Statistics and Its Interface Volume 13 (2020) 335–346

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

Nathalia L. Chaves, Caio L. N. Azevedo*,
Filidor Vilca-Labra, and Juvêncio S. Nobre

In this paper we introduce a new regression model for positive and skewed data, a log Birnbaum-Saunders model based on the centred skew-normal distribution, also presenting several inference tools for this model. Initially, we developed a new version of the skew-sinh-normal distribution, describing some of its properties. For the proposed regression model, we carry out, through the expectation conditional maximization (ECM) algorithm, parameter estimation, model fit assessment, model comparison and residual analysis. Finally, our model accommodates more suitably the asymmetry of the data, compared with the usual log Birnbaum-Saunders model, which is illustrated through a real data analysis.

AMS 2000 subject classifications: 62F10, 62J12, 62J20, 62P99.
Keywords and phrases: Birnbaum-Saunders distribution, Skew-sinh-normal distribution, Frequentist inference, ECM algorithm.

1. INTRODUCTION

It is well known, in the data analysis, that to consider some transformation (as, the natural logarithm, square root among others) on the response variable along the use of traditional regression models, as the normal linear one, has some drawbacks, see [1] and [16]. Indeed, such approach can lead to some problems as: the difficulty in interpreting the parameters (in the original scale of the response), the increasing in the bias of the estimates, and the impossibility in obtaining of symmetry and/or homoscedasticity. However, when the transformed response does not present none of those problems, such approach can allow the use of well developed statistical tools for data analysis with no issues. Furthermore, such transformation(s) can be necessary due to requirements of the problem under investigation/researcher and/or even due to the nature of the experiment/survey.

When the response variable is positive, the most common transformation is the log (ln), which can lead to at least of one the aforementioned problems. Indeed, when the original response can be modeled by a log-normal or a Birnbaum-Saunders distribution, the respective log-transformation will follow a normal and a sinh-normal distribution (which are symmetric random variables), respectively. However, sometimes, some trace of asymmetry still remains, as when the original response can be suitably modeled by a Birnbaum-Saunders based on the skew normal distributions, see [14]. Therefore, when a log transformation is needed/suitable/required, and the transformed variable still presents an asymmetric behavior, a random variable that account this feature can be considered. This is the case of the real data set here analyzed. In this work we consider a log regression model where the response is modeled by a Birnbaum-Saunders (BS) based on the centred parameterization skew normal distribution, see [14]. This implies that the respective log response follow a sinh-skew normal distribution, see [14]. More details are provided in the next sections. In the following, we present a literature review related to the BS regression models.

Regression models based on the Birnbaum-Saunders (BS) and the correspondent log-Birnbaum-Saunders (log-BS) distributions, which are related to the family of sinh-normal distributions, see [24], have been receiving considerable attention in the past few years. These regression models are based on the BS or a log-BS distribution which, in their turn, are based on a random variable different from the standard normal. Examples of these distributions are: skew-eliptical BS [32], Student-t BS [10], scale-mixture of normals BS [7] and skew scale-mixture Birnbaum-Saunders [8]. In terms of log-BS regression models, we can cite: Student-t BS model [9], skew-normal BS model [27] and scale-mixture of normals BS model [35].

In this paper, we develop a set of statistical analysis tools for the log Birnbaum-Saunders regression model based on the skew-normal (SN) distribution under the centred parameterization (also named centred skew-normal distribution) [6], named log-SNBS regression model. In the work of [14], the authors provided empirical evidences that their centred skew-normal BS (SNBS) distribution has advantages, in terms of inference, over the skew-normal BS distribution proposed by [34], similarly to those of the centred SN compared with the usual SN [5], see [23] and [4]. In this paper,

*Corresponding author.

we show that these advantages are inherited by the respective log-SNBS regression model.

The aforementioned inference tools are: parameter estimation, residual analysis and statistics for model comparison, which are developed through the Expectation conditional maximization (ECM) algorithm. Also, the impact of some factors of interest (sample size, asymmetry level of the log-SNBS distribution and the value of the shape parameter) on the estimates, are measured through suitable simulation studies. In addition, the performance of two usual statistics of model comparison is studied concerning the selection between our model and the log-BS regression model proposed by Rieck and Nedelman [25], using simulated data.

The paper is outlined as follows. In Section 2, we present the log-SNBS distribution along with some of its properties. In Section 3, we introduce the log-SNBS regression model, discussing the ECM algorithm for maximum-likelihood (ML) estimation. In Sections 4, 5 and 6, we develope the residual analysis, the statistics for model comparison and simulation studies, respectively. In Section 7, a real data analysis is discussed and finally, in Section 8, the concluding remarks are given.

2. THE LOG-SNBS DISTRIBUTION

2.1 The centred skew-normal BS distribution

A random variable (r.v.) T follows the centred skew-normal BS (SNBS) distribution, denoted by $T \sim \text{SNBS}(\alpha, \eta, \gamma), \alpha, \eta \in \mathbb{R}, \gamma \in (-.99527, .99527)$, where α is the shape parameter, η is the location parameter and γ is the asymmetry parameter, if its density is given by

$$\begin{aligned} f_T(t) &= 2\phi\left[a_{t;\mu,\sigma}(\alpha,\eta)\right]\Phi\left\{\lambda(\gamma)\, a_{t;\mu,\sigma}(\alpha,\eta)\right\} \\ &\times \quad A_{t;\sigma}(\alpha,\eta), t > 0, \end{aligned}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and distribution functions of the standard normal distribution, respectively. Also, $a_{t;\mu,\sigma}(\alpha,\eta) = \mu_z + \sigma_z\, a_t(\alpha,\eta)$ and $A_{t;\sigma}(\alpha,\eta) = \sigma_z\, A_t(\alpha,\eta)$, with $a_t(\alpha,\eta) = (\sqrt{t/\eta} - \sqrt{\eta/t})/\alpha$ and $A_t(\alpha,\eta) = \frac{d}{dt}a_t(\alpha,\eta) = \frac{t^{-3/2}(t+\eta)}{2\alpha\eta^{1/2}}$. Finally, $\mu_z = r\delta(\gamma)$, $\sigma_z = \sqrt{1-\mu_z^2}$, $r = \sqrt{2/\pi}$, $\delta(\gamma) = \lambda(\gamma)/\sqrt{1+\lambda(\gamma)^2}$, $\lambda(\gamma) = \gamma^{1/3}s/\sqrt{r^2+s^2\gamma^{2/3}(r^2-1)}$ and $s = [2/(4-\pi)]^{1/3}$. For simplicity, we will refer to $\delta(\gamma)$ only as δ. The parameters are $(\alpha,\eta,\gamma)^\top$ and it will be called *centred parameters*, while the parameters based the usual SN distribution, $(\alpha,\eta,\lambda)^\top$, are called as *direct parameters*. Note that when $\gamma = 0$, we have the usual BS distribution.

More details about this distribution are presented in [14]. In short, we have symmetry around η for $\gamma = 0$ and small values of α. The positive asymmetry is observed as α increases, η decreases and/or γ takes positive values, whereas negative asymmetry is observed as α decreases, η increases and/or γ assumes negative values. Also, the smaller α and η, the smaller the variability. The higher η, the more shifted to the right is the distribution. Another interesting feature of this distribution is that it can model properly positive random variables with a negatively skewed behavior. In the data set presented by [17], for example, the response variable (the failure times of high-speed turbine engine bearings made out of five different compounds) is positive and presents a negatively skewed behavior for some of the five compounds. The same can be observed for the data set present by [20], which is related to football matches of the UEFA Champions League (more details will be presented in Section 7).

2.2 The centred skew-sinh-normal distribution

A r.v. Y is said to have a centred skew-sinh-normal distribution (SSN), denoted by $Y \sim \text{SSN}(\alpha,\rho,\sigma,\gamma)$, where α, ρ and σ are the shape, location and scale parameters and γ is the Pearson's skewness coefficient, respectively, if its probability density function is given by:

$$\begin{aligned} f_Y(y) &= \frac{4\,\sigma_z}{\alpha\,\sigma}\phi\left[\mu_z + \frac{2\,\sigma_z}{\alpha}\sinh\left(\frac{y-\rho}{\sigma}\right)\right]\Phi\Bigg\{\lambda(\gamma)\Bigg[\mu_z \\ &\quad + \frac{2\,\sigma_z}{\alpha}\sinh\left(\frac{y-\rho}{\sigma}\right)\Bigg]\Bigg\}, y \in \mathbb{R}, \end{aligned} \tag{1}$$

where all quantities are as defined before. Figure 1 presents the density of the SSN distribution for different values of α and γ, fixing ρ and σ in suitable values. We can notice that α and γ affects the kurtosis and skewness, respectively. Positive and negative asymmetry are observed when γ assumes positive and negative values, respectively. Note that for $\gamma = 0$, we have the sinh-normal (SHN) distribution developed by [25].

