A log Birnbaum-Saunders regression 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 regression model for positive and skewed data, a log Birnbaum-Saunders model based on the centred skew-normal distribution, and we present a several inference tools for this model. Initially, we developed a new version of skew-sinh-normal distribution and we describe some of its properties. For the proposed regression model, we carry out, through of the expectation conditional maximization (ECM) algorithm, the 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 real data analysis.

keywords: Birnbaum-Saunders distribution; Skew normal distribution; Skew sinh-normal distribution; Frequentist inference; ECM algorithm.

1 Introduction

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 Rieck (1989), have been receiving considerable attention in the past few years. These regression models are built using a 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 (Vilca and Leiva, 2006), Student-t BS (Barros et al., 2009), scale-mixture of normals BS (Balakrishnan et al., 2009) and skew scale-mixture Birnbaum-Saunders (Balakrishnan et al., 2017). In terms of log-BS regression models, we

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

can cite: Student-t BS model (Barros et al., 2008), skew-normal BS model (Santana et al., 2011) and scale-mixture of normals BS model Vilca et al. (2015).

In this paper, we develop a set of tools of statistical analysis for the log Birnbaum-Saunders regression model based on the skew-normal (SN) distribution under the centred parameterization (CP), see Azzalini (2013), named log-SNBS regression model. In the work Chaves et al. (2018a), 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 Vilca et al. (2011), similarly to the advantages of the SN (centred parameterization) compared with the usual SN (direct parameterization), see Pewsey (2000) and Azevedo et al. (2011). In this paper, we will show that these advantages are inherited by the respective log-SNBS regression model.

The aforementioned inference tools comprise of: parameter estimation, residual analysis and statistics for model comparison. Expectation conditional maximization (ECM) algorithm was used to develop these tools. 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 this study. 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 (1991), using simulated data.

The paper is outlined as follows. In Section 2, we present the log-SNBS distribution and we developed some of its properties. In Section 3, we introduce the log-SNBS regression model and discuss the ECM algorithm for the maximum-likelihood (ML) estimation of the model parameters. In Sections 4, 5 and 6, we carry out 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, denote 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

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

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and distribution functions of the standard normal distribution, respectively. Moreover, $a_{t;\mu,\sigma}(\alpha,\eta) = \mu_z + \sigma_z\, a_t(\alpha,\eta)$ and $A_{t;\sigma}(\alpha,\eta) = \sigma_z\, A_t(\alpha,\eta)$, with $a_t(\alpha,\eta) = \left(\sqrt{t/\eta} - \sqrt{\eta/t}\right)/\alpha$ and $A_t(\alpha,\eta) = \frac{d}{dt}a_t(\alpha,\eta) = \frac{t^{-3/2}(t+\eta)}{2\alpha\eta^{1/2}}$. Also, $\mu_z = r\delta$, $\sigma_z = \sqrt{1-\mu_z^2}$, $r = \sqrt{2/\pi}$, $\delta = \lambda/\sqrt{1+\lambda^2}$, $\lambda = \gamma^{1/3}s/\sqrt{r^2 + s^2\gamma^{2/3}(r^2-1)}$ and $s = [2/(4-\pi)]^{1/3}$. The model parameters are $(\alpha,\eta,\gamma)^\top$ and it will be called *centred parameters*, while the model parameters based the usual SN distribution Azzalini (1985), $(\alpha,\eta,\lambda)^\top$, are called as *direct parameters*. Note that for $\gamma = 0$, we have the usual BS distribution.

The construction of this r.v. and its behavior according the values of its parameters are presented in Chaves et al. (2018a). 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 the value of parameter α and η are, the smaller the variability. The higher the value of parameter η is, 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 Lawless (2011), 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 Meintanis (2007), wich 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:

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

where all quantities are as defined before. Figure 1 present the density of the SSN distribution for different values of α and γ. As mentioned, ρ and σ are the location and scale parameters, respectively, so that we fix the values of these parameters on the plots. We can notice that α and γ affects the kurtosis and symmetry of the SSN distribution, 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 Rieck and Nedelman (1991).

2.3 The proposed distribution

Here, we develop a generalization of the usual log-BS distribution of Rieck and Nedelman (1991) and Leiva et al. (2010) based on the centred parameterization of the SN distribution.

A logarithmic version of the SNBS model, called the log-SNBS distribution, can be obtained considering the r.v. $Y = \log(T)$, where $T \sim \text{SNBS}(\alpha, \eta, \gamma)$, whose pdf is given by

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

where $\xi_{2y;\mu,\sigma} = \mu_z + \sigma_z\,\xi_{2y}$ and $\xi_{1y;\sigma} = \sigma_z\xi_{1y}$, with $\xi_{2y} = \xi_2(y;\alpha,\rho) = \frac{2}{\alpha}\text{sinh}(\frac{y-\rho}{2})$, $\xi_{1y} = \xi_1(y;\alpha,\rho) = \frac{2}{\alpha}\text{cosh}(\frac{y-\rho}{2})$ and $\rho = \log(\eta)$. Moreover, μ_z, σ_z and λ are as defined before.

