The multivariate Birnbaum-Saunders distribution based on a asymmetric distribution: EM-estimation

Filidor Vilca[1]*, Camila Borelli Zeller[2], N. Balakrishnan [3]

[1]Departamento de Estatística, Universidade Estadual de Campinas, São Paulo, Brazil
[2]Departamento de Estatística, Universidade Federal de Juiz de Fora, Minas Gerais, Brazil
[3] Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada

Abstract

We derive here a multivariate generalization of the bivariate Birnbaum-Saunders (BS) distribution of Kundu et al. (2010) by basing it on the multivariate skew-normal (SN) distribution. The resulting multivariate Birnbaum-Saunders type distribution is an absolutely continuous distribution whose marginals are in the form of univariate Birnbaum-Saunders type distributions discussed by Vilca et al. (2011). We then study its characteristics and properties, such as the joint distribution function, marginal and conditional distributions. Next, we introduce a non-central multivariate BS distribution in order to present analytically a simple EM-algorithm for iteratively computing the maximum likelihood estimates of the model parameters, and compare the performance of this method with the estimation approach of Jamalizadeh and Kundu (2015). Moreover, the observed Fisher information matrix is analytically derived under the bivariate case, and some simulation studies and an application to a real data set are finally presented for the propose of illustrating the model and inferential results developed here.

Keywords Birnbaum-Saunders distribution; Skew-normal distribution; Marginal and Conditional distributions; EM-algorithm.

*Corresponding author: Filidor Vilca, Departamento de Estatística, Universidade Estadual de Campinas, Caixa Postal 6065, SP, Brazil. Email: *fily@ime.unicamp.br; filyps@yahoo.com.br.*

1 Introduction

Several methods of constructing multivariate continuous distributions have been proposed and studied in the last few decades in the literature; see Kotz et al. (2000) and Balakrishnan and Lai (2009) for elaborate details in this regard. For life distributions and lifetime analysis, interested readers may refer to Marshall and Olkin (2007) and the references therein.

The univariate Birnbaum-Saunders (BS) distribution, which is closely related to the normal distribution, has received considerable attention recently in the literature. This distribution has found many important applications in such diverse fields as engineering, industry, business, reliability, survival analysis, and medical sciences. It fits well the low or high percentiles of the lifetime distribution, a region where little data are usually found. This distribution is obtained as a transformation of the standard normal distribution, and it is related to the normal model through the stochastic representation

$$T = \frac{\beta}{4}\left(\alpha\, Z + \sqrt{(\alpha\, Z)^2 + 4}\right)^2, \qquad (1)$$

where $Z \sim \mathrm{N}(0,1)$, $\alpha > 0$ and $\beta > 0$. The random variable T in (1) is said to have a BS distribution with α and β as shape and scale parameters, respectively, and is usually denoted by $T \sim \mathrm{BS}(\alpha, \beta)$; for more details, see Birnbaum and Saunders (1969 a,b). A slight extension of the BS distribution (1) is obtained by following two aspects: (i) The standard normal assumption could be relaxed by considering $N(0, \sigma^2)$ instead of $N(0,1)$, and in this case the resulting BS distribution is equivalent to (1), with just $\sigma\alpha$ in place of α; (ii) The BS distribution was constructed under restrictive conditions that may not be valid for certain applications; see Mann et al. (1974, p. 152) and Desmond (1985). An idea to overcome this problem was presented by Guiraud et al. (2009) who constructed a non-central Birnbaum-Saunders (NBS) distribution by assuming $N(\nu, 1)$ in (1) in place of $N(0,1)$. So, combining (i) and (ii), and assuming that Z follows $N(\nu, \sigma^2)$ distribution, the resulting probability density function (pdf) of T in (1) becomes

$$f_T(t) = \phi\big(a_t(\alpha, \beta) - \nu; 0, \sigma^2\big)\, A_t(\alpha, \beta), \quad t > 0, \qquad (2)$$

where $\phi(\cdot; 0, \sigma^2)$ is the pdf of $N(0, \sigma^2)$ distribution, $a_t(\alpha, \beta) = \big(\sqrt{t/\beta} - \sqrt{\beta/t}\big)/\alpha$ and $A_t(\alpha, \beta) = t^{-3/2}(t+\beta)/(2\,\alpha\,\beta^{1/2})$. This non-central Birnbaum-Saunders (NBS) distribution is denoted by $T \sim \mathrm{NBS}(\alpha, \beta, \nu, \sigma^2)$; see Guiraud et al. (2009) when $\sigma^2 = 1$.

Much work has been done on univariate BS distribution and its extensions, many of which have been applied with success, mainly due to their robust parameter estimation; see, for example, Barros et al. (2008) and Balakishnan et al. (2009). But, these models may still be inadequate to fit data that are highly concentrated on the left-tail of the distribution. This problem can be overcome by considering the skew-normal (SN) distribution (Azzalini, 1985) instead of the normal distribution in (1); see Vilca and Leiva (2006) and Vilca et al. (2011). Such a BS distribution based on the SN distribution, referred to as the skew-normal Birnbaum-Saunders (SNBS) distribution, is able to fit the extreme lower percentiles very well. In fact, if the distribution of data follows a SNBS distribution with negative asymmetry parameter λ (associated with the SN distribution), and instead we fit these data by the usual BS distribution, or by a BS model based on a symmetric distribution (Leiva et al., 2008), we will end up overestimating the lower percentiles. By the way, the SNBS distribution is defined as follows: Let Z in (1) follow a standard skew-normal distribution, $SN(0, 1, \lambda)$, with asymmetry parameter $\lambda \in \mathbb{R}$. Then, the positive random variable T in (1) is said to have a SNBS distribution, and its pdf is given by (Vilca et al., 2011)

$$f_T(t; \lambda) = 2\phi\big(a_t(\alpha, \beta)\big)\,\Phi\big(\lambda a_t(\alpha, \beta)\big)\, A_t(\alpha, \beta), \quad t > 0, \tag{3}$$

where $\phi(\cdot)$ and $\Phi(.)$ denote the pdf and the cumulative distribution function (cdf) of $N(0, 1)$ distribution, respectively, and $a_t(\alpha, \beta)$ and $A_t(\alpha, \beta)$ are as given in (2). This distribution is denoted by $T \sim \text{SNBS}(\alpha, \beta, \lambda)$.

In the bivariate context, Kundu et al. (2010) constructed a bivariate BS distribution with dependence structure and established several properties. In addition to possessing a close relationship with the bivariate normal distribution, this distribution also has its marginal distributions as univariate BS distributions, and the conditional distribution as non-central Birnbaum-Saunders (NBS) distributions, as pointed out in Vilca et al. (2014). But, the bivariate or multivariate BS distributions of Kundu et al. (2010, 2013) may not be suitable for modelling data which are quite skewed, as for example when there are data highly concentrated around the zero.

The aim of this paper is to provide a form of multivariate Birnbaum-Saunders distribution that is based on the multivariate skew-normal distribution, inspired by the idea of Vilca et al. (2011) and **the motivating example in Section 2 arising from reliability properties**. The resulting family of distributions has its marginal and conditional distributions in closed-forms. **MAS COMENTARIOS**. This new resulting family of multivariate BS-type distributions will be referred to as the multivariate skew-normal Birnbaum-Saunders (SNBS) distributions.

The rest of this paper is organized as follows. In Section 2, we describe briefly the bivariate and multivariate BS distributions of Kundu et al. (2010, 2013) and present motivating examples that shows the presence of the proposed distribution. In Section 3, we construct the multivariate BS distribution based on the multivariate skew-normal distribution, and present its pdf and some other properties. In Section 4, a non-central bivariate BS distribution, an estimation method based on the EM algorithm for the maximum likelihood estimation and the observed Fisher information matrix are derived. In Section 5, numerical examples using both simulated and real data are presented to illustrate the proposed methodology. Finally, some concluding remarks are made in Section 6.

2 Motivating Examples

Recently, Kundu et al. (2013) introduced a multivariate Birnbaum-Saunders distribution by using a multivariate symmetric distribution. When the multivariate normal distribution is used as a base kernel for the transformation, the cumulative distribution function of the multivariate BS random vector $\mathbf{T} = (T_1, \ldots, T_p)^\top$ is of the form

$$F_{\mathbf{T}}(\mathbf{t}) = \Phi_p\big(a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}); \boldsymbol{\Sigma}\big), \tag{4}$$

where $\Phi_p(.; \mathbf{0}, \boldsymbol{\Sigma}\big)$ is the cdf of $\mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ distribution, with $\boldsymbol{\Sigma}$ being a $p \times p$ positive-definite correlation matrix, with parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ in $\mathbb{R}_+^p$, which is the positive part of $\mathbb{R}^p$, and $a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (a_{t_1}(\alpha_1, \beta_1), \ldots, a_{t_p}(\alpha_p, \beta_p))^\top$, with $a_{t_j}(\alpha_j, \beta_j) = (1/\alpha_j)\big(\sqrt{t_j/\beta_j} - \sqrt{\beta_j/t_j}\big)$, $j = 1, \ldots, p$. For the distribution in (4), we use the notation $\mathbf{T} \sim \mathrm{BS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$. The bivariate BS distribution developed in Kundu et al. (2010), in which a random vector $\mathbf{T} = (T_1, T_2)^\top$ is said to have a bivariate BS distribution with parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$ and $-1 < \rho < 1$, has the joint probability density function (pdf) of T_1 and T_2 as

$$f_{\mathbf{T}}(\mathbf{t}) \;=\; \phi_2\big(a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}); \rho\big)\, A_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}),\ \mathbf{t} \in \mathbb{R}_+^2, \tag{5}$$

where

$$\phi_2\big(\mathbf{z}; \rho\big) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big\{ -\frac{1}{2}\mathbf{z}^\top \boldsymbol{\Sigma}^{-1}\mathbf{z}\Big\}$$

is the pdf of $\mathbf{Z} = (Z_1, Z_2)^\top \sim \mathrm{N}_2(\mathbf{0}, \boldsymbol{\Sigma})$, with $\sigma_{11} = \sigma_{22} = 1$, and correlation coefficient ρ, $a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ as in (4), $A_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = A_{t_1}(\alpha_1, \beta_1)\, A_{t_2}(\alpha_2, \beta_2)$, with $A_{t_j}(\alpha_j, \beta_j) = (t_j + \beta_j)/\big(2\alpha_j\sqrt{\beta_j}\sqrt{t_j^3}\big)$ for $j = 1, 2$. For the distribution in (5), we use the notation $\mathbf{T} \sim \mathrm{BS}_2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \rho)$.

Applications of bivariate and multivariate BS distributions in reliability analysis have not been discussed yet in detail as in the case of some other lifetime distributions such as the multivariate exponential distribution, which could be useful for identifying appropriate models. For the bivariate case, some results can be found in Vilca et al. (2014). For example, they have presented the reliability function of $\mathbf{T} = (T_1, T_2)^\top$ and the conditional reliability function of T_2, given $T_1 = t_1$, which involves the univariate non-central BS distribution (Guiraud et al., 2009). Another subject is the conditional reliability properties of T_2, given $T_1 > t_1$ (based on conditional hazard functions) as pointed out by Navarro and Sarabia (2013). Next, we present a special case when $t_1 = \beta_1$, which provides a motivation for our study of the BS distribution based on the skew-normal distribution for the univariate case as well as for the multivariate case.

(i) For $p = 2$, the bivariate BS distribution possesses several interesting properties as detailed in Kundu el al. (2010) and Vilca et al. (2014). Suppose we are interested in the shape of the hazard function of T_2, given $T_1 > \beta_1$; see the works of Gupta (2006) and Navarro and Sarabia (2013) in this regard. The first step for this is to obtain the conditional pdf of T_2, given $T_1 > \beta_1$, which is given by

$$\begin{aligned}
f_{T_2|T_1}(t_2|T_1>\beta_1) &= \frac{1}{P(T_1>\beta_1)}\int_{\beta_1}^{\infty} f_{T_1,T_2}(u,t_2)du,\\
&= \frac{1}{P(T_1>\beta_1)}\int_{\beta_1}^{\infty} f_{T_1|T_2}(u|t_2)f_{T_2}(t_2)du\\
&= \frac{f_{T_2}(t_2)}{P(T_1>\beta_1)}P\big(Z_1>-\lambda a_{t_2}(\alpha_2,\beta_2)\big),\\
&= 2\phi\big(a_{t_2}(\alpha_2,\beta_2)\big)\Phi\big(\lambda a_{t_2}(\alpha_2,\beta_2)\big)\,A_{t_2}(\alpha_2,\beta_2), \quad t_2>0,
\end{aligned}$$

where $\lambda = \rho/\sqrt{1-\rho^2}$. It is evident that this just corresponds to the BS distribution based on the skew-normal distribution, referred to as the skew-normal Birnbaum-Saunders (SNBS) distribution by Vilca et al. (2011), that was derived by considering the standard skew-normal distribution in place of $N(0,1)$ distribution in (1).

(ii) For $p = 3$, we can get a result similar to the one in (i), which is obtained by following the idea of Azzalini and Dalla-Valle (1996) and Gupta et al. (2004). Let $\mathbf{T} = (T_1, T_2, T_3)^\top \sim \mathrm{BS}_3(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \boldsymbol{\delta}^\top \\ \boldsymbol{\delta} & \boldsymbol{\Omega} \end{pmatrix}, \text{ with } \boldsymbol{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Then, according to Kundu et al. (2013),

$$\mathbf{V} = a_{\mathbf{T}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (a_{T_1}(\alpha_1,\beta_1), a_{T_2}(\alpha_2,\beta_2), a_{T_3}(\alpha_3,\beta_3))^\top \sim \mathrm{N}_3(\mathbf{0}, \boldsymbol{\Sigma}).$$

In this case, the conditional joint pdf of T_2 and T_3, given $T_1 > \beta_1$, is of the form

$$f_{T_2,T_3|T_1}(t_2,t_3|T_1>\beta_1) = 2\phi_2\big(a_{\mathbf{t}_{2,3}};\rho\big)\Phi\big(\boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{-1/2}a_{\mathbf{t}_{2,3}}\big)\,A_{\mathbf{t}_{2,3}},\ t_2>0, t_3>0, \quad (6)$$

where $\phi_2(.;\rho)$ is the pdf of $\mathrm{N}_2(\mathbf{0}, \boldsymbol{\Omega})$ distribution, $a_{\mathbf{t}_{2,3}} = (a_{t_2}(\alpha_2,\beta_2), a_{t_3}(\alpha_3,\beta_3))^\top$, $A_{\mathbf{t}_{2,3}} = A_{t_2}(\alpha_2,\beta_2)A_{t_3}(\alpha_3,\beta_3)$ and $\boldsymbol{\lambda} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{\delta}/(1-\boldsymbol{\delta}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\delta})^{1/2}$.