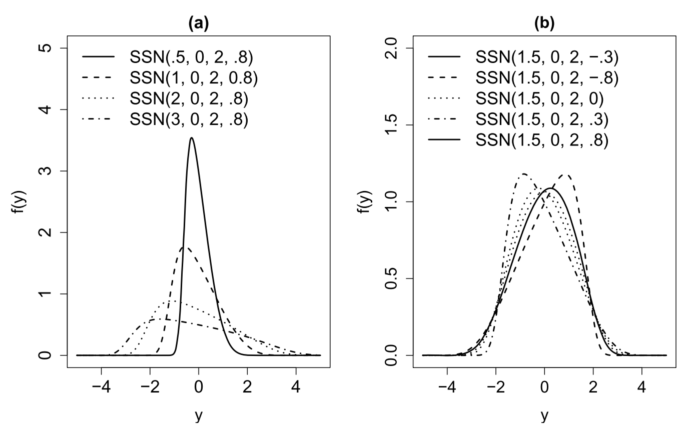

Figure 1. Density of the SSN distribution for different values of α (a) and different values of γ (b).

2.3 The proposed distribution

Here, we develop a generalization of the usual log-BS distribution (see [25] and [18]) based on the centred SN distribution.

A logarithmic version of the SNBS model, called the log-SNBS distribution, can be obtained by considering $Y = \log(T)$, where $T \sim \text{SNBS}(\alpha, \eta, \gamma)$, which lead the following pdf

$$(2) \quad f_Y(y) = \phi(\xi_{2y;\mu,\sigma}) \, \Phi\{\lambda(\gamma)\xi_{2y;\mu,\sigma}\} \, \xi_{1y;\sigma}, y \in \mathbb{R},$$

where $\xi_{2y;\mu,\sigma} = \mu_z + \sigma_z \xi_{2y}$, $\xi_{1y;\sigma} = \sigma_z \xi_{1y}$, with $\xi_{2y} = \xi_2(y; \alpha, \rho) = \frac{2}{\alpha}\sinh(\frac{y-\rho}{2})$, $\xi_{1y} = \xi_1(y; \alpha, \rho) = \frac{2}{\alpha}\cosh(\frac{y-\rho}{2})$ and $\rho = \log(\eta)$. Furthermore, μ_z, σ_z and $\lambda(\gamma)$ are as defined before. When $\gamma = 0$, that is, $f_Y(y) = \phi(\xi_{2y}) \, \xi_{1y}; y \in \mathbb{R}$, we have the log-BS distribution.

We denote this distribution by $Y \sim \text{SSN}(\alpha, \rho, \sigma = 2, \gamma)$. We use this notation, including a specific value for the parameter σ, since the log-SNBS distribution is a particular case of the SSN distribution when $\sigma = 2$ in (1). This former distribution can be also defined directly in terms of a stochastic representation of the centred SN distribution. That is, Y may be written as

$$(3) \quad Y = \rho + 2\,\text{arcsinh}\left(\frac{\alpha[\delta H + \sqrt{1-\delta^2}X - \mu_z]}{2\sigma_z}\right),$$

where μ_z and δ were defined before, H has a standard half-normal (HN) distribution, denoted by $H \sim \text{HN}(0,1)$, and $X \sim N(0,1)$, and $H \perp X$.

Figure 2 presents the density of the SNBS distribution for different values of γ and α. In short, the distribution is symmetric around ρ, for $\gamma = 0$ (in this case we have the log-BS distribution) and for small values of α. Positive asymmetry is observed as α increases, ρ decreases and/or γ assumes positive values. On the other hand, negative asymmetry is observed as α decreases, ρ increases and/or γ assumes negative values.

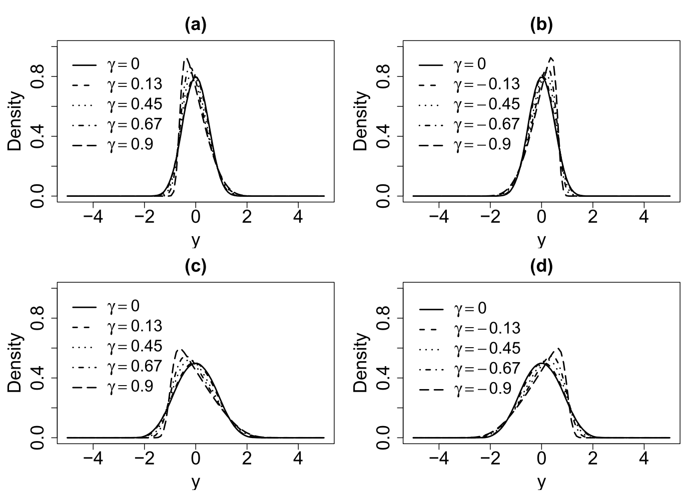

Figure 2. Density of the log-SNBS distribution for different values of γ, with $\rho = 0$, (a)–(b) $\alpha = .5$ and (c)–(d) $\alpha = .8$.

The following theorem is very useful for obtaining the conditional expectations, which will be used in the implementation of the ECM algorithm for ML estimation.

Theorem 1. *Let $Y \sim SSN(\alpha, \rho, \sigma = 2, \gamma)$ as in (3). Then,*

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

$$(4) \quad f_{Y|H}(y) = \frac{1}{2}\phi[\nu_h + \xi_2(y; \alpha_\delta, \rho)] \times \xi_1(y; \alpha_\delta, \rho), y \in \mathbb{R},$$

where $\alpha_\delta = \alpha\frac{\sqrt{1-\delta^2}}{\sigma_z}$ and $\nu_h = -\frac{\mu_z + \delta h}{\sqrt{1-\delta^2}}$.

ii) $f_{H|Y}(h) = \dfrac{\phi\left[h \middle| \delta\,\xi_{2y;\mu,\sigma}; 1-\delta^2\right]}{\Phi(\lambda(\gamma)\,\xi_{2y;\mu,\sigma})}, h > 0$, where $\phi(\cdot|\mu, \sigma^2)$ denotes the density of the $N(\mu, \sigma^2$. Also, we have that

$$\mathbb{E}(H|Y = y) = \eta_y + W_\Phi\left(\frac{\eta_y}{\tau}\right)\tau,$$
$$\mathbb{E}(H^2|Y = y) = \eta_y^2 + \tau^2 + W_\Phi\left(\frac{\eta_y}{\tau}\right)(\eta_y\tau)$$

where $\eta_y = \delta\,\sigma_z\left(\xi_{2y} + \frac{\mu_z}{\sigma_z}\right)$, $\tau = \sqrt{1-\delta^2}$ and $W_\Phi\left(\frac{\eta_y}{\tau}\right) = \frac{\phi\left(\frac{\eta_y}{\tau}\right)}{\Phi\left(\frac{\eta_y}{\tau}\right)}$.

The density in Theorem 1 corresponds to the four-parameter sinh-normal (SHN) distribution proposed by [18]. The proof of this theorem can be found in Appendix A.

3. SNBS REGRESSION MODELS AND ECM ALGORITHM

In this section, we introduce the SNBS regression model. Also, we use a modification of the expectation maximization (EM) algorithm, called the ECM algorithm, proposed by [21], to perform maximum likelihood estimation (MLE). In Appendix A, we present some useful results.

Consider $Y_i \overset{\text{ind}}{\sim} \text{SSN}(\alpha, \boldsymbol{x}_i^\top\boldsymbol{\beta}, \sigma = 2, \gamma)$, $i = 1, \ldots, n$. Associated with the i-th individual, we assume a know $p \times 1$ vector of covariates $\boldsymbol{x}_i$, which we use to specify the linear predictor $\boldsymbol{x}_i^\top\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression coefficients. Thus, the response Y_i can be represented as

$$(5) \quad Y_i = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \epsilon_i, \; \epsilon_i \sim \text{SSN}(\alpha, 0, \sigma = 2, \gamma), \; i = 1, \ldots, n.$$

Note that, when $\gamma = 0$, the log-BS regression model developed by [25] is obtained.

Considering $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top, \gamma)^\top$, the respective log-likelihood based on observed data $\mathbf{y} = (y_1, \ldots, y_n)^\top$ is $\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}|y_i)$, where

$$(6) \quad \ell_i(\boldsymbol{\theta}|y_i) = \log[\phi(\xi_{2i;\mu,\sigma})] + \log[\Phi(\lambda(\gamma)\xi_{2i;\mu,\sigma})] + \log(\xi_{1i;\sigma}),$$

with $\xi_{2i;\mu,\sigma} = \mu_z + \sigma_z\xi_{2i}$, $\xi_{1i;\sigma} = \sigma_z\xi_{1i}$, such that $\xi_{1i} = \xi_1(y_i;\alpha,\mathbf{x}_i^\top\boldsymbol{\beta}) = \frac{2}{\alpha}\cosh\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{2}\right)$, $\xi_{2i} = \xi_2(y_i;\alpha,\mathbf{x}_i^\top\boldsymbol{\beta}) = \frac{2}{\alpha}\sinh\left(\frac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{2}\right)$. Since the observed log-likelihood involves complex expressions, it is difficult to work directly with it. Thus, instead consider the direct maximization of (6) we will obtain the ML estimates through a modification of the EM algorithm, called the ECM algorithm. In this case, we need to work with the so-called augmented likelihood, which is obtained from the joint distribution of the observable variables and non-observable variables, leading to more tractable expressions, both analytically and computationally, under suitable choices of the augmented variables, see [33]. Also, instead estimating $\boldsymbol{\theta} = (\alpha,\boldsymbol{\beta}^\top,\gamma)^\top$, we will estimate $\boldsymbol{\theta} = (\alpha,\boldsymbol{\beta}^\top,\delta)^\top$, where δ was defined before, which facilitates the related algebra and the optimization process.

