Nielsen [5], is one of the prominent distributions in this class. This class, in addition, includes several important distributions such as skew Laplace [6] and normal inverse Gaussian (NIG) [3] as special cases. The random variable with GH distribution can be represented as a NMVM variable when W in (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

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

*Corresponding author. Email: tamandi@vru.ac.ir

where $K_\kappa(\cdot)$ (with $\kappa \in \mathbb{R}$) denotes the modified Bessel function of the third kind with the 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 density is unimodal and contains the gamma and inverse gamma densities as special cases when $\chi = 0$ and $\psi = 0$, respectively.

Let the random variable $\boldsymbol{Y}$ be as in (1), where $W \sim GIG(\kappa, \chi, \psi)$. Then, the PDF of $\boldsymbol{Y}$ is given by

$$f_{GH_p}(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Sigma}, \lambda, \kappa, \chi, \psi) = \frac{\left(\frac{\psi}{\chi}\right)^{\kappa/2}\left(\lambda^\top \boldsymbol{\Sigma}^{-1}\lambda + \psi\right)^{\frac{p}{2}-\kappa}}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}K_\kappa\left(\sqrt{\chi\psi}\right)} \exp\left\{(y-\xi)^\top \boldsymbol{\Sigma}^{-1}\lambda_1\right\}$$
$$\times \frac{K_{\kappa-p/2}\left(\sqrt{\left(\chi + (y-\xi)^\top \boldsymbol{\Sigma}^{-1}(y-\xi)\right)\left(\lambda^\top \boldsymbol{\Sigma}^{-1}\lambda + \psi\right)}\right)}{\left(\sqrt{\left(\chi + (y-\xi)^\top \boldsymbol{\Sigma}^{-1}(y-\xi)\right)\left(\lambda^\top \boldsymbol{\Sigma}^{-1}\lambda + \psi\right)}\right)^{\frac{p}{2}-\kappa}}.$$

for $\boldsymbol{y} \in \mathbb{R}^p$. The corresponding cumulative distribution function (CDF) is denoted by $F_{GH_p}(\cdot; \boldsymbol{\xi}, \boldsymbol{\Sigma}, \lambda, \kappa, \chi, \psi)$. For further details about GIG and GH distributions, see [7].

The other members of NMVM class of distributions can be introduced similarly by choosing other suitable mixing random variable W. For example, let W be a Birnbaum–Saunders (BS) [8] random variable with shape parameter α and scale parameter β, denoted by $W \sim BS(\alpha, \beta)$. The PDF of W is given by

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

Now, assuming $W \sim BS(\alpha, 1)$ in (1), Pourmousa *et al.* [9] presented a NMVM of BS (NMVBS) distribution. They studied some properties of this model and illustrated its application in autoregressive conditional heteroskedastic (ARCH) models.

To shift from symmetric distributions to asymmetric ones, Azzalini [4] proposed the univariate skew-normal (SN) distribution, which can be used as a substitute for normal distribution in the modeling of data displaying some asymmetry. The multivariate version of SN (MSN) distribution was introduced by Azzalini and Dalla Valle [10] and Arellano-Valle and Genton [11], 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 matrix $\boldsymbol{\Sigma}$, and skewness parameter vector $\lambda \in \mathbb{R}^p$, denoted by $\mathbf{Y} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \lambda)$ if it has density

$$f_{MSN}(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Sigma}, \lambda) = 2\phi_p(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Sigma})\,\Phi\left(\lambda^\top \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y}-\boldsymbol{\xi})\right), \tag{3}$$

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

According to Arellano-Valle and Genton [11], the MSN distribution has a convenient stochastic representation as

$$\mathbf{Y} \overset{d}{=} \xi + \boldsymbol{\Sigma}^{1/2}\left\{\delta|U_0| + \left(\mathbf{I}_p - \delta\delta^\top\right)^{1/2}\mathbf{U}_1\right\}, \tag{4}$$

where $\boldsymbol{\delta} = \lambda/\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$, $U_0 \sim N(0,1)$ and $\mathbf{U}_1 \sim N_p\left(\mathbf{0}, \mathrm{I}_{\boldsymbol{p}}\right)$ independently.

If $\mathbf{Z} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ is used in (1), then a skew-normal mean-variance mixture distribution is attained. Arslan [12] assumed that W is a GIG distribution and studied properties of the corresponding model.

The aim of this paper is to introduce a multivariate skew-normal mean-variance mixture based on BS (SNMVBS) distribution. The random variable $\mathbf{Y} \in \mathbb{R}^p$ is said to have an SNMVBS distribution if

$$\mathbf{Y} = \xi + W\lambda_1 + W^{1/2}\mathbf{Z}, \tag{5}$$

where $\mathbf{Z} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \lambda_2)$, $W \sim BS(\alpha, 1)$, and W and $\mathbf{Z}$ are independent. The parameter vectors $\boldsymbol{\xi}$, λ_1 and λ_2 are in $\mathbb{R}^p$, and $\boldsymbol{\Sigma}$ is a positive definite matrix.

Tamandi *et al.* [13] studied the univariate SNMVBS and discussed some of its properties. The modeling of correlated data in the presence of outliers provides a main motivation for extending the univariate SNMVBS to multivariate setup. On the other hand, the SNMVBS distribution serves as an alternative to skew-t (ST) model introduced by Azzalini and Capitanio [14]. The ST distribution is a SN mean-variance mixture distribution, where the mixing distribution is Gamma $(\nu/2, \nu/2)$. But, it is well known that there is not a closed-form estimator for ν in the ST model. Interestingly, the maximum likelihood (ML) estimates of the parameters in SNMVBS model can be obtained by solving some simple linear equations. Moreover, the SNMVBS model extends the NMVBS distribution, introduced by Pourmousa *et al.* [9].

The rest of this paper proceeds as follows. Section 2 introduces the multivariate SNMVBS distribution and then discusses some of its characteristics. Section 3 develops an ECM algorithm for estimating the parameters of SNMVBS distribution. In Section 4, asymptotic properties of the ML estimators are investigated using simulations. Section 5 illustrates some applications of the SNMVBS distribution in modeling environmental data. Finally, Section 6 provides some concluding remarks.

2. MULTIVARIATE SNMVBS DISTRIBUTION

Let $\mathbf{Y}$ be a random variable following the representation in (5), where $W \sim BS(\alpha, 1)$ with $\alpha > 0$. Then, we say that $\mathbf{Y}$ follows an SNMVBS distribution, and then write $\mathbf{Y} \sim SNMVBS(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \lambda_1, \lambda_2, \alpha)$ in short. It must be noted that if we set $W \sim BS(\alpha, \beta)$, with a general β, then the corresponding model becomes nonidentifiable. For this reason, we assume $\beta = 1$ to have an identifiable model with desirable properties. Using (5) and the known properties of SN distribution, another stochastic representation of SNMVBS random variable can be obtained as follows:

$$\begin{aligned} \mathbf{Y}|W = w &\sim SN_p(\boldsymbol{\xi} + w\boldsymbol{\lambda}_1, w\boldsymbol{\Sigma}, \boldsymbol{\lambda}_2), \\ W &\sim BS(\alpha, 1). \end{aligned} \tag{6}$$

Theorem 2.1. *If the random vector $\mathbf{Y}$ follows the representation in (6), then the PDF of $\mathbf{Y}$ is given by*

$$f_{SNMVBS}(\boldsymbol{y}) = f_{\left\{-\frac{1}{2}\right\}}(\boldsymbol{y})\, H_{\left\{-\frac{p+1}{2}\right\}}(a) + f_{\left\{\frac{1}{2}\right\}}(\boldsymbol{y})\, H_{\left\{-\frac{p-1}{2}\right\}}(a), \quad \mathbf{y} \in \mathbb{R}^p, \tag{7}$$

where $f_{\{\kappa\}}(\boldsymbol{x}) = f_{GH_p}\left(\boldsymbol{x}; \boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}_1, \kappa, \alpha^{-2}, \alpha^{-2}\right)$ and $H_{\{\kappa\}}(x) = F_{GH_1}(x; 0, 1, b, \kappa, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$. In addition $\boldsymbol{\delta}_1 = \left(\boldsymbol{y} - \boldsymbol{\xi}\right)^\top \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y} - \boldsymbol{\xi}\right) + \alpha^{-2}$, $\boldsymbol{\delta}_2 = \boldsymbol{\lambda}_1^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_1 + \alpha^{-2}$, $a = \boldsymbol{\lambda}_2^\top \boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{y} - \boldsymbol{\xi}\right)$, and $b = \boldsymbol{\lambda}_2^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}_1$.

A detailed proof of this theorem is given in Appendix A. The notations in Theorem 2.1 will be used throughout the paper. Figure 1 displays some examples of the density function in (7). These figures present four different densities and the corresponding contour plots for $\boldsymbol{\xi} = 0$, $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ and different values of λ_1, λ_2, and α. We find that the skewness of contours increases as a function of λ_2. Also, these figures depict that the SNMVBS distribution can be an useful model in the presence of outliers when α becomes large.

