A new Birnbaum-Saunders model based on the skew-normal distribution under the centred parameterization

Nathalia L. Chaves [1], Caio L N Azevedo[1]*, Filidor Vilca-Labra[1], Juvêncio S. Nobre[2]

[1] Department of Statistics, State University of Campinas, Brazil
[2] Department of Statistics and Applied Math, Federal University of Ceará, Brazil

Abstract

In this paper we introduce a new distribution for positive and skewed data by combining the Birnbaum-Saunders (BS) distribution and the centred skew-normal distribution. Several of its properties are developed. Our model accommodates both positively and negatively skewed positive data. Also, we show that our model circumvents some problems related to another BS distribution, based on the skew-normal distribution under the direct parameterization, previously presented in the literature. We developed both maximum likelihood (ML) and Bayesian estimation procedures, comparing them through a suitable simulation study. The convergence of the expectation conditional maximization (ECM) (for ML inference) and MCMC algorithms (for Bayesian inference) were verified and several factors of interest were compared in the parameter recovery study. In general, as the sample size increases, the results indicated that the Bayesian approach provided the most accurate estimates. Finally, our model accommodates the asymmetry of the data, compared with the usual BS model, which is illustrated through real data analysis.

keywords: Birnbaum-Saunders distribution; Skew-normal distribution; Frequentist inference; ECM algorithm; Bayesian inference; MCMC algorithms.

1 Introduction

The Birnbaum-Saunders distribution (BS), which is characterized by two parameters and defined in terms of the standard normal distribution, has been receiving considerable attention over the past few years, since it has been used quite effectively to model positively

*Corresponding author: Caio L N Azevedo, Department of Statistics, State University of Campinas, Mailbox 6065, SP, Brazil. Email: cnaber@ime.unicamp.br

skewed data, especially lifetime data and crack growth data. Since the pioneering work of Birnbaum and Saunders (1969 a, b) was published, several extensions of the BS distribution have been proposed in the literature, under both a frequentist viewpoint and a Bayesian perspective. A random variable T is said to have a BS distribution, if it can be expressed as

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

where Z is a standard normal random variable. The notation $T \sim BS(\alpha, \beta)$ is used in this case, where $\alpha > 0$ and $\beta > 0$ are the shape and scale parameters, respectively.

From a frequentist viewpoint, Birnbaum and Saunders (1969b) presented a discussion on the maximum likelihood estimation of the parameters of the model. Mann et. al. (1974) showed that the BS distribution is unimodal. Engelhardt et. al. (1981) developed confidence intervals and hypothesis tests for each one of the two parameters. Desmond (1985) developed a BS-type distribution based on a biological model. Desmond (1986) investigated the relationship between the BS distribution and the inverse Gaussian distribution. Lu and Chang (1997) used bootstrap methods to construct prediction intervals for future observations. In the linear regression context, Rieck and Nedelman (1991) developed a related log-linear model and showed that it can be used for modeling accelerated life tests or to compare average life time of different populations.

Most of the generalizations of the BS distribution are based on the elliptical and skew-elliptical distributions, in order to obtain more robust and flexible models. Some works developed under symmetric distribution are Diaz-Garcia and Leiva (2005) who generalized the BS distribution using the elliptical distributions that includes the Cauchy, Laplace, Logistic, Normal and Student-t distributions as particular cases. Other works on the BS distributions are: the generalized BS distribution (Leiva et al., 2007b), the Student-t BS distribution (Barros et al., 2008), and the scale-mixture of normal BS distribution (Balakrishnan et al., 2009), among others. More information can be found in Leiva et al. (2009) that is a review of the BS distribution. Other generalizations have been obtained in different ways to those mentioned before, as for example, Owen and Padgett (1999) developed a three-parameter BS distribution and the β-BS distribution (Cordeiro and Lemonte, 2011). Finally, extensions of the BS distribution based on the skew-ellipticals can be found in (Vilca and Leiva, 2006), Leiva et al. (2007b), Leiva et al. (2008) and Vilca et al. (2011). In these works theoretical results were obtained, that extend the properties of the BS and log-BS distributions.

Now, from Bayesian perspective there are few works on BS distribution. The first one is due to Achcar (1993) who developed Bayesian estimation using numerical approximations for the marginal posterior distributions of interest based on the Laplace approximation, Xu and Tang (2011) presented a Bayesian study with partial information, Wang et al. (2016) assumed that the model parameters follow inverse gamma distributions. All these results were studied under a normal distribution for generating the BS distribution.

Research related to studies from a Bayesian perspective for the BS distributions based on skew-normal distribution is a problem which does not appear to have been considered in the literature, which will be one of goals of our proposal. Vilca et al. (2011) considered, from frequentist perspective, the BS distribution based on the standard skew-normal (SN) distribution of Azzalini (1985), whose probability density function (pdf) is given by $f(z) = 2\phi(z)\Phi(\lambda z)$, $z \in \mathbb{R}$. However, even though the SN distribution has been applied with success in several fields, under this distribution there is a problem of singularity of the information matrix, when the asymmetric parameter (λ) is equal to zero. Recently, to overcome this problem, Arellano-Valle and Azzalini (2008) and Azzalini (2013) explored a skew-normal distribution under a convenient parametrization (proposed by Azzalini (1985) and deeper explored by Pewsey (2000), *the centred parametrization* (CP), which allows us to have the information matrix to be non-singular. Moreover, the relative profile log-likelihood function (RPLL) for the the Pearsons index of skewness (γ) exhibits a more regular behaviour, much closer to quadratic functions, and without a stationary point at $\gamma = 0$. The resulting empirical distributions of the estimators under the CP, called CP estimators, are much closer to normality than those obtained under the usual skew-normal distribution, which is called the direct parametrization (DP) estimators. All these good properties derived under the CP may be transferred to a BS distribution based on the SN distribution, under the CP. Dupuis and Mills (1998) developed robust estimation methods in the presence of outliers. They showed that this robust procedure is a powerful alternative technique, when the data present discrepant observations. Rieck (1999) obtained a moment generating function for the Sinh-Normal (normal hyperbolic sine) distribution that can be used to obtain integer or fractional moments of the BS distribution. Barros et al. (2007) considered the Student-t distribution, instead of the normal distribution, to generate a robust BS distribution. Based on this extension, Leiva et al. (2007a), presented inference and diagnostic studies that corresponds to an extension of some results developed by Galea et al. (2004). Diaz-Garcia and Leiva (2005) generalized the BS distribution using a class of distributions with elliptic contours that includes the Cauchy, Laplace, Logistic, Normal and Student-t distributions as particular cases.

The main goal of this work is to propose an alternative to the skew-normal Birnbaum-Saunders (SNBS) distribution proposed by Vilca et al. (2011), considering the skew-normal distribution under the centred parameterization. The resulting BS-type distribution has advantages, in inferential terms over the SNBS distribution, similarly to those properties obtained by using the SN distribution under the CP. The specific goals of this work are: to develop and present the Birnbaum-Saunders distribution based on the SN distribution, highlighting their respective advantages (over the available SNBS distribution proposed by Vilca et al. (2011) and properties. Specifically, estimation procedures under both frenquentist and Bayesian approaches are developed, considering different scenarios based on the levels of some factors of interest, comparing the frequentist and Bayesian estimates. Finally, two real data sets are analyzed, where it is shown some advantages of our model compared with the usual BS distribution.

The paper is outlined as follows. In Section 2, we present our distribution and some motivation for its development. In Section 3, the estimation methods are proposed. In Section 4, some simulation studies are presented. In Section 5, two real data sets are analyzed and, finally, in Section 6, some additional comments and conclusions are provided.

2 The centred skew-normal BS distribution

2.1 The centred skew-normal distribution

First, we recall the centred parameterization of the skew-normal distribution of Azzalini (1985). A random variable Y is said to have a skew-normal distribution of Azzalini (2013) under the CP, denoted by $Y \sim \mathrm{SN}(\mu, \sigma, \gamma)$, where μ, σ and γ are the mean, the standard deviation and is the Pearson's skewness coefficient, respectively, if its probability density function is given by:

$$\begin{aligned} f_Y(y) &= 2\frac{\sigma_z}{\sigma}\phi\Big(\mu_z + \frac{\sigma_z}{\sigma}(y-\mu)\Big)\Phi\Big[\lambda\Big(\mu_z + \frac{\sigma_z}{\sigma}(y-\mu)\Big)\Big] \\ &= \frac{2}{\omega}\phi\Big(\frac{y-\xi}{\omega}\Big)\Phi\Big[\lambda\Big(\frac{y-\xi}{\omega}\Big)\Big], \end{aligned}$$

where $\mu_z = r\delta$, $\sigma_z^2 = 1 - \mu_z^2$, $\lambda = \frac{\gamma^{1/3}s}{\sqrt{r^2+s^2\gamma^{2/3}(r^2-1)}}$, $r = \sqrt{\frac{2}{\pi}}$, $\gamma = r\delta^3(4/\pi - 1)(1-\mu_z^2)^{-3/2}$, $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$, $\xi = \mu - \sigma\gamma^{1/3}s$, $\omega = \sigma\sqrt{1+\gamma^{2/3}s^2}$, and $s = \left(\frac{2}{4-\pi}\right)^{1/3}$. The parameter λ is the asymmetry parameter, see Azzalini (1985). For $\mu = 0$ and $\sigma = 1$, we have the standard SN distribution, denoted by $Y \sim \mathrm{SN}(0, 1, \gamma)$, and its pdf is given by

$$f_Y(y) = \frac{2}{\omega}\phi\Big(\frac{y-\xi}{\omega}\Big)\Phi\Big[\lambda\Big(\frac{y-\xi}{\omega}\Big)\Big],$$

where $\xi = -\gamma^{1/3}s$ and $\omega = \sqrt{1+\gamma^{2/3}s^2}$. For inferential purposes, a useful stochastic representation of Y is given by

$$Y = \frac{1}{\sigma_z}\left\{\delta\,|X_0| + (1-\delta^2)X_1 - \mu_z\right\}, \tag{2}$$

where $X_i \sim \mathrm{N}(0,1), i = 0, 1$ are independent and so $H = |X_0|$ follows a half-normal distribution.

2.2 The proposed distribution

Here, we present a generalization of the usual BS distribution of Birnbaum and Saunders (1969 a, b) and Vilca et al. (2011) based on the centred parameterization of the SN distribution, which is defined similarly to the original BS and the SNBS distributions:

$$T = \beta\left[\frac{\alpha Y}{2} + \sqrt{\left(\frac{\alpha Y}{2}\right)^2 + 1}\right]^2, \tag{3}$$

where $Y \sim \mathrm{SN}(0, 1, \gamma)$, with γ being the Pearson's skewness coefficient in the centred parameterization (CP) of the SN distribution. This resulting distribution is denoted, by $T \sim BS(\alpha, \beta, \gamma)$, similarly to Vilca et al. (2011), but here γ is the Pearson's skewness coefficient in the skew-normal distribution under the CP.