From (3), we have the following hierarchical representation

$$\begin{aligned} Y_i|(H_i = h_i) &\overset{\text{ind}}{\sim} SHN(\alpha_\delta, \boldsymbol{x}_i^\top\boldsymbol{\beta}, \sigma = 2, \nu_{h_i}),\\ H_i &\overset{\text{ind}}{\sim} HN(0,1), i = 1,\dots,n, \end{aligned}$$

where $\alpha_\delta = \alpha\frac{\sqrt{1-\delta^2}}{\sigma_z}$ and $\nu_{h_i} = \frac{\mu_z-\delta h_i}{\sqrt{1-\delta^2}}$. Then, defining $\mathbf{y}_c = (\mathbf{y},\mathbf{h}^\top)^\top$, where $\mathbf{h} = (h_1,..,h_n)^\top$, the augmented likelihood is given by

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{y}_c) &= \sum_{i=1}^n \log f_{Y|H}(y_i) + \sum_{i=1}^n \log f_H(h_i)\\ &= n\left[\log(\sqrt{2/\pi}) - \log(2)\right] + \sum_{i=1}^n \log\Big\{\phi\,[\nu_{h_i}\\ &\quad + \xi_2(y_i;\alpha_\delta,\mathbf{x}_i^\top\boldsymbol{\beta})]\Big\} + \sum_{i=1}^n \log\left[\xi_1(y_i;\alpha_\delta,\mathbf{x}_i^\top\boldsymbol{\beta})\right]\\ &\quad - \frac{1}{2}\sum_{i=1}^n h_i^2. \end{aligned}$$

For a current value $\boldsymbol{\theta}$, says $\widehat{\boldsymbol{\theta}}$, the E-step requires the evaluation of $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathbb{E}\left[\ell(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y},\widehat{\boldsymbol{\theta}}\right]$, where the expectation is taken with respect to the conditional distribution $H|(Y = y)$ and evaluated at $\widehat{\boldsymbol{\theta}}$. Given a estimate of $\boldsymbol{\theta}$ at r-th iteration, says $\widehat{\boldsymbol{\theta}}^{(r)} = (\widehat{\alpha}^{(r)},\widehat{\boldsymbol{\beta}}^{(r)},\delta^{(r)})^\top$, defines $\widehat{h}_i^{(r)} = \mathbb{E}[H_i|y_i,\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(r)}]$ and $\widehat{h}_i^{2(r)} = \mathbb{E}[H_i^2|y_i,\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(r)}]$, which are obtained by using the conditional expectation given in Theorem 1, which become

$$\begin{aligned} \widehat{h}_i &= \widehat{\eta}_{y_i} + W_\Phi\left(\frac{\widehat{\eta}_{y_i}}{\widehat{\tau}}\right)\widehat{\tau}, \quad \text{and}\\ \widehat{h}_i^2 &= \widehat{\eta}_{y_i}^2 + \widehat{\tau}^2 + W_\Phi\left(\frac{\widehat{\eta}_{y_i}}{\widehat{\tau}}\right)(\widehat{\eta}_{y_i}\widehat{\tau}), \end{aligned} \tag{7}$$

where, $\widehat{\eta}_{y_i} = \widehat{\delta}\sqrt{1-r^2\widehat{\delta}^2}\left(\xi_2(y_i;\widehat{\alpha},\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}) + \frac{r\widehat{\delta}}{1-r^2\widehat{\delta}^2}\right)$, $\widehat{\tau} = \sqrt{1-\widehat{\delta}^2}$ and $W_\Phi(z) = \phi(z)/\Phi(z)$, $z\in\mathbb{R}$.

After some algebra, it follows that the conditional expectation of the augmented log-likelihood is

$$\begin{aligned} Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}\right) &= \mathbb{E}\left[\ell(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y},\widehat{\boldsymbol{\theta}}^{(r)}\right]\\ &= c - \frac{\delta^{2(r)}}{2(1-\delta^{2(r)})}\sum_{i=1}^n\left(r^2 - 2\,r\,\widehat{h}_i^{(r)} + \widehat{h}_i^{2(r)}\right)\\ &\quad - \frac{\delta^{(r)}}{\sqrt{1-\delta^{2(r)}}}\sum_{i=1}^n\left[\left(r-\widehat{h}_i^{(r)}\right)\xi_2\left(y_i;\alpha_\delta^{(r)},\boldsymbol{x}_i^\top\boldsymbol{\beta}^{(r)}\right)\right]\\ &\quad - \frac{1}{2}\sum_{i=1}^n\left\{\xi_2\left(y_i;\alpha_\delta^{(r)},\boldsymbol{x}_i^\top\boldsymbol{\beta}^{(r)}\right)\right\}^2 + \sum_{i=1}^n\log\Big[\xi_1\Big(y_i;\\ &\quad \alpha_\delta^{(r)},\boldsymbol{x}_i^\top\boldsymbol{\beta}^{(r)}\Big)\Big] - \frac{1}{2}\sum_{i=1}^n\widehat{h}_i^{2(r)}. \end{aligned}$$

Hence, the ECM algorithm corresponds to iterate the following steps:

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(r)}$, to compute $\widehat{h}_i$ and $\widehat{h}_i^2$, for $i = 1,\dots,n$ using results in (7);

CM-step 1: To fix $\widehat{\boldsymbol{\beta}}^{(r)}$ and $\widehat{\delta}^{(r)}$ and to 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

$$\begin{aligned} \hat{b}^{(r)} &= \frac{2\widehat{\delta}^{(r)}\sqrt{1-r^2\widehat{\delta}^{2(r)}}}{n(1-\widehat{\delta}^{2(r)})}\Big[\sum_{i=1}^n\sinh\left(\frac{\mathbf{y}_i-\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}^{(r)}}{2}\right)\widehat{h}_i^{(r)}\\ &\quad - r\sum_{i=1}^n\sinh\left(\frac{\mathbf{y}_i-\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}^{(r)}}{2}\right)\Big],\\ \hat{c}^{(r)} &= -\frac{4\left(1-r^2\widehat{\delta}^{2(r)}\right)}{n(1-\widehat{\delta}^{2(r)})}\sum_{i=1}^n\left\{\sinh\left(\frac{\mathbf{y}_i-\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}^{(r)}}{2}\right)\right\}^2. \end{aligned}$$

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

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

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

The updating of $\widehat{\boldsymbol{\beta}}^{(r+1)}$ and $\widehat{\delta}^{(r+1)}$ need to be done through some numerical optimization method. In this work we use the function `optim`, available at software `R` [26], considering the L-BFGS-B optimization algorithm [12].

We start the ECM algorithm with initial values $\widehat{\alpha}^{(0)}$, $\widehat{\boldsymbol{\beta}}^{(0)}$ and $\widehat{\delta}^{(0)}$. The estimates $\widehat{\boldsymbol{\beta}}^{(0)}$ can be obtained through ordinary least squares estimates of the log-SNBS regression model. The value $\widehat{\alpha}^{(0)}$ can be obtained from $\widehat{\alpha}^{(0)} = \left\{(4/n)\sum_{i=1}^n\left[\sinh\left(\mathbf{y}_i-\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}^{(r)}/2\right)\right]^2\right\}^{1/2}$, see [19] for details. Once we have $\widehat{\alpha}^{(0)}$ and $\widehat{\boldsymbol{\beta}}^{(0)}$, we can calculate $z_i =$

$\left(2/\widehat{\alpha}^{(0)}\right)\sinh\left(\mathbf{y}_i - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{(r)}/2\right)$; $i = 1,\ldots,n$, which represent observations from the SN distribution. 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

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

According to [34], for ensuring that the true ML estimates are obtained, it is recommended to run the ECM algorithm using a range of different starting values, checking whether all of them result in similar estimates. The steps of the ECM algorithm are repeated until a suitable convergence criterion is attained, for example, until $\left\|\boldsymbol{\theta}^{(r)} - \boldsymbol{\theta}^{(r-1)}\right\| < \varepsilon, \varepsilon > 0$.