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

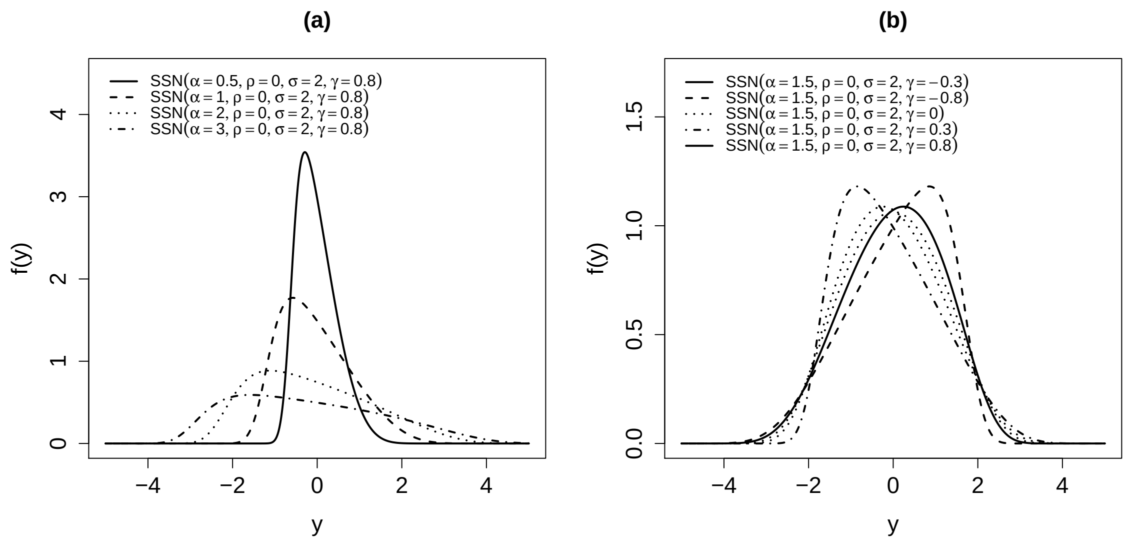

Figure 1: Density of the SSN distribution for different values of α (a) and different values of γ (b).

the SSN distribution when $\sigma = 2$ in (1). This former distribution can be also defined directly in terms of the SN r.v., similarly to the log-BS distribution, see Rieck and Nedelman (1991). That is, the distribution (2) may be stochastically represented as

$$Y = \rho + 2\,\mathrm{arcsinh}\big(\alpha\, Z/2\big), \tag{3}$$

where, in our case, $Z \sim \mathrm{SN}(0, 1, \gamma)$, with $\mathrm{SN}(0, 1, \gamma)$ standing for a SN distribution with zero mean, variance one and asymmetry parameter γ.

Figure 2 present the density of the SNBS distribution for different values of γ and α. In short, in terms of the three parameters, 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.

The following theorem is very useful to implement the ECM algorithm for ML estimation of the SNBS regression models.

Theorem 1. *Let $Y \sim SSN(\alpha, \rho, \sigma = 2, \gamma)$ as in (3) with Z having the representation given by $Z = \dfrac{1}{\sigma_z}\left[\delta|X_0| + \sqrt{1-\delta^2}X_1 - \mu_z\right]$, where $X_i \sim N(0,1); i = 0, 1$ are independent and $H = |X_0| \sim HN(0,1)$, where $HN(0,1)$ stands for a standard half normal distribution. Then,*

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

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

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

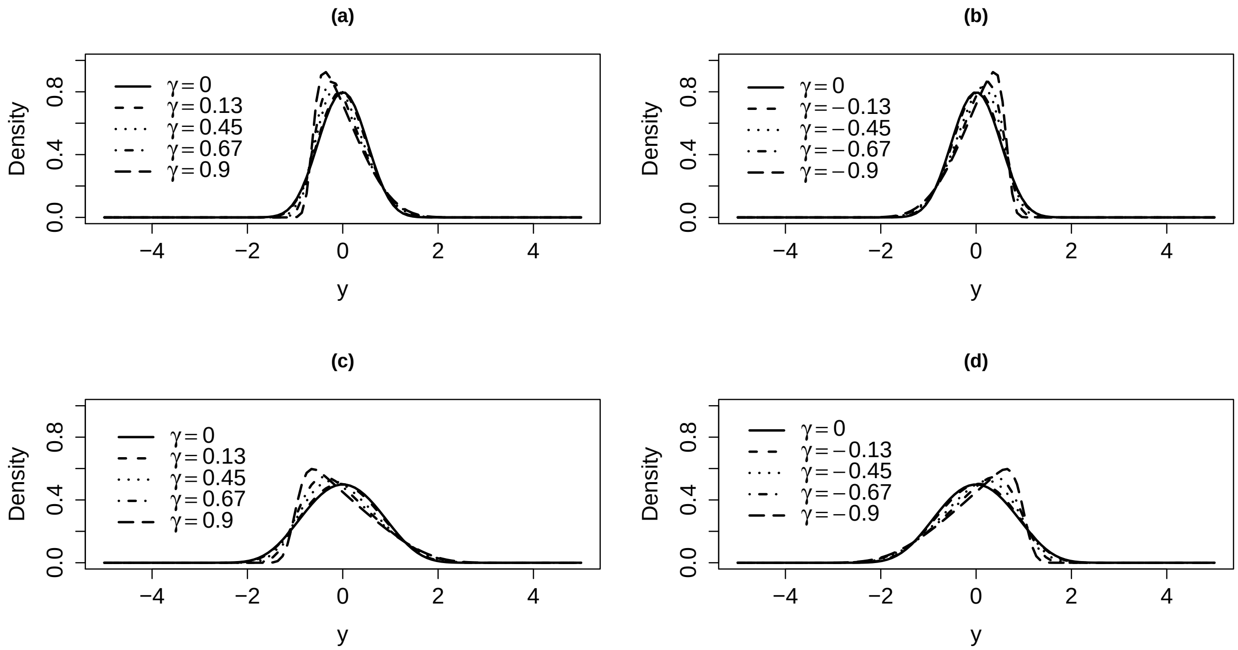

Figure 2: Density of the log-SNBS distribution for different values of γ, with $\rho = 0$, (a)$-$(b) $\alpha = .2$ and (c)$-$(d) $\alpha = .8$.

ii) $$f_{H|Y}(h) = \frac{\phi\left\{h \,\middle|\, \delta\,\sigma_z\left(\xi_{2y} + \dfrac{\mu_z}{\sigma_z}\right); 1 - \delta^2\right\}}{\Phi\left\{\lambda\left[\sigma_z\left(\xi_{2y} + \dfrac{\mu_z}{\sigma_z}\right)\right]\right\}}\mathbb{1}(h > 0),$$

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

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

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

The density in Theorem 1 corresponds to the four-parameter sinh-normal (SHN) distribution proposed by Leiva et al. (2010). The proof of the theorem is 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 Meng and Rubin (1993), which we implement for the maximum likelihood estimation (MLE) of the proposed model. In Appendix A, we present some results that are useful to obtain the maximum likelihood estimators.