Following the same steps as in the traditional BS distribution, it can be shown that the pdf of proposed distribution is given by

$$f_T(t) = 2\phi\left[a_{t;\mu,\sigma}(\alpha,\beta)\right]\Phi\left[\lambda\, a_{t;\mu,\sigma}(\alpha,\beta)\right]A_{t;\sigma}(\alpha,\beta), t>0, \tag{4}$$

where $a_{t;\mu,\sigma}(\alpha,\beta) = \mu_z + \sigma_z\, a_t(\alpha,\beta)$, $A_{t;\sigma}(\alpha,\beta) = \sigma_z A_t(\alpha,\beta)$, $a_t(\alpha,\beta) = \left(\sqrt{t/\beta} - \sqrt{\beta/t}\right)/\alpha$ and $A_t(\alpha,\beta) = \frac{d}{dt}a_t(\alpha,\beta) = t^{-3/2}(t+\beta)/(2\alpha\beta^{1/2})$, with μ_z and σ_z use the same definition as the SN distribution under the CP, and λ is the asymmetry parameter in the SN distribution of Azzalini (1985). Note that for $\gamma = 0$, we have the usual BS distribution. The mean and variance of T (see Appendix A for more details) are given, respectively, by,

$$\mathbb{E}(T) = \mathbb{E}(T_{BS}) + \Delta_1 \quad \text{and} \quad \text{Var}(T) = \text{Var}(T_{BS}) + \frac{1}{2}\alpha^4\beta)^2\big(\Delta_2 - 3\big) + \Delta_3, \tag{5}$$

where $\Delta_1 = \beta(\alpha U_1)$, $\Delta_2 = \frac{1}{\sigma_z^4}\left(3 - 6\mu_z^2 + 4\mu_z^2\delta^2 - 3\mu_z^4\right)$ and $\Delta_3 = (\alpha\beta)^2[\alpha(U_3 - U_1) - U_1^2]$. Moreover, $T_{BS} \sim BS(\alpha,\beta)$, with $\mathbb{E}(T_{BS}) = \beta\Big(1 + \frac{\alpha^2}{2}\Big)$ and $\text{Var}(T_{BS}) = (\alpha\beta)^2\big(1 + \frac{5\alpha^2}{4}\big)$. To obtain the expected values of $U_k = \mathbb{E}\left(Y^k\sqrt{\alpha^2Y^2+4}\right), k = 1,2,3$, the related integrals need to be solved through some numerical method. This can be done by using the function "integrate" available in the `R` program, see R Development Core Team (2017).

The model parameters are $(\alpha,\beta,\gamma)^\top$ and it will be called *centred parameters* (CP) while the model parameters based the SN distribution of Azzalini (1985), $(\alpha,\beta,\lambda)^\top$ are called *direct parameters* (DP). Figures 1-3 present the density of the SNBS distribution for different values of α, β e γ. Specifically, from Figure 1, we have that for $\alpha = .2$ the density is concentrated around β ($\beta = 1$), and for $\alpha = .8$ the density is more asymmetric with a higher variability. As α increases, the density becomes more flat, positively skewed and more dispersed, as it can be seen in Figure 2 for different values of α, fixing the other parameters. In order to know about the parameter β, Figure 3 shows densities more concentrated around β for different values of α and β, with $\gamma = .9$. It is possible to see that for large values of β, the density is more negatively skewed.

In short, in terms of the three parameters, the distribution tends to be symmetric around β, for $\gamma = 0$ (the usual BS distribution) and for small values of α. Positive asymmetry is observed as α increases, β decreases and/or γ assumes positive values. On the other hand, negative asymmetry is observed as α decreases, β increases and/or γ assumes negative values. Another interesting point is that the SNBS distribution may be negatively skewed, which is an unusual behavior for positive random variables. This feature makes our distribution a very useful alternative for modeling positive skewed data.

The following theorem is very useful to develop both classical and Bayesian approaches. The conditional distributions allow us to implement the EM algorithm, and simplify the Bayesian development. Due to these results, the proposed model can be written hierarchically, as we will see later. This facilitates the implementation of the estimation methods which allows, for example, the use of standard MCMC softwares, such as WinBUGS, OpenBUGS, JAGS or Stan, see Luun et. al. (2000, 2009); Depaoli et. al. (2016); Carpenter et al. (2016).

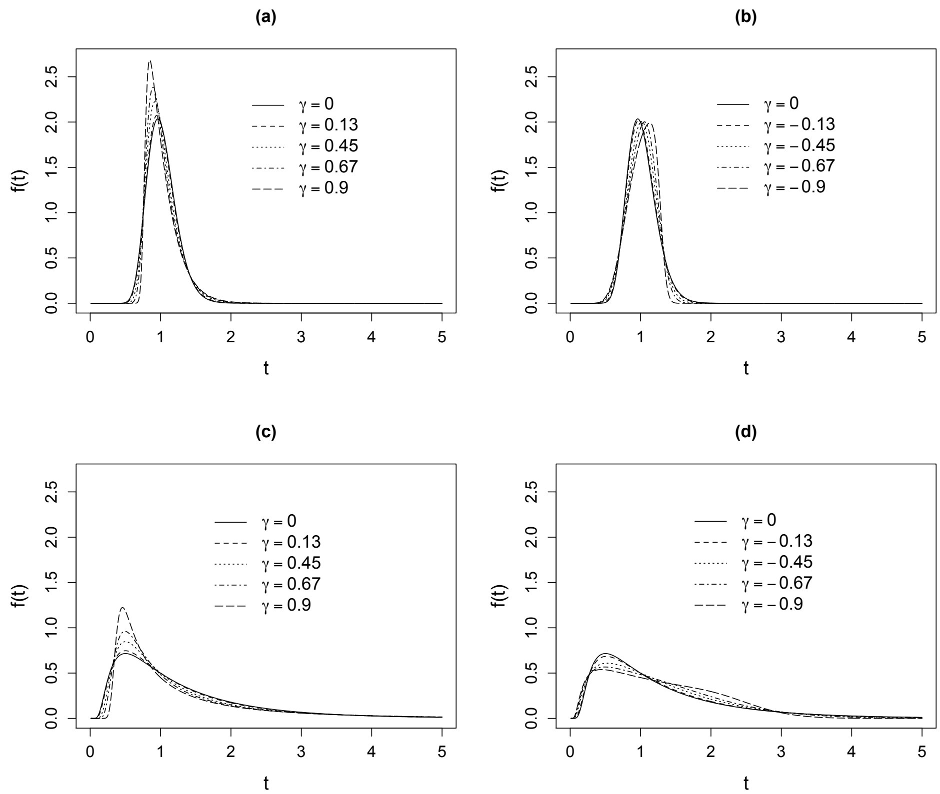


Figure 1: Density of the SNBS distribution for different values of γ, with $\beta = 1$, (a)−(b) $\alpha = .2$ and (c)−(d) $\alpha = .8$.

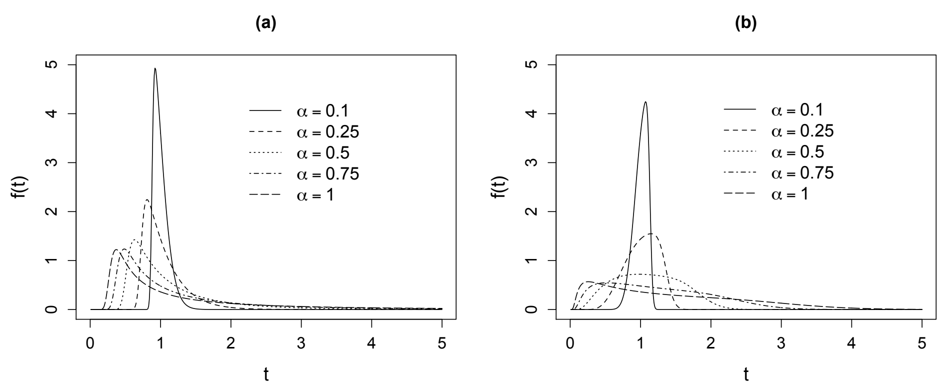


Figure 2: Density of the SNBS distribution for different values of α, with $\beta = 1$, (a) $\gamma = .9$ and (b) $\gamma = -.9$.

Theorem 1. *Let $T \sim SNBS(\alpha, \beta, \gamma)$ as in (3) with Y having the representation given by $Y = \dfrac{1}{\sigma_z}\left[\delta H + \sqrt{1-\delta^2}X_1 - \mu_z\right]$, where H is as in (2). Then,*

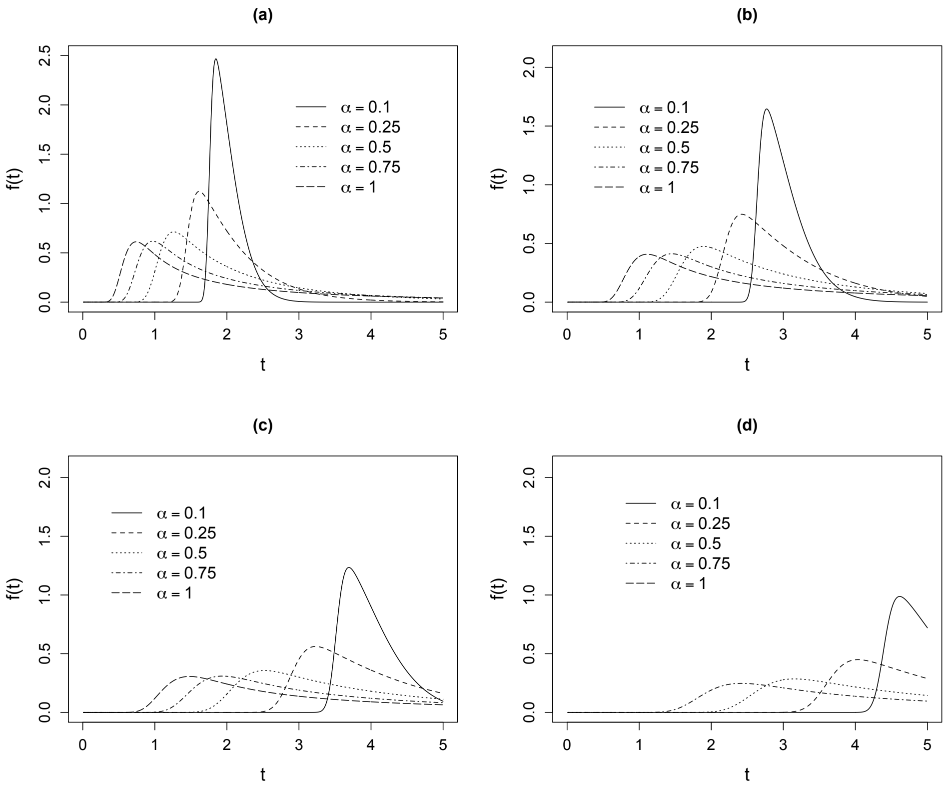


Figure 3: Density of the SNBS distribution with (a) $\beta = 2$, (b) $\beta = 3$, (c) $\beta = 4$, (d) $\beta = 5$; and the same values of α from Figure 2 and $\gamma = .9$

i) The conditional density of T, given $H = h$, can be expressed by

$$f_{T|H}(t|h) = \phi(\nu_h + a_t(\alpha_\delta, \beta))A_t(\alpha_\delta, \beta),$$

where $\alpha_\delta = \alpha\sqrt{\frac{1-\delta^2}{1-r^2\delta^2}}$ and $\nu_h = -\frac{\delta(h-r)}{\sqrt{1-\delta^2}}$.