So, analogous to the construction of the multivariate skew-normal distribution proposed by Azzalini and Dalla-Valle (1996), the conditional joint pdf of T_2 and T_3, given $T_1 > \beta_1$, is obtained from the conditional joint pdf of $\big(a_{T_2}(\alpha_2,\beta_2), a_{T_3}(\alpha_3,\beta_3)\big)$, given $a_{T_1}(\alpha_1,\beta_1) > 0$, which is the bivariate skew-normal distribution, $\mathrm{SN}_2(\mathbf{0}, \boldsymbol{\Omega})$. The remaining steps to obtain $f_{T_2|T_1}(t_2|T_1 > \beta_1)$ follow the same line as in Part (i). In these examples we can see the presence on of a new extension of the bivariate or multivariate BS distribution based on the skew-normal distribution. This motivation was not observed when the univariate BS distribution based on the skew-normal distribution was proposed.

Finally, the non-central version of the multivariate BS distribution will be useful to obtain the conditional distribution and it will also facilitate the calculation of the maximum likelihood estimates by the use of an EM-algorithm. The details for the univariate non-central BS case can be found in Guiraud et al. (2009) and Vilca et al. (2011).

3 The multivariate SNBS distribution

In this section, we present the multivariate skew-normal Birnbaum-Saunders (SNBS) distribution by basing it on the multivariate skew-normal (SN) distribution, and then by proceeding along the lines of Kundu et al. (2010) and Vilca et al. (2011), we derive several interesting properties of this distribution.

First, we start with the definition of a multivariate skew-normal distribution. We say that a $p \times 1$ random vector $\mathbf{Z}$ follows a p-variate SN distribution with a $p \times 1$ location vector $\boldsymbol{\mu}$, $p \times p$ positive-definite dispersion matrix $\boldsymbol{\Sigma}$, and a $p \times 1$ skewness parameter vector $\boldsymbol{\lambda}$, if its pdf is given by (Azzalini and Dalla-Valle, 1996)

$$f(\mathbf{z}) = 2\phi_p(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi\big(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\mu})\big), \ \mathbf{z} \in \mathbb{R}^p, \tag{7}$$

where $\phi_p(.; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the pdf of the p–variate normal distribution with mean vector $\boldsymbol{\mu}$, covariate matrix $\boldsymbol{\Sigma}$, and $\Phi(.)$ denotes the cdf of the standard univariate normal distribution. This distribution is usually denoted by $\mathbf{Z} \sim \mathrm{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, and the following stochastic representation holds for $\mathbf{Z}$:

$$\mathbf{Z} \ \stackrel{d}{=} \ \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\{\boldsymbol{\delta}|Y_0| + (\mathbf{I}_p - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{Y}_1\}, \tag{8}$$

where $\boldsymbol{\delta} = \boldsymbol{\lambda}/\sqrt{1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda}}$, $Y_0 \sim \mathrm{N}_1(0, 1)$ and $\mathbf{Y}_1 \sim \mathrm{N}_p(\mathbf{0}, \mathbf{I}_p)$ are independent. It is well-known that marginal distributions of $\mathbf{Z}$ are SN distributions of the form $Z_j \sim \mathrm{SN}(\mu_j, \tau_{jj}, \eta_j)$ for $j = 1, \ldots, p$.

Analogous to the multivariate BS distribution of Kundu et al. (2013) and the univariate SNBS distribution proposed by Vilca et al. (2011) and Vilca et al. (2014), we now introduce an multivariate BS distribution, based on the multivariate skew-normal distribution, through the following stochastic representation for $T_1, \ldots, T_p$:

$$T_j = \frac{\beta_j}{4}\left[\alpha_j\, Z_j + \sqrt{\{\alpha_j\, Z_j\}^2 + 4}\right]^2, \ \text{for } j = 1, \ldots, p, \tag{9}$$

where $\mathbf{Z} = (Z_1, \ldots, Z_p)^\top \sim \mathrm{SN}_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ as in (8), with $\boldsymbol{\Sigma}$ being a $p \times p$ positive-definite matrix with ones at the diagonal. Thus, the random vector $\mathbf{T} = (T_1, \ldots, T_p)^\top$ is said to have a multivariate skew-normal Birnbaum-Saunders (SNBS) distribution with parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\lambda}$, and it is denoted by $\mathbf{T} \sim \mathrm{SNBS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. For $p = 2$, we use the notation $\mathbf{T} \sim \mathrm{SNBS}_2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \rho, \boldsymbol{\lambda})$.

The individual representation of T_j in (9) corresponds to the one in Vilca et al. (2011), whose pdf has the form given in (3). Since each Z_j has the stochastic representation of the form $Z_j = a_j H_0 + b_j W_{0j}$, where H_0 and W_{0j} are independent for $j = 1, \ldots, p$, and have half-normal and normal distributions, respectively. So, we have the following stochastic representation for T_j given in (9), which can therefore be expressed as

$$T_j = \frac{\beta_j}{4}\left[\alpha_j\big(a_j H_0 + b_j W_{0j}\big) + \sqrt{\alpha_j^2\big(a_j H_0 + b_j W_{0j}\big)^2 + 4}\right]^2, \qquad (10)$$

for $j = 1, \ldots, p$. Hence, from Vilca et al. (2011), the conditional distribution of T_j, given $H_{0j} = h_{0j}$, follows a noncentral BS distribution denoted by $\mathrm{NBS}(\alpha_{bj}, \beta_j, \nu_{0j}) = \mathrm{EBS}(\alpha_{bj}, \beta_j, \sigma = 2, -\nu_{0j})$ distribution, where $\alpha_{bj} = \alpha_j b_j$ and $\nu_{0j} = a_j h_{0j}/b_j$ for $j = 1, \ldots, p$. Due to this property, the distribution of $\mathbf{T}$, given $|Y_0| = h$, becomes a multivariate moncentral BS distribution, as we shall see later.

Theorem 1. *Let* $\mathbf{T} \sim SNBS_p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$*. Then, the pdf of* $\mathbf{T}$ *is*

$$f_{\mathbf{T}}(\mathbf{t}) \;=\; 2\phi_p\left(a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}); \boldsymbol{\Sigma}\right)\Phi\big(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2} a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta})\big)\, A_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}),\ \mathbf{t} \in \mathbb{R}_+^p, \qquad (11)$$

where $\phi_p\big(.; \boldsymbol{\Sigma}\big) = \phi_p\big(.; \mathbf{0}, \boldsymbol{\Sigma}\big)$ *is as in (7), and* $a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ *is as (4) and* $A_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{j=1}^p A_{t_j}(\alpha_j, \beta_j)$ *with* $A_{t_j}(\alpha_j, \beta_j)$ *is as in and (5).*

Proof. From the stochastic representation in (9), we can write the joint cdf of $\mathbf{T}$ as

$$F_{\mathbf{T}}(\mathbf{t}) = \Phi_p\big(a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}); \boldsymbol{\Sigma}, \boldsymbol{\lambda}\big),$$

where $\Phi_p\big(\cdot; \boldsymbol{\Sigma}, \boldsymbol{\lambda}\big)$ is the cdf of $\mathbf{Z} \sim \mathrm{SN}_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. The required result then readily follows when the above expression is differentiated with respect to $\mathbf{t}$. ■

Remark 1. *By considering some especial cases or slight generalizations of the multivariate skew-normal distribution, we can be derived others versions of the SNBS distribution:*

i) We emphasize two special cases associated with this new distribution. Firstly, for $\boldsymbol{\lambda} = \mathbf{0}$*, we obtain the multivariate BS distribution of Kundu et al. (2013); see also Kundu et al. (2010) for* $p = 2$*. Secondly, for* $p = 2$ *and* $\boldsymbol{\Sigma} = \mathbf{I}_p$*, the pdf in (11) reduces to*

$$f_{\mathbf{T}}(\mathbf{t}) \;=\; 2\phi\left(a_{t_1}(\alpha_1, \beta_1)\right)\phi\left(a_{t_2}(\alpha_2, \beta_2)\right)\Phi\big(\lambda_1 a_{t_1}(\alpha_1, \beta_1) + \lambda_2\, a_{t_2}(\alpha_2, \beta_2)\big)\, A_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}),$$

which is a bivariate BS-type distribution based on the bivariate skew-normal distribution obtained using the conditional specification proposed by Azzalini and Dalla-Valle (1996) and Arnold et al. (2002) **ADICIONAR** *ii) Now, letting $\boldsymbol{\lambda} = \boldsymbol{\delta}/\sqrt{1-\boldsymbol{\delta}^\top\boldsymbol{\delta}}$, with $\delta_1 = \delta_2 = \alpha$, we observe that the above pdf can be written as*

$$f_{\mathbf{T}}(\mathbf{t}) = 2\phi\left(a_{t_1}(\alpha_1,\beta_1)\right)\phi\left(a_{t_2}(\alpha_2,\beta_2)\right)\Phi\Big(\frac{\alpha}{\sqrt{1-2\alpha^2}}\{a_{t_1}(\alpha_1,\beta_1)+a_{t_2}(\alpha_2,\beta_2)\}\Big)\,A_{\mathbf{t}}(\boldsymbol{\alpha},\boldsymbol{\beta})$$

which is a bivariate SNBS distribution based on a special case of the bivariate skew-normal of Azzalini and Dalla-Valle (1996) and Sahu et al. (2003, page 134).

(a) (b)

(c) (d)

Figure 1: The joint pdf's of (T_1, T_2) when $\alpha_1 = \alpha_2 = 0.5, \beta_1 = \beta_2 = 1, \rho = 0.5$ and (a) $\lambda_1 = \lambda_2 = 0$, (b) $\lambda_1 = \lambda_2 = 4$, (c) $\lambda_1 = \lambda_2 = -4$, (d) $\lambda_1 = 4$ and $\lambda_2 = -4$.

Before presenting some properties of the bivariate SNBS distribution, we discuss the shape of its density. Figures 1 and 2 display joint bivariate SNBS densities and their contours for the two cases i) $\rho = 0.5$ and ii) $\rho = -0.5$, and in both cases $\alpha_1 = \alpha_2 = 0.5, \beta_1 = \beta_2 = 1$. Moreover, we have made the following choices for λ_1 and λ_2: (a) $\lambda_1 = \lambda_2 = 0$, (b) $\lambda_1 = \lambda_2 = 4$, (c) $\lambda_1 = \lambda_2 = -4$, (d) $\lambda_1 = 4$ and $\lambda_2 = -4$. From these figures, it can be noted that for λ_1 and λ_2 positive (see Figure 1 (b)), the bivariate SNBS distribution has heavier tails than the bivariate BS distribution toward large values of

T_1 and T_2. On the other hand, for λ_1 or λ_2 negative, the resulting distribution may be suitable for fitting small values (close to zeros) of T_1 or T_2, respectively. For example, the distribution in Figure 2 (d) fits small values of T_2 very well.

(a) (b)

(c) (d)

Figure 2: The joint pdf's of (T_1, T_2) when $\alpha_1 = \alpha_2 = 0.5, \beta_1 = \beta_2 = 1, \rho = -0.5$ and (a) $\lambda_1 = \lambda_2 = 0$, (b) $\lambda_1 = \lambda_2 = 4$, (c) $\lambda_1 = \lambda_2 = -4$, (d) $\lambda_1 = 4$ and $\lambda_2 = -4$.
st;snbs2

3.1 Properties of the multivariate SNBS distribution

In this section, we discuss some properties of the multivariate SBS distribution. According to (9), we have the distribution of $\mathbf{T}$ is directly related to $\mathbf{Z} \sim \mathrm{SN}_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. This relationship is extremely useful to derive properties of the multivariate SNBS distribution as will see in the following theorems. Let $\mathbf{T} = (\mathbf{T}_1^\top, \mathbf{T}_2^\top)^\top \sim \mathrm{SNBS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ be such that $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top)^\top$,

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \ \boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \end{pmatrix}, \ \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \text{ and } \boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & -\boldsymbol{\Sigma}_{12} \\ -\boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \qquad (12)$$

where $\mathbf{T}_1$, $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ are all $q \times 1$ vectors and $\boldsymbol{\Sigma}_{11}$ is $q \times q$ correlation matrix, with the remaining elements all defined suitably so that the corresponding orders match. As in the multivariate SN distribution, the conditional distribution of $\mathbf{T}_1$, given $\mathbf{T}_2 = \mathbf{t}_2$, also depends on the matrix $\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ that is not a correlation matrix. Hence, the conditional distribution of $\mathbf{T}_1$, given $\mathbf{T}_2 = \mathbf{t}_2$, will not be closed under conditioning, even though for the case when $p = 2$, the conditional distribution is the univariate non-central BS distribution, as shown below.