In the special case when $\lambda_1 = \mathbf{0}$, the PDF in (7) reduces to a scale mixture of two multivariate SN distributions with the BS model as the mixing distribution. Also, when $\lambda_2 = \mathbf{0}$, then $\mathbf{Y}$ has NMVBS distribution studied by Pourmousa *et al.* [9]. As mentioned in the introduction, Arslan [12] introduced a mean-variance mixture of the SN distribution with $W \sim GIG(\kappa, \chi, \psi)$ as the mixing distribution. Our motivation for introducing SNMVBS distribution here is that the SNMVBS model serves as an extension of Arslan's model under proper settings of parameters since the BS distribution is a mixture of two GIG distributions with special parameters [15]. Moreover, estimates of all parameters of SNMVBS distribution can be obtained through some simple linear equations in closed-form, but in Arslan's model, we must obtain the estimates of GIG distribution using numerical methods.

The mean vector and covariance matrix of $\mathbf{Y} \sim SNMVBS(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \lambda_1, \lambda_2, \alpha)$ are obtained by iterated expectations on representation (6) and Part *(ii)* of Lemma Appendix A.2, presented in Appendix A. These are given by

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

Table 1 represents the upper bounds of the measures of the skewness and kurtosis coefficients for the univariate SNMVBS under $\xi = 0$, $\sigma = 1$, and several different values of α. It can be observed that the ranges of these measures for the SNMVBS model grows when the parameter α increases. However, the multivariate skewness and kurtosis measures [16] for the multivariate SNMVBS cannot be obtained in closed forms, but the numerical computations show that the results are in agreement with the univariate version. Moreover, the upper bounds of skewness and kurtosis in SNMVBS model are greater than that of the SN model. According to Azzalini and Capitanio [17], the approximate maximal values of the skewness and kurtosis coefficients of the SN model are 0.9905 and 0.869, respectively, while from Table 1, we find these values for the SNMVBS model when $\alpha = 40$ to be 4.7859 and 30.349, respectively. It is useful to note that the skewness and kurtosis coefficients for ST model with $\nu = 5$ are 0.9905 and 23.108, respectively. Thus, the SNMVBS model proposed here provides considerably larger ranges for skewness and kurtosis than both the SN and ST models.

From (4) and (5), we get

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\xi} + W\boldsymbol{\lambda}_1 + W^{1/2}\boldsymbol{\Sigma}^{1/2}\left\{\boldsymbol{\lambda}_2\left(1 + \boldsymbol{\lambda}_2^\top\boldsymbol{\lambda}_2\right)^{-1/2}|U_0| + \left(\boldsymbol{I}_p + \boldsymbol{\lambda}_2\boldsymbol{\lambda}_2^\top\right)^{-1/2}\boldsymbol{U}_1\right\}.$$

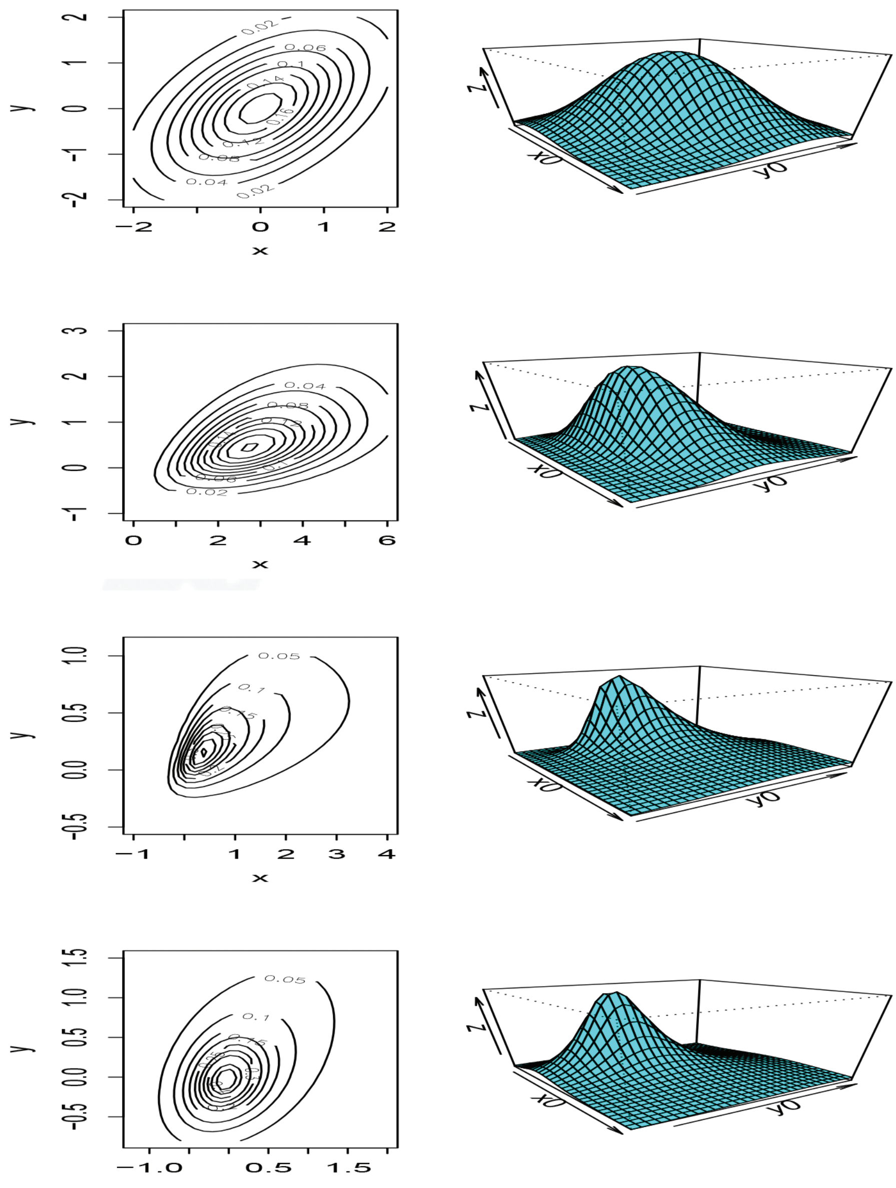

Figure 1 Density functions and corresponding contours of skew-normal mean-variance mixture based on Birnbaum–Saunders (SNMVBS) distribution. From top to bottom: $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2 = (0,0), \alpha = 0.3; \boldsymbol{\lambda}_1 = (3,0), \boldsymbol{\lambda}_2 = (0,3), \alpha = 0.3; \boldsymbol{\lambda}_1 = (3,0), \boldsymbol{\lambda}_2 = (0,3), \alpha = 1.7; \boldsymbol{\lambda}_1 = (1,1), \boldsymbol{\lambda}_2 = (-1,-1), \alpha = 1.7$, respectively.

Table 1 Upper bounds of the measures of skewness and kurtosis coefficients for the univariate SNMVBS under $\xi = 0, \sigma = 1$, and different values of α.

	α									
Measure	**0.5**	**1**	**2**	**4**	**8**	**10**	**15**	**20**	**30**	**40**
Skewness	1.4547	2.6579	3.8459	4.4786	4.6996	4.7301	4.7619	4.7738	4.7826	4.7859
Kurtosis	6.8981	10.789	20.910	27.107	29.409	29.736	30.082	30.213	30.312	30.349

Denoting by $\gamma = w^{1/2}\left(1 + \boldsymbol{\lambda}_2^\top \boldsymbol{\lambda}_2\right)^{1/2} |U_0|$, a further hierarchical representation of SNMVBS can be given as

$$\begin{aligned} \mathbf{Y}|W &= w, \gamma \sim N_p\left(\boldsymbol{\xi} + w\boldsymbol{\lambda}_1 + \frac{\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}_2\gamma}{\left(1 + \boldsymbol{\lambda}_2^\top \boldsymbol{\lambda}_2\right)}, w\boldsymbol{\Sigma}^*\right), \\ \gamma|W &= w \sim HN\left(0, w\left(1 + \boldsymbol{\lambda}_2^\top \boldsymbol{\lambda}_2\right)\right), \\ &\quad W \sim BS(\alpha, 1), \end{aligned} \tag{8}$$

where $HN\left(\mu,\sigma^2\right)$ denotes the univariate half-normal distribution and $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}^{1/2}\left(\boldsymbol{I}+\lambda_2\lambda_2^\top\right)^{-1}\boldsymbol{\Sigma}^{1/2}$.