ii) $f_{H|T}(h|t) = \dfrac{\phi\left(h \mid \delta\sqrt{1-r^2\delta^2}\left(a_t(\alpha,\beta)+\frac{r\delta}{\sqrt{1-r^2\delta^2}}\right), 1-\delta^2\right)}{\Phi\left(\lambda\sigma_z\left(a_t(\alpha,\beta)+\frac{r\delta}{\sqrt{1-r^2\delta^2}}\right)\right)}$*, $h > 0$. Moreover*

$$\mathbb{E}(H|T=t) = \eta_t + W_\Phi\left(\frac{\eta_t}{\tau}\right)\tau \text{ and } \mathbb{E}(H^2|T=t) = \eta_t^2 + \tau^2 + W_\Phi\left(\frac{\eta_t}{\tau}\right)\eta_t\tau,$$

where $\eta_t = \delta\sqrt{1-r^2\delta^2}\big(a_{t_i}(\alpha,\beta) + \frac{r\delta}{\sqrt{1-r^2\delta^2}}\big)$,

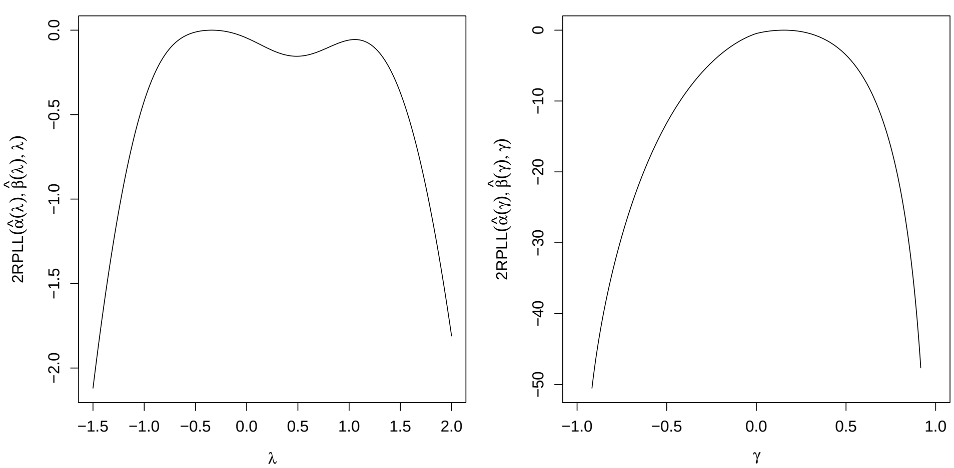
The density in Theorem 1 corresponds to the extended Birnbaum-Saunders (EBS) discussed in Leiva et al. (2008) and is denoted by $\text{EBS}(\alpha_\delta, \beta, \sigma = 2, \nu_h)$ following the parameters of our model. The proof of Theorem 1 is in Appendix B.

2.3 Some motivational remarks on the proposal

i) It is well known that there is some difficulty in estimating the parameters of the SN distribution by the maximum likelihood approach when the asymmetry parameter is close to zero. Some problems seem to persist even if one switched to the Bayesian approach, unless a strongly informative prior is considered, as pointed out by Arellano-Valle and Azzalini (2008).

   The SNBS distribution based on the skew-normal of Azzalini (1985) seems to inherit such problems in the estimation. So, the proposed SNBS distribution, which makes use of the CP of the SN distribution, circumvents problems inherited of the BS distribution obtained by using the SN distribution of Azzalini (1985).

ii) The centred parameterization removes the singularity of the expected Fisher information matrix, which occurs when the asymmetry parameter is equals to 0. Moreover, it circumvents the problem concerning the existence of an inflection point in the relative profiled log-likelihood of this parameter, as it can be seen in Figure 2.3, which refers to the plots of relative profiled log-likelihood (RPLL) function for the asymmetry parameter in the DP for the SNBS distribution (left panel) and the SNBS distribution based on the CP (right panel) (ahead, we present more details).



The RPLL, see Arellano-Valle and Azzalini (2008), corresponds to $\ell(\widehat{\alpha}(\gamma), \widehat{\beta}(\gamma), \gamma) - \ell(\widehat{\alpha}(\gamma), \widehat{\beta}(\gamma), \widehat{\gamma})$, where $\ell(\cdot)$ represents the log-likelihood. The respective plots were constructed by considering a random sample of both distributions, under suitable values of the parameters α, β and γ. We can notice a non-quadratic form of the log-likelihood of the SNBS model, induced by the existence of an inflection point when the asymmetry parameter is very close to zero, making it difficult to obtaining the maximum likelihood estimators. However, the log-likelihood in the CP is well-behaved and it presents a concave shape. Also, there is no inflection point when the asymmetry parameter equals zero.

3 Estimation methods

We will present the maximum likelihood estimation, based on the Expectation Conditional Maximization (ECM) algorithm (Meng and Rubin, 1993), and the Bayesian approach, through MCMC algorithms. Let $T \sim \text{SNBS}(\alpha, \beta, \gamma)$ with representation as in (3). Then, from Theorem 1, $T|(H = h) \sim \text{EBS}(\alpha_\delta, \beta, \sigma = 2, \nu_h)$, where $H = |X_0|$, $\alpha_\delta = \alpha\sqrt{\frac{1-\delta^2}{1-r^2\delta^2}}$ and $\nu_h = -\frac{\delta(h-r)}{\sqrt{1-\delta^2}}$. In Appendix B, we present some results that are useful for obtaining the maximum likelihood estimators.

3.1 The ECM algorithm

In this section, we discuss the maximum likelihood (ML) estimation for the unknown model parameters based on a random sample $T_1, \ldots, T_n$ from $T \sim \text{SNBS}(\alpha, \beta, \gamma), i = 1, \ldots, n$. The log-likelihood function for $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^\top$, is given by $\ell(\boldsymbol{\theta}|\mathbf{t}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}|t_i)$, where

$$\ell_i(\boldsymbol{\theta}|t_i) = \log(2) + \log\left\{\phi\left[a_{t_i;\mu,\sigma}(\alpha, \beta)\right]\right\} + \log\left\{\Phi\left[\lambda\, a_{t_i;\mu,\sigma}(\alpha, \beta)\right]\right\} + \log\left[A_{t_i;\sigma}(\alpha, \beta)\right], \tag{6}$$

with $a_{t_i;\mu,\sigma}(\alpha, \beta)$ and $A_{t_i;\sigma}(\alpha, \beta)$ are as in (4). Instead of considering the direct maximization of (6), we will obtain the ML estimates through the ECM algorithm, since it allows for a more tractable optimization process. In this case, we need to work with the so-called complete likelihood function. Also, instead of estimating $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^\top$, we will estimate $\boldsymbol{\theta} = (\alpha, \beta, \delta)^\top$, where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. This will be done since the related expressions for the implementation of the ECM algorithm are more tractable, both analytically and computationally.

From the Theorem 1, we have the following hierarchical representation

$$\begin{aligned} T_i|H_i = h_i &\overset{ind}{\sim} \text{EBS}(\alpha_\delta, \beta, \sigma = 2, \nu_{h_i}) \\ H_i &\overset{ind}{\sim} \text{HN}(0, 1); i = 1, \ldots, n, \end{aligned}$$

where $\alpha_\delta = \alpha\sqrt{\frac{1-\delta^2}{1-r^2\delta^2}}$ and $\nu_{h_i} = -\frac{\delta(h_i-r)}{\sqrt{1-\delta^2}}$. Then, defining $\mathbf{t}_c = (\mathbf{t}^\top, \boldsymbol{h}^\top)^\top$, where $\mathbf{t} = (t_1, .., t_n)^\top$ and $\boldsymbol{h} = (h_1, \ldots, h_n)^\top$, the complete log-likelihood function for $\boldsymbol{\theta}$ is given by

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{t}_c) &= \sum_{i=1}^n \log f_{T|H}(t_i|h_i) + \sum_{i=1}^n f_H(h_i) \\ &= c - \frac{\delta^2}{2(1-\delta^2)}\sum_{i=1}^n h_i^2 + \frac{r\delta^2}{(1-\delta^2)}\sum_{i=1}^n h_i - \frac{nr^2\delta^2}{2(1-\delta^2)} + \\ &\quad + \frac{\delta\sqrt{1-r^2\delta^2}}{1-\delta^2}\sum_{i=1}^n h_i a_{t_i}(\alpha, \beta) - \frac{r\delta\sqrt{1-r^2\delta^2}}{1-\delta^2}\sum_{i=1}^n a_{t_i}(\alpha, \beta) - \\ &\quad - \frac{1-r^2\delta^2}{2(1-\delta^2)}\sum_{i=1}^n a_{t_i}^2(\alpha, \beta) + \frac{n}{2}\log(1-r^2\delta^2) + \sum_{i=1}^n \log(t_i + \beta) - \\ &\quad - \frac{n}{2}\log(1-\delta^2) - n\log(\alpha) - \frac{n}{2}\log\beta. \end{aligned}$$

For the current value $\boldsymbol{\theta}$, the E-step of the ECM algorithm requires the evaluation of $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathbb{E}\left[\ell(\boldsymbol{\theta}|\mathbf{t}_c)|\mathbf{t},\widehat{\boldsymbol{\theta}}\right]$, where the expectation is taken with respect to the conditional distribution $H|(T=t)$ and evaluated at $\widehat{\boldsymbol{\theta}}$. For the estimate of $\boldsymbol{\theta}$ at $r-$th iteration, say $\boldsymbol{\theta}^{(r)} = (\alpha^{(r)},\beta^{(r)},\delta^{(r)})^\top$, consider $\widehat{h}_i = \mathbb{E}[H_i|\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},t_i]$ and $\widehat{h}_i^2 = \mathbb{E}[H_i^2|\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},t_i]$ that are obtained by using the conditional expectation given in Theorem 1, and are given by

$$\widehat{h}_i \;=\; \widehat{\eta}_{t_i} + W_\Phi\left(\frac{\widehat{\eta}_{t_i}}{\widehat{\tau}}\right)\widehat{\tau} \text{ and } \widehat{h}_i^2 = \widehat{\eta}_{t_i}^2 + \widehat{\tau}^2 + W_\Phi\left(\frac{\widehat{\eta}_{t_i}}{\widehat{\tau}}\right)(\widehat{\eta}_{t_i}\widehat{\tau}), \tag{7}$$

where $\widehat{\eta}_{t_i} = \widehat{\delta}\sqrt{1-r^2\widehat{\delta}^2}\left(a_{t_i}(\widehat{\alpha},\widehat{\beta}) + \frac{r\widehat{\delta}}{\sqrt{1-r^2\widehat{\delta}^2}}\right)$, $\widehat{\tau} = \sqrt{1-\widehat{\delta}^2}$ and $W_\Phi(z) = \phi(z)/\Phi(z)$, $z \in \mathbb{R}$.