Theorem 2. *Let* $\mathbf{T} \sim \mathrm{SNBS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. *Then:*

(i) $\mathbf{c} \odot \mathbf{T} \sim \mathrm{SNBS}_p(\boldsymbol{\alpha}, \mathbf{c} \odot \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, *where* $\mathbf{c} = (c_1, \ldots, c_p)^\top \in \mathbb{R}_+^p$ *and* $\odot$ *denotes the Hadamard product;*

(ii) $\mathbf{T}^{-1} = (T_1^{-1}, \ldots, T_p^{-1})^\top \sim \mathrm{SNBS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}^{-1}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, *where* $\boldsymbol{\beta}^{-1} = (1/\beta_1, \ldots, 1/\beta_p)^\top$*;*

(iii) $(\mathbf{T}_1^{-1}, \mathbf{T}_2)^\top \sim \mathrm{SNBS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}_{12}^{-1}, \boldsymbol{\Psi}, \boldsymbol{\lambda}_{12})$, *where* $\boldsymbol{\beta}_{12}^{-1} = (\boldsymbol{\beta}_1^{-1}, \boldsymbol{\beta}_2)^\top$ *and* $\boldsymbol{\lambda}_{12} = (-\boldsymbol{\lambda}_1^\top, \boldsymbol{\lambda}_2^\top)^\top$, *with* $\boldsymbol{\beta}_2$ *and* $\boldsymbol{\Psi}$ *are as in (12) and* $\boldsymbol{\beta}_1^{-1} = (1/\beta_1, \ldots, 1/\beta_q)^\top$*;*

(iv) $(\mathbf{T}_1, \mathbf{T}_2^{-1})^\top \sim \mathrm{SNBS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}_{21}^{-1}, \boldsymbol{\Psi}, \boldsymbol{\lambda}_{21})$ *where* $\boldsymbol{\beta}_{21}^{-1} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2^{-1})^\top$ *and* $\boldsymbol{\lambda}_{21} = (\boldsymbol{\lambda}_1^\top, -\boldsymbol{\lambda}_2^\top)^\top$ *, with* $\boldsymbol{\beta}_1$ *and* $\boldsymbol{\Psi}$ *are as in (12) and* $\boldsymbol{\beta}_2^{-1} = (1/\beta_{q+1}, \ldots, 1/\beta_p)^\top$*;*

v) *Let* $\boldsymbol{V} = (T_1^{\eta_1}, \ldots, T_p^{\eta_p})$ *be a transformation of* $\mathbf{T}$, *with* $\eta_j > 0$, $j = 1, \ldots, p$. *Then the pdf of* $\boldsymbol{V}$ *is given by*

$$f_{\mathbf{V}}(\mathbf{v}) = 2\phi_p\big(b_{\mathbf{v}}(\boldsymbol{\alpha}, \boldsymbol{\beta}); \boldsymbol{\Sigma}\big)\Phi\big(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2} b_{\mathbf{v}}(\boldsymbol{\alpha}, \boldsymbol{\beta})\big)\, B_{\mathbf{v}}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \quad \mathbf{v} \in \mathbb{R}_+^p,$$

where $b_{\mathbf{v}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (b_{v_1}, \ldots, b_{v_p})^\top$ *and* $B_{\mathbf{v}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \Pi_{i=1}^p B_{v_i}$, *with*

$$b_{v_j} = \frac{1}{\alpha_j}\left(\left[\frac{v_j}{\delta_j}\right]^{\frac{1}{\sigma_j}} - \left[\frac{\delta_j}{v_j}\right]^{\frac{1}{\sigma_j}}\right) \text{ and } B_{v_j} = \frac{1}{\alpha_j\, \sigma_j\, v_j}\left(\left[\frac{v_j}{\delta_j}\right]^{\frac{1}{\sigma_j}} + \left[\frac{\delta_j}{v_j}\right]^{\frac{1}{\sigma_j}}\right),$$

$\delta_j = \beta_j^{\eta_j}$ *and* $\sigma_j = 2\eta_j > 0$, $j = 1, \ldots, p$

Proof. Parts (i)-(iv) are directly obtained from the change-of-variable method and then by following the lines of Kundu et al. (2010) and Vilca et al. (2014). Part (v) is obtained by following the same lines as in Vilca et al. (2014). ■

Remark 2. *Some conclusions that can be readily obtained from Theorem 2 are:*

i) Part (i) states that the SNBS distribution belongs to a scale type family, and it preserves the same property of the usual univariate BS distribution of Birnbaum and Saunders (1969). Parts (ii)-(iv) state that these distributions are closed under reciprocation for both components or for one of them, similar to the bivariate case discussed in Vilca et al. (2014);

ii) *From Theorem 2 (i), we can obtain a re-parameterized form of the SNBS distribution. Consider* $\mathbf{c} = (\alpha_1/\beta_1, \dots, \alpha_p/\beta_p)^\top$, *then* $\mathbf{c} \odot \mathbf{T} \sim \mathrm{SNBS}_p(\boldsymbol{\alpha}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, *and for* $\mathbf{c} = (1/\beta_1, \dots, 1/\beta_p)^\top$, $\mathbf{c} \odot \mathbf{T} \sim \mathrm{SBS}_p(\boldsymbol{\alpha}, \boldsymbol{1}_p, \boldsymbol{\Sigma}; H)$, *where* $\boldsymbol{1}_p$ *is the p-dimensional vector with all elements as one. Hence, the distribution of any function of* T_j/β_j *and* T_k/β_k *not depend on* β_j *and* β_k, *as well as the random variables* $T_j^a T_k^b/(\beta_j^a \beta_k^b)$, *with* a *and* b *in* $\mathbb{R}$;

iii) *For the bivariate case,* $\mathbf{T} \sim \mathrm{SNBS}_2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \rho, \boldsymbol{\lambda})$, *we have* $\mathbf{T}_1^{-1} = (T_1^{-1}, T_2)^\top \sim \mathrm{SNBS}_2(\boldsymbol{\alpha}, \boldsymbol{\beta}_1^{-1}, -\rho, \boldsymbol{\lambda}_1)$, *where* $\boldsymbol{\beta}_1^{-1} = (1/\beta_1, \beta_2)^\top$ *and* $\boldsymbol{\lambda}_1 = (-\lambda_1, \lambda_2)^\top$.

On the moments of the multivariate SNBS distribution are complicated to be obtained, a effort for $p = 2$, is given in the following theorem, which are obtained by using the relationship between $\mathbf{T}$ and $\mathbf{Z}$ given in (9).

Theorem 3. *Let* $\mathbf{T} = (T_1, T_2)^\top \sim \mathrm{SNBS}_2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \rho, \boldsymbol{\lambda})$. *Then, the elements of the mean vector and the covariance matrix are as follows:*

$$
\begin{aligned}
\mathbb{E}[T_j] &= \frac{\beta_j}{2}(2 + \alpha_j^2) + \frac{\alpha_j \beta_j}{2} w_{1j}, \\
\mathrm{Var}[T_j] &= \frac{\beta_j^2 \alpha_j^2}{4}\Big(4 + 5\alpha_j^2\Big) + \frac{\alpha_j^2 \beta_j^2}{4}\alpha_{wj}, \\
\mathbb{E}\Big[\frac{T_1 T_2}{\beta_1 \beta_2}\Big] &= 1 + \mathrm{tr}\big(D^2(\boldsymbol{\alpha})\boldsymbol{\Sigma}\big) + \frac{\alpha_1^2 \alpha_2^2}{16}\Big\{2\mathrm{tr}\big[(\boldsymbol{A}\boldsymbol{\Sigma})^2\big] + \big(\mathrm{tr}\big(\boldsymbol{A}\boldsymbol{\Sigma}\big)\big)^2\Big\} + \frac{1}{2}\big(\alpha_1 w_{11} + \alpha_2 w_{12}\big) \\
&\quad + \frac{\alpha_1 \alpha_2}{4} c(\boldsymbol{\alpha}),
\end{aligned}
$$

where $\alpha_{wj} = 2\alpha_j(w_{3j} - w_{1j}) - w_{1j}^2$, $w_{kj} = \mathbb{E}\big[Z_j^k \sqrt{\alpha_j^2 Z_j^2 + 4}\big]$, $k = 1, 3$, $j = 1, 2$, *and* $c(\boldsymbol{\alpha}) = \alpha_2 \mathbb{E}[Z_2^2 G_1] + \alpha_1 \mathbb{E}[Z_1^2 G_2] + \mathbb{E}[G_1 G_2]$, *with* $G_j = Z_j \sqrt{\alpha_j^2 Z_j^2 + 4}$ *and* $\mathbf{Z} = (Z_1, Z_2)^\top \sim \mathrm{SN}_2(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$.

The following theorem presents the marginal and conditional distributions of the bivariate SNBS distribution.

Theorem 4. *Let* $\mathbf{T} = (T_1, \dots, T_p)^\top \sim \mathrm{SNBS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, *with* $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top)^\top$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ *and* $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \boldsymbol{\gamma}_2^\top)^\top = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}$. *Then:*

(i) $\mathbf{T}_1 \sim \mathrm{SNBS}_p(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1, \boldsymbol{\eta}_1)$ *and* $\mathbf{T}_2 \sim \mathrm{SNBS}_q(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2, \boldsymbol{\eta}_2)$ *(Vilca et al., 2011), where*

$$
\boldsymbol{\eta}_1 = \boldsymbol{\Sigma}_{11}^{1/2}\Big(\frac{\boldsymbol{\gamma}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}_2}{\sqrt{1 + \boldsymbol{\gamma}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\gamma}_2}}\Big) \text{ and } \boldsymbol{\eta}_2 = \boldsymbol{\Sigma}_{22}^{1/2}\Big(\frac{\boldsymbol{\gamma}_2 + \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\gamma}_1}{\sqrt{1 + \boldsymbol{\gamma}_1^\top \boldsymbol{\Sigma}_{11.2}\boldsymbol{\gamma}_1}}\Big);
$$

(ii) *The conditional pdf of* $\mathbf{T}_2$*, given* $\mathbf{T}_1 = \mathbf{t}_1$*, is given by*

$$f_{\mathbf{T}_2|\mathbf{T}_1}(\mathbf{t}_2|\mathbf{t}_1) = \phi_q\Big(a_{\mathbf{t}_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2); \boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}\Big) \frac{\Phi\big(\boldsymbol{\gamma}^\top a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta})\big)}{\Phi\big(\widetilde{\boldsymbol{\gamma}}^\top a_{\mathbf{t}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1)\big)} \, \Pi A_{\mathbf{t}_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2),$$

where $\boldsymbol{\mu}_{2.1} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} a_{\mathbf{t}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1)$, $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ *and* $\widetilde{\boldsymbol{\gamma}} = \frac{\boldsymbol{\gamma}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}_2}{\sqrt{1+\boldsymbol{\gamma}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\gamma}_2}}$;

(iii) *The conditional pdf of* $\mathbf{T}_2$*, given* $\mathbf{T}_1 = \boldsymbol{\beta}_1$*, is given by*

$$f_{\mathbf{T}_2|\mathbf{T}_1}(\mathbf{t}_2|\boldsymbol{\beta}_1) = 2\phi_q\big(a_{\mathbf{t}_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2); \boldsymbol{\Sigma}_{22.1}\big)\Phi\big(\boldsymbol{\gamma}_2^\top a_{\mathbf{t}_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2)\big)\, \Pi A_{\mathbf{t}_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2).$$

Proof. Part (i) is obtained directly by using the stochastic representation in (9) and the marginal distribution of $\text{SN}_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Parts (ii) and (iii) are derived as a by-product of the conditional distribution of the multivariate $\text{SN}_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ distribution and by following the same steps as in Kundu et al. (2010). ■

Corollary 1. *Let* $\mathbf{T} = (T_1, T_2)^\top \sim \text{SNBS}_2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \rho, \boldsymbol{\lambda})$*, with* $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top$*,* $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$ *and* $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^\top = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}$*. Then:*

(i) $T_1 \sim \text{SNBS}(\alpha_1, \beta_1, \eta_1)$ *and* $T_2 \sim \text{SNBS}(\alpha_2, \beta_2, \eta_2)$ *(Vilca et al., 2011), where*

$$\eta_1 = \frac{\gamma_1 + \rho\gamma_2}{\sqrt{1 + \gamma_2^2(1-\rho^2)}} \text{ and } \eta_2 = \frac{\gamma_2 + \rho\gamma_1}{\sqrt{1 + \gamma_1^2(1-\rho^2)}};$$

(ii) *The conditional pdf of* T_2*, given* $T_1 = t_1$*, is given by*

$$f_{T_2|T_1}(t_2|t_1) = \phi\Big(a_{t_2}(\alpha_{2\rho}, \beta_2) - \mu_{2.1\rho}\Big) \frac{\Phi\big(\boldsymbol{\gamma}^\top a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta})\big)}{\Phi\big(\eta_1 a_{t_1}(\alpha_1, \beta_1)\big)} \, A_{t_2}(\alpha_{2\rho}, \beta_2),$$

where $\mu_{2.1\rho} = \rho a_{t_1}(\alpha_{1\rho}, \beta_1)$*;*

(iii) *The conditional pdf of* T_2*, given* $T_1 = \beta_1$*, is given by*

$$f_{T_2|T_1}(t_2|\beta_1) = 2\phi\big(a_{t_2}(\alpha_{2\rho}, \beta_2)\big)\Phi\big(\gamma_2 a_{t_2}(\alpha_2, \beta_2)\big)\, A_{t_2}(\alpha_{2\rho}, \beta_2).$$

Figure 3 displays plots of the marginal pdf's of T_1 and T_2 following the results in Part (i) of Corollary 1 when $\alpha_1 = \alpha_2 = 0.5, \beta_1 = \beta_2 = 1, \rho = 0.5$ for Cases (a) $\lambda_1 = \lambda_2 = 4$, and (b) $\lambda_1 = \lambda_2 = -4$.

(a) (b)

Figure 3: The marginal pdf's of T_1 and T_2 when (a) $\lambda_1 = \lambda_2 = 4$, (b) $\lambda_1 = \lambda_2 = -4$ and density of BS distribution for the indicated parameters (solid line).

4 Maximum likelihood estimation

In this section, we discuss the ML estimation of the parameters of the SNBS distribution by using an EM algorithm and also derive the observed information matrix associated with this distribution. Before discussing the EM algorithm, we present a non-central multivariate BS distribution of Kundu et al. (2010, 2013) that is useful to obtain the maximum likelihood estimates via the EM algorithm.