The observed information matrix is obtained as $I(\boldsymbol{\theta}) = -\ddot{\ell}$. Here, $\ddot{\ell} = [\ddot{\ell}_{\theta_1\theta_2}]$, where $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \equiv \alpha, \boldsymbol{\beta}, \gamma$, is the Hessian matrix, where $\ddot{\ell}_{\theta_1\theta_2} = \ddot{\ell}_{\theta_2\theta_1} = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2^\top = \sum_{i=1}^{n}\partial^2\ell_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2^\top$. The second derivatives of $\ell_i(\boldsymbol{\theta})$ are provided in Appendix B. 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 the respective ML estimates $\widehat{\boldsymbol{\theta}}$.

3.1 Some advantages of the proposed model

i) It is well known that there is some difficulty in estimating the parameters of the usual SN distribution by maximum likelihood, when the asymmetry parameter is near to zero. The log-SNBS regression model, based on the usual SN, inherits such problem. On the other hand, the log-SNBS regression model (which is based on the centred SN distribution) circumvents problems inherited of the log-BS regression model, based on the usual SN distribution.

ii) When the asymmetry parameter is equals to zero, the expected Fisher information is singular, even if all parameters are identifiable. This fact affects the asymptotic properties of the maximum likelihood estimators (MLEs). To illustrate such behavior, we have run a little simulation experiment generating 5,000 samples of size $n = 200$, from the log-SNBS model, based on the usual SN distribution. For each sample the MLEs $(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\lambda})^\top$ have been computed. In this case, we fix $\alpha = .8$, $\boldsymbol{\beta} = (\beta_0, \beta_1) = (1, 2)^\top$ and $\lambda = 1$. Figure 3 displays the corresponding empirical distributions of $\widehat{\alpha}$ and $(\widehat{\alpha}, \widehat{\beta}_0)^\top$, left panel and a right panel, respectively. Furthermore, it was generated 5,000 samples of size $n = 200$, from the log-SNBS model based on the centred SN distribution. For each sample the MLEs $(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\gamma})^\top$ have been computed. In this case, we fix $\alpha = .8$, $\boldsymbol{\beta} = (1, 2)^\top$ and $\gamma = .137$. The values of λ and γ were the same ones used by [2]. The empirical distribution of the estimates of α is shown in the left panel of Figure 4, while that of $(\widehat{\alpha}, \widehat{\beta}_0)^\top$ is presented in the right panel. Clearly, these empirical distributions are much closer to normality than those in Figure 3. In fact, it can be shown that the singularity of the expected Fisher information matrix, when the skewness parameter is null, does not occur any longer.

iii) Some advantages of the BSCP distribution can be observed in terms of parameter recovery. A short simulation study, for the original skew normal distributions (both direct and centered ones) can be found in [28].

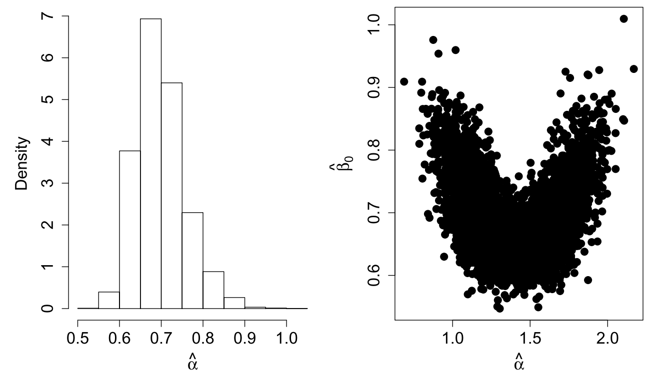

Figure 3. Estimated distributions of the MLEs when samples of size $n = 200$ are drawn from log-SNBS based on the usual SN distribution; the left panel displays the histogram of $\widehat{\alpha}$, the right panel displays the scatter plot of $(\widehat{\alpha}, \widehat{\beta}_0)^\top$.

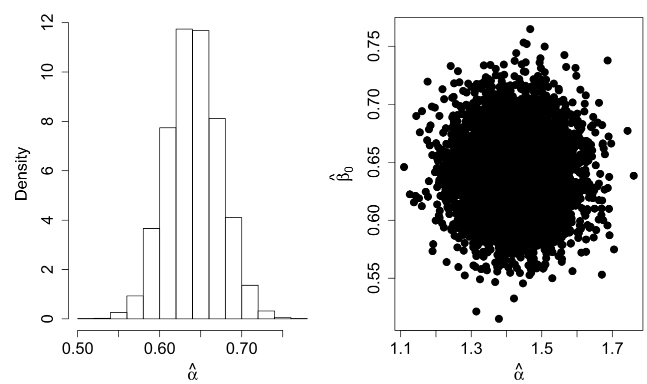

Figure 4. Estimated distributions of the MLEs when samples of size $n = 200$ are drawn from log-SNBS based on the centred SN distribution; the left panel displays the histogram of $\widehat{\alpha}$, the right panel displays the scatter plot of $(\widehat{\alpha}, \widehat{\beta}_0)^\top$.

4. RESIDUAL ANALYSIS

Residual analysis is an important tool for model fit assessment. It allows for checking the presence of outliers, as well as the departing from model assumptions. Following the methodology proposed by [15], we consider the quantile residual.

Let $Y_i|\boldsymbol{\theta} \sim SSN(\alpha, \boldsymbol{x}_i^\top\boldsymbol{\beta}, \sigma = 2, \gamma)$ be a r.v. with a cumulative distribution function (cdf) given by $F_{Y_i}(y_i) = \Phi(\mu_z + \sigma_z\xi_{2i}) - 2T(\mu_z + \sigma_z\xi_{2i}, \lambda(\gamma))$, where $T(a, \lambda(\gamma)) = \dfrac{1}{2\pi}\displaystyle\int_0^{\lambda(\gamma)} \frac{\exp[-0.5\, a^2(1+x^2)]}{1+x^2}dx$ was defined by [22]. Therefore we can define the quantile residual as

$$R_i = \Phi^{-1}\Big\{\Phi\left[\mu_z + \sigma_z\xi_2\left(y_i;\widehat{\alpha}, \boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}\right)\right]$$
$$-2\,T\left[\mu_z + \sigma_z\xi_2\left(y_i;\widehat{\alpha}, \boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}\right), \lambda(\gamma)\right]\Big\}, \tag{8}$$

where $\widehat{(\cdot)}$ is the respective ML estimator. Therefore, with $\widehat{\alpha}$, $\widehat{\gamma}$ and $\widehat{\boldsymbol{\beta}}$ being consistent estimators of α, γ and $\boldsymbol{\beta}$, respectively, we have that R_i converges in distribution to a standard normal distribution.

5. STATISTICS FOR MODEL COMPARISON

There exist a variety of methodologies to compare several competing models, for a given data set. We consider model selection tools which can be easily computed using the available ECM algorithm output, namely: the Akaike's information criterion (AIC) proposed by [3] and Bayesian information criterion (BIC) proposed by [30]. The AIC is based on the likelihood, penalized by the number of parameters. On the other hand, BIC, considers both the number of parameters and the sample size. The respective formulas are given by: $AIC = -2\ell(\boldsymbol{\theta}|\mathbf{y}) + 2k$ and $BIC = -2\ell(\boldsymbol{\theta}|\mathbf{y}) + k\log(n)$, where $\ell(\boldsymbol{\theta}|\mathbf{y})$ is the likelihood defined in (6), k is the total number of parameters and n is the number of observations. Lower values of AIC or BIC indicate models with a better quality of fit.

6. SIMULATION STUDIES

In this section we present three simulation studies, namely: parameter recovery (PRC), behavior of the proposed residuals (R) and performance of the statistics of model comparison (SMC). Several relevant scenarios were considered, which correspond to the combination of the levels of some factors of interest. The factors (with the respective levels within parenthesis) are: sample size (n) $(10, 50, 200)$, that is, small, medium and large sample sizes, values of α $(.5, 1.5)$, that is, low and moderate variability, and value of γ $(-.67, -.45, 0, .45, .67)$, that is high and medium negative skewness, symmetry and high and medium positive skewness, respectively. Therefore, we have a total of 30 scenarios. For the PRC and SMC studies, all scenarios and $R = 1{,}000$ replicas (simulated responses from the model) were considered. For the other study (R), only one replica and only scenario is presented. Specifically for the PRC study, we here present only the results for nine scenarios, since for the other scenarios, the patterns were similar and they can be found in the supplementary material, http://intlpress.com/site/pub/files/_supp/sii/2020/0013/0003/SII-2020-0013-0003-s002.pdf. Additional details are presented in the following subsections.

The general structure of the model here used is

$$Y_i = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \epsilon_i,\ \epsilon_i \sim \text{SSN}(\alpha, 0, \sigma = 2, \gamma),\ i = 1, \ldots, n,$$

where $\boldsymbol{\beta} = (1, 2)^\top$.