Let $\boldsymbol{\theta}^{(r)} = (\alpha^{(r)},\beta^{(r)},\delta^{(r)})^\top$ be the estimate of $\boldsymbol{\theta}$ at the $k-$th iteration. By considering the conditional expectation T_i, given $H_i = h_i$, we have that the complete log-likelihood is given by

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) &= \mathbb{E}[\ell(\boldsymbol{\theta}|\mathbf{t}_c)|\mathbf{t},\widehat{\boldsymbol{\theta}}^{(r)}] \\
&= c - \frac{\delta^{2(r)}}{2\left(1-\delta^{2(r)}\right)}\sum_{i=1}^n \widehat{h}_i^{2(r)} + \frac{r\delta^{2(r)}}{\left(1-\delta^{2(r)}\right)}\sum_{i=1}^n \widehat{h}_i^{(r)} - \frac{nr^2\delta^{2(r)}}{2\delta^{2(r)}} \\
&\quad + \frac{\delta^{(r)}\sqrt{1-r^2\delta^{2(r)}}}{\alpha^{(r)}\left(\delta^{2(r)}\right)}\sum_{i=1}^n \widehat{h}_i^{(r)} a_{t_i}(1,\beta^{(r)}) - \frac{r\delta^{(r)}\sqrt{1-r^2\delta^{2(r)}}}{\alpha^{(r)}\left(1-\delta^{2(r)}\right)}\sum_{i=1}^n a_{t_i}(1,\beta^{(r)}) \\
&\quad - \frac{1-r^2\delta^{2(r)}}{2\alpha^{2(r)}\left(1-\delta^{2(r)}\right)}\sum_{i=1}^n \left[a_{t_i}(1,\beta^{(r)})\right]^2 + \frac{n}{2}\log\left(1-\delta^{2(r)}\right) \\
&\quad + \sum_{i=1}^n \log\left(t_i+\beta^{(r)}\right) - \frac{n}{2}\log\left(1-\delta^{2(r)}\right) - n\log\left(\alpha^{(r)}\right) \\
&\quad - \frac{n}{2}\log\left(\beta^{(r)}\right).
\end{aligned}$$

Hence, the ECM algorithm corresponds to iterate the following steps:
E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(r)}$, compute $\widehat{h}_i$ and $\widehat{h}_i^2$, for $i = 1,\ldots,n$ using results in (7);

CM-step 1: Fix $\widehat{\beta}^{(r)}$ and $\widehat{\delta}^{(r)}$ and update $\widehat{\alpha}^{(r)}$ through the positive root of the following quadratic equation

$$\widehat{\alpha}^2 + \hat{b}^{(r)}\widehat{\alpha} + \hat{c}^{(r)} = 0,$$

where

$$\begin{aligned}
\hat{b}^{(r)} &= \frac{1}{(1-\widehat{\delta}^{2(r)})}\left[\widehat{\delta}^{(r)}\sqrt{1-r^2\widehat{\delta}^{2(r)}}\frac{1}{n}\sum_{i=1}^n \widehat{h}_i a_{t_i}(1,\widehat{\beta}^{(r)}) - r\widehat{\delta}^{(r)}\sqrt{1-r^2\widehat{\delta}^{2(r)}}\frac{1}{n}\sum_{i=1}^n a_{t_i}(1,\widehat{\beta}^{(r)})\right], \\
\hat{c}^{(r)} &= -\frac{(1-r^2\widehat{\delta}^{2(r)})}{(1-\widehat{\delta}^{2(r)})}\frac{1}{n}\sum_{i=1}^n \widehat{h}_i\left[a_{t_i}(1,\widehat{\beta}^{(r)})\right]^2.
\end{aligned}$$

That is, $\widehat{\alpha}^{(r+1)} = \frac{-b+\sqrt{b^2-4c}}{2}$;

CM-step 2: Fix $\widehat{\alpha}^{(r+1)}$ and update $\widehat{\beta}^{(r)}$ and $\widehat{\delta}^{(r)}$ using

$$\widehat{\beta}^{(r+1)} = \underset{\beta}{\text{argmax}} \quad Q\left(\widehat{\alpha}^{(r+1)}, \beta, \widehat{\delta}^{(r)}\right) \text{ and } \widehat{\delta}^{(r+1)} = \underset{\delta}{\text{argmax}} \quad Q\left(\widehat{\alpha}^{(r+1)}, \widehat{\beta}^{(r+1)}, \delta\right).$$

The updating of $\widehat{\beta}^{(r+1)}$ and $\widehat{\delta}^{(r+1)}$ need to be done through some numerical optimization method. In this work we use the function `optim`, available on the software `R` (see R Development Core Team (2017)), considering the L-BFGS-B optimization algorithm (Byrd et al., 1995).

We start the ECM algorithm with initial values, say, $\widehat{\alpha}^{(0)}$, $\widehat{\beta}^{(0)}$ and $\widehat{\delta}^{(0)}$. We may consider

$$\widehat{\alpha}^{(0)} = [2(s/v) - 1]^{1/2} \text{ and } \widehat{\beta}^{(0)} = (sv)^{1/2},$$

where $s = \frac{1}{n}\sum_{i=1}^{n} t_i$ and $v = \left[\frac{1}{n}\sum_{i=1}^{n} 1/t_i\right]^{-1}$, as in Vilca et al. (2011). After getting $\widehat{\alpha}^{(0)}$ and $\widehat{\beta}^{(0)}$, we define $z_i = \frac{1}{\widehat{\alpha}^{(0)}}\left[\sqrt{\frac{t_i}{\widehat{\beta}^{(0)}}} - \sqrt{\frac{\widehat{\beta}^{(0)}}{t_i}}\right]$, $i = 1, 2, ..., n$, which are observations of the SN distribution under the CP. Thus, $\widehat{\delta}^{(0)}$ can be obtained by maximizing (numerically) the log-likelihood function of SN distribution with respect to δ, which is given by

$$\ell(\theta) = \sum_{i=1}^{n}\Big[\log(2) + \log(\sigma_z) + \log\left[\phi\left(\mu_z + \sigma_z y_i\right)\right] + \log\Phi\left[\lambda(\mu_z + \sigma_z y_i)\right]\Big].$$

According to Vilca et al. (2011), for ensuring that the true ML estimates are obtained, it is recommended to run the ECM algorithm using a range of different starting values and checking whether all of them result in similar estimates. The steps of the ECM algorithm are repeated until a suitable convergence is attained, for example, using $\left\|\boldsymbol{\theta}^{(r)} - \boldsymbol{\theta}^{(r-1)}\right\| < \varepsilon, \varepsilon > 0$.

3.1.1 The Observed information matrix

As seen in (5), the calculation of the moments of T requires the use of numerical integration methods, which makes it difficult the obtaining of the expected Fisher information. Therefore, we shall work with the observed Fisher information. It is worthwhile to mention that, under certain regularity conditions $\widehat{\boldsymbol{\theta}} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_3(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}})$. We approximate $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}$ by $I^{-1}(\boldsymbol{\theta})$, with $I(\boldsymbol{\theta}) = -\ddot{\ell}$. Here, $\ddot{\ell} = [\ddot{\ell}_{\theta_1\theta_2}]$, $\theta_1, \theta_2 = \alpha, \beta$ or γ is the Hessian matrix, where $\ddot{\ell}_{\theta_1\theta_2} = \ddot{\ell}_{\theta_2\theta_1} = \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\theta_1\theta_2} = \sum_{i=1}^{n}\frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\theta_1\theta_2}$. The second derivatives of $\ell_i(\boldsymbol{\theta})$ are provided in Appendix C. The approximate standard errors (SE) of $\widehat{\boldsymbol{\theta}}$ can be estimated by using the square roots of the diagonal elements of $I^{-1}(\boldsymbol{\theta})$, replacing $\boldsymbol{\theta}$ by their respective ML estimates.

3.1.2 A short simulation study

A short simulation study was conducted to assess the behavior of the ECM algorithm, in terms of parameter recovery, and the accuracy of the standard errors calculated through the observed Fisher information. $N = 1,000$ replicas were generated with a size of $n = 500$, considering $\boldsymbol{\theta}^\top = (\alpha, \beta, \gamma) = (.5, 1, .67)$, which induce a strong positively skewed behavior of the SNBS distribution. In Table 1 we can see the mean of the estimates ($\bar{\widehat{\boldsymbol{\theta}}}$), the mean of the theoretical (asymptotic) standard errors (SE) (SE($\widehat{\boldsymbol{\theta}}$)) and the empirical SE (SE_{emp}). We can notice that the parameters were well recovered and that the empirical SE are close to the theoretical ones, which indicates that the use of the observed information matrix, to obtain the standard errors, is appropriate.

Table 1: Results of the short simulation study.

	$\widehat{\boldsymbol{\theta}}$	SE($\widehat{\boldsymbol{\theta}}$)	SE_{emp}
$\widehat{\alpha}$	.495	.019	.021
$\widehat{\beta}$	1.003	.032	.028
$\widehat{\gamma}$	.667	.015	.012

3.2 Bayesian inference

In this Section we present the developments related to the Bayesian inference through MCMC algorithms.

3.2.1 Prior, posterior and full conditional distributions

For Bayesian inference we consider both original and complete likelihood. The first is given by

$$L(\boldsymbol{\theta}|\mathbf{t}) = \prod_{i=1}^{n} 2\phi\left[a_{t_i;\mu,\sigma}(\alpha,\beta)\right] \Phi\left[\lambda\, a_{t_i;\mu,\sigma}(\alpha,\beta)\right] A_{t_i;\sigma}(\alpha,\beta).$$

We assume the following prior distributions: $\alpha \sim \text{gamma}(\mathrm{r}_\alpha; \lambda_\alpha)$, $\beta \sim \text{gamma}(r_\beta; \lambda_\beta)$ and $\gamma \sim \text{U}(a; b)$, are mutually independent, where $\alpha \sim \text{gamma}(r, \lambda)$ stands for a gamma distribution such that $\mathbb{E}(\alpha) = \dfrac{r}{\lambda}$ and $\text{Var}(\alpha) = \dfrac{r}{\lambda^2}$ and $U(a; b)$ stands for a continuous uniform distribution over the interval $[a, b]$. Combining the likelihood with the prior distribution, we have that the joint posterior distribution is given by:

$$\pi(\boldsymbol{\theta}|\mathbf{t}) \quad \propto \quad \alpha^{\mathrm{r}_\alpha - 1}\beta^{\mathrm{r}_\beta - 1} \exp\left\{-(\alpha\lambda_\alpha + \beta\lambda_\beta)\right\} \prod_{i=1}^{n} \phi\left[a_{t_i;\mu,\sigma}(\alpha,\beta)\right] \Phi\left[\lambda\, a_{t_i;\mu,\sigma}(\alpha,\beta)\right] A_{t_i;\sigma}(\alpha,\beta)$$

and the full conditional distributions are given by:

$$
\begin{aligned}
\pi(\alpha|\beta,\gamma,\mathbf{t}) &\propto \alpha^{\mathrm{r}_\alpha-1}e^{-\alpha\lambda_\alpha}\prod_{i=1}^{n}\phi\left[a_{t_i;\mu,\sigma}(\alpha,\beta)\right]\Phi\left[\lambda\, a_{t_i;\mu,\sigma}(\alpha,\beta)\right]A_{t_i;\sigma}(\alpha,\beta)\\
\pi(\beta|\alpha,\gamma,\boldsymbol{t}) &\propto \beta^{\mathrm{r}_\beta-1}e^{-\beta\lambda_\beta}\prod_{i=1}^{n}\phi\left[a_{t_i;\mu,\sigma}(\alpha,\beta)\right]\Phi\left[\lambda\, a_{t_i;\mu,\sigma}(\alpha,\beta)\right]A_{t_i;\sigma}(\alpha,\beta)\\
\pi(\gamma|\alpha,\beta,\mathbf{t}) &\propto \prod_{i=1}^{n}\phi\left[a_{t_i;\mu,\sigma}(\alpha,\beta)\right]\Phi\left[\lambda\, a_{t_i;\mu,\sigma}(\alpha,\beta)\right]A_{t_i;\sigma}(\alpha,\beta).
\end{aligned}
$$