A slight extension of the multivariate BS distribution is obtained by following the idea of Guiraud et al. (2009), and considering the comments in (i) and (ii) in Section 1. Specifically, we consider the multivariate $\mathrm{N}_p(\boldsymbol{\nu}, \boldsymbol{\Sigma})$ distribution in place of $\mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ distribution, where $\boldsymbol{\nu}$ is the $p \times 1$ mean vector and $\boldsymbol{\Sigma}$ is any $p \times p$ positive definite matrix. Then, the joint cumulative distribution function (cdf) of $T_1, \ldots, T_p$ is given by

$$F_{\mathbf{T}}(\mathbf{t}) = \Phi_p\big(a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \boldsymbol{\nu}; \mathbf{0}, \boldsymbol{\Sigma}\big),$$

where $\Phi_p(.,.; \mathbf{0}, \boldsymbol{\Sigma}\big)$ is the cdf of $\mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ distribution. In this case, the p-variate random vector $\mathbf{T} = (T_1, \ldots, T_p)^\top$ is said to have a multivariate non-central Birnbaum-Saunders (NBS) distribution with parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\nu}$ and $\boldsymbol{\Sigma}$, and is denoted by $\mathbf{T} \sim \mathrm{NBS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\Sigma})$.

Next, we present the joint pdf of $\mathbf{T}$, which can be derived along the same lines as in Kundu et al. (2010, 2013) and Vilca et al. (2014):

$$f_{\mathbf{T}}(\mathbf{t}) \;\; = \;\; \phi_p\left(a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \boldsymbol{\nu}; \mathbf{0}, \boldsymbol{\Sigma}\right) A_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}),\ \mathbf{t} \in \mathbb{R}^p_+, \tag{13}$$

where $\phi_2(.,.; \mathbf{0}, \boldsymbol{\Sigma})$ is the pdf of $\mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ distribution, and $a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $A_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ are as given in (4) and (5), respectively. This distribution presents a slight generalization of the multivariate BS distribution: (i) This distribution is equivalent to $a_{\mathbf{T}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \sim N_p(\boldsymbol{\nu}, \boldsymbol{\Sigma})$; (ii) When $\boldsymbol{\nu} = \mathbf{0}$, the multivariate SNBS distribution reduces to the multivariate BS distribution of Kundu et al. (2013), with $(\alpha_1\sigma_{11}^{1/2}, \ldots, \alpha_p\sigma_{pp}^{1/2})^\top$

in place of $(\alpha_1, \ldots, \alpha_p)^\top$; (iii) When $\sigma_{11} = \cdots = \sigma_{pp} = 1$ in $\mathbf{\Sigma}$, the multivariate NBS distribution has its marginal and conditional distributions in the same class of distributions, as we shall see in the following theorems that present some general properties of the multivariate NBS distribution, which are similar to those of the bivariate BS distribution presented by Kundu et al. (2010) and Vilca et al. (2014). The following properties can be derived easily along the same lines as in the bivariate BS distribution of Kundu et al. (2010) and Romero et al. (2018).

Theorem 5. *Let* $\mathbf{T} = (\mathbf{T}_1^\top, \mathbf{T}_2^\top)^\top \sim \mathrm{NBS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\nu}, \mathbf{\Sigma})$ *be such that* $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ *and* $\mathbf{\Sigma}$ *are partitioned as in (12) and* $\boldsymbol{\nu} = (\boldsymbol{\nu}_1^\top, \boldsymbol{\nu}_2^\top)^\top$*. Then:*

(i) $\mathbf{T}_1 \sim \mathrm{NBS}_q(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1, \boldsymbol{\nu}_1, \mathbf{\Sigma}_{11})$;

(ii) $\mathbf{c} \odot \mathbf{T} \sim \mathrm{NBS}_p(\boldsymbol{\alpha}, \mathbf{c} \odot \boldsymbol{\beta}, \boldsymbol{\nu}, \mathbf{\Sigma})$, *where* $\mathbf{c} \in \mathbb{R}_+^p$;

(iii) $\mathbf{T}^{-1} = (\mathbf{T}_1^{-1}, \mathbf{T}_2^{-1})^\top \sim \mathrm{NBS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}^{-1}, -\boldsymbol{\nu}, \mathbf{\Sigma})$, *where* $\boldsymbol{\beta}^{-1}$ *is as Theorem 3;*

(iv) $\mathbf{T}_1^{-1} = (T_1^{-1}, T_2)^\top \sim \mathrm{NBS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}_1^{-1}, \boldsymbol{\nu}_1, \mathbf{\Psi})$, *where* $\boldsymbol{\beta}_1^{-1}$ *is as Theorem 3,* $\mathbf{\Psi}$ *as in (12) and* $\boldsymbol{\nu}_1 = (-\boldsymbol{\nu}_1^\top, \boldsymbol{\nu}_2^\top)^\top$*.*

Before discussing the steps of the EM algorithm, we present the following theorem that shows the conditional distribution of $\mathbf{T}$, given $|Y_0| = h$, involves the bivariate Non-central BS distribution detailed in Section 2, where an unrestricted matrix $\mathbf{\Sigma}$ has been considered. This result becomes useful in the implementation of the algorithm.

Theorem 6. *Let* $\mathbf{T} \sim \mathrm{SNBS}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{\Sigma}, \boldsymbol{\lambda})$ *with its stochastic representation as in (10), where* $H = |Y_0|$*. Then:*

(i) *The conditional distribution of* $\mathbf{T}$*, given* $H = h$*, follows a multivariate non-central BS distribution, viz.,* $\mathbf{T}|(H = h) \sim \mathrm{NBS}_p\big(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\nu}_h, \mathbf{\Sigma} - \mathbf{\Delta}\mathbf{\Delta}^\top\big)$*, where* $\boldsymbol{\nu}_h = \mathbf{\Delta} h$ *and* $\mathbf{\Delta} = \mathbf{\Sigma}^{1/2}\boldsymbol{\delta}$*, with* $\boldsymbol{\delta} = \boldsymbol{\lambda}/\sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$*. Moreover, its pdf is*

$$f_{\mathbf{T}|H}(\mathbf{t}|h) \;=\; \phi_p\left(a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathbf{\Delta} h; \mathbf{0}, \mathbf{\Sigma} - \mathbf{\Delta}\mathbf{\Delta}^\top\right) A_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \;\; \mathbf{t} \in \mathbb{R}_+^p; \qquad (14)$$

(ii) *The conditional distribution of* H*, given* $\mathbf{T} = \mathbf{t}$*, follows a truncated normal distribution with pdf*

$$f_{H/\mathbf{T}}(h/\mathbf{t}) = \frac{\phi\big(h; \mathbf{\Delta}^\top\mathbf{\Sigma}^{-1} a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}), 1 - \mathbf{\Delta}^\top\mathbf{\Sigma}^{-1}\mathbf{\Delta}\big)}{\Phi\big(\boldsymbol{\lambda}^\top\mathbf{\Sigma}^{-1/2} a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta})\big)}, \; h > 0, \qquad (15)$$

where $\phi(\cdot; \mu, \sigma^2)$ *denotes the normal pdf with mean* μ *and variance* σ^2*. Moreover,*

$$\mathbb{E}[H|(\mathbf{T} = \mathbf{t})] = \mathbf{\Delta}^\top\mathbf{\Sigma}^{-1} a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \sqrt{1 - \mathbf{\Delta}^\top\mathbf{\Sigma}^{-1}\mathbf{\Delta}}\, W_\Phi\left(\frac{\mathbf{\Delta}^\top\mathbf{\Sigma}^{-1} a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\sqrt{1 - \mathbf{\Delta}^\top\mathbf{\Sigma}^{-1}\mathbf{\Delta}}}\right)$$

and

$$\begin{aligned}\mathbb{E}[H^2|(\mathbf{T}=\mathbf{t})] &= (\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\, a_{\boldsymbol{t}}(\boldsymbol{\alpha},\boldsymbol{\beta}))^2 + (1-\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}) \\ &+ W_\Phi\left(\frac{\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\, a_{\boldsymbol{t}}(\boldsymbol{\alpha},\boldsymbol{\beta})}{\sqrt{1-\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}}}\right)\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\, a_{\boldsymbol{t}}(\boldsymbol{\alpha},\boldsymbol{\beta})\sqrt{1-\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}},\end{aligned}$$

where $W_\Phi(u)=\phi(u)/\Phi(u)$, *for* $u\in\mathbb{R}$ *and* $\boldsymbol{t}\in\mathbb{R}_+^2$.

Proof. For Part (i), we consider (8) and (9), and the following relationship between the components of $\mathbf{T}$ and $\mathbf{Z}$:

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_p \end{pmatrix} = \begin{pmatrix} \frac{1}{\alpha_1}\left(\sqrt{\frac{T_1}{\beta_1}}-\sqrt{\frac{\beta_1}{T_1}}\right) \\ \vdots \\ \frac{1}{\alpha_p}\left(\sqrt{\frac{T_p}{\beta_p}}-\sqrt{\frac{\beta_p}{T_p}}\right) \end{pmatrix} = \boldsymbol{\Sigma}^{1/2}\{\boldsymbol{\delta}|Y_0| + (\mathbf{I}_p-\boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{Y}_1\}. \tag{16}$$

Thus, the first member of the above equation, given $|Y_0|=h$, follows $N_p(\boldsymbol{\Delta}h,\boldsymbol{\Sigma}-\boldsymbol{\Delta}\boldsymbol{\Delta}^\top)$ distribution, where $\boldsymbol{\Delta}=\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$. This normal distribution is considered in the non-central BS distribution. So, the rest of the proof is analogous to that of Theorem 5 on the bivariate non-central BS distribution. Part (ii) is proved simply by using the properties of conditional distributions. In fact, the resulting distribution is the well-known half-normal distribution, $\mathbf{T}|(H=h)\sim HN(\eta,\tau^2)$, where $\eta=\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\, a_{\mathbf{t}}(\boldsymbol{\alpha},\boldsymbol{\beta})$ and $\tau^2=1-\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}$. Finally, the conditional moments $\mathbb{E}[H|(\mathbf{T}=\mathbf{t})]$ and $\mathbb{E}[H^2|(\mathbf{T}=\mathbf{t})]$ follow from the moments of the half-normal distribution. Hence, the theorem. ■

4.1 The EM algorithm

Now, based on the results established above, we develop the steps of the EM algorithm due to Meng and Rubin (1993), which is a well-known technique for the ML estimation when unobserved (or missing) data or latent variables are present while modeling. This algorithm facilitates computationally efficient determination of the ML estimates when iterative methods are required. Let $\mathbf{T}_1,\ldots,\mathbf{T}_n$ be a random sample of size n, where $\mathbf{T}_i\sim\text{SNBS}_p(\boldsymbol{\alpha},\boldsymbol{\beta},\rho,\boldsymbol{\lambda})$, for $i=1,\ldots,n$. Then, the log-likelihood function for $\boldsymbol{\theta}=(\boldsymbol{\alpha}^\top,\boldsymbol{\beta}^\top,\boldsymbol{\lambda}^\top)^\top$ is given by

$$\ell(\boldsymbol{\theta})=\sum_{i=1}^{n}\big(\ell_{i\phi}(\boldsymbol{\theta})+\ell_{i\Phi}(\boldsymbol{\theta})\big), \tag{17}$$

where $\ell_{i\phi}(\boldsymbol{\theta})=-\frac{p}{2}\log(2\pi)-(1/2)\log(|\boldsymbol{\Sigma}|)+\log(A_i)-\frac{1}{2}d(\mathbf{t}_i)$ and $\ell_{i\Phi}(\boldsymbol{\theta})=\log(2)+\log\big(\Phi(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}a(\mathbf{t}_i;\boldsymbol{\alpha},\boldsymbol{\beta}))\big)$, with $a(\mathbf{t}_i;\boldsymbol{\alpha},\boldsymbol{\beta})$ and $A_i=A(\mathbf{t}_i;\boldsymbol{\alpha},\boldsymbol{\beta})=\Pi_{j=1}^{p}A(t_i;\alpha_j,\beta_j)$ as given in (5), and $d(\mathbf{t}_i)=a_{\mathbf{t}_i}^\top(\boldsymbol{\alpha},\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}a_{\mathbf{t}_i}(\boldsymbol{\alpha},\boldsymbol{\beta})$. We now develop the EM-algorithm by using the one developed in Vilca et al. (2011) for the univariate skew-normal BS distribution. Specifically, let $\mathbf{t}_o = [\mathbf{t}_1, \ldots, \mathbf{t}_n]^\top$ and $\mathbf{h} = [h_1, \ldots, h_n]^\top$ be observed and unobserved data, respectively. The complete data $\mathbf{t}_c = [\mathbf{t}_o^\top, \mathbf{h}^\top]^\top$ corresponds to the original observed data $\mathbf{t}_o$ augmented with $\mathbf{h}$. We now describe the implementation of the EM-algorithm for the ML estimation of the parameters of the bivariate SNBS distribution. Notice that from Theorem 6, we can write

$$\mathbf{T}_i|(H_i = h_i) \overset{\text{ind}}{\sim} \sim \text{NBS}_p\big(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\nu}_{h_i}, \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top\big), \tag{18}$$

$$H_i \overset{\text{ind}}{\sim} \text{HN}(0,1), \quad i = 1, \ldots, n, \tag{19}$$

where $\boldsymbol{\nu}_{h_i} = \boldsymbol{\Delta} h_i$ and $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$, with $\boldsymbol{\delta} = \boldsymbol{\lambda}/\sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$. Then, under the hierarchical representation given in (18) and (19), it follows that the complete log-likelihood function associated with $\mathbf{t}_c = [\mathbf{t}_o^\top, \mathbf{h}^\top]^\top$ is expressed as $\ell_c(\boldsymbol{\theta}|\mathbf{t}_c) = \sum_{i=1}^n \ell_c(\boldsymbol{\theta}|\mathbf{t}_i, h_i)$, where

$$\ell_c(\boldsymbol{\theta}|\mathbf{t}_i, h_i) = -\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Psi}|) + \log\left(A(\mathbf{t}_i; \boldsymbol{\alpha}, \boldsymbol{\beta})\right) - \frac{1}{2} d(\mathbf{t}_i, h_i, \boldsymbol{\theta}), \tag{20}$$

with $\boldsymbol{\Psi} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$ and $d(\mathbf{t}_i, h_i, \boldsymbol{\theta}) = \big(a_{\mathbf{t}_i}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \boldsymbol{\Delta} h_i\big)^\top \Psi^{-1}\big(a_{\mathbf{t}_i}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \boldsymbol{\Delta} h_i\big)$.