The log-SNBS regression model is defined as follows. Suppose we have a sample of size n, say $Y_1, \ldots, Y_n$, where $Y_i \overset{\text{ind}}{\sim} \text{SSN}(\alpha, \boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma = 2, \gamma)$, for $i = 1, \ldots, n$. 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

$$Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \epsilon_i,\ \epsilon_i \sim \text{SSN}(\alpha, 0, \sigma = 2, \gamma),\ i = 1, \ldots, n. \tag{4}$$

Note that, when $\gamma = 0$, the log-BS regression model developed by Rieck and Nedelman (1991) is obtained.

For the log-SNBS regression model and $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top, \gamma)^\top$, its 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

$$\ell_i(\boldsymbol{\theta}|y_i) = \log\left[\phi\left(\xi_{2i;\mu,\sigma}\right)\right] + \log\left[\Phi\left(\lambda \xi_{2i;\mu,\sigma}\right)\right] + \log\left(\xi_{1i;\sigma}\right), \tag{5}$$

with $\xi_{2i;\mu,\sigma} = \mu_z + \sigma_z \xi_{2i}$, $\xi_{1i;\sigma} = \sigma_z \xi_{1i}$, such that $\xi_{1i} = \xi_1\left(y_i; \alpha, \boldsymbol{x}_i^\top \boldsymbol{\beta}\right) = \frac{2}{\alpha}\cosh\left(\frac{y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}}{2}\right)$, $\xi_{2i} = \xi_2(y_i; \alpha, \boldsymbol{x}_i^\top \boldsymbol{\beta}) = \frac{2}{\alpha}\sinh\left(\frac{y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}}{2}\right)$. Instead consider the direct maximization of (5), 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 augmented likelihood function. Also, instead estimating $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top, \gamma)^\top$, we will estimate $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top, \delta)^\top$, 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

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

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

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

For the current value $\boldsymbol{\theta}$, the E-step of the ECM algorithm requires the evaluation of $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathbb{E}\left[\ell(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}\right]$, where the expectation is taken with respect to the conditional distribution $H|(Y = y)$ and evaluated at $\widehat{\boldsymbol{\theta}}$. For the estimate of $\boldsymbol{\theta}$ at r-th iteration, say $\widehat{\boldsymbol{\theta}}^{(r)} =$

$(\widehat{\alpha}^{(r)}, \widehat{\boldsymbol{\beta}}^{(r)}, \delta^{(r)})^\top$, consider $\widehat{h}_i^{(r)} = \mathbb{E}[H_i|y_i, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(r)}]$ and $\widehat{h}_i^{2(r)} = \mathbb{E}[H_i^2|y_i, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(r)}]$ that are obtained by using the conditional expectation given in Theorem 1, and are given by

$$
\begin{aligned}
\widehat{h}_i &= \widehat{\eta}_{y_i} + W_\Phi\left(\frac{\widehat{\eta}_{y_i}}{\widehat{\tau}}\right)\widehat{\tau}, \quad \text{and} \\
\widehat{h}_i^2 &= \widehat{\eta}_{y_i}^2 + \widehat{\tau}^2 + W_\Phi\left(\frac{\widehat{\eta}_{y_i}}{\widehat{\tau}}\right)(\widehat{\eta}_{y_i}\widehat{\tau}),
\end{aligned} \tag{6}
$$

where, $\widehat{\eta}_{y_i} = \widehat{\delta}\sqrt{1 - r^2\widehat{\delta}^2}\left(\xi_2(y_i; \widehat{\alpha}, \boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}}) + \dfrac{r\widehat{\delta}}{1 - r^2\widehat{\delta}^2}\right)$, $\widehat{\tau} = \sqrt{1 - \widehat{\delta}^2}$ 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 function has the form

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

Hence, the ECM algorithm corresponds to iterate the following steps:
E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(r)}$, compute $\widehat{h}_i$ and $\widehat{h}_i^2$, for $i = 1, \ldots, n$ using results in (6);
CM-step 1: Fix $\widehat{\boldsymbol{\beta}}^{(r)}$ 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

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

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: Fix $\widehat{\alpha}^{(r+1)}$ and update $\widehat{\boldsymbol{\beta}}^{(r)}$ and $\widehat{\delta}^{(r)}$ using

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

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` (R Core Team, 2008), considering the L-BFGS-B optimization algorithm (Byrd et al., 1995).

We start the ECM algorithm with initial values $\widehat{\alpha}^{(0)}$, $\widehat{\boldsymbol{\beta}}^{(0)}$ and $\widehat{\delta}^{(0)}$. The values $\widehat{\boldsymbol{\beta}}^{(0)}$ can be obtained through ordinary least squares estimates of 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 Lemonte and Cordeiro (2010) to details. After getting $\widehat{\alpha}^{(0)}$ and $\widehat{\boldsymbol{\beta}}^{(0)}$, get $z_i = \left(2/\widehat{\alpha}^{(0)}\right)\sinh\left(\mathbf{y}_i - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}^{(r)}/2\right)$; $i = 1, \ldots, n$, observations that have 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

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

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 $\left\|\boldsymbol{\theta}^{(r)} - \boldsymbol{\theta}^{(r-1)}\right\| < \varepsilon, \varepsilon > 0$.

The observed information matrix is obtained as $I(\boldsymbol{\theta}) = -\ddot{\ell}$. Here, $\ddot{\ell} = [\ddot{\ell}_{\theta_1\theta_2}]$, $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \alpha, \boldsymbol{\beta}, \gamma$, 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 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 the maximum likelihood approach when the asymmetry parameter is close to zero. The log-SNBS regression model, based on the skew-normal of Azzalini (1985) seems to inherit such problems in the estimation. On the other hand, the log-SNBS regression which makes use of the CP of the SN distribution (Chaves et al., 2018a), circumvents problems inherited of the log-BS regression obtained by using SN distribution of Azzalini (1985) .

ii) When the asymmetry parameter is equal 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 get a direct perception of the problem, we have run a little simulation experiment generating $5,000$ samples of size $n = 200$ each from log-SNBS based on the skew-normal of Azzalini (1985) and for each sample the MLEs $(\hat{\alpha}, \widehat{\boldsymbol{\beta}}, \hat{\lambda})^\top$ have been computed. In this case, we fix $\alpha = .8$, $\boldsymbol{\beta} = (1, 2)^\top$ and $\lambda = 1$. Figure 3 displays the corresponding empirical distribution of $\hat{\alpha}$ and of $(\hat{\alpha}, \hat{\beta}_0)^\top$, in the form of an histogram (left panel) and a scatter plot (right panel), respectively. Moreover, it was generated $5,000$ samples of size $n = 200$ each from log-SNBS based on the CP and for each sample the MLEs $(\hat{\alpha}, \widehat{\boldsymbol{\beta}}, \hat{\gamma})^\top$ 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 Arellano and Azzalini (2008). The empirical distribution of the estimates $\hat{\alpha}$ of parameter α is as shown in the left panel of Figure 4, while that of $(\hat{\alpha}, \hat{\beta}_0)^\top$

is in the right panel of the same figure. 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.