The conditional distribution of γ, given $\mathbf{Y}=\boldsymbol{y}$ and $W=w$, is given by

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

Using representation (8), after some algebra, we note that

$$\gamma|\left(W=w,\mathbf{Y}=\boldsymbol{y}\right)\sim HN\left(\lambda_2^\top\boldsymbol{u}_2,w\right),$$

where $\boldsymbol{u}_2=\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{y}-\xi-w\lambda_1\right)$ and

$$\mu_\gamma=\lambda_2^\top\boldsymbol{u}_2,\sigma_\gamma^2=w.$$

Thus, we have

$$E\left(\gamma|W=w,\mathbf{Y}=\boldsymbol{y}\right)=\mu_\gamma+\sigma_\gamma\frac{\phi\left(\frac{\mu_\gamma}{\sigma_\gamma}\right)}{\Phi\left(\frac{\mu_\gamma}{\sigma_\gamma}\right)}, \tag{9}$$

and so by known properties of conditional expectations, we have

$$\begin{aligned} E\left(\gamma|\boldsymbol{y}\right) &= E\left(E\left(\gamma|w,\boldsymbol{y}\right)|\boldsymbol{y}\right),\\ E\left(\gamma W^{-1}|\boldsymbol{y}\right) &= E\left(W^{-1}E\left(\gamma|w,\boldsymbol{y}\right)|\boldsymbol{y}\right). \end{aligned} \tag{10}$$

In Appendix A, we have established some theoretical properties and characteristics of SNMVBS distribution, including moment generating function, affine transformation, and conditional distributions. In the following theorem, conditional moments of W, given $\mathbf{Y}=\mathbf{y}$, are presented which will be used in subsequent sections.

Theorem 2.2. *Let* $\mathbf{Y}\sim SNMVBS\left(\xi,\boldsymbol{\Sigma},\lambda_1,\lambda_2,\alpha\right)$ *and* $W\sim BS(\alpha,1)$. *Then*

i. *The conditional PDF of* W, *given* $\mathbf{Y}=\mathbf{y}$, *is*

$$\begin{aligned} f\left(w|\boldsymbol{y}\right) =\ & p\left(\boldsymbol{y}\right)\frac{\Phi\left(w^{-1/2}a-w^{1/2}b\right)}{H_{\left\{-\frac{p+1}{2}\right\}}(a)}f_{GIG}\left(w;-\frac{p+1}{2},\boldsymbol{\delta}_1,\boldsymbol{\delta}_2\right)\\ &+\left(1-p\left(\boldsymbol{y}\right)\right)\frac{\Phi\left(w^{-1/2}a-w^{1/2}b\right)}{H_{\left\{-\frac{p-1}{2}\right\}}(a)}f_{GIG}\left(w;-\frac{p-1}{2},\boldsymbol{\delta}_1,\boldsymbol{\delta}_2\right), \end{aligned} \tag{11}$$

where

$$p\left(\boldsymbol{y}\right)=\frac{1}{f_{SNMVBS}\left(\boldsymbol{y}\right)}f_{\left\{-\frac{1}{2}\right\}}\left(\boldsymbol{y}\right)H_{\left\{-\frac{p+1}{2}\right\}}(a);$$

ii.

$$E\left(W^n|\boldsymbol{y}\right)=\left(\frac{\boldsymbol{\delta}_1}{\boldsymbol{\delta}_2}\right)^{n/2}\left\{p\left(\boldsymbol{y}\right)\frac{H_{\left\{-\frac{p+1}{2}+n\right\}}(a)}{H_{\left\{-\frac{p+1}{2}\right\}}(a)}R_{\left(-\frac{p+1}{2},n\right)}\left(\sqrt{\boldsymbol{\delta}_1\boldsymbol{\delta}_2}\right)+\left(1-p\left(\boldsymbol{y}\right)\right)\frac{H_{\left\{-\frac{p-1}{2}+n\right\}}(a)}{H_{\left\{-\frac{p-1}{2}\right\}}(a)}R_{\left(-\frac{p-1}{2},n\right)}\left(\sqrt{\boldsymbol{\delta}_1\boldsymbol{\delta}_2}\right)\right\}, \tag{12}$$

for $n=\pm1,\pm2,\ldots$;

iii.

$$\begin{aligned} E_{(m)}\left(\boldsymbol{y}\right) &= E\left(W^mq\left(W^{-1/2}\lambda_2^\top\boldsymbol{U}_2\right)|\mathbf{Y}=\mathbf{y}\right)=\frac{1}{\sqrt{2\pi}f_{SNMVBS}\left(\boldsymbol{y}\right)}\left(\frac{|\boldsymbol{\Sigma}^*|}{|\boldsymbol{\Sigma}|}\right)^{1/2}\\ &\times\left\{f^*_{\{m+1/2\}}\left(\boldsymbol{y}\right)R_{\left(\frac{1}{2},m\right)}\left(\alpha^{-2}\right)+f^*_{\{m-1/2\}}\left(\boldsymbol{y}\right)R_{\left(-\frac{1}{2},m\right)}\left(\alpha^{-2}\right)\right\} \end{aligned} \tag{13}$$

for $m=\pm1/2$, *where* $f^*_{\{\kappa\}}\left(\boldsymbol{y}\right)=f_{GH_p}\left(\boldsymbol{y};\xi,\boldsymbol{\Sigma}^*,\lambda_1,\kappa,\alpha^{-2},\alpha^{-2}\right)$ *and* $q(x)=\frac{\phi(x)}{\Phi(x)}$.

Proof. The conditional PDF in (11) can be obtained by using Bayes' rule and some algebraic operations. The conditional expectation in (12) is obtained from Part *(i)* of the theorem and Part *(ii)* of Lemma Appendix A.1 in Appendix A. The conditional expectation in (13) is calculated directly.

To determine the parameter estimates of the SNMVBS distribution, we propose the Expectation-Maximization (EM)-type algorithms in the next section. The conditional expectations given in Theorem 2.2 become very useful for this purpose.

3. PARAMETER ESTIMATION

3.1. ECM Algorithm

Dempster *et al.* [18] introduced the EM algorithm technique as a versatile tool for ML estimation of models when there are missing data or latent variables. Simplicity in the implementation of the algorithm and the monotonic convergence are two most important features of the EM procedure. However, it is not directly applicable for estimating the parameters of the SNMVBS model because the M-step involves intractable computations. To avoid this problem, we use the Expectation-Conditional-Maximization (ECM) algorithm, suggested by Meng and Rubin [19], which replaces the M-step of the EM by a sequence of simpler conditional maximization steps.

Let $\mathbf{y}^\top = \left(\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top\right)$ be n samples, which have been collected randomly, and the corresponding unobserved random values are $\boldsymbol{\gamma}^\top = (\gamma_1, \ldots, \gamma_n)$ and $\boldsymbol{w}^\top = (w_1, \ldots, w_n)$. By using (8), we have the following hierarchical representation:

$$\begin{aligned} \mathbf{Y}_i | (\gamma_i, W_i = w_i) &\sim N_p\left(\boldsymbol{\xi} + w\boldsymbol{\lambda}_1 + \frac{\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}_2\gamma_i}{\left(1 + \boldsymbol{\lambda}_2^\top\boldsymbol{\lambda}_2\right)}, w_i\boldsymbol{\Sigma}^*\right), \\ \gamma_i | (W_i = w_i) &\sim HN\left(0, w_i\left(1 + \boldsymbol{\lambda}_2^\top\boldsymbol{\lambda}_2\right)\right), \\ W_i &\sim BS(\alpha, 1). \end{aligned} \tag{14}$$

Then, under the hierarchical representation in (14), with $\boldsymbol{\nu} = \Sigma^{-1/2}\boldsymbol{\lambda}_2$, the complete log-likelihood function associated with $\mathbf{y}_c^\top = \left(\boldsymbol{y}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{w}^{T6}\right)$ is given by

$$\begin{aligned} \ell_c\left(\boldsymbol{\theta}|\mathbf{y}_c\right) = c &- \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^n w_i^{-1}(\mathbf{y}_i - \boldsymbol{\xi} - w_i\boldsymbol{\lambda}_1)^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\xi} - w_i\boldsymbol{\lambda}_1) \\ &- \frac{1}{2}\sum_{i=1}^n w_i^{-1}\left(\gamma_i - \boldsymbol{\nu}^\top(\mathbf{y}_i - \boldsymbol{\xi} - w_i\boldsymbol{\lambda}_1)\right)^2 \\ &- n\log\alpha - \frac{1}{2\alpha^2}\sum_{i=1}^n(w_i + w_i^{-1} - 2). \end{aligned} \tag{15}$$

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

$$\hat{q}_{1i}^{(k)} = E\left(W_i^{-1}|\boldsymbol{y}_i, \hat{\boldsymbol{\theta}}^{(k)}\right), \quad \hat{q}_{2i}^{(k)} = E\left(W_i|\boldsymbol{y}_i, \hat{\boldsymbol{\theta}}^{(k)}\right),$$

$$\hat{q}_{3i}^{(k)} = E\left(\gamma_i W_i^{-1}|\boldsymbol{y}_i, \hat{\boldsymbol{\theta}}^{(k)}\right), \ \hat{q}_{4i}^{(k)} = E\left(\gamma_i|\boldsymbol{y}_i, \hat{\boldsymbol{\theta}}^{(k)}\right). \tag{16}$$

The $\hat{q}_{1i}^{(k)}$ and $\hat{q}_{2i}^{(k)}$ are obtained directly from (12). For $\hat{q}_{3i}^{(k)}$ and $\hat{q}_{4i}^{(k)}$, one can use the conditional expectations presented in (10). So, the last two quantities in (16) can be obtained as

$$\hat{q}_{3i}^{(k)} = \boldsymbol{\lambda}_2^\top\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{y}_i - \boldsymbol{\xi}\right)\hat{q}_{1i}^{(k)} - \boldsymbol{\lambda}_2^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}_1 + M_1 E_{(-1/2)}\left(\boldsymbol{y}\right),$$

$$\hat{q}_{4i}^{(k)} = \boldsymbol{\lambda}_2^\top\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{y}_i - \boldsymbol{\xi} - \hat{q}_{2i}^{(k)}\boldsymbol{\lambda}_1\right) + M_2,$$

where $M_1 = E_{(-1/2)}\left(\boldsymbol{y}\right)$ and $M_2 = E_{(1/2)}\left(\boldsymbol{y}\right)$ can be obtained from (13).