Let $\widehat{\boldsymbol{\theta}}^{(k)} = [\widehat{\boldsymbol{\alpha}}^{(k)}, \widehat{\boldsymbol{\beta}}^{(k)}, \widehat{\boldsymbol{\lambda}}^{(k)}]^\top$ be the ML estimate of $\boldsymbol{\theta}$ at the k-th iteration, $\widehat{h}_i = \text{E}[H_i|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{t}_i]$ and $\widehat{h_i^2} = \text{E}[H_i^2|\widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{t}_i]$. Then, by using known properties of conditional expectation and Theorem 6, we obtain

$$\widehat{h}_i = \boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1} a_{\mathbf{t}_i}(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \sqrt{1 - \boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}}\, W_\Phi\left(\frac{\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1} a_{\mathbf{t}_i}(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\sqrt{1 - \boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}}}\right), \tag{21}$$

$$\begin{aligned}\widehat{h_i^2} &= (\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1} a_{\mathbf{t}_i}(\boldsymbol{\alpha}, \boldsymbol{\beta}))^2 + (1 - \boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}) \\ &+ W_\Phi\left(\frac{\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1} a_{\mathbf{t}_i}(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\sqrt{1 - \boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}}}\right)\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1} a_{\mathbf{t}_i}(\boldsymbol{\alpha}, \boldsymbol{\beta})\sqrt{1 - \boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}}.\end{aligned} \tag{22}$$

After some algebraic simplification and, using (18) and (19) once again, it follows that the conditional expectation of the complete-data log-likelihood function has the form $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$, where

$$\begin{aligned}Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) &= -\frac{p}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Psi}|) + \log\left(A(\mathbf{t}_i; \boldsymbol{\alpha}, \boldsymbol{\beta})\right) \\ &\quad -\frac{1}{2}\big(a_{\mathbf{t}_i}(\boldsymbol{\alpha}, \boldsymbol{\beta})^\top \Psi^{-1} a_{\mathbf{t}_i}(\boldsymbol{\alpha}, \boldsymbol{\beta}) - 2\widehat{h}_i a_{\mathbf{t}_i}(\boldsymbol{\alpha}, \boldsymbol{\beta})^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\Delta} + \widehat{h^2}_i \boldsymbol{\Delta}^\top\boldsymbol{\Psi}^{-1}\boldsymbol{\Delta}\big).\end{aligned}$$

We now propose the following EM-type algorithm:
E-step. Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$, compute $\widehat{h}_i^{(k)}$ and $\widehat{h_i^2}^{(k)}$, for $i = 1, \ldots, n$, using (21) and (22);

CM-step 1. Fix $\widehat{\boldsymbol{\beta}}^{(k)}$ and then update $\widehat{\boldsymbol{\alpha}}^{(k)} = (\widehat{\alpha}_1^{(k)}, \ldots, \widehat{\alpha}_p^{(k)})^\top$ and $\widehat{\boldsymbol{\Delta}}^{(k)} = (\widehat{\Delta}_1^{(k)}, \ldots, \widehat{\Delta}_p^{(k)})^\top$ by

$$\begin{aligned}
\widehat{\alpha}_j^{2(k+1)} &= \frac{1}{n}\sum_{i=1}^{n} a_{t_{ji}}^2(1, \widehat{\beta}_j^{(k)}) + (1 - \overline{\widehat{h}^2})\left(\frac{\sum_{i=1}^n \widehat{h}_i^{(k)}\, a_{t_{ji}}(1, \widehat{\beta}_j^{(k)})}{n\overline{\widehat{h}^2}}\right)^2, \; j = 1, \ldots, p, \\
\widehat{\boldsymbol{\Psi}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n} \Big\{ a_{\mathbf{t}_i}(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}) a_{\mathbf{t}_i}(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)})^\top - \widehat{h}_i a_{\mathbf{t}_i}(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}) \boldsymbol{\Delta}^{\top(k)} \\
&\quad - \widehat{h}_i \boldsymbol{\Delta} a_{\mathbf{t}_i}(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)})^\top + \widehat{h}^2{}_i \boldsymbol{\Delta}^{(k)} \boldsymbol{\Delta}^{\top(k)} \Big\}, \\
\widehat{\Delta}_j^{(k+1)} &= \frac{1}{\widehat{\alpha}_j^{(k)}} \frac{\sum_{i=1}^n \widehat{h}_i^{(k)} a_{t_{ji}}(1, \widehat{\beta}_j^{(k)})}{n\overline{\widehat{h}^2}}, \; j = 1, \ldots, p,
\end{aligned}$$

where $\overline{\widehat{h}^2} = \sum_{i=1}^n \widehat{h}^2{}_i^{(k)}/n$;

CM-step 2. Fix $\widehat{\boldsymbol{\alpha}}^{(k+1)}$, $\widehat{\boldsymbol{\Delta}}^{(k+1)}$ and $\widehat{\boldsymbol{\Psi}}^{(k+1)}$, and update $\widehat{\boldsymbol{\beta}}^{(k)} = (\widehat{\beta}_1^{(k)}, \ldots, \widehat{\beta}_p^{(k)})^\top$ as

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, Q(\widehat{\boldsymbol{\alpha}}^{(k+1)}, \boldsymbol{\beta}, \widehat{\boldsymbol{\Delta}}^{(k+1)} | \widehat{\boldsymbol{\theta}}^{(k)}).$$

The iterations are repeated until a suitable convergence rule is satisfied, such as

$$\left| \ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1 \right| < 10^{-5}. \tag{23}$$

Remark 3. *For the EM algorithm described above, it is important to mention some points for the iterative optimization procedures that we present next:*

(i) *From the above algorithm, we get an estimate of the parameters* $\boldsymbol{\Delta}$ *and* $\boldsymbol{\Psi}$*. Thus, from the invariance properties of the ML estimators, we have the ML estimates of* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\lambda}$ *to be*

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Psi}} + \widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\Delta}}^\top \text{ and } \widehat{\boldsymbol{\lambda}} = \widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\Delta}} \Big/ \sqrt{1 - \widehat{\boldsymbol{\Delta}}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\Delta}}}, \tag{24}$$

respectively. Moreover, from the **CM-step 1** *the estimates of the vector* $\boldsymbol{\Delta}$ *can be written as*

$$\boldsymbol{\Delta} = \frac{\sum_{i=1}^n \widehat{h}_i^{(k)} a_{\mathbf{t}_i}(\widehat{\boldsymbol{\alpha}}^{(k)}, \widehat{\boldsymbol{\beta}}^{(k)})}{n\overline{\widehat{h}^2}}$$

and the estimates of elements of $\boldsymbol{\Sigma}$*, which is denoted by* σ_{jk}*,* $j \neq k$ *are given by*

$$\widehat{\sigma}_{jk} = \widehat{\psi}_{jk} + \widehat{\Delta}_j \widehat{\Delta}_k, \; j, k = 1, \ldots, p, \tag{25}$$

where $\widehat{\psi}_{jk}$ *is the ML estimate of* ψ_{jk} *which is the off-diagonal element of the matrix* $\boldsymbol{\Psi}$*. The results in (24) and (25) help the EM-algorithm to be readily* *implementable, numerically stable, and very accurate, as will be shown later. For $p=2$, the ML estimate of ρ is given by*

$$\widehat{\rho}=\widehat{\psi}_{12}+\widehat{\Delta}_1\widehat{\Delta}_2. \tag{26}$$

(ii) *Observe that for $\boldsymbol{\lambda}=\mathbf{0}$, the algorithm reduces to the estimating equations*

$$\widehat{\alpha}_j^2=\frac{S_j}{\widehat{\beta}_j}+\frac{\widehat{\beta}_j}{R_j}-2 \quad \text{and} \quad \widehat{\sigma}_{jk}=\frac{\sum_{i=1}^n \xi(t_{ji},\widehat{\beta}_j)\xi(t_{ki},\widehat{\beta}_k)}{\sqrt{\sum_{i=1}^n \xi^2(t_{ji},\widehat{\beta}_j)}\sqrt{\sum_{i=1}^n \xi^2(t_{ji},\widehat{\beta}_k)}},$$

where $\xi(t,a)=\sqrt{t/a}-\sqrt{a/t}$, $S_j=\frac{1}{n}\sum_{i=1}^n t_{ji}$ and $R_j=\frac{1}{\frac{1}{n}\sum_{i=1}^n(1/t_{ji})}$, $j=1,\ldots,p$, in which the ML estimate $\widehat{\boldsymbol{\beta}}$ is obtained as

$$\widehat{\boldsymbol{\beta}}=\underset{\boldsymbol{\beta}}{\operatorname{argmax}}\ \ell(\widehat{\boldsymbol{\alpha}},\boldsymbol{\beta},\widehat{\boldsymbol{\Sigma}});$$

(iii) *Useful starting values are required to implement this ECM-algorithm. These can be easily obtained from the modified moment estimates of Kundu et al. (2013) for the multivariate BS distribution (under normality), which we denote by $\widetilde{\boldsymbol{\alpha}}$, $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\Sigma}}$. From the construction of the multivariate SNBS distribution, $a_{\mathbf{T}}(\boldsymbol{\alpha},\boldsymbol{\beta})=\big(a_{T_1}(\alpha_1,\beta_1),\ldots,a_{T_p}(\alpha_p,\beta_p)\big)^\top\sim SN_p(\mathbf{0},\boldsymbol{\Sigma},\boldsymbol{\lambda})$, we have*

$$\mathbf{Z}_i=\widetilde{\boldsymbol{\Sigma}}^{-1/2}a_{\mathbf{T}_i}(\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}}),\ i=1,\ldots,n,$$

to be observations from the $SN_p(\mathbf{0},\mathbf{I}_p,\boldsymbol{\lambda})$ distribution. Thus, by using the estimation method for the parameters of the multivariate skew-normal distribution, $\boldsymbol{\lambda}$ can be estimated. In order to ensure that the true ML estimates are identified, we recommend running the EM algorithm using a range of different starting values to check for the convergence;

(iv) *According to the construction of the multivariate SNBS distribution, see the result in (16),*

$$d_i=a_{\mathbf{T}_i}^\top(\boldsymbol{\alpha},\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}a_{\mathbf{T}_i}(\boldsymbol{\alpha},\boldsymbol{\beta})$$

has a χ_p^2 distribution and this result is extremely useful for testing the goodness of fit of the model and also for detecting outliers; see Lange and Sinsheimer (1993), Vilca et al. (2014) and Romeiro et al. (2018).

4.2 Estimation under some restrictions

Estimation of the model parameters under certain restrictions can be interesting during the fitting of the model. We consider the following specific problems that are of practical interest: (i) $H_{01}:\boldsymbol{\lambda}=\mathbf{0}$ and (ii) $H_{02}:\alpha_1=\cdots=\alpha_p$. The likelihood ratio

(LR) test statistic can be easily adopted once the ML estimation under restrictions can be obtained from a adaptation of the proposed EM algorithm:
i) *Estimation under* H_{01}: With the proposed estimates in the our EM algorithm and the estimates in Kundu et al. (2013), can be obtained easily the LR statistic;
ii) *Estimation under* H_{02}: Under this hypotheses, the estimative of α^2 in the CM-step 1 is given by

$$\widetilde{\alpha}^{2(k+1)} = \frac{1}{pn}\sum_{i=1}^{n} a_{\mathbf{t}_i}^{\top}(\mathbf{1},\widehat{\boldsymbol{\beta}}^{(k)})a_{\mathbf{t}_i}(\mathbf{1},\widehat{\boldsymbol{\beta}}^{(k)}) + (1-\widehat{\widehat{h^2}})\,\frac{\sum_{i=1}^{n}\widehat{h}_i^{(k)}\,a_{\mathbf{t}_i}^{\top}(\mathbf{1},\widehat{\boldsymbol{\beta}}^{(k)})}{n\widehat{\widehat{h^2}}}\,\frac{\sum_{i=1}^{n}\widehat{h}_i^{(k)}\,a_{\mathbf{t}_i}(\mathbf{1},\widehat{\boldsymbol{\beta}}^{(k)})}{n\widehat{\widehat{h^2}}},$$

and

$$\widetilde{\boldsymbol{\Delta}} = \frac{\sum_{i=1}^{n}\widehat{h}_i^{(k)}a_{\mathbf{t}_i}(\widehat{\alpha}^{(k)}\mathbf{1},\widehat{\boldsymbol{\beta}}^{(k)})}{n\widehat{\widehat{h^2}}}$$

and estimative of $\boldsymbol{\Psi}$ is as in the CM-step 1, with the estimates under H_{02}.