6.1 Parameter recovery

As previously mentioned, we present only the results related to the scenarios where $\alpha = .5$, $\gamma \in (-.67, 0, .45)$, varying the sample size. The values for the sample size were chosen in order to verify the proprieties of the ML estimators, as consistency and accuracy.

We calculated the usual statistics to measure the accuracy of the estimates, that are: standard deviation (SD), coverage probability (CP) of the 95% equi-tailed confidence intervals (CI), bias, 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 respective formulas of the statistics are: SD $= \sqrt{(1/R)\sum_{r=1}^R(\widehat{\theta}_r - \overline{\widehat{\theta}})^2}$, CP $= (1/R)\sum_{r=1}^R \text{I}([\widehat{\theta}_{r,LCL}, \widehat{\theta}_{r,UCL}] \supset \theta)$, where $\widehat{\theta}_{r,LCL}$ and $\widehat{\theta}_{r,UCL}$ are the estimated lower and upper limits of the 95% equi-tailed CI's, respectively; Bias $= \overline{\widehat{\theta}} - \theta$, RMSE $= \sqrt{(1/R)\sum_{r=1}^R(\theta - \widehat{\theta}_r)^2}$, AVRB $= |\overline{\widehat{\theta}} - \theta|/|\theta|$. We considered $(< .001)$ to represent positive values (statistics and/or estimates) and $(> -.001)$ to denote negative values, when they are close to zero.

Tables 1, 2 and 3 present some results. We can notice that the estimates related to α, β_0 and β_1 tend to the correspondent true values in all scenarios, as the sample size increases. On the other hand, γ is always underestimated. This is probably due to that the estimation of the asymmetry parameter is more complicated, than to the other parameters. In general, as the sample size increases, we can notice that standard deviation, bias, RMSE and AVRB decrease. Furthermore, as the sample size increases, the coverage probabilities of the 95% equi-tailed confidence intervals tend to 0.95, regardless the selected true values for α and γ, unless for the parameter γ, for which a larger sample size is required to have desirable the coverage probabilities.

6.2 Behavior of the residuals

Here we considered the scenario where $\alpha = .5$, $\gamma = .67$ and $n = 200$. We simulated only one set of observations under four different models: log-SNBS, log-BS, log-BS-t [13] and log-StBS [8]. The first one is the model given by Equation (5) while the second corresponds to its particular case

Table 1. Results of simulation study (PRC) - $\gamma = -.67$.

Parameter	n	Mean	SD	CP	Bias	RMSE	AVRB
α	10	.610	.091	.631	.110	.142	.220
	50	.475	.076	.967	-.025	.080	.051
	200	.498	.005	1.000	-.002	.005	.005
β_0	10	.982	.255	.621	-.018	.256	.018
	50	.989	.415	.653	-.011	.415	.011
	200	.997	.055	.976	-.003	.055	.003
β_1	10	1.996	.824	.463	-.004	.824	.002
	50	2.009	.459	.486	.009	.459	.005
	200	2.003	.109	.933	.003	.109	.002
γ	10	-.134	.978	$< .001$	.536	1.115	.800
	50	-.383	.910	$< .001$	.287	.954	.428
	200	-.555	.277	.750	.115	.300	.171

Table 2. Results of simulation study (PRC) - $\gamma = 0$.

Parameter	n	Mean	SD	CP	Bias	RMSE	AVRB
α	10	.645	.066	.398	.145	.159	.290
	50	.477	.072	.973	-.023	.075	.045
	200	.498	.002	1.000	-.002	.003	.005
β_0	10	.987	.439	.557	-.013	.439	.013
	50	1.002	.273	.622	.002	.273	.002
	200	.998	.059	.973	-.002	.059	.002
β_1	10	2.040	.864	.455	.040	.865	.020
	50	1.991	.504	.427	-.009	.504	.005
	200	2.003	.117	.954	.003	.117	.002
γ	10	-.016	.987	$< .001$	-.016	.988	-
	50	.002	.988	$< .001$	.002	.988	-
	200	.004	.153	.961	.004	.153	-

Table 3. Results of simulation study (PRC) - $\gamma = .45$.

Parameter	n	Mean	SD	CP	Bias	RMSE	AVRB
α	10	.629	.075	.496	.129	.149	.258
	50	0.475	.076	.969	-.025	.080	.051
	200	0.498	.003	1.000	-.002	.004	.005
β_0	10	1.013	.432	.545	.013	.432	.013
	50	.995	.267	.634	-.005	.267	.005
	200	1.002	.057	.976	.002	.057	.002
β_1	10	1.987	.848	.450	-.013	.848	.006
	50	2.003	.488	.439	.003	.488	.002
	200	1.999	.114	.945	-.001	.114	.001
γ	10	.041	.987	$< .001$	-.409	1.068	.908
	50	.328	.932	.001	-.122	.940	.272
	200	.332	.244	.672	-.118	.271	.262

when $\gamma = 0$. The third and the fourth models correspond to model (5) using, in Equation (3), instead of a centred SN distribution, a Student-t and a skew Student-t distribution, with $\nu = 4$ degrees of freedom and asymmetry parameter $\gamma = .67$, respectively. For each simulated data set we fit a log-SNBS regression model and calculate the residuals presented in (8). Four plots were built for each situation, including an simulated envelope for the residuals, which are

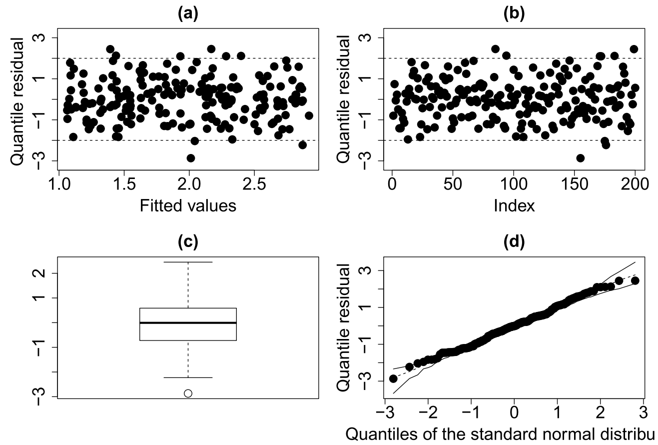

Figure 5. Residual plots for the observations generated from a log-SNBS regression model.

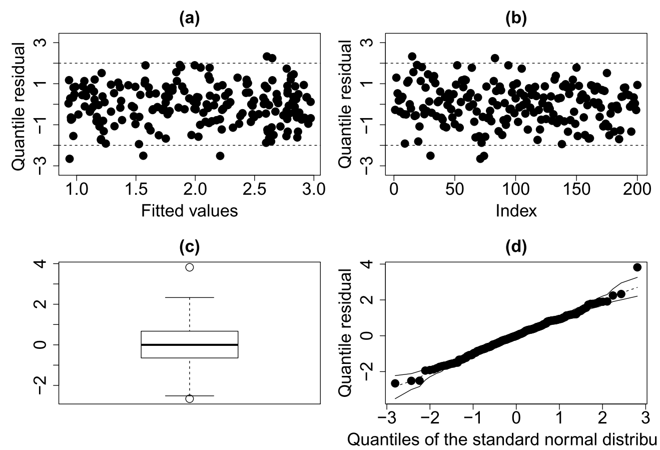

Figure 6. Residual plots for the observations generated from a log-BS regression model.

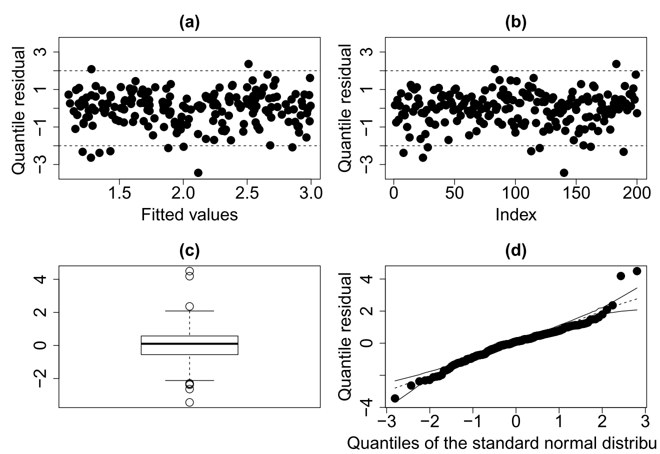

Figure 7. Residual plots for the observations generated from a log-BS-t model.