On the other hand, the complete likelihood is given by:

$$
\begin{aligned}
L(\boldsymbol{\theta}|\mathbf{t}_c) &= \prod_{i=1}^{n} f_{T|H}(t_i|h_i)f_H(h_i)\\
&= \prod_{i=1}^{n}\sqrt{2/\pi}\,\phi\left[\nu_{h_i}+a_{t_i}(\alpha,\beta)\right]A_{t_i}(\alpha,\beta)\exp\left\{-\frac{h_i^2}{2}\right\}.
\end{aligned}
$$

Similarly, combining the complete likelihood with the prior distribution, we have that the posterior density is given by

$$
\pi(\boldsymbol{\theta},\boldsymbol{h}|\mathbf{t}) \propto \alpha^{\mathrm{r}_\alpha-1}\beta^{\mathrm{r}_\beta-1}\prod_{i=1}^{n}\phi\left[a_{t_i,h_i}(\alpha,\beta)\right]A_{t_i}(\alpha,\beta)\exp\left\{-\frac{1}{2}\left[h_i^2+2\alpha\lambda_\alpha+2\beta\lambda_\beta\right]\right\}
$$

and the full conditional distributions, are given by

$$
\begin{aligned}
\pi(\boldsymbol{h}|\alpha,\beta,\gamma,\mathbf{t}_c) &\propto \prod_{i=1}^{n}\phi\left[a_{t_i,h_i}(\alpha,\beta)\right]A_{t_i}(\alpha,\beta)\exp\left\{-\frac{h_i^2}{2}\right\},\\
\pi(\alpha|\beta,\gamma,\mathbf{t}_c) &\propto \alpha^{\mathrm{r}_\alpha-1}\prod_{i=1}^{n}\phi\left[a_{t_i,h_i}(\alpha,\beta)\right]A_{t_i}(\alpha,\beta)\exp\left\{-\frac{1}{2}\left[h_i^2+2\alpha\lambda_\alpha\right]\right\}\\
\pi(\beta|\alpha,\gamma,\mathbf{t}_c) &\propto \beta^{\mathrm{r}_\beta-1}\prod_{i=1}^{n}\phi\left[a_{t_i,h_i}(\alpha,\beta)\right]A_{t_i}(\alpha,\beta)\exp\left\{-\frac{1}{2}\left[h_i^2+2\beta\lambda_\beta\right]\right\}\\
\pi(\gamma|\alpha,\beta,\mathbf{t}_c) &\propto \prod_{i=1}^{n}\phi\left[a_{t_i,h_i}(\alpha,\beta)\right]A_{t_i}(\alpha,\beta)\exp\left\{-\frac{h_i^2}{2}\right\},
\end{aligned}
$$

where $a_{t_i,h_i}(\alpha,\beta)=\nu_{h_i}+a_{t_i}(\alpha,\beta)$.

These full conditional distributions do not correspond to known distributions, but they can be simulated through some auxiliary algorithm such as the Metropolis-Hastings, slice sampling or adaptive rejection. All these algorithms can be easily implemented in `R` program. On the other hand, with is the approach pursued here, we can use a general MCMC computational framework, such as the `OpenBUGS` program, see Luun et. al. (2009). In this case, it is necessary to provide the original or the complete likelihood, along with the prior distributions, such that the full conditional distributions will be simulated through suitable algorithms, following a pre-defined hierarchy available on `OpenBUGS`. We made all simulations using the `R` package `R2OpenBUGS`.

3.3 Model selection criteria

The Akaike information criterion (AIC) proposed by Akaike (1974) is based on the penalized likelihood by the number of model parameters, while for the Bayesian information criterion (BIC), proposed by Schwarz (1978), in addition to the number of parameters, the sample size is also considered in a penalization term. The smaller the values of AIC and BIC, the better the model fits. In the frequentist context, such selection criteria are defined as follows: $AIC = -2\ell(\boldsymbol{\theta}|\mathbf{t}) + 2k$ and $BIC = -2\ell(\boldsymbol{\theta}|\mathbf{t}) + k\log(n)$, where $\ell(\boldsymbol{\theta}|\mathbf{t})$ is defined in (6), k is the number of model parameters and n is the number of observations.

On the other hand, according to Spiegelhalter et al. (2014), the selection criteria, in the Bayesian context, are obtained considering the posterior density of the parameters. To introduce the Bayesian criteria, we first define $\mathrm{D}(\boldsymbol{\theta}) = -2\ell(\boldsymbol{\theta}|\mathbf{t})$. Also, let $\boldsymbol{\theta}^{(m)}$, $m = 1, ..., M$, be the $m-$th value of the valid MCMC sample, that is, the MCMC sample obtained after discarding the burn-in and considering a proper spacing (lag) between the values. Finally, let $\overline{\boldsymbol{\theta}}$ be the vector with the posterior expectation, of each parameter, based on the valid MCMC sample, and $\overline{\mathrm{D}(\boldsymbol{\theta})} = \frac{1}{M}\sum_{m=1}^{M} D(\boldsymbol{\theta}^{(m)})$. Denote also the deviance by $\mathrm{D}(\overline{\boldsymbol{\theta}}) = -2\ell(\overline{\boldsymbol{\theta}}|\mathbf{t})$, and the deviance information criterion (DIC) by $\mathrm{DIC} = \mathrm{D}(\overline{\boldsymbol{\theta}}) + 2p_D$, where $p_D = \overline{\mathrm{D}(\boldsymbol{\theta})} - \mathrm{D}(\overline{\boldsymbol{\theta}})$.

The EAIC (posterior expectation of AIC) and EBIC (posterior expectation of BIC) are given, respectively, by $\mathrm{EAIC} = D(\bar{\boldsymbol{\theta}}) + 2p_{\boldsymbol{\theta}}$ and $\mathrm{EBIC} = \mathrm{D}(\bar{\boldsymbol{\theta}}) + p_{\boldsymbol{\theta}}\log(n)$, where $p_{\boldsymbol{\theta}}$ is the total number of parameters. Finally, let $L(\boldsymbol{\theta}|t_i)$, $i = 1, ..., n$, be the original likelihood related to the ith observation. Then, the LPML (logarithm of the pseudo-marginal likelihood) is calculated as $\mathrm{LPML} = \sum_{i=1}^{n} \ln(\widehat{\mathrm{CPO}}_i)$ where

$$\widehat{\mathrm{CPO}}_i = \left\{ \frac{1}{M} \sum_{m=1}^{M} \frac{1}{L\left(\boldsymbol{\theta}^{(m)}|t_i\right)} \right\}^{-1}.$$

The smaller the values of DIC, EAIC, EBIC, the better the model fit, which occurs the opposite with the LPML.

4 Simulation studies

We considered a total of 30 scenarios, resulting from the combination of the levels of: three different sample sizes (n) (10, 50, 200), under $\alpha \in (.5; 1.5)$, $\beta = 1$ and $\gamma \in (-.67; -.45; 0; .45; .67)$. The sample sizes were chosen in order to verify the proprieties of the estimators, as consistency, and to compare their behavior, in terms of accuracy. The values of α and β were chosen in order to induce different shapes and small variability, whereas the values of γ induce from null to high positive and negative asymmetry. We calculated the usual statistics to measure the accuracy of the estimates: bias, variance (Var), root mean squared error (RMSE) and absolute value of the relative bias (AVRB). Let θ be the parameter of interest, $\widehat{\theta}_r$ be some estimate related to the replica r and $\overline{\widehat{\theta}} = (1/R)\sum_{r=1}^{R} \widehat{\theta}_r$. The adopted statistics are: $\mathrm{BIAS} = \overline{\widehat{\theta}} - \theta$, $\mathrm{Var} = (1/R)\sum_{r=1}^{R}(\widehat{\theta}_r - \overline{\widehat{\theta}})^2$, $\mathrm{RMSE} = \sqrt{(1/R)\sum_{r=1}^{R}(\theta - \widehat{\theta}_r)^2}$, $\mathrm{AVRB} = |\overline{\widehat{\theta}} - \theta|/|\theta|$.

The usual tools for monitoring the convergence of the MCMC algorithms, see Gamerman and Lopes (2006), indicate that a burn-in of 4,000, a thin of 100, simulating a total of 100,000 values, were enough to produce valid MCMC samples of size 1,000 for each parameter.

Since the other results are similar and can be found in the supplementary material, we present only those related to the scenario where $\alpha = .5$, $\beta = 1$, $\gamma = -.67$, varying the value of the sample size. We considered ($< .001$) to represent positive values (statistics and/or estimates) and ($> -.001$) to denote negative values, when they are close to zero. In addition, we will refer to the maximum likelihood estimates by ML and the Bayesian estimates as augmented, when the "augmented" likelihood was used, and "original" when the original likelihood was considered. The selected results can be seen in Table 2.

In a general way, we can see that, as the sample size increases, the estimates obtained by the three approaches tend to be the correspondent true values. When the true α value is equal to .5, the ML estimates are more accurate than the Bayesian ones, especially considering the bias and AVRB metrics. In other scenarios (not shown), when the true α value is equal to 1.5, the opposite occurs for all sample sizes. Concerning β and γ, it is possible to notice that, under the smallest sample size ($n = 10$), the ML approach presents more accurate estimates than the Bayesian ones. On the other hand, for $n = 50$ and $n = 200$, Bayesian estimates of both parameters are closer to the respective true values. In conclusion, as the sample size increases, the results from the simulation studies (including those not shown here) indicate that the Bayesian approach provided the most accurate estimates. Moreover, we can notice that the original and augmented approach performed quite similarly. Therefore, we decide to use the original (likelihood) approach, since it is easier to implement and faster to produce the results.

5 Real data analysis

In this section, we present two real data analysis. The first data set relates to the amount of time spent (in hours) on eating. The second data set concerns the prices (in euros) of bottles (75 cl) of Barolo wine.

5.1 Time spent in having meal

We analyzed the data set originally discussed in Volle (1985) and also analyzed by Lemonte et al. (2015). It concerns the time spent (in hours) in two categories of activities during 100 days in 1976. We analyzed the time spent in having a meal for each one of 28 selected subjects.

In Table 3 and Figure 4, we present some descriptive analysis. It is possible to see that the distribution is positively skewed and more concentrated in the first class [0,200].

We fitted the SNBS and BS distributions using the augmented and the ML methods. The prior distributions were the same used in Section 3. In Table 4, in addition to the respective ML estimates, standard errors and the confidence intervals, we also present model selection criteria. In Table 5, in addition to the posterior expectations (PE), the posterior standard

Table 2: Results of simulation study (PRC) - $\gamma = -.67$.