4.3 The observed information matrix

Suppose $\mathbf{T}_1,\ldots,\mathbf{T}_n$ are n independent observations with $\mathbf{T}_i \sim \mathrm{SNBS}_2(\boldsymbol{\alpha},\boldsymbol{\beta},\rho,\boldsymbol{\lambda})$, for $i=1,\ldots,n$. Then, the log-likelihood function for $\boldsymbol{\theta}$ is $\ell(\boldsymbol{\theta})$ as given in (17). So, the score functions for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, ρ and $\boldsymbol{\lambda}$ have, respectively, the forms

$$U_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) \;=\; \sum_{i=1}^{n}\Big(U_{\rho}(\boldsymbol{\gamma}) + \frac{1}{A_i}\frac{\partial A_i}{\partial\boldsymbol{\gamma}} - \frac{1}{2}\frac{\partial d(\mathbf{t}_i)}{\partial\boldsymbol{\gamma}} + \frac{\partial \ell_{i\Phi}(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}}\Big),\;\; \boldsymbol{\gamma}=\boldsymbol{\alpha},\boldsymbol{\beta},\rho,\boldsymbol{\lambda}, \tag{27}$$

where $U_{\rho}(\boldsymbol{\gamma}) = \frac{\rho}{1-\rho^2}$ if $\boldsymbol{\gamma}=\rho$, and zero otherwise. The required derivatives of A_i, $d(\mathbf{t}_i)$ and $\ell_{i\Phi}(\boldsymbol{\theta})$ are all presented in Appendix A. It is important to mention here that for obtaining the derivative of $\ell_{i\Phi}(\boldsymbol{\theta})$ with respect to ρ, it is necessary to express $\boldsymbol{\Sigma}$ as $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}$, with $\mathbf{B}=\boldsymbol{\Sigma}^{1/2}$ and $\mathbf{B}^{-1}=\boldsymbol{\Sigma}^{-1/2}$. The following notations are used to derive the observed information matrix; see Appendix A; $\dot{\mathbf{B}}(\rho)=\partial\mathbf{B}/\partial\rho$ and $\dot{\mathbf{B}}_{\rho}=\mathbf{B}^{-1}\dot{\mathbf{B}}(\rho)\mathbf{B}^{-1}$, and $\mathbf{B}_{\rho\rho}=\mathbf{B}^{-1}\big[2\dot{\mathbf{B}}(\rho)\mathbf{B}^{-1}\dot{\mathbf{B}}(\rho)-\ddot{\mathbf{B}}(\rho,\rho)\big]\mathbf{B}^{-1}$, where $\ddot{\mathbf{B}}(\rho,\rho)=\partial^2\mathbf{B}/\partial\rho\partial\rho$ (the same notations are valid for the matrix $\boldsymbol{\Sigma}$). Specifically,

$$\mathbf{B} \;=\; \frac{1}{\sqrt{2+2\upsilon}}\begin{pmatrix}1+\upsilon & \rho\\ \rho & 1+\upsilon\end{pmatrix},\;\; \dot{\mathbf{B}}(\rho)=\frac{\rho}{\upsilon(2+2\upsilon)}\mathbf{B}-\frac{1}{\upsilon\sqrt{2+2\upsilon}}\begin{pmatrix}\rho & -\upsilon\\ -\upsilon & \rho\end{pmatrix},$$

$$\ddot{\mathbf{B}}(\rho,\rho) \;=\; \frac{1+\upsilon(1+\rho^2)}{2\upsilon^3(1+\upsilon)^2}\mathbf{B}+\frac{\rho}{\upsilon(2+2\upsilon)}\dot{\mathbf{B}}(\rho)-b_1\begin{pmatrix}\rho & -\upsilon\\ -\upsilon & \rho\end{pmatrix}-b_2\begin{pmatrix}\upsilon & \rho\\ \rho & \upsilon\end{pmatrix},$$

where $b_1=\frac{\rho(2+3\upsilon)}{\upsilon^3(2+2\upsilon)^{3/2}}$ and $b_2=\frac{1}{\upsilon^2\sqrt{2+2\upsilon}}$, with $\upsilon=\sqrt{1-\rho^2}$.

Finally, the matrix of second derivatives with respect to $\boldsymbol{\theta}$ is as follows:

$$\ddot{\mathbf{L}}=\sum_{i=1}^{n}\frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}=\sum_{i=1}^{n}\left(\frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\tau}^{\top}}\right),\quad \boldsymbol{\gamma},\boldsymbol{\tau}=\boldsymbol{\alpha},\boldsymbol{\beta},\rho,\boldsymbol{\lambda}, \tag{28}$$

where

$$\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} = -\frac{1}{2}\rho_{\gamma\tau} - \frac{1}{2}\frac{\partial^2 d(\mathbf{t}_i)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} - \frac{1}{A_i^2}\frac{\partial A_i}{\partial \boldsymbol{\gamma}}\frac{\partial A_i}{\partial \boldsymbol{\tau}^\top} + \frac{1}{A_i}\frac{\partial^2 A_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} + \frac{\partial^2 \ell_{i\Phi}(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top},$$

with $\rho_{\gamma\tau} = -2\frac{1+\rho^2}{(1-\rho^2)^2} I_{\{\gamma=\tau\}}(\rho)$ and the expressions for $\partial^2 \ell_{i\Phi}(\boldsymbol{\theta})/\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top$ are as presented in Appendix A.

Asymptotic confidence intervals and hypothesis tests based on the ML estimates can be obtained by using the matrix $\mathbf{J}_n(\boldsymbol{\theta}) = -\ddot{\mathbf{L}}$, which is the observed information matrix obtained from the log-likelihood function $\ell(\boldsymbol{\theta})$. The inferential procedures are then developed by using the property that the ML estimator $\widehat{\boldsymbol{\theta}}$ has approximately $N_5(\boldsymbol{\theta}, K_{\boldsymbol{\theta}}^{-1})$ distribution, where $K_{\boldsymbol{\theta}}$ is the (expected) Fisher information. In practice, we use the matrix $\mathbf{J}_n(\boldsymbol{\theta})$, evaluated at the ML estimator $\widehat{\boldsymbol{\theta}}$, since we have a closed-form expression for the observed information matrix for $\boldsymbol{\theta}$.

5 Illustrative examples

In this section, a simulation study and a real-life example are presented to illustrate the performance of the method developed in the preceding sections. First, we carry out a numerical illustration by simulated data, where the quality of the estimation method, proposed in Section 4, and the finite-sample performance of these estimates are evaluated by means of Monte Carlo procedure. Finally, we analyze the real-life data given in Meintanis (2007).

5.1 Experiment 1: Parameter recovery

In this section, we consider four scenarios for simulation in order to verify if we can estimate the true parameter values accurately by using the proposed estimation method. For this purpose, we fit the SNBS model to data that were generated from the SNBS model. In this case, we generated 500 random samples of size $n = 200$ from the $\text{SNBS}_3(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ model with the following parameter values:

$$\boldsymbol{\alpha} = \begin{pmatrix} 1.5 \\ 0.75 \\ 0.5 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} 1.2 \\ 1 \\ 0.8 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \sigma_1 & \sigma_2 \\ \sigma_1 & 1 & \sigma_3 \\ \sigma_2 & \sigma_3 & 1 \end{pmatrix},$$

where $\sigma_1 = 0.5$, $\sigma_2 = 0.7$ and $\sigma_3 = 0.9$. In addition, we consider the following scenarios the parameter $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)^\top$; Scenario 1 : $\boldsymbol{\lambda} = (1, 4, 1)^\top$; Scenario 2 : $\boldsymbol{\lambda} = (-1, -4, -1)^\top$; Scenario 3 : $\boldsymbol{\lambda} = (1, 4, -1)^\top$ and Scenario 4 : $\boldsymbol{\lambda} = (-1, 4, -1)^\top$. We used the ML estimation via EM algorithm for each sample, using the stopping criterion in (23). There is no program available for the estimation of the parameters of SNBS directly. Therefore, ML estimation via the EM algorithm was implemented using

Table 1: Mean and standard deviations (SD) for EM estimates based on 500 samples from the SNBS model. True values of parameters are inside parentheses.

	Scenario 1		Scenario 2	
Parameter	Mean	SD	Mean	SD
$\alpha_1(1.5)$	1.4958	0.1063	1.5073	0.1113
$\alpha_2(0.75)$	0.7442	0.0896	0.7546	0.1146
$\alpha_3(0.5)$	0.5010	0.0527	0.5033	0.0576
$\beta_1(1.2)$	1.2180	0.1615	1.2173	0.1552
$\beta_2(1)$	1.0011	0.0675	1.0014	0.0797
$\beta_3(0.8)$	0.7993	0.0452	0.8014	0.0455
$\sigma_1(0.5)$	0.4966	0.0821	0.5017	0.0817
$\sigma_2(0.7)$	0.7018	0.0471	0.7021	0.0496
$\sigma_3(0.9)$	0.9026	0.0443	0.8917	0.0392
λ_1	0.9780	0.3170	-1.0314	0.3240
λ_2	3.9607	0.4072	-4.0146	0.5075
λ_3	0.9948	0.3593	-1.0163	0.3972
	Scenario 3		Scenario 4	
Parameter	Mean	SD	Mean	SD
$\alpha_1(1.5)$	1.5113	0.0947	1.4954	0.0836
$\alpha_2(0.75)$	0.7669	0.0878	0.7728	0.0731
$\alpha_3(0.5)$	0.5058	0.0376	0.5050	0.0296
$\beta_1(1.2)$	1.1842	0.1500	1.1589	0.1400
$\beta_2(1)$	0.9751	0.0930	0.9513	0.0897
$\beta_3(0.8)$	0.7887	0.0492	0.7768	0.0468
$\sigma_1(0.5)$	0.5121	0.0751	0.5121	0.0624
$\sigma_2(0.7)$	0.7088	0.0461	0.7031	0.0399
$\sigma_3(0.9)$	0.9105	0.0373	0.9267	0.0419
λ_1	1.0045	0.3495	-0.9577	0.2903
λ_2	3.9136	0.4987	3.8952	0.4264
λ_3	-0.9220	0.3303	-0.8928	0.2581

MATLAB. The average values and the corresponding standard deviations (SD) of the EM estimates across all samples were computed and these results are presented in Table 1. Moreover, Figure 4 shows boxplots of the parameter estimates for Scenario 2. For the other scenarios, the results are very close, so they are not shown here to save space. Note that all the point estimates are quite accurate in all the considered scenarios, which suggests that the proposed EM algorithm produces satisfactory estimates.

5.2 Experiment 2: Performance of the ML estimates

In this section, we use Monte Carlo simulations to evaluate the finite-sample performance of the ML estimates of the parameters of the SNBS model determined from the EM algorithm described in Section 4. The sample sizes and true values of the parameters considered were $n = 30, 50, 100, 200$ and 600, $\boldsymbol{\alpha} = (1.5, 0.75, 0.5)^\top$, $\boldsymbol{\beta} = (1.2, 1, 0.8)^\top$, $\boldsymbol{\lambda} = (-1, -4, -1)^\top$, $\sigma_1 = 0.5, \sigma_2 = 0.7$ and $\sigma_3 = 0.9$. The number of Monte Carlo replications was taken once again as $M = 500$. In order to examine the performance of the ML estimates, we computed for each sample size and for each estimate, denoted by $\widehat{\boldsymbol{\theta}}_j$, the mean, denoted by $E[\widehat{\boldsymbol{\theta}}_j]$, the relative bias (RB) in absolute value, defined as $RB_j = |(E[\widehat{\boldsymbol{\theta}}_j] - \boldsymbol{\theta}_j)/\boldsymbol{\theta}_j|$, and the root mean square error (MSE),

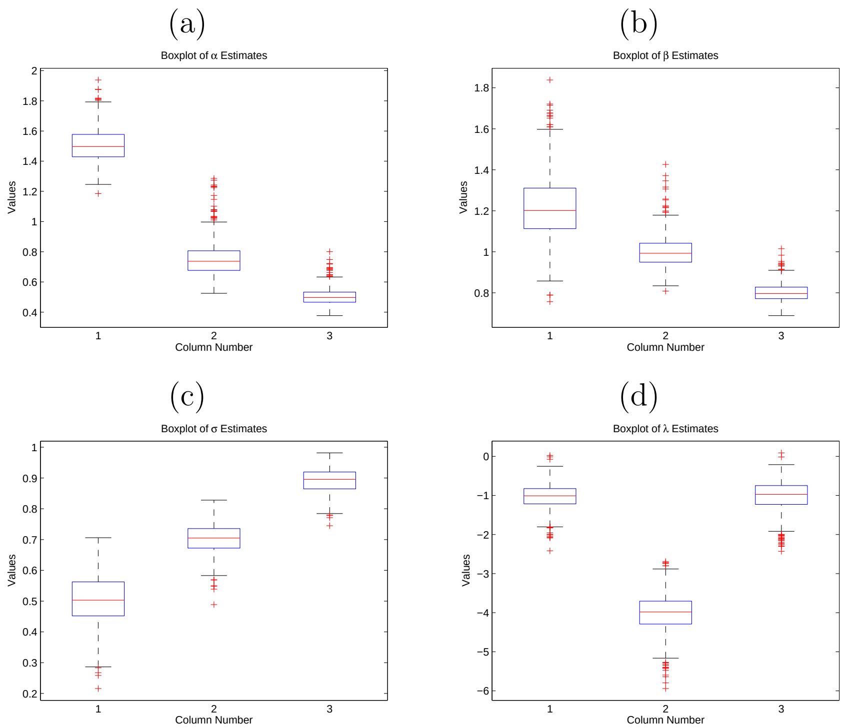

Figure 4: Box plots for estimates of (a) the α parameter, (b) the β parameter, (c) the σ parameter, and (d) the λ parameter. For all panel, the column number corresponds to the index of parameter.

defined by $MSE_j = \sqrt{E[(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)^2]}$, for $j = 1, 2, \ldots, 12$. Here, the MSE is the mean square error computed from the 500 Monte Carlo replications. These quantities are computed to evaluate the performance of the ML estimates individually for each component of the parameter $\boldsymbol{\theta}$. Table 2 and Figures 5-6 present the RB and MSE of the ML estimate of components of the parameter $\boldsymbol{\theta}$. It can be observed that the RB and MSE become smaller as the sample size n increases, as one would expect.

To evaluate the performance of the ML estimates of the parameter $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)^\top$, we use the mean relatives bias (MRB) in absolute value. Suppose that $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$ is a generic vector of parameters and then the MRB is defined as $MRB(\boldsymbol{\phi}) = \dfrac{1}{p}\sum_{k=1}^{p} RB(\phi_k), \phi = \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}$ or $\boldsymbol{\sigma}$. Analogously, the MMSE for $\boldsymbol{\phi}$ is defined as the root of the mean square errors given by $MMSE = \dfrac{1}{p}\sum_{k=1}^{p} MSE(\phi_k)$. It can be observed that estimate of $\boldsymbol{\lambda}$, on average, present the highest RB and MSE, while estimate of $\boldsymbol{\sigma}$, on average, have the smallest values of these measures.

Table 2: Bias and mean squared error for EM estimates based on 500 samples from the SNBS model. True values of parameters are inside parentheses.