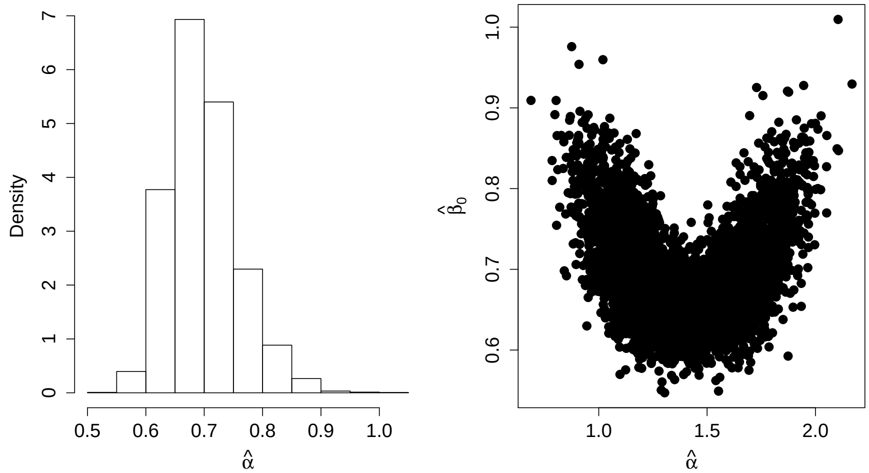

Figure 3: Estimated distributions of the MLEs when samples of size $n = 200$ are drawn from log-SNBS based on the skew-normal of Azzalini (1985); the left panel displays the histogram of $\hat{\alpha}$, the right panel displays the scatter plot of $(\hat{\alpha}, \hat{\beta}_0)^\top$.

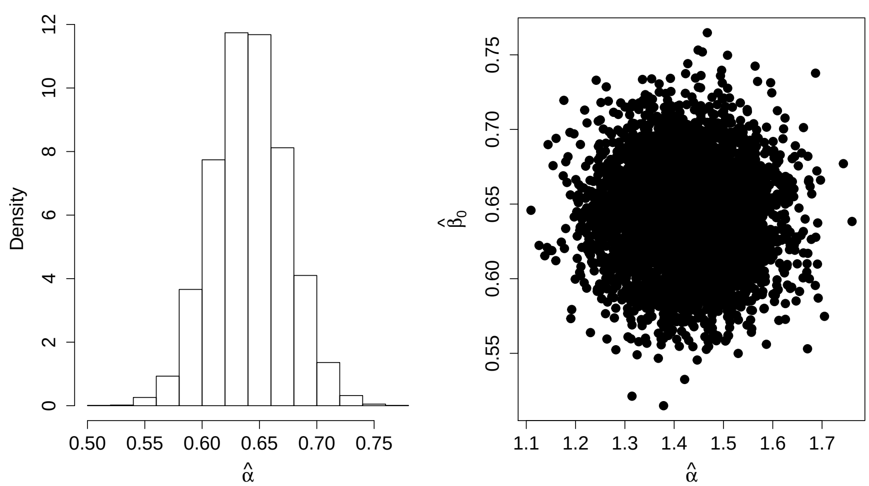

Figure 4: Estimated distributions of the MLEs when samples of size $n = 200$ are drawn from log-SNBS; the left panel displays the histogram of $\hat{\alpha}$, the right panel displays the scatter plot of $(\hat{\alpha}, \hat{\beta}_0)^\top$.

4 Residual analysis

The residual analysis is an important tool for model fit assessment. It is possible, through the residual analysis, checking the presence of outliers, as well as the departing from model assumptions. Following the methodology proposed by Dunn and Smyth (1996), 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 $F_{Y_i}(y_i) = \Phi_\gamma(\xi_{2i})$, where $\Phi_\gamma(\cdot)$ is the cdf of the $\mathrm{SN}(0, 1, \gamma)$, see Azzalini (2013).

Therefore we can define the quantile residual as

$$R_{q,i} = \Phi^{-1}\left\{\Phi_{\widehat{\gamma}}\left[\xi_2\left(y_i; \widehat{\alpha}, \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}\right)\right]\right\}, \tag{7}$$

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 $R_{q,i}$ converges in distribution to the standard normal distribution.

5 Statistics for Model comparison

There exist a variety of methodologies to compare several competing models for a given data set and to select the one that best fits the data. We consider model choice criteria which can be easily computed using the available ECM algorithm output, namely: the Akaike's information criterion (AIC) proposed by Akaike (1974) and Bayesian information criterion (BIC) proposed by Schwarz (1978). The AIC is based on the likelihood penalized by the number of model parameters. The BIC, in addition to the number of parameters, weights the sample size. These are defined as $AIC = -2\ell(\boldsymbol{\theta}|\mathbf{y}) + 2k$ and $BIC = -2\ell(\boldsymbol{\theta}|\mathbf{y}) + k\log(n)$, where $\ell(\boldsymbol{\theta}|\mathbf{y})$ it's the likelihood of the model defined in (5), k is the total number of model parameters and n is the number of observations. Lower AIC or BIC values indicate better fitting models.