Parameter	n	Method	Mean	Variance	Bias	RMSE	AVRB
α	10	Augmented	.577	< **.001**	.077	**.081**	.154
		Original	.578	.001	.078	.082	.156
		ML	**.520**	.071	**.020**	.267	**.040**
	50	Augmented	.511	< **.001**	.011	.016	.022
		Original	.511	< **.001**	.011	**.015**	.021
		ML	**.498**	.001	**-.002**	.033	**.004**
	200	Augmented	**.502**	< **.001**	**.002**	**.005**	**.004**
		Original	**.502**	< **.001**	**.002**	**.005**	**.004**
		ML	.490	< **.001**	-.010	.012	.019
β	10	Augmented	1.006	< **.001**	.006	.023	.006
		Original	**1.004**	< **.001**	**.004**	**.021**	**.004**
		ML	1.105	.214	.105	.474	.105
	50	Augmented	.996	< **.001**	-.004	**.009**	.004
		Original	**.997**	< **.001**	**-.003**	**.009**	**.003**
		ML	1.039	.018	.039	.140	.039
	200	Augmented	**.999**	< **.001**	**-.001**	.005	**.001**
		Original	**.999**	< **.001**	**-.001**	.005	**.001**
		ML	.997	< **.001**	-.003	**.004**	.003
γ	10	Augmented	-.157	.067	.513	.575	.766
		Original	-.182	.054	.488	.540	.728
		ML	**-.603**	**.028**	**.067**	**.179**	**.100**
	50	Augmented	-.493	.059	.177	.301	.264
		Original	-.505	.049	.165	.276	.247
		ML	**-.569**	**.012**	**.101**	**.148**	**.150**
	200	Augmented	**-.614**	.017	**.056**	.142	**.083**
		Original	-.601	.015	.069	**.141**	.103
		ML	-.523	**.002**	.147	.153	.220

Table 3: Descriptive statistics for the time spent in having a meal.

Mean	Median	Minimum	Maximum	SD	Asymmetry	Kurtosis
294.286	130	100	960	330.529	1.317	-.229

deviations (PSD) and the 95% equi-tailed credibility intervals (CI), we also present Bayesian criteria.

Tables 4 and 5 show that the estimates of α and γ (under the SNBS model) indicate that the distribution of the time spent is strongly positively skewed. Even though the original BS distribution can present an asymmetric behavior, the SNBS is more flexible, since it presents an additional parameter that (also) helps to control the asymmetry. Notice also that, under ML or Bayesian approach, we have indications that the asymmetry parameter is different from zero, since the zero does not belong to neither the confidence interval, neither to the credibility interval. Also, we construct QQ plots with simulated envelopes (Atkinson, 1985) using ML estimates (Figure 5) and Bayesian estimates (Figure 6). From both figures, we detect that none of the models provides a good fit for the time spent in having a meal

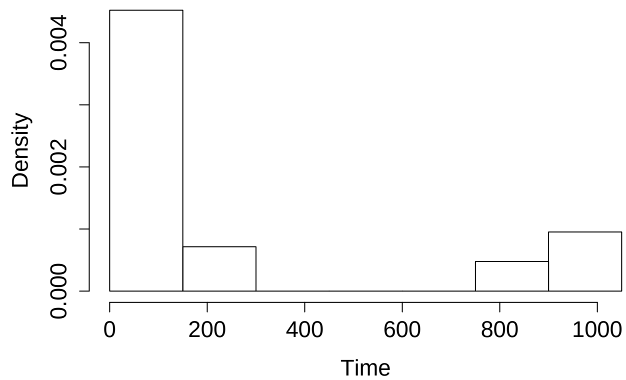


Figure 4: Histogram of the time spent in having a meal.

data set. Specifically, from the QQ plot shown in Figure 5 (right) and Figure 6 (right), we note that there are many observations outside confidence bands, related to the fit of the SNBS distribution. On the other hand, it is noted that all criteria selected the SNBS model. Therefore, in terms of model fit, we have an opposite conclusion, compared with those related to model selection criteria with also an indication that a more suitable model should be considered.

Table 4: ML Estimates, standard errors, 95% confidence intervals for the parameters of the SNBS and BS distribution and model selection criteria.

Parameter	SNBS			BS		
	Estimate	SE	$CI_{95\%}$	Estimate	SE	$CI_{95\%}$
α	.718	.102	[.680; .756]	.899	.126	[.625; 1.146]
β	230.345	26.127	[220.668; 240.023]	214.255	33.543	[148.511; 279.999]
γ	.976	.032	[.964; .988]	-	-	-
AIC		345.835			366.485	
BIC		349.831			369.149	

5.2 Prices of bottles of Barolo wine

We analyzed the data set discussed in Azzalini (2013). It concerns the prices (in euros) of bottles (75 cl) of Barolo wine. The data have been obtained in July 2010 from the websites of four Italian wine resellers, selecting only quotations of Barolo wine, which is produced in the Piedmont region of Italy. The price does not include the delivery charge.

In Table 6 and Figure 7, we present some descriptive analysis. It is possible to see that the distribution is positively skewed and more concentrated in the first class [0,100].

We fitted the SNBS and BS distributions using the Bayesian augmented and the ML methods. The prior distributions were the same used in Section 3. In Table 7, in addition

Table 5: Posterior expectations (PE), posterior standard deviations (PSD), equi-tailed 95% credibility intervals and model selection criteria.

Parameter	SNBS			BS		
	PE	PSD	$CI_{95\%}$	PE	PSD	$CI_{95\%}$
α	.764	.115	[.585; 1.034]	.942	.137	[.721; 1.250]
β	230.009	32.476	[184.493; 310.242]	216.966	35.616	[154.697; 294.925]
γ	.942	.070	[.772; .994]	-	-	-
EAIC		349.785			368.603	
EBIC		353.782			371.268	
DIC		1030.058			1091.813	
LPML		-172.674			-183.194	

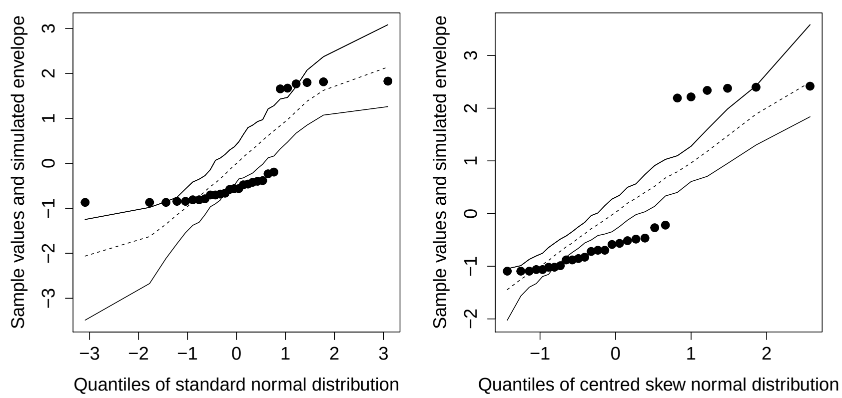


Figure 5: QQ plots with envelopes for BS (left) and SNBS (right) distributions for the time spent in having a meal - ML estimates.

Table 6: Descriptive statistics for the prices of bottles of Barolo wine.

Mean	Median	Minimum	Maximum	SD	Asymmetry	Kurtosis
124.617	72	14	1000	37.041	2.903	12.982

to the respective ML estimates, standard errors and the confidence intervals, we also present model selection criteria. In Table 8, in addition to the posterior expectations (PE), the posterior standard deviations (PSD) and the 95% equi-tailed credibility intervals (CI), we also present Bayesian criteria.

Tables 7 and 8 show that the estimates of α and γ (under the SNBS model) indicate that the distribution of the time spent is strongly positively skewed. Notice also that, under ML or Bayesian approach, we have indications that the asymmetry parameter is different from zero, since the zero does not belong to neither the confidence interval, neither to the credibility interval. Moreover, it is noted that all criteria selected the SNBS model. Also, we built QQ plots with simulated envelopes using ML estimates (Figure 8) and Bayesian

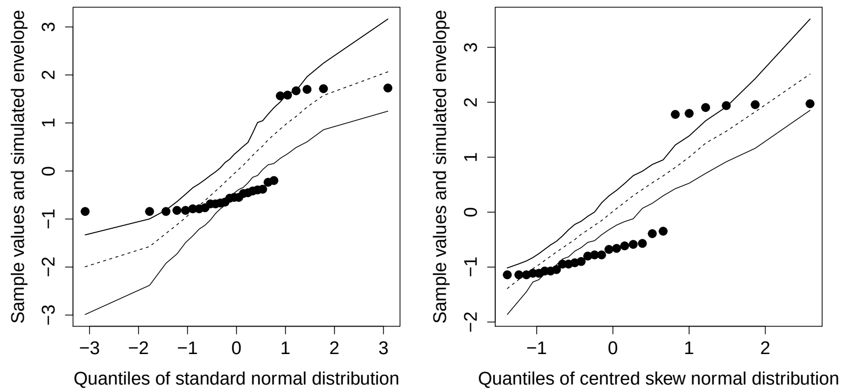


Figure 6: QQ plots with envelopes for BS (left) and SNBS (right) distributions for the time spent in having a meal- Bayesian estimates.

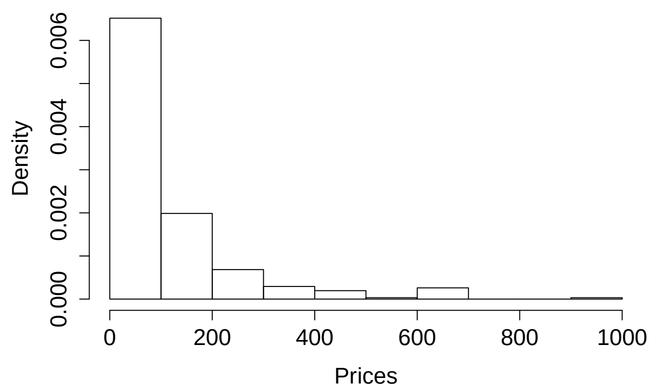


Figure 7: Histogram of the prices of bottles of Barolo wine.

estimates (Figure 9). From both figures, we detect that the SNBS distribution provides a better fit than the BS distribution for the prices of bottles of Barolo wine data. Specifically, from the QQ plot shown in Figure 8 (left) and Figure 9 (left), we note that the observations appear to form a slight upward-facing concave. However, the QQ plot shown in Figure 8 (right) and Figure 9 (right) indicates that the SNBS distribution offers an excellent fit to the prices of bottles of Barolo wine data, provided that the majority of observations are inside of the envelope.

Table 7: ML Estimates, standard errors, 95% confidence intervals for the parameters of the SNBS and BS distribution and model selection criteria.

Parameter	SNBS			BS		
	Estimate	SE	$CI_{95\%}$	Estimate	SE	$CI_{95\%}$
α	.840	.041	[.835; .845]	.853	.034	[.849; .857]
β	89.506	5.088	[88.937; 90.075]	92.291	5.209	[91.708; 92.874]
γ	.710	.067	[.702; .717]	-	-	-
AIC		3434.803			3472.826	
BIC		3445.983			3480.279	

Table 8: Posterior expectations (PE), posterior standard deviations (PSD), equi-tailed 95% credibility intervals and model selection criteria.