In summary, the ECM algorithm proceeds with the following steps:

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

$$\begin{aligned} Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}\right) = {} & c - \frac{n}{2}\log|\boldsymbol{\Sigma}| - n\log\alpha - \frac{1}{2}\sum_{i=1}^{n}\hat{q}_{1i}^{(k)}(\boldsymbol{y}_i-\boldsymbol{\xi})^{\top}\Gamma(\boldsymbol{y}_i-\boldsymbol{\xi}) \\ & -\frac{1}{2}\sum_{i=1}^{n}\hat{q}_{2i}^{(k)}\boldsymbol{\lambda}_1^{\top}\Gamma\boldsymbol{\lambda}_1 + \sum_{i=1}^{n}\hat{q}_{3i}^{(k)}(\boldsymbol{y}_i-\boldsymbol{\xi})^{\top}\boldsymbol{\nu} - \sum_{i=1}^{n}\hat{q}_{4i}^{(k)}\boldsymbol{\lambda}_1^{\top}\boldsymbol{\nu} \\ & +\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\xi})^{\top}\Gamma\boldsymbol{\lambda}_1 + \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\lambda}_1^{\top}\Gamma(\boldsymbol{y}_i-\boldsymbol{\xi}) \\ & -\frac{1}{2\alpha^2}\sum_{i=1}^{n}\left(\hat{q}_{1i}^{(k)}+\hat{q}_{2i}^{(k)}-2\right), \end{aligned} \tag{17}$$

where $\Gamma = \Sigma^{-1} + \boldsymbol{\nu}\boldsymbol{\nu}^{\top}$.

CM-steps: Maximize (17), with respect to $\boldsymbol{\xi}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$ and α, to obtain the following closed-form expressions for the parameters:

$$\begin{aligned} \hat{\alpha}^{(k+1)} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{q}_{1i}^{(k)}+\hat{q}_{2i}^{(k)}-2\right)}, \\ \hat{\boldsymbol{\xi}}^{(k+1)} &= \frac{1}{\sum_{i=1}^{n}\hat{q}_{1i}^{(k)}}\left(\sum_{i=1}^{n}\hat{q}_{1i}^{(k)}\boldsymbol{y}_i - \hat{\boldsymbol{\Gamma}}^{(k)^{-1}}\hat{\boldsymbol{\nu}}^{(k)}\sum_{i=1}^{n}\hat{q}_{3i}^{(k)} - n\hat{\boldsymbol{\lambda}}_1^{(k)}\right), \\ \hat{\boldsymbol{\Sigma}}^{(k+1)} &= \frac{1}{n}\left(\sum_{i=1}^{n}\hat{q}_{1i}^{(k)}\left(\boldsymbol{y}_i-\hat{\boldsymbol{\xi}}^{(k+1)}\right)\left(\boldsymbol{y}_i-\hat{\boldsymbol{\xi}}^{(k+1)}\right)^{\top} + \hat{\boldsymbol{\lambda}}_1^{(k)}\hat{\boldsymbol{\lambda}}_1^{(k)\top}\sum_{i=1}^{n}\hat{q}_{2i}^{(k)}\right. \\ &\qquad \left. -\sum_{i=1}^{n}\left(\boldsymbol{y}_i-\hat{\boldsymbol{\xi}}^{(k+1)}\right)\hat{\boldsymbol{\lambda}}_1^{(k)\top} - \sum_{i=1}^{n}\hat{\boldsymbol{\lambda}}_1^{(k)}\left(\boldsymbol{y}_i-\hat{\boldsymbol{\xi}}^{(k+1)}\right)^{\top}\right), \\ \hat{\boldsymbol{\lambda}}_1^{(k+1)} &= \frac{1}{\sum_{i=1}^{n}\hat{q}_{2i}^{(k)}}\left(\sum_{i=1}^{n}\left(\boldsymbol{y}_i-\hat{\boldsymbol{\xi}}^{(k+1)}\right) - \hat{\boldsymbol{\Gamma}}^{(k)^{-1}}\hat{\boldsymbol{\nu}}^{(k)}\sum_{i=1}^{n}\hat{q}_{4i}^{(k)}\right), \\ \hat{\boldsymbol{\nu}}^{(k+1)} &= \frac{\hat{\boldsymbol{\Sigma}}^{(k+1)^{-1}}}{n}\left(\sum_{i=1}^{n}\hat{q}_{3i}^{(k)}\left(\boldsymbol{y}_i-\hat{\boldsymbol{\xi}}^{(k+1)}\right) - \sum_{i=1}^{n}\hat{q}_{4i}^{(k)}\hat{\boldsymbol{\lambda}}_1^{(k+1)}\right). \end{aligned} \tag{18}$$

We then update $\hat{\boldsymbol{\lambda}}_2^{(k+1)}$ by

$$\hat{\boldsymbol{\lambda}}_2^{(k+1)} = \left(\hat{\boldsymbol{\Sigma}}^{(k+1)}\right)^{\frac{1}{2}}\hat{\boldsymbol{\nu}}^{(k+1)}.$$

The above procedure is repeated until either the maximum number of iterations is reached or a suitable stopping criterion is satisfied. Moreover, Appendix B details an approach for estimating the standard errors of the ML estimates using the observed information matrix.

3.2. Computational Strategies for the Implementation

One of the attractive properties of the EM-type algorithms is that successive iterates monotonically increase the likelihood value. So, a highly useful way to confirm convergence is that the difference between two successive log-likelihood values is less than a user-specified error tolerance. In this paper, we have used the Aitken acceleration method [20] to avoid the lack of process when determining the actual convergence [21]. Given the sequence of observed log-likelihood $\left\{\ell^{(k)}\right\}_{k=0}^{\infty}$ values in successive iterations, we first calculate the Aitken acceleration factor $a^{(k)} = \left(\ell^{(k+1)} - \ell^{(k)}\right) / \left(\ell^{(k)} - \ell^{(k-1)}\right)$. This yields the asymptotic estimate of the log-likelihood $\ell_{\infty}^{(k+1)} = \ell^{(k+1)} + \left(\ell^{(k+1)} - \ell^{(k)}\right) / \left(1 - a^{(k)}\right)$, which can be determined in advance at iteration $k+1$. The algorithm then gets terminated if

$$\ell_{\infty}^{(k+1)} - \ell^{(k)} < \epsilon,$$

where $\epsilon = 10^{-6}$ is the default tolerance employed in our study.

Moreover, the implementation of EM algorithms requires initial values. Essentially, good initial values for the optimization process may speed up or even enable the convergence [22]. In our computations, we used the following method for choosing initial values. For $\boldsymbol{\xi}$ and $\boldsymbol{\Sigma}$, the sample mean and sample covariance matrix were used as initial values. The initial values for $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ were set to be vectors of 1's times a factor d, where d was randomly drawn from a uniform distribution between 0.5 and 2. Lastly, for the initial value of α, we chose an integer randomly from the interval between 1 and 10.

4. SIMULATION STUDY

To examine the performance of the proposed distribution as well as the estimation method, a small simulation study is carried out here. Through this emperical study, we examine the finite-sample properties of ML estimators obtained by the ECM algorithm, described earlier in Section 3. For this purpose, Monte Carlo samples of size $n = 50, 100, 200$, and 500 are generated from the SNMVBS model with true parameters $\boldsymbol{\xi} = (1, 1)$, $Vech(\boldsymbol{\Sigma}) = (1, 0.5, 1)$, $\boldsymbol{\lambda}_1 = (1, 1)$, $\boldsymbol{\lambda}_2 = (-1, -1)$, and $\alpha = 2$, where $Vech(\boldsymbol{D})$ is a vector of $\frac{p(p+1)}{2}$ distinct elements of the $p \times p$ symmetric matrix $\boldsymbol{D}$. Each simulated data set were then fitted under the true model via the ECM algorithm according to the procedure described in Section 3.1. For each sample size, the experiment was replicated 300 times to assess the performance.

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

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

where $\hat{\theta}_i$ denotes the estimate of a specific parameter at the ith replication.

Furthermore, we computed the standard deviations of the ML estimates across 300 Monte Carlo samples (MC.SEs) and compared them with the average values of the approximate standard errors (A.SEs) obtained through the inverse of the observed information matrix, presented in Appendix B. Numerical results displayed in Tables 2 and 3 show the empirical consistency of the ML estimates as the Bias and MSE values both shrink toward zero when n increases. Moreover, in small sample sizes, the differences between A.SEs and MC.SEs of shape parameters are significantly large, but 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, say 200 or more.