6 Simulation studies

In this section we present three simulation studies: parameter recovery of the ECM algorithm (PRC), the behavior of the proposed residuals (R) and the 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, value of the parameter α (.5,1.5), that is, low and moderate variability, and value of the parameter γ (-.67,-.45,0,.45,.67), that is high and medium negative skewness, symmetry and high and medium positive skewness. For the PRC and SMC studies, all scenarios and $R = 100$ replicas (simulated responses from the model) were considered. For the other study, only one replica and only scenario was used. Specifically for the PRC study, we present only the results for three values of $\gamma(-.67, 0, .45)$ and one value of $\alpha(.5)$, since for the others scenarios the patterns were similar and can be found in the supplementary material. More specific details concerning each study are presented in the following subsections.

The general model used was

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

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

6.1 Parameter recovery

As previously mentioned, we present only those related to the scenario where $\alpha = .5$, $\gamma(-.67, 0, .45)$, varying the value of the sample size. The sample sizes were chosen in order to verify the proprieties of the ML estimators, as consistency, and their behavior, in terms of accuracy.

We calculated the usual statistics to measure the accuracy of the estimates: bias, variance (Var), root mean squared error (RMSE) and absolute value of 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: BIAS $= \overline{\widehat{\theta}} - \theta$, Var $= (1/R)\sum_{r=1}^{R}(\widehat{\theta}_r - \overline{\widehat{\theta}})^2$, 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 obtained for α, β_0 and β_1 tend to the correspondent true values in all scenarios. On the other hand, under the sample sizes equal to $n = 10$ and $n = 50$, γ is overestimated, sometimes underestimated. Under the biggest sample size ($n = 200$), this parameter is always overestimated.

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

Parameter	n	Mean	Variance	Bias	REQM	AVRB
	10	.447	.002	-.053	.068	.105
α	50	**.504**	**<.001**	**.004**	**.021**	**.008**
	200	.515	**<.001**	.015	.025	.030
	10	.966	.080	-.034	.285	.034
β_0	50	**.998**	.013	**-.002**	.114	**.002**
	200	.996	**.003**	-.004	**.054**	.004
	10	2.038	.328	.038	.574	.019
β_1	50	**2.013**	.053	**.013**	.230	**.006**
	200	2.019	**.011**	.019	**.104**	.009
	10	**-.575**	.377	.095	.622	.142
γ	50	**-.718**	.080	**-.048**	.287	**.072**
	200	-.778	**.017**	-.108	**.169**	.161

6.2 Behavior of the residuals

For this study we considered the scenario where $\alpha = .5$, $\gamma = .67$ and $n = 200$. Here we simulated only one set of observations for four different models: log-SNBS, log-BS, log-BS-t (Cancho et al., 2010) and log-StBS (Balakrishnan et al., 2017). The first one is the model given by (4) while the second corresponds to its particular case when $\gamma = 0$. The third and the fourth models correspond to the model (4) using in (3) instead of a centred SN distribution, an 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 (7). Four plots were built for each situation, including an simulated 95% confidence envelope for

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

Parameter	n	Mean	Variance	Bias	REQM	AVRB
	10	.445	.001	-.055	.064	.110
α	50	.488	<**.001**	-.012	.016	.024
	200	**.498**	<**.001**	**-.002**	**.004**	**.005**
	10	1.017	.082	.017	.286	.017
β_0	50	.998	.018	-.002	.136	.002
	200	**1.001**	**.004**	**.001**	**.061**	**.001**
	10	1.983	.315	-.017	.562	.008
β_1	50	**2.002**	.072	**.002**	.268	**.001**
	200	1.996	**.015**	-.004	**.122**	.002
	10	-.016	**.011**	-.016	**.108**	-
γ	50	-.039	.026	-.039	.166	-
	200	**.006**	.012	**.006**	.109	-

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

Parameter	n	Mean	Variance	Bias	REQM	AVRB
	10	.439	.003	-.061	.081	.123
α	50	**.506**	.001	**.006**	.028	**.012**
	200	.513	<**.001**	.013	.025	.026
	10	1.051	.128	.051	.361	.051
β_0	50	.992	.016	-.008	.128	.008
	200	**1.001**	**.003**	**.001**	**.056**	**.001**
	10	1.907	.426	-.093	.659	.047
β_1	50	**2.001**	.067	**.001**	.258	**.001**
	200	1.992	**.012**	-.008	**.112**	.004
	10	**.432**	.461	**-.018**	.679	**.040**
γ	50	.525	.156	.075	.401	.166
	200	.614	**.033**	.164	**.245**	.364

the residuals, and they are 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 of them within the simulated 95% confidence envelope within the interval (-2,2), with no systematic behavior. On the other hand, when the underlying model is the log-BS-t, we observe some outliers and many observations that tend to outside the simulated 95% confidence envelope, which, in its turn, presents a behavior compatible with a heavy tails distributions. 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 95% confidence envelope, which, in its turn, presents a behavior compatible with a skewed heavy tails distributions. In conclusion, we can say that the proposed residuals are appropriate to detect when the model does not fit properly to the data, concerning the generating distribution, identifying how this distribution differs from the SN (the generating distribution).

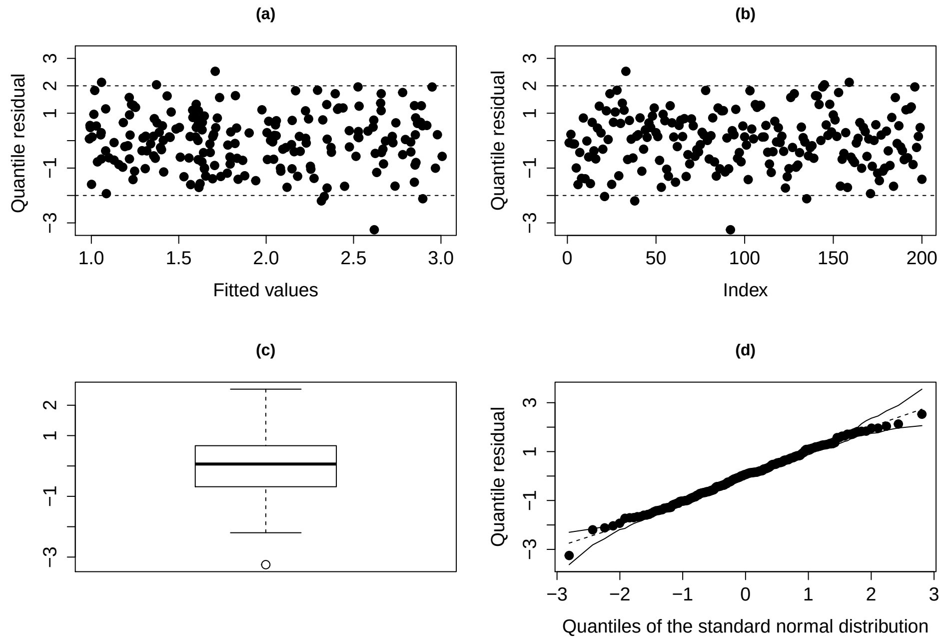

Figure 5: Residual plots for the observations generated from a log-SNBS regression model.