Parameter	SNBS			BS		
	PE	PSD	$CI_{95\%}$	PE	PSD	$CI_{95\%}$
α	.844	.037	[.775; .917]	.858	.035	[.794; .929]
β	89.576	3.911	[82.260; 97.871]	92.444	4.264	[84.778; 101.302]
γ	.690	.070	[.541; .809]	-	-	-
EAIC		3437.879			3474.893	
EBIC		3449.060			3482.346	
DIC		10292.690			10410.620	
LPML		-1718.110			-1736.669	

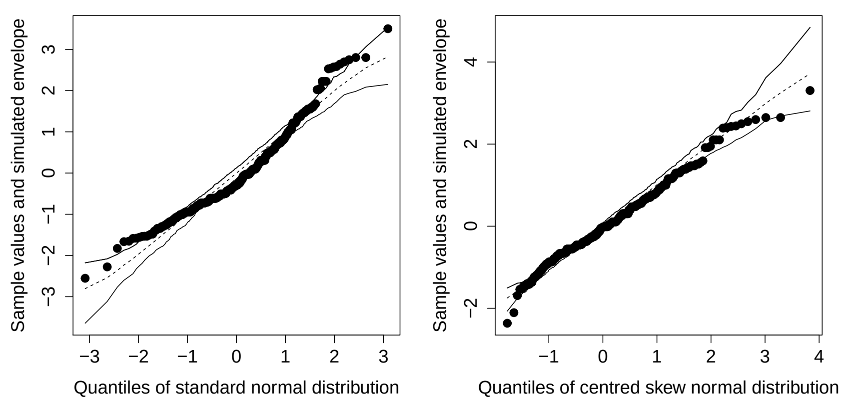


Figure 8: QQ plots with envelopes for BS (left) and SNBS (right) distributions for the prices of bottles of Barolo wine.

6 Concluding Remarks

In this paper we introduce a new distribution for positive data with positive and negative asymmetric behavior, by combining the BS and the skew-normal distributions under the CP. We developed both ML and Bayesian estimation, comparing them through a suitable simulation study. The convergence of the conditional expectation maximization and MCMC

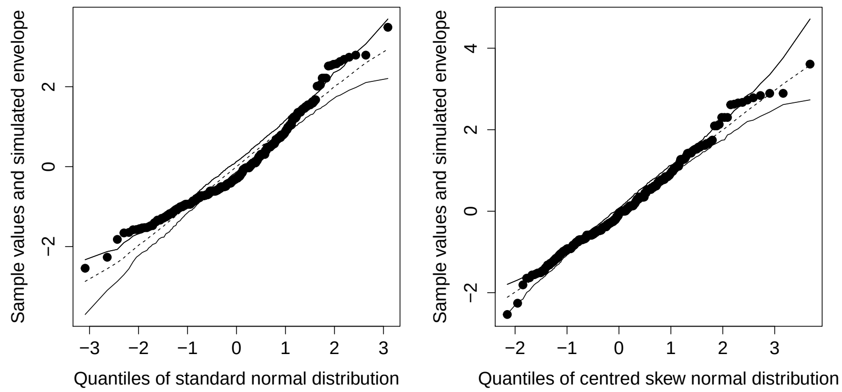


Figure 9: QQ plots with envelopes for BS (left) and SNBS (right) distributions for the prices of bottles of Barolo wine.

algorithms were verified and several factors of interest were compared in the parameter recovery study. In general, as the sample size increases, the results from the simulation studies indicate that the Bayesian approach provided the most accurate estimates. In future developments, we suggest the use of Jeffreys-rule prior and independence Jeffreys prior. Other auxiliary algorithms as the Hamiltonian Monte Carlo (see Homan and Gelman (2014) and Carpenter et al. (2016)), adaptive reject sampling and slice sampling (see Gamerman and Lopes (2006)) could be used and compared, considering also the Metropolis-Hastings algorithm. Other family of distributions could be used instead of the skew-normal distribution under CP, as the scale mixture of the skew-normal distributions, to generate new family of BS-type distributions.

Finally, other numerical methods to obtain approximation for the marginal posterior distributions, such as the INLA algorithm, can be considered, see Rue and Martino (2009).

Acknowledgements

The authors would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the financial support through a Master Scholarship granted to the first author under the guidance of the second and also for Conselho Nacional de Desenvolvimento Científico e Tecnológico, grant number 308339/2015-0, for a research scholarship granted to the second author.

References

Achcar, J. A. (1993). Inferences for the Birnbaum-Saunders fatigue life model using bayesian methods. *Computational statistics & data analysis*, 15(4), 367-380.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.

Atkinson, A. C. (1985). *Plots, Transformations and Regressions*. Oxford University Press, Oxford.

Arellano-Valle, R. B.; Azzalini, A. (2008). The centred parameterization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, 99(7), 1362-1382.

Azzalini, A. (1985). A class of distribution which includes the normal ones. *Scandinavian journal of statistics*, 12(2), 171-178.

Azzalini, A. (2013). *The skew-normal and related families*. IMS monographs, Cambridge University Press, Cambridge.

Balakrishnan, N.; Leiva, V.; Sanhueza, A.; Vilca-Labra, F. (2009). Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm. *SORT*, 33(2), 171-192.

Barros, M.; Paula, G.A.; Leiva, V. (2007). *Log-Birnbaum-Saunders Student-t regression models, In Portuguese*, Brazilian Statistical Association.

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(3), 1-17.

Birnbaum, Z. W.; Saunders, S. C. (1969a). A new family of life distributions. *Journal of Applied Probability*, 6(2), 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(2), 328-347.

Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 15(5), 1190-1208.

Carpenter, B.; Gelman, A.; Hoffman, M.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M. A.; Guo, J.; Li, P.; Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, 20, 1-37.

Coolidge, J. L. (1949). The story of the binomial theorem. *American Mathematical Monthly*, 56(3),147-157.

Cordeiro, G. M.; Lemonte, A. J. (2011). The β-Birnbaum-Saunders distribution: An improved distribution for fatigue life modeling. *Computational Statistics & Data Analysis*, 55(3), 116-124.

Desmond, A. F. (1985). Stochastic models of failure in random environments. *Canadian Journal of Statistics*, 13(3), 171-183.

Depaoli, S.; Clifton, J. P.; Cobb, P. R. (2016). Just Another Gibbs Sampler (JAGS) Flexible Software for MCMC Implementation. *Journal of Educational and Behavioral Statistics*, 41, 6, 628-649.

Desmond, A. F. (1986). On the relationship between two fatigue-life models. *EEE Transactions on Reliability*, 35(2), 167-169.

Diaz-Garcia, J. A.; Leiva, V. (2005). New family of life distributions based on the elliptically contoured distributions. *Journal of Statistical Planning and Inference*, 128(2), 445-457.

Dupuis, D. J.; Mills, J. E. (1998). Robust estimation of the Birnbaum-Saunders distribution. *EEE Transactions on Reliability*, 47(1), 88-95.

Engelhardt, M.; Bain, L. J.; Wright, F. T. (1981). Inferences on the parameters of the Birnbaum-Saunders fatigue life distribution based on maximum likelihood estimation. *Technometrics*, 23(3), 251-256.

Galea, M.; Leiva, V.; Paula, G. (2004). Influence diagnostics in log-Birnbaum-Saunders regression models. *Journal of Applied Statistics*, 31(9), 1049-1064.

Gamerman, D.; Lopes, H. (2006). *MCMC-Stochastic Simulation for Bayesian Inference*, Chapman & Hall/CRC.

Homan, M. D.; Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, 15(1), 1593-1623.

Leiva, V.; Barros, M.; Paula, G. A.; Galea, M. (2007a). Influence diagnostics in log-Birnbaum-Saunders regression models with censored data. *Computational Statistics & Data Analysis*, 51(12), 5694-5707.

Leiva, V.; Barros, M.; Paula, G. A.; Sanhueza, A. (2007b). Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. *Environmetrics*, 19(3), 235-249.

Leiva, V.; Vilca, F.; Balakrishnan, N.; Sanhueza, A. (2008). A skewed sinh-normal distribution and its properties and application to air pollution. *Communications in Statistics Theory and Methods*, 39(3), 426-443.

Leiva, V.; Barros, M.; Paula, G. A.; Galea, M. (2009). Generalized Birnbaum-Saunders Models using R. *Brazilian Statistical Association, São Paulo, Brazil (in English).* http://staff.deuv.cl/leiva/archivos/books/leiva_barros_paula_2009.*pdf*

Lemonte, A. J.; Martínez-Florez, G.; Moreno-Arenas, G. (2015). Multivariate Birnbaum-Saunders distribution: Properties and associated inference. *Journal of Statistical Computation and Simulation*, 85(2), 374-392.

Lu, M. C.; Chang, D. S. (1997). Bootstrap prediction intervals for the Birnbaum-Saunders distribution. *Microelectronics Reliability*, 37(8), 1213-1216.

Lunn, D. J.; Thomas, A.; Best, N.; Spiegelhalter, D. (2000). WinBUGS a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, 325-337.

Lunn, D.; Spiegelhalter, D.; Thomas, A.; Best, N. (2009). The BUGS project: Evolution, critique and future directions (with discussion). *Statistics in Medicine*, 28, 3049–3082.

Mann, N. R.; Singpurwalla, N. D.; Schafer, R. E. (1974). *Methods for statistical analysis of reliability and life data*. Wiley, New York.

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

Owen, W. J.; Padgett, W. J. (1999). Accelerated test models for system strength based on Birnbaum-Saunders distributions. *Lifetime Data Analysis*, 5(2), 133-147.

Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of applied statistics*, 27(7), 859-870.

R Development Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org`

Rieck, J. R.; Nedelman, J. R. (1991). A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, 33(1), 51-60.

Rieck, J. R. (1999). A moment-generating function with application to the Birnbaum-Saunders distribution, *Communications in Statistics-Theory and Methods*, 28(9), 2213-2222.

Rue, H.; Martino, S. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B*, 71(2), 319-392.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.

Spiegelhalter, D. J.; Thomas, A.; Best, N.; Lunn, D. (2005). OpenBUGS version 2.10, user manual. *MRC Biostatistics Unit, Cambridge, United Kingdom.*

Spiegelhalter, D. J.; Best, N. G.; Carlin, B. P.; Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3), 485-493.

Vilca, F.; Leiva, V. (2006). A New Fatigue Life Model Based on the Family of Skew-Elliptical Distributions. *Communications in Statistics: Theory and Methods* 35(2), 229-244.

Vilca, F.; Santana, L.; Leiva, V. (2011). Estimation of extreme percentiles in Birnbaum-Saunders distributions. *Computational Statistics & Data Analysis*, 55(5), 1665-1678.

Volle, M. (1985). *Analyse des Données* (3rd edn). Economica, Paris.

Wang, M.; Sun, X.; Park, C. (2016). Bayesian analysis of BirnbaumSaunders distribution via the generalized ratio-of-uniforms method. *Comput. Stat.*, 31, 207-225.

Xu, A.; Tang, Y. (2011). Bayesian analysis of Birnbaum-Saunders distribution with partial information. *Computational Statistics & Data Analysis*, 55, 2324-2333

7 Appendices

A The pdf of the SNBS distribution and its moments

If $Y \sim \mathrm{SN}(0,1,\gamma)$, then cumulative distribution function of T in (3) is given by

$$\begin{aligned} F_T(t) &= P(T \leq t) = P\left[Y \leq a_t(\alpha,\beta)\right] \\ &= F_Y\left[a_t(\alpha,\beta)\right]. \end{aligned}$$

Therefore, by deriving the expression above, we have the density of T given by

$$\begin{aligned} f_T(t) &= 2\sigma_z\phi\left[\mu_z + \sigma_z a_t(\alpha,\beta)\right]\Phi\left\{\lambda\left[\mu_z + \sigma_z a_t(\alpha,\beta)\right]\right\}A_t(\alpha,\beta) \\ &= 2\phi\left[a_{t;\mu,\sigma}(\alpha,\beta)\right]\Phi\left[\lambda\, a_{t;\mu,\sigma}(\alpha,\beta)\right]A_{t;\sigma}(\alpha,\beta), \end{aligned}$$

where $a_{t;\mu,\sigma}(\alpha,\beta) = \mu_z + \sigma_z\, a_t(\alpha,\beta)$ and $A_{t;\sigma}(\alpha,\beta) = \sigma_z A_t(\alpha,\beta)$.

