

Bayesian inference for a Birnbaum-Saunders regression model based on the centered skew normal distribution

Nathalia L. Chaves¹, Caio L N Azevedo^{1*}, Filidor Vilca-Labra¹,
Juvêncio S. Nobre²

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 developed Bayesian inference tools for a log Birnbaum-Saunders regression model based on the skew normal distribution under the centered parameterization. Parameter estimation, model fit assessment, model comparison, residual analysis and Bayesian case influence diagnostics were developed through MCMC algorithms. Also, a comparison with the maximum likelihood, previously proposed in the literature, was performed, in terms of parameter recovery. We noticed that the results are quite similar, but the Bayesian approach is more easily implemented and for developing influence diagnostics tools, which also allows incorporating prior information. Finally, a real data set is analyzed. The results indicate that our model outperforms the usual log Birnbaum-Saunders regression model in terms of model fit.

keywords: Birnbaum-Saunders distribution; Skew normal distribution; Skew sinh-normal distribution; Bayesian inference; MCMC algorithms.

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

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

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, see Leiva et al. (2009). Examples of these distributions are: skew-elliptical BS Vilca and Leiva (2006), Student-t BS, Barros et al. (2008), scale-mixture of normals BS, Balakrishnan et al. (2009) and skew normal BS, Vilca et al. (2011). In terms of log-BS regression models, we 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 Bayesian statistical analysis for the skew normal log-BS regression model proposed by Chaves et al. (2016b), which is based on the skew normal distribution under the centered parameterization (SNCP) BS distribution developed by Chaves et al. (2016a), named SNCP log-BS regression model. In their former work, the authors provided empirical evidences that their SNCPPBS (skew normal under the centered parameterization Birnbaum-Saunders) distribution has advantages, in terms of inference, both frequentist and Bayesian, over the skew normal BS distribution proposed by Vilca et al. (2011), similarly to the advantages of the SNCP compared with the usual skew normal (direct parameterization), see Pewsey (2000) and Azevedo et al. (2011). Therefore, we can conjecture that these advantages are inherited by the respective skew normal log-BS regression models.

The aforementioned inference tools comprise of: parameter estimation, residual analysis, statistics for model comparison and case influence diagnostics. MCMC algorithms are used to develop these tools. We present a comparison, in terms of parameter recovery, between the Bayesian inference developed and the maximum likelihood inference developed by Chaves et al. (2016b). Also, the impact of some factors of interest (sample size, asymmetry level of the log SNCPPBS distribution and the value of the parameter α , the shape parameter) on the estimates, are measured through this study. In addition, the performance of some 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. In addition, the results obtained through the MCMC algorithm developed using the original likelihood and another developed by using an augmented likelihood are compared. Also, Bayesian influence diagnostics tools are proposed and studied. Since the log-BS regression model is a particular case of our model, all developments presented here are directly applicable to that model. The paper is outlined as follows. In Section 2, we present the log-BS regression model based on the skew normal distribution under the centered parameterization along with the original and augmented likelihoods. In Section 3, Bayesian inference (parameter estimation, residual analysis, model comparison and case influence diagnostics) through MCMC algorithms are developed. In Section 4, simulation studies are presented. In Section 5, a real data analysis is discussed and finally, in Section 6, the concluding remarks are given.

2 Regression model and related likelihoods

2.1 The SNCP log-BS distribution

A positive random variable, says T , is said to follow a SNCPBS distribution, denote by $T \sim \text{SNCPBS}(\alpha, \eta, \gamma)$, $\alpha, \eta \in \mathbb{R}, \gamma \in (-0.99527, 0.99527)$, where α is the shape parameter, η is the location parameter and γ is the asymmetry parameter, if its density is given by

$$f_T(t|\alpha, \eta, \delta) = 2\phi[a_{t;\mu,\sigma}(\alpha, \eta)] \Phi\{\lambda a_{t;\mu,\sigma}(\alpha, \eta)\} A_{t;\sigma}(\alpha, \eta), t > 0, \quad (1)$$

where $a_{t;\mu,\sigma}(\alpha, \eta) = \mu_z + \sigma_z a_t(\alpha, \eta)$ 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 $A_t(\alpha, \eta) = \frac{d}{dt} a_t(\alpha, \eta) = \frac{t^{-3/2}(t+\eta)}{2\alpha\eta^{1/2}}$. Moreover $\mu_z = r\delta$, $\sigma_z^2 = 1 - \mu_z^2$, $r = \sqrt{\frac{2}{\pi}}$, $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ and $\gamma = r\delta^3(4/\pi - 1)(1 - \mu_z^2)^{-3/2}$, which represents the Pearson's asymmetry coefficient of the SNCP distribution and λ is the respective original asymmetry parameter, see Azzalini (1985). As mentioned before, γ is the asymmetry parameter (not the Pearson's asymmetry coefficient) of the SNCPBS. We will work with parameter γ since its parameter space is bounded and it presents an easy interpretation.