	$n = 30$		$n = 50$	
Parameter	RB	MSE	RB	MSE
$\alpha_1(1.5)$	0.0331	0.3919	0.0239	0.2793
$\alpha_2(0.75)$	0.0951	0.4275	0.0649	0.3103
$\alpha_3(0.5)$	0.0822	0.3866	0.0408	0.1360
$\beta_1(1.2)$	0.0918	0.5649	0.0571	0.4021
$\beta_2(1)$	0.0757	0.4184	0.0344	0.2819
$\beta_3(0.8)$	0.0231	0.1496	0.0116	0.1065
$\sigma_1(0.5)$	0.0355	0.2124	0.0142	0.1698
$\sigma_2(0.7)$	0.0158	0.1368	0.0071	0.1060
$\sigma_3(0.9)$	0.0268	0.0914	0.0217	0.0794
$\lambda_1(-1)$	0.1528	1.0775	0.0779	0.7522
$\lambda_2(-4)$	0.0146	1.1764	0.0063	0.9210
$\lambda_3(-1)$	0.0721	1.0616	0.0462	0.7577

	$n = 100$		$n = 600$	
Parameter	RB	MSE	RB	MSE
$\alpha_1(1.5)$	0.0109	0.1789	0.0023	0.0635
$\alpha_2(0.75)$	0.0337	0.1905	0.0004	0.0662
$\alpha_3(0.5)$	0.0217	0.0936	0.0018	0.0344
$\beta_1(1.2)$	0.0285	0.2328	0.0067	0.0893
$\beta_2(1)$	0.0145	0.1399	0.0002	0.0441
$\beta_3(0.8)$	0.0066	0.0703	0.0004	0.0263
$\sigma_1(0.5)$	0.0043	0.1175	0.0029	0.0503
$\sigma_2(0.7)$	0.0031	0.0723	0.0019	0.0311
$\sigma_3(0.9)$	0.0122	0.0551	0.0047	0.0261
$\lambda_1(-1)$	0.0474	0.4808	0.0137	0.1790
$\lambda_2(-4)$	0.0023	0.6884	0.0002	0.2803
$\lambda_3(-1)$	0.0376	0.5588	0.0013	0.2366

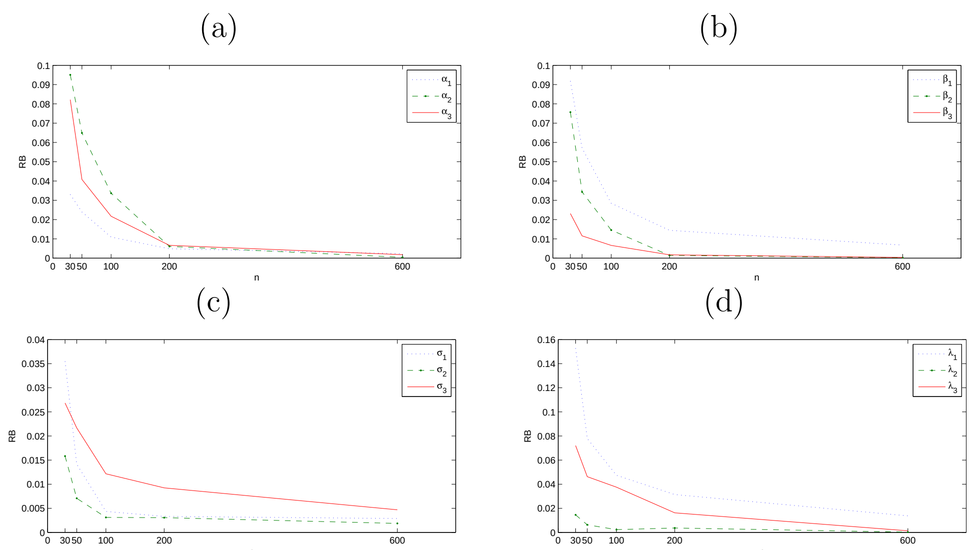

Figure 5: RB for estimates of (a) the α parameter, (b) the β parameter, (c) the σ parameter, and (d) the λ parameter.

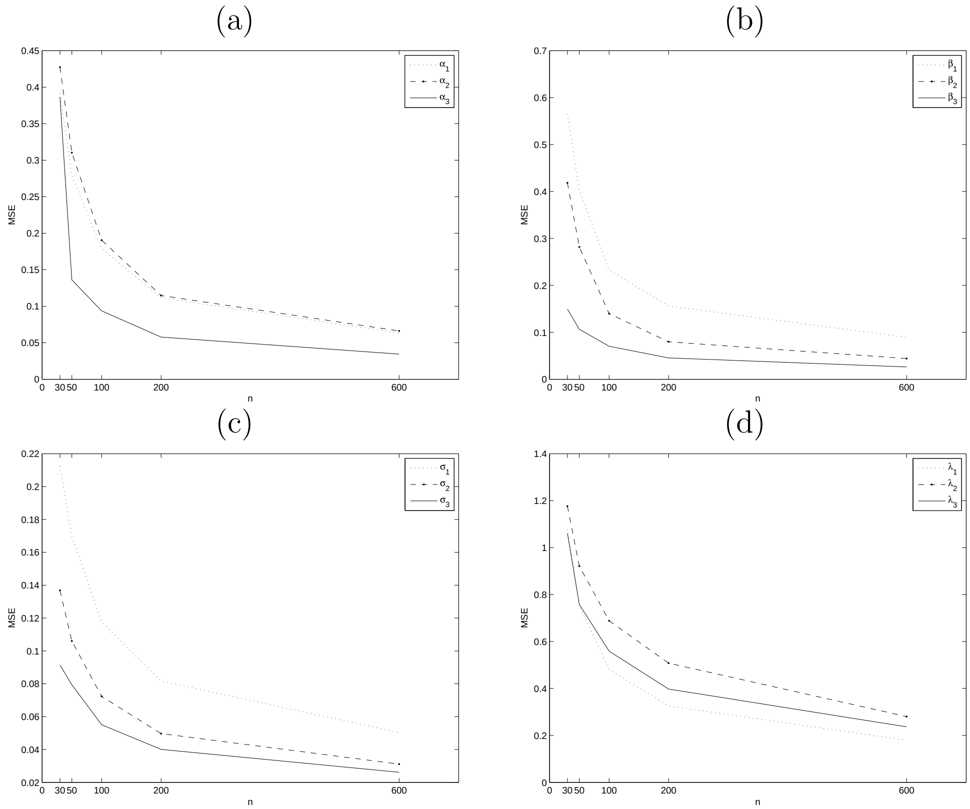

Figure 6: MSE for estimates of (a) the α parameter, (b) the β parameter, (c) the σ parameter, and (d) the λ parameter.

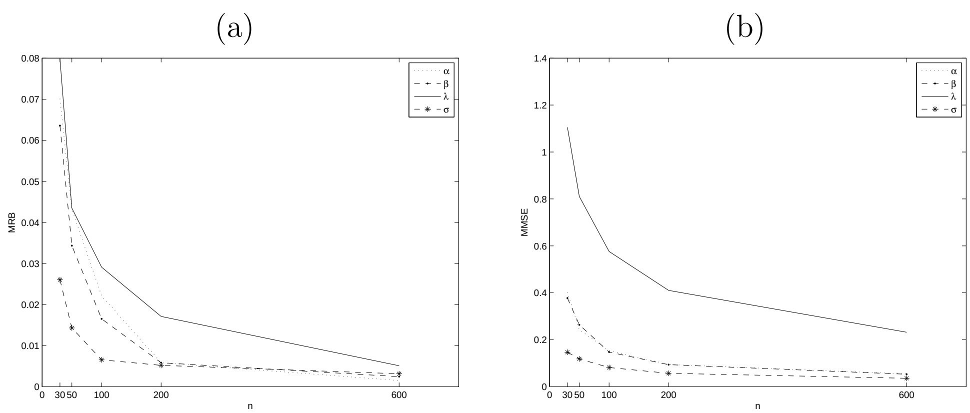

Figure 7: The mean of RB and MSE of components of the parameter $\boldsymbol{\theta}$.

5.3 Performance of the EM-algorithm

Through simulation studies, we compare the performance of the EM-algorithm of Jamalizadeh and Kundu (2015) (denoted here by EM_{JK}-algorithm) with the proposed

Table 3: Bias and MSE for the ρ, λ_1 and λ_2 under various settings of sample sizes.

		Bias		MSE	
Sample sizes	Parameters	EM_{JK}	EM	EM_{JK}	EM
100	ρ	-0.0219	-0.0045	0.0359	0.0048
	λ_1	1.0449	0.1605	29.4508	0.8708
	λ_2	2.5541	0.4483	115.4106	4.8059
200	ρ	-0.0021	-0.0011	0.0131	0.0012
	λ_1	0.0917	0.0614	0.3096	0.2256
	λ_2	0.3389	0.2328	1.7686	1.2297
500	ρ	-0.0026	-0.0014	0.0052	0.0015
	λ_1	0.0122	0.0132	0.0760	0.0694
	λ_2	0.0904	0.0770	0.3728	0.3366
1000	ρ	-0.0008	-0.0001	0.0026	0.0004
	λ_1	0.0060	0.0113	0.0325	0.0333
	λ_2	0.0485	0.0596	0.1602	0.1696

Table 4: Percentiles of the run times (seconds) of algorithms based on a sample of size n. Q_1: 25th, Q_2: 50th and Q_3: 75th percentiles.

Samples sizes	Q_1		Q_2		Q_3	
n	EM_{JK}	EM	EM_{JK}	EM	EM_{JK}	EM
100	9.4941	6.4875	14.1451	9.0335	21.2015	18.1023
200	17.8581	9.9103	25.8925	14.5837	39.4561	25.7338
500	42.4850	18.2352	60.8875	25.5275	85.0265	41.4165
1000	97.6591	35.1701	132.4094	55.7041	196.4541	74.8028

EM-algorithm under the following parameter values: $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top = (1.5, 0.5)^\top$, $\rho = 0.9$, and $\lambda_1 = 1$ and $\lambda_2 = 4$. . Specifically, we compare both approaches based on RB and MSE and we note that both procedures have a similar performance for the estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (which is omitted), but the estimation based on the proposed EM-algorithm of ρ, λ_1 and λ_2 is better than the EM_{JK}-algorithm as shown in Table 3, where it can see that the RB and MSE for the parameters tend to approach zero with increasing sample size (n). On the other hand, we compare the mean run times of the two estimation methods based on mean values of the 25th percentile (Q_1), the 50th percentile (Q_2) and the 75th percentile (Q_3). The mean run time for the proposed EM-algorithm is lower than the mean run time for the EM_{JK}-algorithm. These results are reported in Table 4. Thus, we can conclude that the proposed EM-algorithm produces more precise estimates and furthermore its implementation is computationally simpler.

5.4 Real dataset

We illustrate the proposed methods with a dataset from Meintanis (2007). These data represent the football (soccer) data where at least one goal was scored by the home team and at least one goal was scored directly from a penalty kick, foul kick or any other direct kick by any team. In this case, T_1 represents the time in minutes of the first kick goal scored by any team and T_2 represents the first goal of any type scored by the home team. The Pearson correlation between T_1 and T_2 is given by $r = 0.4698$. These data were analyzed earlier by Kundu and Dey (2009), leading them to propose a different bivariate exponential distribution under the Bayesian paradigm. Histograms of T_1 and T_2 are also provided in Figure 8 which reveal that both T_1 and T_2 are right skewed. Now, we revisit this data set with the aim of expanding the inferential results to the bivariate SNBS distribution. Table 5 presents the ML estimates of the

(a) (b)

Figure 8: Histograms of (a) T_1 and (b) T_2.

parameters of the BS and SNBS models, along with their corresponding standard errors (SE) calculated via the observed information matrix. We note that the estimates of $\widehat{\boldsymbol{\theta}}$ show much difference between the two selected models. We compare the BS and

Table 5: ML estimation results for the two selected models. SE are the estimated standard errors based on observed information matrix.

	SNBS model		BS model	
Parameters	Estimates	SE	Estimates	SE
α_1	1.0610	0.2556	0.8304	0.0994
α_2	1.5059	0.4265	1.0475	0.1163
β_1	17.3109	6.6650	28.8902	3.6325
β_2	9.1115	3.8332	19.6607	3.0820
ρ	0.7422	0.1360	0.5462	0.1173
λ_1	0.8056	0.3742	-	-
λ_2	1.0833	0.3875	-	-
log-likelihood	-703.1851		-706.8277	
AIC	1420.3702		1423.6554	
SIC	1417.3476		1421.4964	

SNBS models through some information selection criteria. We use AIC and SIC for this purpose, and they are defined as $-2\,\ell(\widehat{\boldsymbol{\theta}}) + p\,c_n$, where $\ell(\cdot)$ is the log-likelihood, $\widehat{\boldsymbol{\theta}}$ is the EM estimate of $\boldsymbol{\theta}$, p is the number of free parameters that have to be estimated under the model, and the penalty term c_n is a suitable sequence of positive numbers. We have $c_n = 2$ for AIC and $c_n = \log n$ for SIC, where n is the sample size. Comparing the models by looking at the values of the information criteria presented in Table 5, we observe that the SNBS model presents a better fit than the BS model. In addition,

Table 6: 95% confidence interval (CI) based on the normal approximation of the ML estimates.

Parameters	SNBS model CI	BS model CI
α_1	(0.5498; 1.5722)	(0.6316; 1.0292)
α_2	(0.6529; 2.3589)	(0.8149; 1.2801)
β_1	(3.9809; 30.6409)	(21.6252; 36.1552)
β_2	(1.4486; 16.7814)	(13.4967; 25.8247)
ρ	(0.4702; 1.0000)	(0.3116; 0.7808)
λ_1	(0.0572; 1.5540)	- -
λ_2	(0.3083; 1.8583)	- -

Table 6 presents the 95% confidence interval (CI) for each parameter of the model based on the normal approximation of the ML estimates given by $\hat{\alpha}_j \pm 1.96 \times SE, j = 1, 2,$ for α_j.

The distribution of Mahalanobis distance, $d(\mathbf{t}) = a_{\mathbf{t}}^{\top}(\boldsymbol{\alpha}, \boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$, is used to check the validity of the model, and according to Lange and Sinsheimer (1993) and Vilca et al. (2014), $d(\mathbf{t}) \sim \chi^2_2$. By using the ML estimate of $\boldsymbol{\theta}$ in $d(\mathbf{t}_i)$, $i = 1, \ldots, n$, Figure 9 shows simulated envelopes (lines representing the $5th$ percentile, the mean and the $95th$ percentile of 100 simulated points for each observation). From this figure, we see once again that the SNBS model provides a better fit to the considered data set than the BS model, containing all the observations well inside the envelope.

Figure 9: Simulated envelopes, under the two models.