5. APPLICATIONS

In this section, two applications of the SNMVBS distribution are presented. These examples come from environmental studies in a multivariate setting and possessing some outliers. In these cases, the SNMVBS distribution provides a better fit to these data than the other skewed models mentioned earlier.

Table 2 Simulation results for assessing the asymptotic properties of parameter estimates and standard errors, when $n = 50$ and 100.

	n = 50				n = 100			
Parameter	**Bias**	**MSE**	**MC.SE**	**A.SE**	**Bias**	**MSE**	**MC.SE**	**A.SE**
ξ_1	−0.0531	0.2074	0.2599	0.2959	−0.0189	0.1211	0.1704	0.1882
ξ_2	−0.0532	0.1905	0.2896	0.2918	−0.0262	0.1237	0.1864	0.1890
d_{11}	−0.0603	0.3948	0.2709	0.2945	−0.0321	0.2962	0.1943	0.1835
d_{12}	0.1352	0.3642	0.1975	0.2705	−0.0863	0.2424	0.1379	0.1697
d_{22}	−0.0728	0.4889	0.2821	0.2943	−0.0616	0.3016	0.2029	0.1840
λ_{11}	−0.0625	0.2375	0.3359	0.4935	−0.0404	0.1784	0.2244	0.2886
λ_{12}	−0.0715	0.2524	0.3028	0.4782	−0.0574	0.1801	0.2442	0.2874
λ_{21}	0.0059	0.7948	0.7967	2.1740	−0.0279	0.5528	0.5978	1.2770
λ_{22}	0.1124	0.8508	0.8944	2.1805	0.0856	0.5820	0.6199	1.2425
α	0.0331	0.3653	0.4025	0.4939	0.0431	0.2364	0.2341	0.3085

Table 3 Simulation results for assessing the asymptotic properties of parameter estimates and standard errors, when $n = 200$ and 500.

	n = 200				n = 500			
Parameter	**Bias**	**MSE**	**MC.SE**	**A.SE**	**Bias**	**MSE**	**MC.SE**	**A.SE**
ξ_1	−0.0086	0.0820	0.1135	0.1159	−0.0041	0.0498	0.0762	0.0807
ξ_2	−0.0114	0.0882	0.1178	0.1183	−0.0059	0.0530	0.0706	0.0807
d_{11}	−0.0121	0.2122	0.1293	0.1093	−0.0147	0.1352	0.0903	0.0789
d_{12}	−0.0361	0.1682	0.0786	0.0958	−0.0255	0.1009	0.0544	0.0617
d_{22}	−0.0160	0.2220	0.1407	0.1096	−0.0146	0.1424	0.0973	0.0786
λ_{11}	−0.0137	0.1298	0.1396	0.1633	−0.0074	0.0745	0.0955	0.1098
λ_{12}	−0.0247	0.1229	0.1493	0.1635	−0.0156	0.0737	0.1018	0.1093
λ_{21}	−0.0244	0.3842	0.3746	0.7025	−0.0063	0.2333	0.2594	0.4891
λ_{22}	0.0361	0.4014	0.3862	0.6913	0.0252	0.2653	0.2497	0.4874
α	0.0175	0.1667	0.1469	0.1792	0.0042	0.1016	0.1059	0.1229

5.1. Wind Speed Data

In this example, we consider the wind speed dataset and then fit the SNMVBS distribution in its multivariate form. The wind speed dataset consists of $n = 278$ hourly average wind speeds collected at three meteorological towers: Goodnoe Hills (*gh*), Kennewick (*kw*), and Vansycle (*vs*). Those towers are approximately located on a line and ordered from West to East. Azzalini and Genton [23] considered these data and fitted a trivariate-ST model. They treated the observations as being independent and identically distributed because the tests demonstrated a weak serial correlation between the three stations. The heavy tail behavior indicates the presence of extreme wind speeds. Arslan [6] fitted for these data the multivariate skew-Laplace distribution. She studied two-dimensional vector of wind speed recorded at the locations *gh* and *kw* as a bivariate random sample and noted that the estimates for the skewness parameter reveal apparent skewness in the data.

Here, we fit the SNMVBS distribution for these data as an alternative model. Figure 2 shows classical and robust Mahalanobis distances of this bivariate data. The plots show some outliers in these data and consequently a symmetric distribution will not be a good choice in this case. We assume (*gh, kw*) as a two-dimensional vector and fit a bivariate SNMVBS to that. Similarly, we also fit bivariate NIG, ST, and NMVBS distributions for the sake of comparison. As recommended by an anonymous referee, we also fitted the skew-Laplace distribution, known as shifted asymmetric Laplace (SAL), introduced by Franczak *et al.* [24]. The SAL random variable can be represented by (1) when W is a standard exponential random variable with mean 1. For fitting bivariate ST, we used the package "*sn*" and for NIG we used the package "*ghyp*" built-in in R 3.5.1 statistical software. Table 4 shows the obtained results, assuming $\alpha = \nu$ and $\overline{\alpha}$ for ST and NIG, respectively.

Performance assessments for studied models were made on the adequacy of overall fitness in terms of the Akaike Information Criterion (AIC; [25]), Bayesian Information Criterion (BIC; [26]) and Hannan–Quinn information Criterion (HQIC; [27]), defined as

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

and

$$\text{HQIC} = 2m\log\left(\log\ n\right) - 2\ell_{max},$$

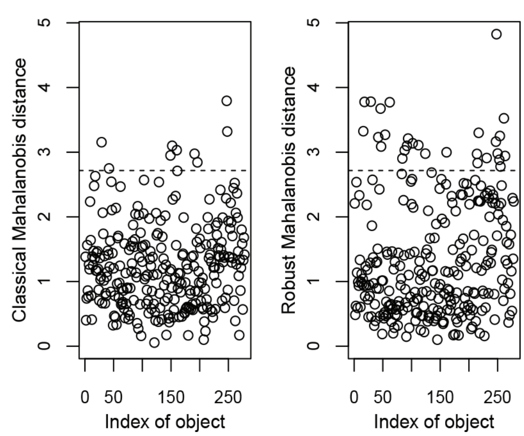

Figure 2 Wind speed data: Classical and robust Mahalanobis distances.

Table 4 Estimation results and information criteria for wind speed data (*gh, kw*). Standard errors of ML estimates are given in parentheses.

Par	SNMVBS	ST	NMVBS	NIG	SAL
ξ_1	25.41 (2.639)	28.48 (0.652)	32.68 (10.49)	33.80 (4.526)	23.07 (1.145)
ξ_2	40.94 (7.584)	24.09 (1.018)	40.09 (14.06)	41.10 (5.527)	22.01 (2.730)
σ_{11}	424.25 (41.25)	387.45 (5.397)	97.97 (31.09)	94.10 (24.03)	136.26 (18.47)
σ_{12}	213.68 (40.57)	241.45 (3.590)	30.36 (39.15)	20.07 (25.26)	85.24 (19.01)
σ_{22}	259.45 (43.36)	349.09 (2.822)	183.84 (52.54)	188.35 (38.59)	345.65 (35.35)
λ_{11}	3.24 (2.521)	-	−18.33 (10.42)	−21.06 (4.573)	−10.32 (1.285)
λ_{12}	−18.37 (7.180)	-	−23.95 (14.15)	−27.07 (5.731)	−7.97 (2.759)
λ_{21}	−5.54 (0.448)	−4.48 (0.180)	-	-	-
λ_{22}	−1.16 (0.347)	−0.10 (0.299)	-	-	-
α	0.38 (0.073)	17.62 (0.121)	0.41 (0.154)	5.08 (0.006)	-
$\ell(\hat{\theta})$	−2228.43	−2237.54	−2244.43	−2244.35	−2257.33
AIC	4476.86	4491.08	4504.86	4504.71	4528.66
BIC	4513.14	4520.11	4533.88	4533.73	4554.05
HQIC	4491.37	4502.73	4516.55	4516.36	4538.85

where m is the number of parameters and $\ell_{\max}$ is the maximized log-likelihood value. As a general rule, lower values of information criteria indicate a better-fitting model.

As can be seen from Table 4, it is evident from the AIC, BIC, and HQIC values that the SNMVBS model provides the best fit for these data. Figure 3 presents the scatter plots of wind data along with the contour plots of the fitted density and the fitted density function of the SNMVBS distribution. The plots reveal that the SNMVBS distribution can effectively capture the skewness and heavy-tails present in the data.

5.2. Water Quality Data

Cook and Johnson [28] introduced a new distribution and investigated its application in water quality. These data consist of log concentrations of seven chemical elements in 655 water samples collected near Grand Junction in Western Colorado. Concentrations were measured for the following elements: Uranium (U), Lithium (Li), Cobalt (Co), Potassium (K), Cesium (Cs), Scandium (Sc), and Titanium (Ti). The data are presented in package "*copula*." Figure 4 shows that there are some potential outliers in these data. The estimates of parameters are not shown here, but Figures. 5 and 6 show the contour plots of pairs of variables with multivariate SNMVBS contours. Clearly, the SNMVBS distribution provides a good fit for these data. Table 5 presents the information criteria of some alternate models. From this table, we observe once again that the AIC, BIC, and HQIC of SNMVBS are the lowest among all considered models.