6.3 Statistical 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 scenarios, 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 ($n = 50$, $n = 200$) and we fit two competing models, the log-SNBS and log-BS regression models. The last two scenarios are equivalent to the two first, but the replicas were simulated from the log-BS regression model.

In the first two scenarios, for $n = 50$, the criteria AIC and BIC chose the log-SNBS regression model (the underlying model) in 97% and 95% of the replicas, respectively, whereas, for $n = 200$, both statistics chosen the true model in 98% of the scenarios. On the other hand, for the two last scenarios, under $n = 50$, the criteria AIC and BIC chose the log-BS model (the underlying model) in 97% and 99% of the scenarios, respectively, whereas, under $n = 200$, these percentages were 100% and 100%, respectively. Table 4 presents the averaged criteria for the four scenarios. It can be seen that the underlying model is chosen, with a high probability, in any situation, even under a small sample size.

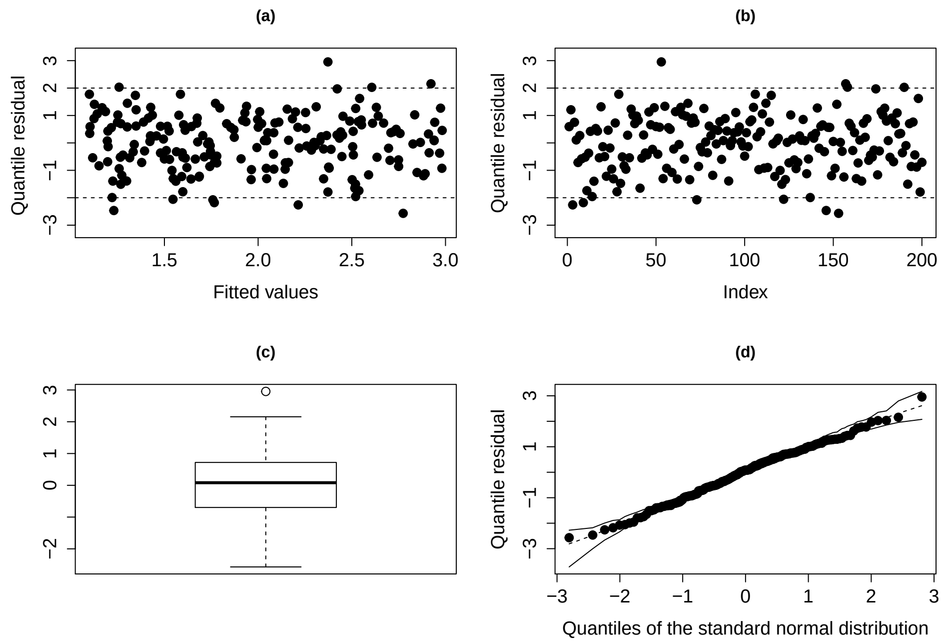

Figure 6: Residual plots for the observations generated from a log-BS regression model.

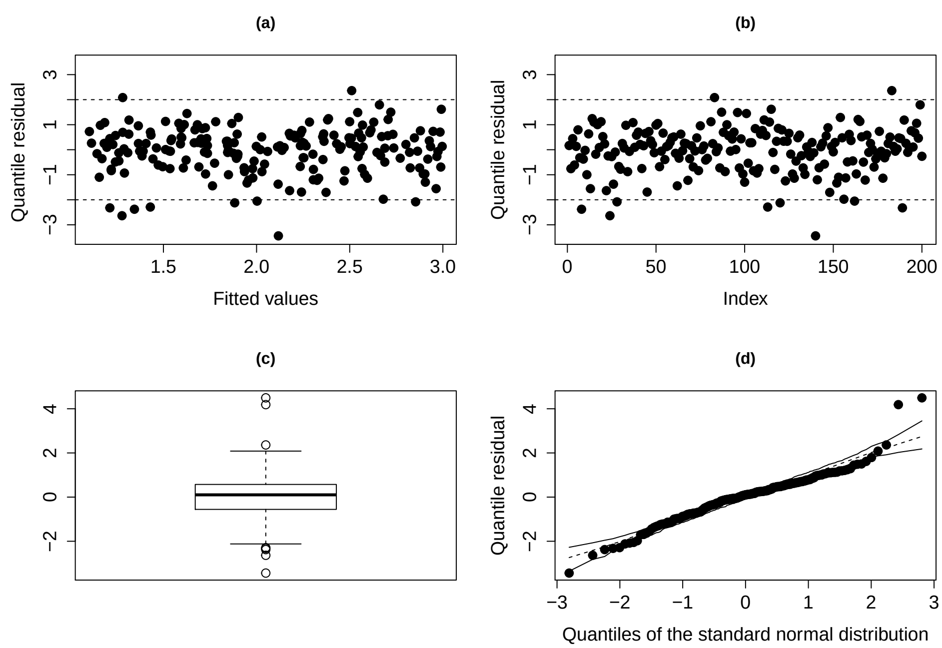

Figure 7: Residual plots for the observations generated from a log-BS-t model.