The construction of this random variable and its behavior according the values of its parameters are presented in Chaves et al. (2016b). 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 higher the value of parameter α is, the smaller the variability and the smaller/higher the value of the parameter η , and the more shifted to the right/left is the distribution. Another interesting feature of this distribution is that it can model properly positive random variables with a negatively skewed behavior. When $\gamma = 0$, our model corresponds to the usual BS distribution, see Birnbaum and Saunders (1969).

The correspondent SNCP log-BS distribution is defined as $Y = \log(T)$, which is a particular case of the skew normal hyperbolic-sine distribution under the centered parameterization (SNHSCP), see Chaves et al. (2016b), and its density is given by

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

where $\xi_{2y;\mu,\sigma} = \mu_z + \sigma_z \xi_{2y}$ and $\xi_{1y;\sigma} = \sigma_z \xi_{1y}$, with $\xi_{2y} = \xi_2(y; \alpha, \rho) = \frac{2}{\alpha} \sinh(\frac{y-\rho}{2})$, $\xi_{1y} = \xi_1(y; \alpha, \rho) = \frac{2}{\alpha} \cosh(\frac{y-\rho}{2})$ and $\rho = \log(\eta)$.

We denote this distribution by $Y \sim \text{SNHSCP}(\alpha, \mu, \sigma = 2, \gamma)$. In short, in terms of the three parameters, the distribution is symmetric around ρ , for $\gamma = 0$ (in this case we have the sinh-normal 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. We use this notation, including a specific value for the parameter σ , since the SNCP log-BS distribution is a particular case of the SNHSCP distribution

when $\sigma = 2$, see Chaves et al. (2016b). This former distribution can be also defined directly in terms of the SNCP random variable, similarly to the sinh-normal (symmetric) distribution, see Rieck and Nedelman (1991). That is, the distribution (2) may be stochastically represented as

$$Y = \rho + \sigma \operatorname{arcsinh}(\alpha Z/2), \quad (2)$$

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

2.2 The SNCP log-BS regression model and the likelihoods

Under the Bayesian paradigm, the SNPC log-BS regression model is given by

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \varepsilon_i | (\boldsymbol{\beta}, \alpha, \gamma) \stackrel{i.i.d.}{\sim} \text{SNHSCP}(\alpha, 0, \sigma = 2, \gamma), i = 1, \dots, n,$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is the vector of the regression parameters and $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ is the vector of the (non-stochastic) co-variables of observation i . Equivalently, $\varepsilon_i = \log(T_i), T_i | (\boldsymbol{\beta}, \alpha, \gamma) \stackrel{iid}{\sim} \text{SCNPBS}(\alpha, \eta = 1, \gamma)$. Also, from a property of the SNHSCP distribution presented in Chaves et al. (2016b), we have that $Y_i | \boldsymbol{\theta} \stackrel{ind.}{\sim} \text{SNHSCP}(\alpha, \mathbf{x}_i^\top \boldsymbol{\beta}, \sigma = 2, \gamma)$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \alpha, \gamma)$. Therefore, we have that the original likelihood is given by

$$L(\boldsymbol{\theta} | \mathbf{y}) = \prod_{i=1}^n \phi(\xi_{2i;\mu,\sigma}) \Phi(\lambda \xi_{2i;\mu,\sigma}) \xi_{1i;\sigma}, \quad (3)$$

where $\xi_{2i;\mu,\sigma} = \mu_z + \sigma_z \xi_{2i}$ and $\xi_{1i;\sigma} = \sigma_z \xi_{1i}$, with $\mathbf{y} = (y_1, \dots, y_n)^\top$, $\xi_{1i} = \xi_1(y_i; \alpha, \mathbf{x}_i^\top \boldsymbol{\beta}) = \frac{2}{\alpha} \cosh\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right)$, $\xi_{2i} = \xi_2(y_i; \alpha, \mathbf{x}_i^\top \boldsymbol{\beta}) = \frac{2}{\alpha} \sinh\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right)$ are as in Equation (2) and the other quantities are as before defined.