Next, we discuss the moments of the SNBS distribution. By using the binomial theorem, see Coolidge (1949). The kth moment of the SNBS distribution is related to the correspondent moment of the SN distribution under the CP.

Theorem 2. *Let $T \sim SNBS(\alpha,\beta,\gamma)$ and $Y \sim SN(0,1,\gamma)$. Then, $\mathbb{E}(T^n)$ exists and depends on $\mathbb{E}\left\{(\alpha Y/2)^{k+l}\left[(\alpha Y/2)^2+1\right]^{\frac{k-l}{2}}\right\}$, where $k=0,\ldots,n$ and $l=0,\ldots,k$. Moreover,*

$$\mathbb{E}(T^n) = \beta^n \sum_{k=0}^{n}\binom{n}{k}\sum_{l=0}^{k}\binom{k}{l}2^k\mathbb{E}\left\{(\alpha Y/2)^{k+l}\left[(\alpha Y/2)^2+1\right]^{\frac{k-l}{2}}\right\}.$$

Proof: From (3), we have that

$$\mathbb{E}\left(T/\beta\right) = \mathbb{E}\left\{\left[\left(\alpha Y/2 + \sqrt{(\alpha Y/2)^2+1}\right)^2\right]^n\right\} = \mathbb{E}\left\{\left[1 + \left(\frac{\alpha^2}{2}Y^2 + \alpha Y\sqrt{(\alpha Y/2)^2+1}\right)\right]^n\right\}.$$

Using the binomial theorem, that is, $(a+b)^m = \sum_{k=0}^{m}\binom{m}{k}a^{m-k}b^k$, we have that

$$\mathbb{E}\left(T/\beta\right)^n = \sum_{k=0}^{n} \binom{n}{k} \mathbb{E}\left\{\left[\frac{\alpha^2}{2}Y^2 + \alpha Y\sqrt{\left(\alpha Y/2\right)^2+1}\right]^k\right\}.$$

Expanding the Newton's binomial, it comes to that

$$\mathbb{E}\left(T/\beta\right)^n = \sum_{k=0}^{n} \binom{n}{k} \sum_{l=0}^{k} \binom{k}{l} 2^k \mathbb{E}\left\{\left(\alpha Y/2\right)^{k+l}\left[\left(\alpha Y/2\right)^2+1\right]^{\frac{k-l}{2}}\right\}. \tag{8}$$

Then, $\mathbb{E}\left(T/\beta\right)^n$ exists and depends on $\mathbb{E}\Big\{\left(\alpha Y/2\right)^{k+l}\left[\left(\alpha Y/2\right)^2+1\right]^{\frac{k-l}{2}}\Big\}$, $k = 0, \ldots, n$ and $l = 0, \ldots, k$, which also exist, as we will see next. Let

$$U_{kl} = \left(\alpha Y/2\right)^{k+l}\left[\left(\alpha Y/2\right)^2+1\right]^{\frac{k-l}{2}}.$$

Then using the Markov inequality, it comes to that

$$\begin{aligned}\mathbb{E}(|U_{kl}|) &\leq \mathbb{E}^{1/2}\left[\left(\alpha Y/2\right)^{2(k+l)}\right]\mathbb{E}^{1/2}\left\{\left[\left(\alpha Y/2\right)^2+1\right]^{k-l}\right\}\\ &< \infty,\end{aligned}$$

since all the moments of Y are finite.

B The ECM algorithm

The following result is used in the proof of Theorem 1.

Lemma 1. *Let $X \sim \mathcal{N}(\eta, \tau^2)$, thus $\forall a \in \mathbb{R}$*

$$\mathbb{E}(X|X>a) = \eta + \frac{\phi\left(\frac{a-\eta}{\tau}\right)}{1-\Phi\left(\frac{a-\eta}{\tau}\right)}\tau;\ \mathbb{E}(X^2|X>a) = \eta^2 + \tau^2 + \frac{\phi\left(\frac{a-\eta}{\tau}\right)}{1-\Phi\left(\frac{a-\eta}{\tau}\right)}(\eta+a)\tau.$$

Proof of Theorem 1

i) Since $Y \sim \text{SNCP}(0, 1, \gamma)$, using the stochastic representation given by (2) we can define Y as $Y = \frac{1}{\sigma_z}\left[\delta H + \sqrt{1-\delta^2}X_1 - \mu_z\right] = \frac{1}{\alpha}\left[\sqrt{T/\beta} - \sqrt{\beta/T}\right]$. Therefore, $Y|(H=h) = \frac{1}{\alpha}\left(\sqrt{T/\beta} - \sqrt{\beta/T}\right)\Big|(H=h) \sim \mathcal{N}(\mu_h, \sigma^2)$, where $\mu_h = \frac{\delta(h-r)}{\sqrt{1-r^2\delta^2}}$ and $\sigma^2 = \frac{1-\delta^2}{1-r^2\delta^2}$. Then,

$$W|(H=h) = -\frac{\mu_h}{\sigma} + \frac{1}{\sigma\alpha}\left(\sqrt{T/\beta} - \sqrt{\beta/T}\right)\Big|(H=h) \sim N(0,1).$$

$$T = \beta\left[\frac{\alpha}{2}\left(\sigma W + \mu_h\right) + \sqrt{\left[\frac{\alpha}{2}\left(\sigma W + \mu_h\right)\right]^2 + 1}\right].$$

From the above result, the proof is concluded.

ii) As $f_H(h) = 2\phi(h|0,1), h > 0$ and

$$\phi(\nu_h + a_t(\alpha_{\delta,\beta})) = \frac{\sqrt{1-\delta^2}}{\sqrt{1-r^2\delta^2}}\phi\left(a_t(\alpha,\beta)\left|\frac{\delta(h-r)}{\sqrt{1-r^2\delta^2}}, \frac{1-\delta^2}{1-r^2\delta^2}\right.\right).$$

Then, we have

$$\begin{aligned}\phi\left(a_t(\alpha,\beta)\Big|\frac{\delta(h-r)}{\sqrt{1-r^2\delta^2}}, \frac{1-\delta^2}{1-r^2\delta^2}.\right)\phi(h|0,1) &= \phi\Big(a_t(\alpha,\beta)\Big| - \frac{r\delta}{\sqrt{1-r^2\delta^2}}; \frac{1}{1-r^2\delta^2}\Big) \\ &\times \phi\Big(h\Big|\delta\sqrt{1-r^2\delta^2}\big(a_t(\alpha,\beta) + \frac{r\delta}{\sqrt{1-r^2\delta^2}}\big), 1-\delta^2\Big).\end{aligned}$$

Therefore, the proof of i) follows directly from $f_{H|T}(h|t) = f_{T|H}(t|h)f_H(h)/f_T(t)$. For proving ii)−iii), notice that, for $k = 1, 2$, we have that

$$\begin{aligned}\mathbb{E}\left[H^k|T\right] &= \frac{1}{\Phi\Big(\lambda\sigma_z\big(a_t(\alpha,\beta) + \frac{r\delta}{\sqrt{1-r^2\delta^2}}\big)\Big)}\int_0^\infty h^k\phi\left\{h\big|\eta_t, 1-\delta^2\right\}dh \\ &= \mathbb{E}(X^k|X > 0).\end{aligned}$$

Then, using some proprieties of the half-normal (HN) distribution from Lemma 1, the proof is concluded.

C The Observed Fisher information matrix

The necessary expressions are given below. For the sake of simplicity, we consider the following notation to obtain the necessary expressions, $a_{t_i;\mu,\sigma} = a_{t_i;\mu,\sigma}(\alpha,\beta)$ and $A_{t_i;\sigma} = A_{t_i;\sigma}(\alpha,\beta)$.

$$\begin{aligned}\frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2} &= -\frac{1}{A^2_{t_i;\sigma}}\frac{\partial A_{t_i;\sigma}}{\partial\theta_1}\frac{\partial A_{t_i;\sigma}}{\partial\theta_2} + \frac{1}{A_{t_i;\sigma}}\frac{\partial^2 A_{t_i;\sigma}}{\partial\theta_1\partial\theta_2} - \frac{1}{2}\frac{\partial^2 a^2_{t_i;\mu,\sigma}}{\partial\theta_1\partial\theta_2} + \lambda^2 W'_\Phi[\lambda a_{t_i;\mu,\sigma}]\frac{\partial a_{t_i;\mu,\sigma}}{\partial\theta_1}\frac{\partial a_{t_i;\mu,\sigma}}{\partial\theta_2} \\ &\quad +\lambda W_\Phi[\lambda a_{t_i;\mu,\sigma}]\frac{\partial^2 a_{t_i;\mu,\sigma}}{\partial\theta_1\partial\theta_2}, \quad \theta_1,\theta_2 = \alpha,\beta \\ \frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\theta_3\partial\gamma} &= -\frac{1}{A^2_{t_i;\sigma}}\frac{\partial A_{t_i;\sigma}}{\partial\theta_3}\frac{\partial A_{t_i;\sigma}}{\partial\gamma} + \frac{1}{A_{t_i;\sigma}}\frac{\partial^2 A_{t_i;\sigma}}{\partial\theta_3\partial\gamma} - \frac{1}{2}\frac{\partial^2 a^2_{t_i;\mu,\sigma}}{\partial\theta_3\partial\gamma} + \lambda W_\Phi[\lambda a_{t_i;\mu,\sigma}]\frac{\partial^2 a_{t_i;\mu,\sigma}}{\partial\theta_3\partial\gamma} \\ &\quad + \left\{\lambda W'_\Phi[\lambda a_{t_i;\mu,\sigma}]\frac{\partial\lambda a_{t_i;\mu,\sigma}}{\partial\gamma} + W_\Phi[\lambda a_{t_i;\mu,\sigma}]\frac{\partial\lambda}{\partial\gamma}\right\}\frac{\partial a_{t_i;\mu,\sigma}}{\partial\theta_3}, \quad \theta_3 = \alpha,\beta \\ \frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\gamma^2} &= -\frac{1}{A^2_{t_i;\sigma}(\alpha,\beta)}\frac{\partial^2 A_{t_i;\sigma}}{\partial\gamma^2} + \frac{1}{A_{t_i;\sigma}(\alpha,\beta)}\frac{\partial^2 A_{t_i;\sigma}}{\partial\gamma^2} - \frac{1}{2}\frac{\partial^2 a^2_{t_i;\mu,\sigma}}{\partial\gamma^2} + W'_\Phi[\lambda a_{t_i;\mu,\sigma}]\left(\frac{\partial\lambda a_{t_i;\mu,\sigma}}{\partial\gamma}\right)^2 \\ &\quad +W_\Phi[\lambda a_{t_i;\mu,\sigma}]\frac{\partial^2\lambda a_{t_i;\mu,\sigma}}{\partial\gamma^2},\end{aligned}$$

where $W'_\Phi(x) = -W_\Phi(x)[x + W_\Phi(x)]$ is the derivative of $W_\Phi(x)$ with respect to x, see Vilca et al. (2011), and the other quantities are as before defined.