7 Real data analysis

We considered the data set analyzed by Meintanis (2007), which is related to football matches of the UEFA Champions League (*Union of European Football Associations*). It is related to football

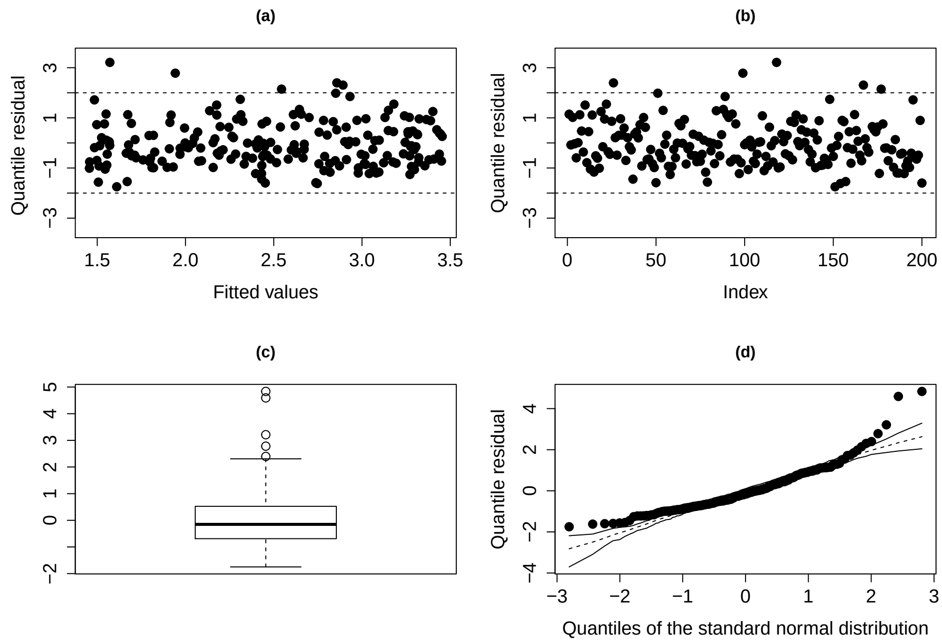

Figure 8: Residual plots for the observations generated from a log-StBS model.

Table 4: Averaged criteria for the simulation study (SMC).

True underlying model: log-SNBS			
Model	n	AIC	BIC
log-SNBS	50	69.481	77.129
	200	270.105	283.299
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	79.681
	200	284.139	294.034

matches where (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 eas 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.

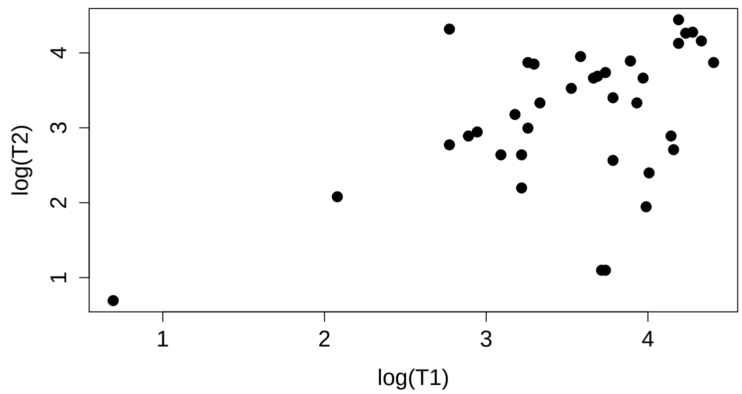

Figure 9: Scatter plot between the natural logarithm of the T1 and T2.

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

The two proposed model are (the log-SNBS model):

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

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, $\varepsilon_i|\boldsymbol{\theta} \overset{i.i.d.}{\sim} SSN(\alpha, 0, \sigma = 2, \gamma)$ and the log-BS model (i.e., considering $\gamma = 0$). Figures 10 and 11 present the residual analysis for the two models. We detect that the log-SNBS model provides a better fit than the log-BS model for the UEFA Champions League data. Specifically, from the simulated 95% confidence envelope shown in Figure 11(d), we can notice that the observations appear to form a slight downward-facing. Also, there are observations absolutelying outside the simulated 95% confidence envelope for the log-BS model. However, the simulated 95% confidence envelope in Figure 10(d) indicates that the log-SNBS model offers an excellent fit to the UEFA Champions League data, providing that all the observations are inside of the simulated 95% confidence envelope, without show any systematic behavior.

Table 5 presents the estimates of the parameters, standard error (SE) and the 95% equi-tailed confidence intervals for the two models. We have indications that the asymmetry parameter is different from zero, since zero does not belong 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, it is noted that both criteria selected the log-SNBS model.

8 Concluding Remarks

In this paper, we introduce a new log-SNBS regression model and develop several inference tools for this model. Parameter estimation, model fit assessment and model comparison were developed through ECM algorithm. The results from the simulation studies indicated that the ML method recovered all parameters properly. Also, the tools for model comparison and model fit assessment indicated that the log-SNBS regression model fitted to the data well and better than the usual

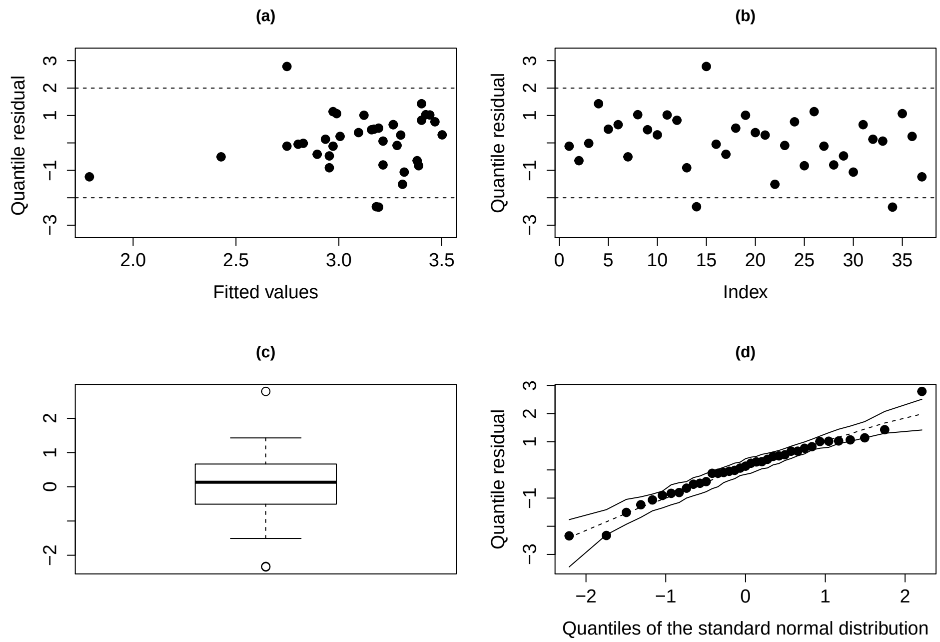

Figure 10: Residual analysis for the log-SNBS model.