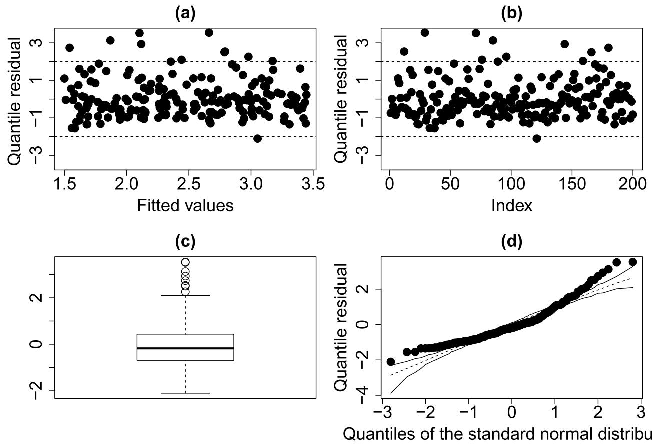

Figure 8. Residual plots for the observations generated from a log-StBS model.

presented in Figures 5, 6, 7 and 8. To simulate from the skew Student-t distribution we used the function *rst* from the R package *sn*.

We can notice that, when the log-SNBS regression model (or its particular case, the log-BS regression model) is the underlying one, the residuals present a symmetric behavior, resembling a standard normal distribution, with all respective values within the simulated envelope, with no systematic behavior. On the other hand, when the underlying model is the log-BS-t, we observe some outliers and many observations lying outside the simulated envelope, which, in its turn, presents a behavior compatible with a heavy tailed distribution. Finally, when the underlying model is the log-StBS, we observe some outliers, a skewed behavior of the residuals, with many observations lying outside the simulated envelope, which, in its turn, presents a behavior compatible with a skewed heavy tailed distribution. In conclusion, we can say that the proposed residuals are appropriate to detect whether or not the model fit properly to the data, concerning the true underlying distribution, identifying, when it is the case, how this distribution differs from the SN (the generating distribution).

6.3 Statistics of model comparison

In order to verify the performance of the statistics of model comparison, we conducted a simulation study considering four different scenarios. In the first two, we simulated $R = 100$ replicas (observations) of the log-SNBS regression model with $\alpha = .5$, $\boldsymbol{\beta} = (1, 2)^\top$, $\gamma = .67$, considering two samples sizes ($n = 50$, $n = 200$), fitting two competing models, the log-SNBS and log-BS regression ones. The last two scenarios are equivalent to the two first, but the replicas were simulated from the log-BS regression model. Table 4 presents the averaged criteria and the number of times (in percentage) that the underlying model was selected (within parenthesis) for the four scenarios. It can be seen that the true underlying model is chosen, with a high probability, in any situation, even under a small sample size.

Table 4. Averaged criteria and the number of times (in percentage) that the underlying model was selected for the simulation study (SMC).

True underlying model: log-SNBS			
Model	n	AIC	BIC
log-SNBS	50	69.481 (97%)	77.129 (95%)
	200	270.105 (98%)	283.299 (98%)
log-BS	50	73.389	79.125
	200	282.964	292.859
True underlying model: log-BS			
Model	n	AIC	BIC
log-SNBS	50	74.930	82.578
	200	287.320	300.514
log-BS	50	73.945 (97%)	79.681 (99%)
	200	284.139 (100%)	294.034 (100%)

7. REAL DATA ANALYSIS

We considered the data set analyzed by [20], which is related to football matches of the UEFA Champions League (*Union of European Football Associations*). These football matches are such that: (i) there was at least one goal scored by the home team, and (ii) there was at least one goal scored by either team from the penalty spot, lack of kick, or any other direct bid. Let T_1 be the time in minutes that the first goal was scored by either team and let T_2 be the time in minutes that the first goal of any sort, was scored by the home team. The objective is to predict the time in minutes for the first goal be scored by the home team based on the time in minutes the first goal scored by either team. From Figure 9 it can be seen that a linear model can be suitable to link the natural logarithm of these two variables.

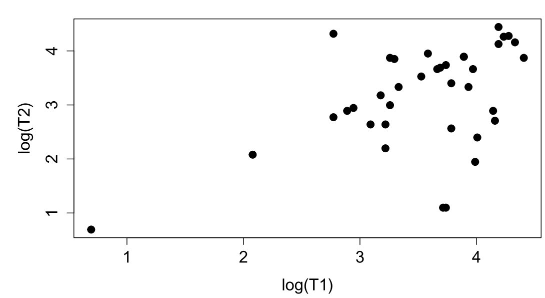

Figure 9. Scatter plot between the natural logarithm of the T1 and T2.

We assume that the response variable, in its original scale, can be modeled by a SN distribution. Therefore, the correspondent natural logarithm can be modeled by a SSN distribution.

The first proposed model is the log-SNBS model:

$$\mathrm{Y}_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ \epsilon_i \sim \mathrm{SSN}(\alpha, \boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma = 2, \gamma),$$
$$i = 1, \ldots, 37,$$

where $\mathrm{Y}_i = \log(T_{2_i})$, $x_i = \log(T_{1i})$, $T_{ji}, j = 1, 2$, is the value of the variable j for the team i, $\varepsilon_i \overset{i.i.d.}{\sim} SSN(\alpha, 0, \sigma = 2, \gamma)$. The second model is the log-BS model (i.e., considering $\gamma = 0$). Figures 10 and 11 present the residual analysis for both models. We can see that the log-SNBS model provides a better fit than the log-BS model. Indeed, from the simulated envelope shown in Figure 11(d), we can notice that the observations appear to form a slight downward-facing. Also, there are observations lying outside the simulated envelope for the log-BS model. However, the simulated envelope in Figure 10(d) indicates that the log-SNBS model offers an excellent fit to the data, since most of the observations are completely within the bands, without show any systematic behavior.

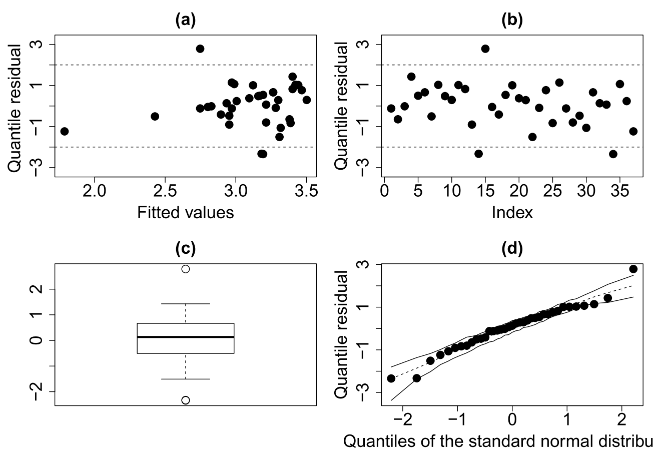

Figure 10. Residual analysis for the log-SNBS model.

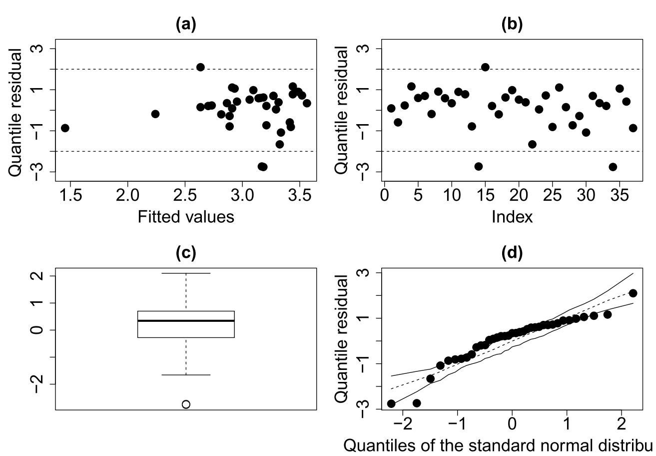

Figure 11. Residual analysis for the log-BS model.

Table 5 presents the estimates of the parameters, standard error (SE) and the 95% equi-tailed confidence intervals for both models. We have indications that the asymmetry parameter is different from zero, since such does not belong to the confidence interval. Also the larger is the time to the first goal be scored by either team, the higher is the time to a goal of any sort be scored. Moreover, both information criteria selected the log-SNBS model. Also, since the log-BS model did not fit properly to the data, we can conclude that the the respective SE are not well estimated, which could lead to wrong conclusions, if we would choose it. Also, the estimates of the regression parameters, between the two models, indicate that the respective true values are not equal. Therefore, different conclusions could be drawn from the two models.

Table 5. Estimates, standard error, 95% confidence intervals for the parameters of the the log-SNBS and log-BS models and model selection criteria.