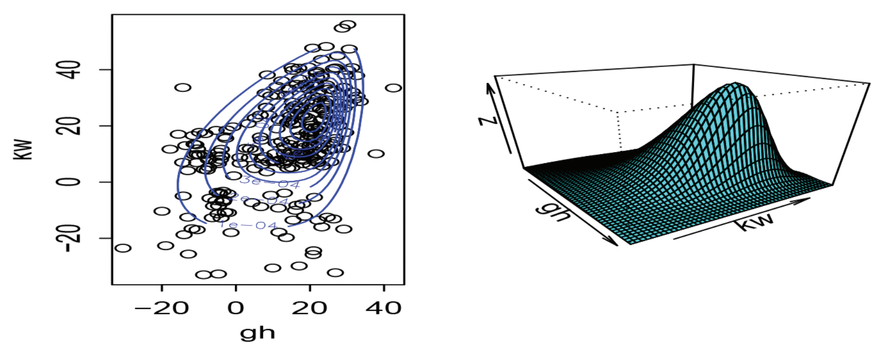

Figure 3 Wind speed data: Scatter plot (*gh, kw*) with fitted density contours and the fitted density of the skew-normal mean-variance mixture based on Birnbaum–Saunders (SNMVBS) distribution.

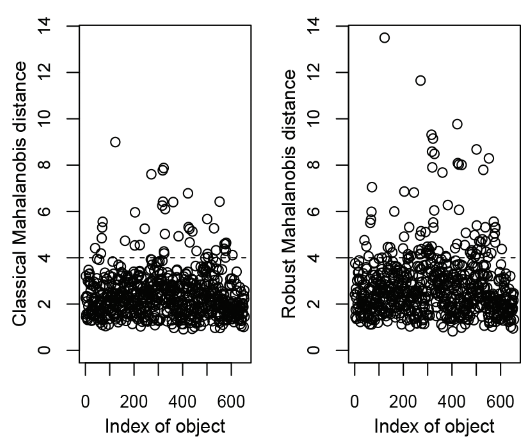

Figure 4 Water quality data: Classical and robust Mahalanobis distances.

Table 5 Water quality data: maximum log-likelihood, AIC, and BIC for different fitted models.

Criterion	SNMVBS	ST	NMVBS	NIG	SAL
$\ell(\hat{\theta})$	1969.61	1923.66	1925.58	1926.93	1839.19
AIC	−3839.20	−3761.33	−3765.16	−3767.86	−3594.39
BIC	−3614.96	−3568.49	−3572.32	−3575.02	−3406.04
HQIC	−3752.26	−3686.57	−3690.75	−3693.09	−3521.36

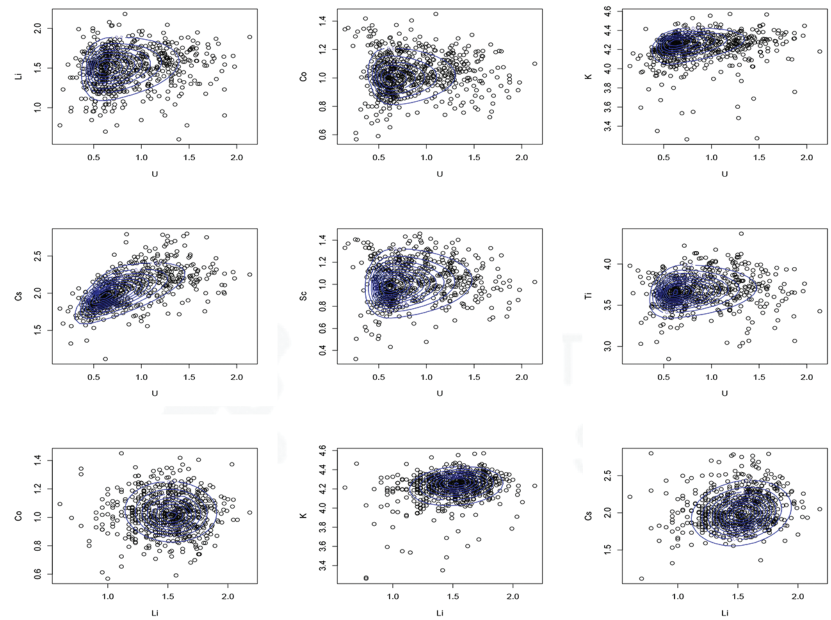

Figure 5 | Water quality data: Scatter plots with superimposed fitted skew-normal mean-variance mixture based on Birnbaum–Saunders (SNMVBS) density contours for some concentrations.

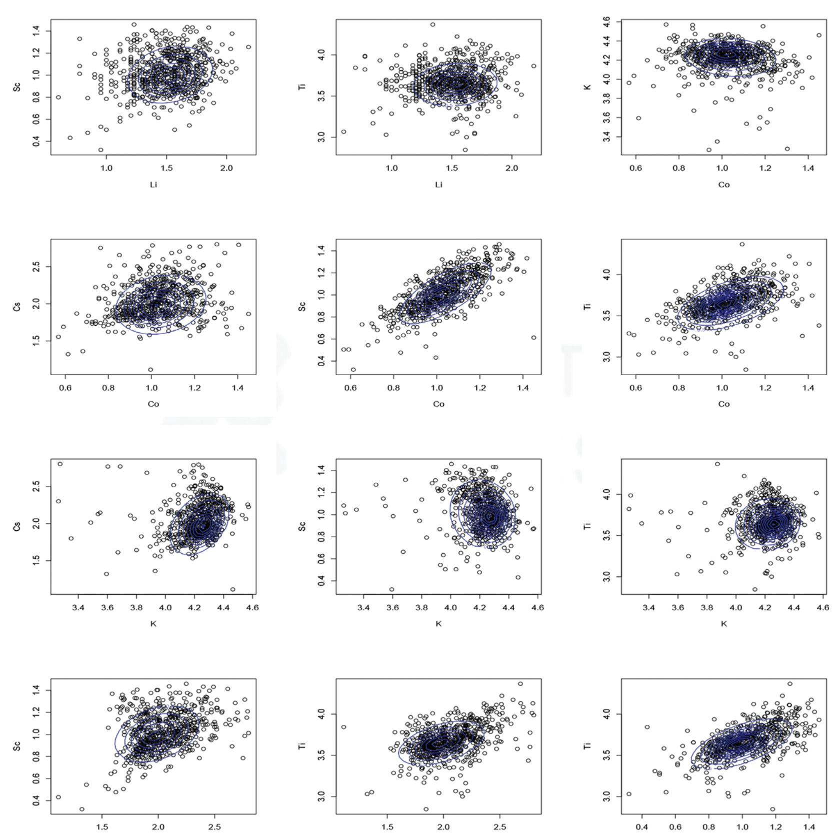

Figure 6 | Water quality data: (Continued).

6. CONCLUDING REMARKS

We have introduced in this work a multivariate skew-normal mean-variance mixture distribution obtained by employing BS as a mixing distribution. The properties of the SNMVBS distribution are studied, and a feasible EM-type algorithm for estimating parameters has been developed. We have demonstrated our methodology with a simulation study and have shown that the SNMVBS model outperforms other skew models in two data sets. The proposed distribution has some desirable properties. It offers a flexible class of distributions for modeling skewed and heavy-tailed data. It also provides a good fit to data consisting of outliers. Moreover, in contrast to the ST model, estimates of all parameters of the SNMVBS distribution can be obtained by solving some simple linear equations in the proposed ECM algorithm.

ACKNOWLEDGMENTS

We gratefully acknowledge the associate editor and three anonymous referees for their valuable comments and suggestions, which led to this greatly improved version of the article. We are also grateful to Prof. Tsung-I Lin for his help on the earlier version of the paper.