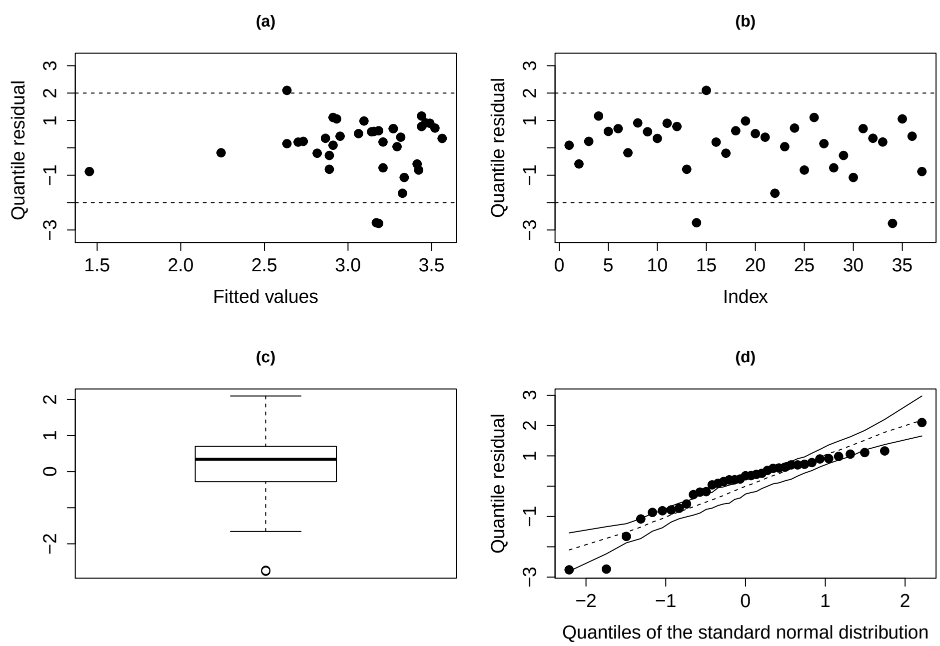

Figure 11: Residual analysis for the log-BS model.

log-BS model. As future developments, we suggest the use of other family distributions, as the skew scale-mixture of normals distributions, to generate new family of BS-type distributions.

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

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.

References

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

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

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.

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.

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

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.

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

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

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.

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.

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. *manuscript under preparation*.

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

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

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.

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

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

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

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

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

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

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

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.

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

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.

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.

Appendix

Appendix A: The ECM algorithm

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

Lemma 1. *Let $X \sim N(\eta, \tau^2)$, 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 $Z \sim \mathrm{SN}(0, 1, \gamma)$, using the stochastic representation, we can define Z as $Z = \frac{1}{\sigma_z}\left[\delta H + \sqrt{1-\delta^2}X_1 - \mu_z\right] = \frac{2}{\alpha}\sinh\left(\frac{z-\rho}{2}\right)$.
Therefore, $Z|(H = h) = \frac{2}{\alpha}\sinh\left(\frac{z-\rho}{2}\right)\Big|(H = h) \sim N(\mu_h, \sigma^2)$, where $\mu_h = \frac{\delta h - \mu_z}{\sigma_z}$ and $\sigma_h^2 = \frac{1-\delta^2}{\sigma_z^2}$. Then,

$$W|(H = h) = -\frac{\mu_h}{\sigma} + \frac{2}{\sigma\alpha}\sinh\left(\frac{z-\rho}{2}\right)\Big|(H = h) \sim N(0, 1).$$

$$Y = \rho + 2\mathrm{arcsinh}\left[\frac{\alpha}{2}(\mu_h + \sigma_h W)\right].$$

From the above result, the proof is concluded.

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

$$\phi[\nu_h + \xi_2(y; \alpha_\delta, \rho)] = \frac{\sqrt{1-\delta^2}}{\sigma_z}\phi\left(\xi_{2y}\Big|\frac{\delta h - \mu_z}{\sigma_z}; \frac{1-\delta^2}{\sigma_z^2}\right).$$

Then, we have

$$\phi\left(\xi_{2y}\Big|\frac{\delta h - \mu_z}{\sigma_z}; \frac{1-\delta^2}{\sigma_z^2}\right)\phi(h|0, 1) = \phi\left(\xi_{2y}\Big| - \frac{\mu_z}{\sigma_z}; \frac{1}{\sigma_z^2}\right)\phi\left[h\Big|\delta\xi_{2y;\mu,\sigma}; 1 - \delta^2\right],$$

where $\phi(\cdot|\mu, \sigma^2)$ denotes the density of normal with mean μ and variance σ^2.

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

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

Then, using some proprieties of the half-normal (HN) distribution from Lemma 1, 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) + \lambda\left[W_\Phi(\lambda\xi_{2i;\mu,\sigma})\left(\frac{\partial^2\xi_{2i;\mu,\sigma}}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2^\top}\right)\right.\\
&\left. +\ \lambda W_\Phi^{'}(\lambda\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] + \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)\right.\\
&\left. -\ \left(\frac{\partial\xi_{1i;\sigma}}{\partial\boldsymbol{\theta}_1}\right)\left(\frac{\partial\xi_{1i;\sigma}}{\partial\boldsymbol{\theta}_2}\right)\right]; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \alpha, \boldsymbol{\beta},\\
\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) + \lambda W_\Phi(\lambda\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\{\lambda 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) + W_\Phi(\lambda\xi_{2i;\mu,\sigma})\left(\frac{\partial\lambda}{\partial\gamma}\right)\right]\right\}\\
&+ \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]; \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\xi_{2i;\mu,\sigma})\left[\lambda\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}{\partial\gamma}\right)\right.\\
&\left. +\ \xi_{2i;\mu,\sigma}\left(\frac{\partial^2\lambda}{\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],
\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.