Parameter		log-SNBS		
		Estimate	SE	$\mathrm{CI}_{95\%}$
α		.877	.113	[.841; .914]
β_0		1.468	.863	[1.189; 1.746]
β_1		.462	.238	[.385; .538]
γ		-.748	.211	[-.816; -.680]
AIC	92.839			
BIC	99.282			
Parameter		log-BS		
		Estimate	SE	$\mathrm{CI}_{95\%}$
α		.900	.105	[.866; .933]
β_0		1.060	.724	[.827; 1.293]
β_1		.568	.200	[.503; .632]
AIC	97.468			
BIC	102.301			

8. CONCLUDING REMARKS

In this paper we introduce a new regression model suitable to analyze data where the log response can be wither symmetric or (left/right) asymmetric. Frequentist inference is performed by a suitable ECM algorithm and tools for model fit assessment are proposed. The simulation studies indicated that the proposed methodologies perform very well. Some advantages of our proposal, over that based on the direct parameterization of the skew normal distribution, are illustrated. In the real data analysis, it is clear that: the log transformation did not lead to a symmetric behavior of the response variable and that our model is more appropriate than the usual log BS. As future research we suggest to consider other distributions to generate a more general class of Birnbaum-Saunders type distribution, as the skew student t. Also, nonlinear regression structures can be considered. In addition, instead of modeling the log response, the modeling of the original response, in terms of its mean, can be considered as in [29].

APPENDIX A. THE ECM ALGORITHM

The following result is used in the proof of Theorem 1 (which are helpful to obtain the conditional expectations in the EM algorithm).

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

$$\begin{aligned}
\mathbb{E}(X|X>a) &= \int_a^\infty x f(x|x>a)dx \\
&= \frac{1}{P(X<a)} \int_a^\infty x f(x)dx \\
&= \frac{1}{\Phi\left(\frac{a-\eta}{\tau}\right)} \left\{ \eta \left[1 - \Phi\left(\frac{a-\eta}{\tau}\right)\right] \right. \\
&\quad \left. + \tau \phi\left(\frac{a-\eta}{\tau}\right) \right\} \\
&= \eta + \frac{\phi\left(\frac{a-\eta}{\tau}\right)}{1-\Phi\left(\frac{a-\eta}{\tau}\right)} \tau.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(X^2|X>a) &= \int_a^\infty x^2 f(x|x>a)dx \\
&= \frac{1}{P(X<a)} \int_a^\infty x^2 f(x)dx \\
&= \frac{1}{\Phi\left(\frac{a-\eta}{\tau}\right)} \left\{ \eta^2 \left[1 - \Phi\left(\frac{a-\eta}{\tau}\right)\right] + \tau^2 \right. \\
&\quad \left. + \eta\tau\phi\left(\frac{a-\eta}{\tau}\right) + a\tau\phi\left(\frac{a-\eta}{\tau}\right) \right\} \\
&= \eta^2 + \tau^2 + \frac{\phi\left(\frac{a-\eta}{\tau}\right)}{1-\Phi\left(\frac{a-\eta}{\tau}\right)} (\eta + a)\tau.
\end{aligned}$$

Proof of Theorem 1. i) Let $Y \sim \text{SSN}(\alpha, \rho, \sigma = 2, \gamma)$ as in (3). Then,

$$\begin{aligned}
F_{Y|H}(y) &= P\left\{ \rho + 2\,\text{arcsinh}\left(\frac{\alpha[\delta H + \sqrt{1-\delta^2} X - \mu_z]}{2\sigma_z} \right) \right. \\
&\quad \left. \leq y \right\} \\
&= P\left[X \leq -\frac{\mu_z + \delta h}{\sqrt{1-\delta^2}} + \frac{\alpha\sigma_z}{2\sqrt{1-\delta^2}} \sinh\left(\frac{y-\rho}{2}\right) \right] \\
&= \Phi[X \leq \nu_h + \xi_2(y;\alpha_\delta,\rho)]
\end{aligned}$$

Deriving $F_{Y|H}(y|h)$ with respect to y, we obtain

$$f_{Y|H}(y) = \frac{1}{2}\phi\left[\nu_h + \xi_2(y;\alpha_\delta,\rho)\right] \xi_1(y;\alpha_\delta,\rho), y \in \mathbb{R},$$

where $\alpha_\delta = \alpha \frac{\sqrt{1-\delta^2}}{\sigma_z}$ and $\nu_h = -\frac{\mu_z + \delta h}{\sqrt{1-\delta^2}}$. Thus, the proof is concluded.

ii) The density of $H|(Y = y)$ can be obtained through the following relation $f_{H|T}(h) = \frac{f_{Y|H}(y) f_H(h)}{f_Y(y)}$, where $f_H(h) = 2\phi(h|0,1), h > 0$, $f_{Y|H}(y)$ was defined in (4), and $f_Y(y)$ was defined in (1). After some algebra, we obtain that

$$f_{H|Y}(h) = \frac{\Phi\left(\lambda(\gamma)\,\xi_{2y;\mu,\sigma}\right)}{\phi\left[h\middle|\delta\,\xi_{2y;\mu,\sigma}; 1-\delta^2\right]}, h > 0,$$

where $\phi(\cdot|\mu,\sigma^2)$. The conditional expectations are obtained from

$$\begin{aligned}
\mathbb{E}\left[H^k|Y\right] &= \frac{1}{\Phi\left(\lambda(\gamma)\,\xi_{2y;\mu,\sigma}\right)} \int_0^\infty h^k \phi\left[h\middle|\delta\,\xi_{2y;\mu,\sigma};\right. \\
&\quad \left. 1-\delta^2\right] dh \\
&= E(H^k|H>0).
\end{aligned}$$

Now, using the conditional expectations for truncated distributions presented in Lemma 1, we obtain

$$\begin{aligned}
\mathbb{E}(H|Y=y) &= \eta_y + W_\Phi\left(\frac{\eta_y}{\tau}\right)\tau, \\
\mathbb{E}(H^2|Y=y) &= \eta_y^2 + \tau^2 + W_\Phi\left(\frac{\eta_y}{\tau}\right)(\eta_y\tau)
\end{aligned}$$

where $\eta_y = \delta\,\sigma_z\left(\xi_{2y} + \frac{\mu_z}{\sigma_z}\right)$, $\tau = \sqrt{1-\delta^2}$ and $W_\Phi\left(\frac{\eta_y}{\tau}\right) = \frac{\phi\left(\frac{\eta_y}{\tau}\right)}{\Phi\left(\frac{\eta_y}{\tau}\right)}$. The proof is concluded.

APPENDIX B. THE OBSERVED FISHER INFORMATION MATRIX

$$\begin{aligned}
&\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_1 \partial\boldsymbol{\theta}_2^\top} \\
&= \xi_{2i;\mu,\sigma}\left(\frac{\partial^2\xi_{2i;\mu,\sigma}}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2^\top}\right) + \left(\frac{\partial\xi_{2i;\mu,\sigma}}{\partial\boldsymbol{\theta}_1}\right)\left(\frac{\partial\xi_{2i;\mu,\sigma}}{\partial\boldsymbol{\theta}_2}\right) \\
&\quad + \lambda(\gamma)\left[W_\Phi(\lambda(\gamma)\xi_{2i;\mu,\sigma})\left(\frac{\partial^2\xi_{2i;\mu,\sigma}}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2^\top}\right)\right. \\
&\quad \left. + \lambda(\gamma)W'_\Phi(\lambda(\gamma)\xi_{2i;\mu,\sigma})\left(\frac{\partial\xi_{2i;\mu,\sigma}}{\partial\boldsymbol{\theta}_1}\right)\left(\frac{\partial\xi_{2i;\mu,\sigma}}{\partial\boldsymbol{\theta}_2}\right)\right] + \\
&\quad \frac{1}{\xi_{1i;\sigma}^2}\left[\xi_{1i;\sigma}\left(\frac{\partial^2\xi_{1i;\sigma}}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2^\top}\right) - \left(\frac{\partial\xi_{1i;\sigma}}{\partial\boldsymbol{\theta}_1}\right)\left(\frac{\partial\xi_{1i;\sigma}}{\partial\boldsymbol{\theta}_2}\right)\right]; \\
&\quad \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \alpha, \boldsymbol{\beta},
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_3 \partial\gamma} \\
&= \xi_{2i;\mu,\sigma}\left(\frac{\partial^2\xi_{2i;\mu,\sigma}}{\partial\boldsymbol{\theta}_3\partial\gamma}\right) + \left(\frac{\partial\xi_{2i;\mu,\sigma}}{\partial\boldsymbol{\theta}_3}\right)\left(\frac{\partial\xi_{2i;\mu,\sigma}}{\partial\gamma}\right) \\
&\quad + \lambda(\gamma)W_\Phi(\lambda(\gamma)\xi_{2i;\mu,\sigma})\left(\frac{\partial^2\xi_{2i;\mu,\sigma}}{\partial\boldsymbol{\theta}_3\partial\gamma}\right) + \left(\frac{\partial\xi_{2i;\mu,\sigma}}{\partial\boldsymbol{\theta}_3}\right) \\
&\quad \times\left\{\lambda(\gamma)W'_\Phi(\lambda(\gamma)\xi_{2i;\mu,\sigma})\left[\lambda(\gamma)\left(\frac{\partial\xi_{2i;\mu,\sigma}}{\partial\gamma}\right) + \xi_{2i;\mu,\sigma}\right.\right. \\
&\quad \left.\left. \times\left(\frac{\partial\lambda(\gamma)}{\partial\gamma}\right) + W_\Phi(\lambda(\gamma)\xi_{2i;\mu,\sigma})\left(\frac{\partial\lambda(\gamma)}{\partial\gamma}\right)\right]\right\} \\
&\quad + \frac{1}{\xi_{1i;\sigma}^2}\left[\xi_{1i;\sigma}\left(\frac{\partial^2\xi_{1i;\sigma}}{\partial\boldsymbol{\theta}_3\partial\gamma}\right) - \left(\frac{\partial\xi_{1i;\sigma}}{\partial\boldsymbol{\theta}_3}\right)\left(\frac{\partial\xi_{1i;\sigma}}{\partial\gamma}\right)\right];
\end{aligned}$$