REFERENCES

1. I. Alameddine, M.A. Kenney, R.J. Gosnell, K.H. Reckhow, J. Environ. Eng. 136 (2010), 1299–1304.
2. C. Adcock, M. Eling, N. Loperfido, Eur. J. Finance. 21 (2015), 1253–1281.
3. O. Barndorff-Nielsen, Proc. R. Soc. Land. A. 353 (1977), 401–419.
4. A. Azzalini, Scand. J. Stat. 12 (1985), 171–178.
5. O. Barndorff-Nielsen, Scand. J. Stat. 9 (1978), 43–46.
6. O. Arslan, Stat. Pap. 51 (2010), 865–887.
7. W. Hu, PhD Thesis, Florida State University, 2005. https://diginole.lib.fsu.edu/islandora/object/fsu%3A181953
8. Z.W. Birnbaum, S.C. Saunders, J. Appl. Probab. 6 (1969), 319–327.
9. R. Pourmousa, A. Jamalizadeh, M. Rezapour, J. Stat. Comput. Simul. 85 (2015), 2736–2749.
10. A. Azzalini, A. Dalla Valle, Biometrika. 83 (1996), 715–726.
11. R.B. Arellano-Valle, M.G. Genton, J. Multivariate Anal. 96 (2005), 93–116.
12. O. Arslan, Stat. Pap. 56 (2015), 353–378.
13. M. Tamandi, H. Negarestani, A. Jamalizadeh, M. Amiri, JIRSS. (to appear). http://jirss.irstat.ir/article-1-499-en.html
14. A. Azzalini, A. Capitaino, J. R. Stat. Soc. B. 65 (2003), 367–389.
15. V. Leiva, The Birnbaum-Saunders Distribution, Elsevier Science, London, 2016.
16. K.V. Mardia, Biometrika. 57 (1970), 519–530.
17. A. Azzalini, A. Capitanio, J. R. Stat. Soc. B. 61 (1999), 579–602.
18. A.P. Dempster, N.M. Laird, D.B. Rubin, J. R. Stat. Soc. B. 39 (1977), 1–38.
19. X.L. Meng, D.B. Rubin, Biometrika. 80 (1993), 267–278.
20. A.C. Aitken, Proc. R. Soc. Edinb. 46 (1927), 289–305.
21. P.D. McNicholas, T.B. Murphy, A.F. McDaid, D. Frost, Comput. Stat Data. Anal. 54 (2010), 711–723.
22. T.I. Lin, H.J. Ho, C.R. Lee, Stat. Comput. 24 (2014), 531–546.
23. A. Azzalini, M.G. Genton, Int. Stat. Rev. 76 (2008), 106–129.
24. B.C. Franczak, R.P. Browne, P.D. McNicholas, IEEE. Trans. Pattern Anal. 36 (2014), 1149–1157.
25. H. Akaike, in: B.N. Petrov, F. Csaki, 2nd International Symposium on Information Theory, Akadémiai Kiadó, Budapest, 1973, pp. 267–281.
26. G. Schwarz, Ann. Stat. 6 (1978), 461–464.
27. E.J. Hannan, B.G. Quinn, J. R. Stat. Soc. B. 41 (1979), 190–195.
28. R.D. Cook, M.E. Johnson, Technometrics. 28 (1986), 123–131.
29. K.E. Basford, D.R. Greenway, G.J. McLachlan, D. Peel, Comput. Stat. 12 (1997), 1–17.

APPENDIX A. THEORETICAL PROPERTIES

The following lemmas are useful in establishing some properties of the skew-normal mean-variance mixture based on Birnbaum–Saunders (SNMVBS) distribution.

Lemma A.1. If $W \sim GIG(\kappa, \chi, \psi)$, then

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

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

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

where $R_{(\kappa,r)}(x) = \frac{K_{\kappa+r}(x)}{K_\kappa(x)}$, $a, b \in \mathbb{R}$,, $\Phi(\cdot)$ *is the cumulative distribution function (CDF) of the univariate standard normal distribution, and F is the CDF of* $GH_1(0, 1, b, \chi, \psi)$.

Parts *(i)* and *(ii)* of the above lemma are well-known properties of generalized inverse Gaussian (GIG) distribution; see [7] for details. Part *(iii)* has been proved by [12].

Lemma A.2. Let $W \sim BS(\alpha, 1)$. Then

i. *The probability density function (PDF) of W can be expressed as*

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

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

In particular, 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 A.2 and some other useful properties of BS can be found in Leiva [15], for example.

We now present the proof of Theorem 2.1.

Proof. From (6), we have

$$\begin{aligned}
f(\boldsymbol{Y}) &= \int_0^\infty f_{MSN}(\mathbf{y}|w) f_{BS}(w) dw \\
&= 2\int_0^\infty \phi_p(\mathbf{y}; \xi + w\lambda_1, w\boldsymbol{\Sigma})\Phi\left(w^{-1/2}\lambda_2^\top \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \xi - w\lambda_1)\right) f_{BS}(w) dw \\
&= \int_0^\infty \frac{w^{-\frac{p+3}{2}}(1+w)}{\alpha(2\pi)^{\frac{p+1}{2}}|\Sigma|^{1/2}} \Phi\left(w^{-1/2}\lambda_2^\top \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \xi - w\lambda_1)\right) \\
&\quad \times \exp\left\{-\frac{w^{-1}}{2}\left(\boldsymbol{y} - \xi - w\lambda_1\right)^\top \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y} - \xi - w\lambda_1\right) - \frac{1}{2\alpha^2}\left(\sqrt{w} - \frac{1}{\sqrt{w}}\right)^2\right\} dw \\
&= \frac{\exp\left\{\left(\boldsymbol{y} - \xi\right)^\top \boldsymbol{\Sigma}^{-1}\lambda_1\right\}}{2(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2} K_{1/2}(\alpha^{-2})} \\
&\quad \times \int_0^\infty w^{-\frac{p+3}{2}}(1+w)\Phi(w^{-1/2}a - w^{1/2}b)\exp\left\{-\frac{1}{2}(\delta_1 w^{-1} + \delta_2 w)\right\} dw \\
&= f_{GH_p}\left(\boldsymbol{y}; \xi, \boldsymbol{\Sigma}, \lambda_1, -1/2, \alpha^{-2}, \alpha^{-2}\right) E_{W_1}\left(\Phi\left(W_1^{-1/2}a - W_1^{1/2}b\right)\right) \\
&\quad + f_{GH_p}\left(\boldsymbol{y}; \xi, \boldsymbol{\Sigma}, \lambda_1, 1/2, \alpha^{-2}, \alpha^{-2}\right) E_{W_2}\left(\Phi\left(W_2^{-1/2}a - W_2^{1/2}b\right)\right),
\end{aligned}$$

where $W_1 \sim GIG\left(-\left(p+1\right)/2, \delta_1, \delta_2\right)$ and $W_2 \sim GIG\left(-\left(p-1\right)/2, \delta_1, \delta_2\right)$. The proof is then completed using Part *(iii)* of Lemma A.1.

The following proposition presents the moment generating function (mgf) of $\mathbf{Y}$.

Proposition A.1. *The mgf of* $\mathbf{Y} \sim SNMVBS(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \lambda_1, \lambda_2, \alpha)$ *is given by*

$$\begin{aligned} M_{\mathbf{Y}}(\boldsymbol{s}) &= \frac{e^{\boldsymbol{s}^\top \boldsymbol{\xi}}}{K_{1/2}\left(\alpha^{-2}\right)} K_{1/2}\left(\left(\alpha^{-2} - 2\boldsymbol{s}^\top \lambda_1 - \boldsymbol{s}^\top \boldsymbol{\Sigma} \boldsymbol{s}\right)\alpha^{-2}\right) \\ &\times \left\{ \left(\sqrt{\frac{\left(\alpha^{-2} - 2\boldsymbol{s}^\top \lambda_1 - \boldsymbol{s}^\top \boldsymbol{\Sigma}\boldsymbol{s}\right)}{\alpha^{-2}}}\right)^{-1/2} H^*_{\{0\}}\left(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{1/2}\boldsymbol{s}\right) \right. \\ &\left. + \left(\sqrt{\frac{\left(\alpha^{-2} - 2\boldsymbol{s}^\top \lambda_1 - \boldsymbol{s}^\top \boldsymbol{\Sigma}\boldsymbol{s}\right)}{\alpha^{-2}}}\right)^{1/2} H^*_{\{1\}}\left(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{1/2}\boldsymbol{s}\right) \right\}, \end{aligned} \tag{A.1}$$

where $H^*_{\{\kappa\}}(x) = F_{GH_1}\left(x; 0, 1, 0, \kappa, \alpha^{-2} - 2\boldsymbol{s}^\top \lambda_1 - \boldsymbol{s}^\top \boldsymbol{\Sigma}\boldsymbol{s}, \alpha^{-2}\right)$.

Proof. $\mathbf{Z} \sim SN_p\left(\xi, \boldsymbol{\Sigma}, \lambda\right)$ implies $M_{\mathbf{Z}}(\mathbf{s}) = 2e^{s^\top \xi + \frac{1}{2} s^\top \Sigma s}\Phi\left(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{1/2}\boldsymbol{s}\right)$. So using (6) we have

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{s}) = E_W\left(M_{\mathbf{Y}|W}(\mathbf{s})\right) &= \frac{e^{\boldsymbol{s}^\top \boldsymbol{\xi}}}{2K_{1/2}\left(\alpha^{-2}\right)} \int_0^\infty w^{-\frac{3}{2}}(1+w)\Phi(w^{1/2}\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{1/2}\boldsymbol{s}) \\ &\times \exp\left\{-\frac{1}{2}\left(\alpha^{-2}w^{-1} + \left(\alpha^{-2} - 2\boldsymbol{s}^\top \lambda_1 - \boldsymbol{s}^\top \boldsymbol{\Sigma}\boldsymbol{s}\right)w\right)\right\} dw. \end{aligned} \tag{A.2}$$

The proof is completed by performing some algebraic calculations.