6 Concluding remarks

We have developed a bivariate BS-type distribution based on the skew-normal distribution, referred to as the bivariate SNBS distribution. This new class of distributions, in addition to including the bivariate BS distribution introduced recently by Kundu et al. (2010), has its marginal distributions as the univariate SNBS distribution of Vilca et al. (2011), and its conditional distributions as the BS-type distribution based on the extended skew-normal of Capitanio et al. (2003). We have established some important characteristics and properties of this family of distributions as well as the observed information matrix. The ML estimates of the model parameters can be achieved efficiently by using the EM-algorithm. The proposed EM-algorithm has several advantages over direct maximization of the likelihood function since it is easily implementable, numerically stable, and very accurate, as is demonstrated through simulation studies. Moreover, empirical results clearly indicate that the proposed EM-estimates outperform the estimates obtained via the EM_{JK}-algorithm for the parameters ρ, λ_1 and λ_2.

The bivariate SNBS distribution introduced here may be suitable for representing highly skewed data. Moreover, this distribution can be used along the same lines as the univariate BS distribution and its generalizations, and also in the regression model setup as in Rieck and Nedelman (1991) and Santana et al. (2011).

Appendix A: The observed information matrix

Let $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^\top$ and $\boldsymbol{\tau} = (\tau_1, \tau_2)^\top$. Here, we present the first-derivatives and second-derivatives of $A_i = A_{t_{1i}}(\alpha_1, \beta_1)\, A_{t_{2i}}(\alpha_2, \beta_2)$, $d(\mathbf{t}_i)$ and $\ell_{i\Phi}(\boldsymbol{\theta})$ which are necessary to obtain the Hessian matrix. For simplicity in notation, we omit the index i in the expressions without causing any confusion. Let us use $u_{\mathbf{t}} = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2} a_{\mathbf{t}}$ in order to shorten the formulas for derivatives. Then, the first-derivatives are as follows:

$$
\begin{aligned}
\frac{\partial A}{\partial \boldsymbol{\tau}} &= \left(\frac{\partial A_{t_1}}{\partial \tau_1} A_{t_2}, A_{t_1}\frac{\partial A_{t_2}}{\partial \tau_2}\right)^\top, \quad \frac{\partial a_{\mathbf{t}}^\top}{\partial \boldsymbol{\tau}} = \begin{pmatrix} \frac{\partial a_{t_1}}{\partial \tau_1} & 0 \\ 0 & \frac{\partial a_{t_2}}{\partial \tau_2} \end{pmatrix}, \boldsymbol{\tau} = \boldsymbol{\alpha}, \boldsymbol{\beta}, \\
\frac{\partial d(\mathbf{t})}{\partial \boldsymbol{\tau}} &= 2\frac{\partial a_{\mathbf{t}}^\top}{\partial \boldsymbol{\tau}} \boldsymbol{\Sigma}^{-1} a_{\mathbf{t}},\ \boldsymbol{\tau} = \boldsymbol{\alpha}, \boldsymbol{\beta}, \quad \frac{\partial d(\mathbf{t})}{\partial \rho} = -a_{\mathbf{t}}^\top \dot{\boldsymbol{\Sigma}}_\rho a_{\mathbf{t}}, \\
\frac{\partial \ell_{i\Phi}(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}} &= W_\Phi(u_{\mathbf{t}}) \frac{\partial a_{\mathbf{t}}^\top}{\partial \boldsymbol{\tau}} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\lambda},\ \boldsymbol{\tau} = \boldsymbol{\alpha}, \boldsymbol{\beta}, \\
\frac{\partial \ell_{i\Phi}(\boldsymbol{\theta})}{\partial \rho} &= -W_\Phi(u_{\mathbf{t}}) \boldsymbol{\lambda}^\top \dot{\mathbf{B}}_\rho a_{\mathbf{t}}, \quad \frac{\partial \ell_{i\Phi}(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} = W_\Phi(u_{\mathbf{t}}) \boldsymbol{\Sigma}^{-1/2} a_{\mathbf{t}_i},
\end{aligned}
$$

where $\dot{\boldsymbol{\Sigma}}_\rho$ and $\dot{\mathbf{B}}_\rho$ are as in Section 4.1. The second derivatives are as follows:

$$\frac{\partial^2 A}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} = \begin{pmatrix} \frac{\partial^2 A_{t_1}}{\partial \gamma_1 \partial \tau_1} A_{t_2} & \frac{\partial A_{t_1}}{\partial \gamma_1}\frac{\partial A_{t_2}}{\partial \tau_2} \\ \frac{\partial A_{t_1}}{\partial \tau_1}\frac{\partial A_{t_1}}{\partial \gamma_2} & A_{t_1}\frac{\partial^2 A_{t_2}}{\partial \gamma_2 \partial \tau_2} \end{pmatrix}, \boldsymbol{\gamma}, \boldsymbol{\tau} = \boldsymbol{\alpha}, \boldsymbol{\beta},$$

$$\frac{\partial^2 d(\mathbf{t})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} = 2\Big[\frac{\partial a_{\mathbf{t}}^\top}{\partial \boldsymbol{\gamma}} \boldsymbol{\Sigma}^{-1}\Big(\frac{\partial a_{\mathbf{t}}^\top}{\partial \boldsymbol{\tau}}\Big)^\top + D(\mathbf{b}_{\gamma\tau})D(c_{\mathbf{t}})\Big], \boldsymbol{\gamma}, \boldsymbol{\tau} = \boldsymbol{\alpha}, \boldsymbol{\beta},$$

$$\frac{\partial^2 d(\mathbf{t})}{\partial \rho \partial \boldsymbol{\tau}^\top} = -2a_{\mathbf{t}}^\top \dot{\boldsymbol{\Sigma}}_\rho \Big(\frac{\partial a_{\mathbf{t}}^\top}{\partial \boldsymbol{\tau}}\Big)^\top, \boldsymbol{\tau} = \boldsymbol{\alpha}, \boldsymbol{\beta},$$

$$\frac{\partial^2 d(\mathbf{t})}{\partial \rho \partial \rho} = a_{\mathbf{t}}^\top \boldsymbol{\Sigma}_{\rho\rho} a_{\mathbf{t}},$$

$$\frac{\partial^2 \ell_{i\Phi}(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\tau}^\top} = W_\Phi(u_{\mathbf{t}}) D(\mathbf{b}_{\gamma\tau})D(\mathbf{c}) + W_\Phi^{'}(u_{\mathbf{t}})\frac{\partial a_{\mathbf{t}}^\top}{\partial \boldsymbol{\gamma}} \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}\frac{\partial a_{\mathbf{t}}^\top}{\partial \boldsymbol{\tau}}, \boldsymbol{\gamma}, \boldsymbol{\tau} = \boldsymbol{\alpha}, \boldsymbol{\beta},$$

$$\frac{\partial^2 \ell_{i\Phi}(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \rho} = -W_\Phi(u_{\mathbf{t}})\frac{\partial a_{\mathbf{t}}^\top}{\partial \boldsymbol{\gamma}} \dot{\mathbf{B}}_\rho \boldsymbol{\lambda} - W_\Phi^{'}(u_{\mathbf{t}})\frac{\partial a_{\mathbf{t}}^\top}{\partial \boldsymbol{\gamma}} \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top \dot{\mathbf{B}}_\rho a_{\mathbf{t}}, \boldsymbol{\gamma} = \boldsymbol{\alpha}, \boldsymbol{\beta},$$

$$\frac{\partial^2 \ell_{i\Phi}(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\lambda}^\top} = W_\Phi(u_{\mathbf{t}})\frac{\partial a_{\mathbf{t}}^\top}{\partial \boldsymbol{\gamma}} \boldsymbol{\Sigma}^{-1/2} + W_\Phi^{'}(u_{\mathbf{t}})\frac{\partial a_{\mathbf{t}_i}^\top}{\partial \boldsymbol{\gamma}} \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda} a_{\mathbf{t}_i}^\top \boldsymbol{\Sigma}^{-1/2}, \boldsymbol{\gamma} = \boldsymbol{\alpha}, \boldsymbol{\beta},$$

$$\frac{\partial^2 \ell_{i}(\boldsymbol{\theta})}{\partial \rho \partial \rho} = W_\Phi(u_{\mathbf{t}})\boldsymbol{\lambda}^\top \mathbf{B}_{\rho\rho} a_{\mathbf{t}} + W_\Phi^{'}(u_{\mathbf{t}})\big(\boldsymbol{\lambda}^\top \dot{\mathbf{B}}_\rho a_{\mathbf{t}}\big)^2,$$

$$\frac{\partial^2 \ell_{i}(\boldsymbol{\theta})}{\partial \rho \partial \boldsymbol{\lambda}^\top} = -W_\Phi(u_{\mathbf{t}}) a_{\mathbf{t}}^\top \dot{\mathbf{B}}_\rho - W_\Phi^{'}(u_{\mathbf{t}})\boldsymbol{\lambda}^\top \dot{\mathbf{B}}_\rho a_{\mathbf{t}} a_{\mathbf{t}}^\top \boldsymbol{\Sigma}^{-1/2},$$

$$\frac{\partial^2 \ell_{i}(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^\top} = W_\Phi^{'}(u_{\mathbf{t}})\boldsymbol{\Sigma}^{-1/2} a_{\mathbf{t}} a_{\mathbf{t}}^\top \boldsymbol{\Sigma}^{-1/2},$$

where $W_\Phi'(u) = -W_\Phi(u)(u + W_\Phi(u))$, $\mathbf{c} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}$, $c_{\mathbf{t}} = \boldsymbol{\Sigma}^{-1} a_{\mathbf{t}}$ and $\mathbf{b}_{\gamma\tau} = \Big(\frac{\partial^2 a_{t_1}}{\partial \gamma_1 \partial \tau_1}, \frac{\partial^2 a_{t_2}}{\partial \gamma_2 \partial \tau_2}\Big)$. Finally, the derivatives of a_t and A_t can be found in Vilca et al. (2014).

References

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

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

Balakrishnan, N.; Lai, C. D. (2009). *Continuous Bivariate Distributions*, Second edition. Springer, New York.

Balakrishnan, N.; Leiva, V.; Sanhueza, A.; Vilca, F. (2009). Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm. *Statistics and Operations Research Transactions*, 33, 171-192.

Barros, M.; Paula, G. A.; Leiva, V. (2008). A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. *Lifetime Data Analysis*, 14, 316-332.

Birnbaum, Z. W.; Saunders, S. C. (1969a). A new family of life distributions. *Journal of Applied Probability*, 6, 637-652.

Birnbaum, Z. W.; Saunders, S.C. (1969b). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, 6, 328-347.

Cabral, C. R. B.; Lachos, V. H.; Zeller, C. B. (2014). Multivariate measurement error models using finite mixtures of skew-Student distributions. *Journal of Multivariate Analysis*, 124, 179-198.

Capitanio, A.; Azzalini, A.; Stanghellini, E. (2003). Graphical models for skew-normal variates. *Scandinavian Journal of Statistics*, 30, 129-144.

Desmond, A. (1985). Stochastic models of failure in random environments. *Canadian Journal of Statistics*, 13, 171-183.

Guiraud, P.; Leiva, V.; Fierro R. (2009). A non-central version of the Birnbaum-Saunders distribution for reliability analysis. *IEEE Transactions on Reliability*, 58, 152-160.

Gupta, A. K.; González-Farías, G.; Domínguez-Molina, J. A. (2004). A multivariate skew normal distribution. *Journal of Multivariate Analysis*, 89, 181-190.

Gupta, R. C. (2006). Reliability studies of bivariate distributions with Pearson type VII conditionals, *Metron*, 64, 239-251.

Jamalizadeh, A; Kundu, D. (2015). A multivariate Birnbaum-Saunders distribution based on multivariate skew normal distribution. *Journal of Japan Statistical Society.*, 45, 1-20.

Kotz, S.; Balakrishnan, N.; Johnson, N. L. (2000). *Continuous Multivariate Distributions - Vol. 1*, Second edition. Wiley, New York.

Kundu, D.; Dey, A. K. (2009). Estimating the parameters of the Marshall-Olkin bivariate Weibull distribution by EM algorithm. *Computational Statistics & Data Analysis*, 53, 956-965.

Kundu, D.; Balakrishnan, N.; Jamalizadeh, A. (2010). Bivariate Birnbaum-Saunders distribution and associated inference. *Journal of Multivariate Analysis*, 101, 113-125.

Kundu, D.; Balakrishnan, N.; Jamalizadeh, A. (2013). Generalized multivariate Birnbaum-Saunders distributions and related inferential issues. *Journal of Multivariate Analysis*, 116, 230-244.

Lange, K. L.; Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, 2, 175-198.

Leiva, V.; Riquelme, M.; Balakrishnan, N.; Sanhueza, A. (2008). Lifetime analysis based on the generalized Birnbaum-Saunders distribution. *Computational Statistics & Data Analysis*, 52, 2079-2097.

Mann, N. R.; Schafer, R. E.; Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*. Wiley, New York.

Marshall, A. W.; Olkin, I. (2007). *Life Distributions*. Springer, New York.

Meintanis, S. G. (2007). Test of fit for Marshall-Olkin distributions with applications. *Journal Statistical Planning and Inference*, 137, 3954-3963.

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

Navarro, J.; Sarabia, J. M. (2013). Reliability properties of bivariate conditional proportional hazard rate models. *Journal of Multivariate Analysis*, 113, 116-127.

Rieck, J. R.; Nedelman, J. R. (1991). A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, 33, 51-60.

Santana, L.; Vilca, L. F.; Leiva, V. (2011). Influence analysis in skew-Birnbaum-Saunders regression models and applications. *Journal of Applied Statistics*, 38, 1633-1649.

Sahu, S. K.; Dey, D. K.; Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics*, 31, 129-150.

Vilca, L. F.; Leiva, V. (2006). A new fatigue life model based on the family of skew-elliptical distributions. *Communications in Statistics-Theory and Methods*, 35, 229-244.

Vilca, L. F.; Santana, L.; Leiva, V.; Balakrishnan, N. (2011). Estimation of extreme percentiles in Birnbaum-Saunders distributions. *Computational Statistics & Data Analysis*, 55, 1665-1678.

Vilca, F.; Balakrishnan, N.; Zeller, C. B. (2014). A robust extension of the bivariate Birnbaum-Saunders distribution and associated inference. *Journal of Multivariate Analysis*, 124, 418-435.