$\boldsymbol{\theta}_3 = \alpha, \boldsymbol{\beta}$,

$$\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \gamma^2}$$
$$= \xi_{2i;\mu,\sigma}\left(\frac{\partial^2 \xi_{2i;\mu,\sigma}}{\partial \gamma^2}\right) + \left(\frac{\partial \xi_{2i;\mu,\sigma}}{\partial \gamma}\right)^2 + W_\Phi(\lambda(\gamma)\xi_{2i;\mu,\sigma})$$
$$\times \left[\lambda(\gamma)\left(\frac{\partial^2 \xi_{2i;\mu,\sigma}}{\partial \gamma^2}\right) + \left(\frac{\partial \xi_{2i;\mu,\sigma}}{\partial \gamma}\right)\left(\frac{\partial \lambda(\gamma)}{\partial \gamma}\right)\right.$$
$$\left. + \xi_{2i;\mu,\sigma}\left(\frac{\partial^2 \lambda(\gamma)}{\partial \gamma^2}\right) + \left(\frac{\partial \lambda}{\partial \gamma}\right)\left(\frac{\partial \xi_{2i;\mu,\sigma}}{\partial \gamma}\right)\right]$$
$$+ W_\Phi'(\lambda \xi_{2i;\mu,\sigma})\left[\lambda\left(\frac{\partial \xi_{2i;\mu,\sigma}}{\partial \gamma}\right) + \xi_{2i;\mu,\sigma}\left(\frac{\partial \lambda}{\partial \gamma}\right)\right]^2$$
$$+ \frac{1}{\xi_{1i;\sigma}^2}\left[\xi_{1i;\sigma}\left(\frac{\partial^2 \xi_{2i;\mu,\sigma}}{\partial \gamma^2}\right) - \left(\frac{\partial \xi_{2i;\mu,\sigma}}{\partial \gamma}\right)^2\right],$$

where $W_\Phi'(x) = -W_\Phi(x)[x + W_\Phi(x)]$ is the derivative of $W_\Phi(x)$ with respect to x, see [34], and the other quantities are as before defined.

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 to Conselho Nacional Desenvolvimento Científico e Tecnológico for a research scholarship, grant number (308339/2015-0).

Received 27 April 2018

REFERENCES

[1] Atkinson, A. C. (1987). *Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. Oxford University Press.

[2] Arellano-Valle, R. B.; Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, 99(7), 1362–1382. MR2424355

[3] Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716–723. MR0423716

[4] Azevedo, C. L. N.; Bolfarine H.; Andrade, D. F. (2011). Bayesian inference for a skew-normal IRT model under the centred parameterization. *Comput. Stat. Data Anal*, 55(1), 353–365. MR2736560

[5] Azzalini, A. (1985). A class of distribution which includes the normal ones, *Scandinavian journal of statistics*, 12(2), 171–178. MR0808153

[6] Azzalini, A. (2013). *The skew-normal and related families*. IMS monographs, Cambridge University Press, Cambridge. MR3468021

[7] 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. MR2643505

[8] Balakrishnan, N.; Saulo, H.; Leão, J. (2017). On a new class of skewed Birnbaum-Saunders models. *Journal of Statistical Theory and Practice*, 11(4), 573–593. MR3683767

[9] 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. MR2516848

[10] Barros, M.; Paula, G. A.; Leiva, V. (2009). An R implementation for generalized Birnbaum-Saunders distributions. *Computational Statistics & Data Analysis*, 53(4), 1511–1528. MR2657109

[11] Birnbaum, Z. W.; Saunders, S. C. (1969a). A new family of life distributions. *Journal of Applied Probability*, 6(2), 637–652. MR0253493

[12] Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5), 1190–1208. MR1346301

[13] Cancho, V. G.; Ortega, E. M.; Paula, G. A. (2010). On estimation and influence diagnostics for log-Birnbaum-Saunders Student-t regression models: Full Bayesian analysis. *Journal of Statistical Planning and Inference*, 140(9), 2486–2496. MR2644070

[14] Chaves, N. L.; Azevedo, C. L. N.; Vilca-Labra, F.; Nobre, J. S. N. (2018a). A new Birnbaum-Saunders model based on the skew normal distribution under the centred parameterization. *Submitted to Chilean Journal of Statistics*. MR3946268

[15] Dunn, P. K.; Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3), 236–244.

[16] Huang, S.; Qu, Y. (2006). The Loss in Power When the Test of Differential Expression Is Performed under a Wrong Scale. *Journal of Computational Biology*, 13 (3), 786–797. MR2255443

[17] Lawless, J. F. (2011). *Statistical models and methods for lifetime data* . John Wiley & Sons. MR0640866

[18] Leiva, V.; Vilca, F.; Balakrishnan, N.; Sanhueza, A. (2010). A skewed sinh-normal distribution and its properties and application to air pollution. Communications in Statistics-Theory and Methods, 39(3), 426–443. MR2745286

[19] Lemonte, A. J.; Cordeiro, G. M. (2010). Asymptotic skewness in Birnbaum-Saunders nonlinear regression models. *Statistics & probability letters 80*, 9, 892–898. MR2608830

[20] Meintanis, S. G. (2007). Test of fit for marshall-olkin distributions with applications. *Journal of Statistical Planning and inference*, 137(12), 3954–3963. MR2368540

[21] Meng, X. L.; Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80(2), 267–278. MR1243503

[22] Owen, D. B. (1956). Tables for computing bivariate normal probabilities. *The Annals of Mathematical Statistics*, 1075–1090. MR0127562

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

[24] Rieck, J. R. (1989). *Statistical analysis for the Birnbaum-Saunders fatigue life distribution*, Ph.D dissertation, Clemson University, South Carolina, USA. MR2637462

[25] Rieck, J. R.; Nedelman, J. R. (1991). A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, 33(1), 51–60. MR2637462

[26] R Development Core Team (2008). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria*. http://www.R-project.org.

[27] Santana, L.; Vilca-Labra, F.; Leiva, V. (2011). Influence analysis in skew-Birnbaum-Saunders regression models and applications. *Journal of Applied Statistics*, 38(8), 1633–1649. MR2819378

[28] Santos, J. R. S. (2012). *An Item Response multiple groups model with centralized skew normal distributions (In Portuguese). Master's Dissertation, Department of Statistics, State University of Campinas*, http://repositorio.unicamp.br/bitstream/REPOSIP/306791/1/Santos_JoseRobertoSilvados_M.pdf.

[29] Santos-Neto, M.; Cysneiros, F. J. A.; Leiva, V.; Ahmed, S. E. (2012). On new parameterizations of the Birnbaum-Saunderes distribution. *Pakistan Journal of Statistics*, 38(8), 1633–1649. MR2931825

[30] Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461–464. MR0468014

[31] Singer, J. M.; Nobre, J. S.; Sef, H. C. (2004). Regression models for pretest/posttest data in blocks. *Statistical Modelling*, 4(4),

324–338. MR2086495

[32] 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. MR2274046

[33] van Dyk, D. A.; Meng, Xiao-Li (2006). The Art of Data Augmentation. *Journal of Computational and Graphical Statistics* 10(1), 1–50. MR1936358

[34] Vilca, F.; Santana, L.; Leiva, V. (2011). Estimation of extreme percentiles in Birnbaum-Saunders distributions. *Computational Statistics & Data Analysis*, 55(5), 1665–1678. MR2748670

[35] Vilca, F.; Zeller, C. B.; Cordeiro, G. M. (2015). The sinh-normal/independent nonlinear regression model. *Journal of Applied Statistics*, 42(8), 1659–1676, DOI:10.1080/02664763.2015.1005059. MR3350457

Nathalia L. Chaves
Department of Statistics
State University of Campinas
Brazil
E-mail address: nathalialimachaves@gmail.com

Caio L. N. Azevedo
Department of Statistics
State University of Campinas
Brazil
E-mail address: cnaber@ime.unicamp.br

Filidor Vilca-Labra
Department of Statistics
State University of Campinas
Brazil
E-mail address: fily@ime.unicamp.br

Juvêncio S. Nobre
Department of Statistics and Applied Mathematics
Federal University of Ceará
Brazil
E-mail address: juvenciosantos@gmail.com