The next theorem shows that SNMVBS random vector is invariant under linear transformations. Also, the marginal distributions of SNMVBS are presented.

Theorem A.1. Let $\mathbf{Y} \sim SNMVBS\left(\xi, \boldsymbol{\Sigma}, \lambda_1, \lambda_2, \alpha\right)$. Then

i. *Foy any* $\mathbf{A} \in \mathbb{R}^{q\times p}$ *and* $\mathbf{b} \in \mathbb{R}^q$, *the q-dimensional random vector* $\mathbf{V} = \mathbf{AY} + \mathbf{b}$ *is distributed as* $SNMVBS\left(\boldsymbol{A\xi} + \boldsymbol{b}, \boldsymbol{A\Sigma A}^\top, A\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2^*, \alpha\right)$, *where* $\boldsymbol{\lambda}_2^* = \frac{\boldsymbol{\delta}^*}{\left(1 - \boldsymbol{\delta}^{*\top}\boldsymbol{\delta}^*\right)^{1/2}}$ *with* $\boldsymbol{\delta}^* = \left(\boldsymbol{A\Sigma A}^\top\right)^{-1/2}\boldsymbol{A\Sigma}^{1/2}\boldsymbol{\delta}$. *Moreover, if* $q = p$ *and* $\mathbf{A}$ *is non-singular, then we get* $\boldsymbol{\lambda}_2^* = \boldsymbol{\lambda}_2$;

ii. *Suppose* $\mathbf{Y}$ *is partitioned as* $\mathbf{Y}^\top = \left(\mathbf{Y}_1^\top, \mathbf{Y}_2^\top\right)$ *of dimensions* p_1 *and* $p_2 = p - p_1$, *respectively. Accordingly, the mean, skewness vector and covariance matrix are partitioned as*

$$\boldsymbol{\xi}^\top = \left(\boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top\right), \quad \boldsymbol{\lambda}_1^\top = \left(\boldsymbol{\lambda}_{11}^\top, \boldsymbol{\lambda}_{12}^\top\right), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

Then, $\mathbf{Y}_1 \sim SNMVBS\left(\boldsymbol{\xi}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\lambda}_{11}, \boldsymbol{\Sigma}_{11}^{1/2}\tilde{\boldsymbol{\nu}}, \alpha\right)$, *where* $\tilde{\boldsymbol{\nu}} = \frac{\boldsymbol{\nu}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\nu}_2}{\sqrt{1 + \boldsymbol{\nu}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\nu}_2}}$, *with* $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ *and* $\boldsymbol{\nu} = \left(\boldsymbol{\nu}_1^\top, \boldsymbol{\nu}_2^\top\right) = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}_2$.

Proof. The proof of Part *(i)* is straightforward by obtaining the mgf of $\mathbf{V}$ from Proposition A.1. Applying Part *(i)* to $\mathbf{A} = \left[\boldsymbol{I}_{p_1} \quad \mathbf{0}_{p_2}\right]$ with $p = p_1 + p_2$, Part *(ii)* can be proved.

Although there is no closed-form for the conditional distribution of SNMVBS distribution, it is clear that the conditional distribution of $\boldsymbol{Y}_2$, given $\boldsymbol{Y}_1$, is not in the same class of distributions.

APPENDIX B. PROVISION OF STANDARD ERRORS

To compute the asymptotic covariance of the maximum likelihood (ML) estimates, we use the method suggested by Basford *et al.* [29]. Let $\ell_{ci}\left(\boldsymbol{\theta}|\boldsymbol{y}_i, \boldsymbol{\gamma}_i, \boldsymbol{w}_i\right)$ be the complete data log-likelihood contributed from the single observation $\mathbf{y}_i$. Then, the empirical information matrix is given by $\hat{\boldsymbol{I}} = \sum_{i=1}^{n} \hat{\mathbf{s}}_i \hat{\mathbf{s}}_i^\top$, where $\hat{\mathbf{s}}_i$ is the individual score and defined by

$$\hat{\boldsymbol{s}}_i = \boldsymbol{E}_{\hat{\boldsymbol{\theta}}}\left(\frac{\partial \ell_{ci}\left(\boldsymbol{\theta}|\boldsymbol{y}_i, \boldsymbol{\gamma}_i, \boldsymbol{w}_i\right)}{\partial \boldsymbol{\theta}}|\boldsymbol{y}_i\right).$$

The standard errors of ML estimates can then be found by calculating the square root of the diagonal elements of $\hat{I}^{-1}$.

The $\ell_{ci}(\boldsymbol{\theta}|\boldsymbol{y}_i, \boldsymbol{\gamma}_i, \boldsymbol{w}_i)$ can be obtained from (15). Let $\boldsymbol{d} = Vech(\boldsymbol{D})$ be a vector of $\frac{p(p+1)}{2}$ distinct elements of the symmetric matrix $\boldsymbol{D}$, where $\boldsymbol{D} = \hat{\boldsymbol{\Sigma}}^{-1/2}$. Then, we have

$$\begin{aligned}
\hat{\boldsymbol{s}}_{i,\alpha} &= -\frac{1}{\hat{\alpha}} + \frac{1}{\hat{\alpha}^3}\left(\hat{q}_{1i} + \hat{q}_{2i} - 2\right), \quad \hat{\boldsymbol{s}}_{i,\xi} = \hat{\boldsymbol{\Gamma}}\left(\hat{q}_{1i}\left(\boldsymbol{y}_i - \hat{\xi}\right) - \hat{\lambda}_1\right) - \hat{q}_{3i}\hat{\boldsymbol{v}}, \\
\hat{\boldsymbol{s}}_{i,d} &= Vech\left\{2D^{-1} - Diag\left(D^{-1}\right) + \hat{\boldsymbol{A}}_i + \hat{\boldsymbol{A}}_i^\top - Diag\left(\hat{\boldsymbol{A}}_i\right) + \hat{\boldsymbol{B}}_i + \hat{\boldsymbol{B}}_i^\top - Diag\left(\hat{\boldsymbol{B}}_i\right)\right\}, \\
\hat{\boldsymbol{s}}_{i,\boldsymbol{\lambda}_1} &= \hat{\boldsymbol{\Gamma}}\left(\boldsymbol{y}_i - \hat{\xi} - \hat{q}_{2i}\hat{\boldsymbol{\lambda}}_1\right) - \hat{q}_{4i}\hat{\boldsymbol{v}}, \\
\hat{\boldsymbol{s}}_{i,\boldsymbol{v}} &= \hat{q}_{3i}\left(\boldsymbol{y}_i - \hat{\xi}\right) - \hat{q}_{4i}\hat{\lambda}_1 - \left(\hat{q}_{1i}\left(\boldsymbol{y}_i - \hat{\xi}\right)\left(\boldsymbol{y}_i - \hat{\xi}\right)^\top + \hat{q}_{2i}\hat{\boldsymbol{\lambda}}_1\hat{\boldsymbol{\lambda}}_1^\top \right. \\
&\quad \left. - \left(\boldsymbol{y}_i - \hat{\xi}\right)\hat{\lambda}_1^\top - \hat{\lambda}_1\left(\boldsymbol{y}_i - \hat{\xi}\right)^\top\right)\hat{\boldsymbol{v}},
\end{aligned}$$

where

$$\begin{aligned}
\hat{\boldsymbol{A}}_i &= \boldsymbol{D}^{-1}\left(\hat{q}_{1i}\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}_i^\top + \hat{q}_{2i}\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}^\top - \hat{\boldsymbol{u}}_i\hat{\boldsymbol{v}}^\top + \hat{q}_{1i}\hat{\boldsymbol{\lambda}}_2\hat{\boldsymbol{\lambda}}_2^\top\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}_i^\top + \hat{q}_{2i}\hat{\boldsymbol{\lambda}}_2\hat{\boldsymbol{\lambda}}_2^\top\hat{\boldsymbol{v}}\hat{\boldsymbol{v}}_i^\top - \hat{\boldsymbol{\lambda}}_2\hat{\boldsymbol{\lambda}}_2^\top\hat{\boldsymbol{u}}_i\hat{\boldsymbol{v}}^\top\right), \\
\hat{\boldsymbol{B}}_i &= \left(\hat{q}_{4i}\hat{\boldsymbol{v}}\hat{\boldsymbol{\lambda}}_2^\top - \hat{\boldsymbol{u}}_i\hat{\boldsymbol{v}}^\top - \hat{q}_{3i}\hat{\boldsymbol{u}}_i\hat{\lambda}_2^\top - \hat{\boldsymbol{u}}_i\hat{\boldsymbol{v}}^\top\hat{\lambda}_2\hat{\lambda}_2^\top\right)\boldsymbol{D}^{-1},
\end{aligned}$$

$\hat{\boldsymbol{u}}_i = \boldsymbol{D}^{-1}\left(\boldsymbol{y}_i - \hat{\xi}\right)$ and $\hat{v} = \boldsymbol{D}^{-1}\hat{\lambda}_1$. $Diag(\boldsymbol{A})$ is a diagonal matrix created by extracting the main diagonal elements of the square matrix $\boldsymbol{A}$.