Also, it is possible to consider a hierarchical representation of the SCNP as in Azevedo et al. (2012), which is useful for stochastic simulation and parameter estimation, through EM and MCMC algorithms. Therefore, the following hierarchical structure can be defined

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

where $\alpha_\delta = \alpha \sqrt{\frac{1-\delta^2}{1-r^2\delta^2}}$, $\nu_{h_i} = -\frac{\delta(h_i-r)}{\sqrt{1-\delta^2}}$ and $H_i, i = 1, \dots, n$ are augmented (non-observable) variables and the other quantities are as defined before. Then, defining $\mathbf{y}_c = (\mathbf{y}, \mathbf{h}^\top)^\top$, where $\mathbf{h} = (h_1, \dots, h_n)^\top$, the augmented likelihood is given by

$$L(\boldsymbol{\theta} | \mathbf{y}_c) = \prod_{i=1}^n f_{Y|H}(y_i | h_i) f_H(h_i) = \prod_{i=1}^n \sqrt{2/\pi} \phi(\xi_{2i}^*) \xi_{1i}^* \exp\left\{-\frac{1}{2} h_i^2\right\}, \quad (4)$$

where $\xi_{1i}^* = \xi_1(y_i; \alpha_\delta, \mathbf{x}_i^\top \boldsymbol{\beta}) = \frac{2}{\alpha_\delta} \cosh\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right)$ and $\xi_{2i}^* = \nu_{h_i} + \xi_2(y_i; \alpha_\delta, \mathbf{x}_i^\top \boldsymbol{\beta})$ are as in Equation (2).

These two likelihoods will be explored with the obtaining of the marginal posterior distributions of interest, through suitable MCMC algorithms, in the next section.

3 Prior and posterior distributions and MCMC algorithms

The marginal posterior distributions comprise the main tool to perform Bayesian inference. Unfortunately, it is not possible to obtain closed-form expressions of the marginal posterior distributions for our model, regardless the prior structure adopted and/or the likelihood considered. MCMC algorithms will be used to obtain samples from the marginal posteriors, see Gamerman and Lopes (2006). Also, since none of the so-called full conditional distributions are known (as we will show), some auxiliary algorithm needs to be used to sample from them, such as the Metropolis-Hastings, slice sampling or adaptive rejection sampling. We made all implementations in the R program, see R Core Team (2008), through the package `R2OpenBUGS`, see Sturz et al. (2005). That is, the auxiliary algorithms used to simulate from the full conditional distributions follow the predefined hierarchy implemented in `OpenBUGS`, see Spiegelhalter et al. (2014). The respective codes are available from the authors upon request.

Regardless the likelihood considered, the prior structure adopted here, for $\boldsymbol{\theta}$, is $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\alpha)\pi(\gamma)$, where $\boldsymbol{\beta} \sim \mathcal{N}_p(\boldsymbol{\mu}_\beta, \boldsymbol{\Psi})$, $\alpha \sim \text{gamma}(r_\alpha, \lambda_\alpha)$ and $\gamma \sim U(a, b)$, that is

$$\pi(\boldsymbol{\theta}) \propto \alpha^{(r_\alpha-1)} \exp\left\{-\frac{1}{2}(2\alpha\lambda_\alpha + Q(\boldsymbol{\beta}))\right\}, \quad (5)$$

where $Q(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)$. Combining the original likelihood (3) and the prior distribution (5), the joint posterior distribution is given by

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \prod_{i=1}^n \phi(\xi_{2i;\mu,\sigma}) \Phi(\lambda \xi_{2i;\mu,\sigma}) \xi_{1i;\sigma} \alpha^{(r_\alpha-1)} \exp\left\{-\frac{1}{2}(2\alpha\lambda_\alpha + Q(\boldsymbol{\beta}))\right\}. \quad (6)$$

On the other hand, using the augmented likelihood (4), the respective posterior distribution is given by

$$\pi(\boldsymbol{\theta}, h|\mathbf{y}) \propto \prod_{i=1}^n \phi(\xi_{2i}^*) \xi_{1i}^* \alpha^{(r_\alpha-1)} \exp\left\{-\frac{1}{2}(h_i^2 + 2\alpha\lambda_\alpha + Q(\boldsymbol{\beta}))\right\} \quad (7)$$

The joint posterior distributions (6) and (7) have intractable forms but, the full conditionals are easy to sample from even though they are not known (do not correspond

to a particular distribution), which are:

$$\begin{aligned}\pi(\alpha|\boldsymbol{\beta}, \gamma, \mathbf{y}) &\propto \prod_{i=1}^n \phi(\xi_{2i;\mu,\sigma})\Phi(\lambda\xi_{2i;\mu,\sigma})\xi_{1i;\sigma} \alpha^{r_\alpha-1} \exp\{-\alpha\lambda_\alpha\}, \\ \pi(\boldsymbol{\beta}|\alpha, \gamma, \mathbf{y}) &\propto \prod_{i=1}^n \phi(\xi_{2i;\mu,\sigma})\Phi(\lambda\xi_{2i;\mu,\sigma})\xi_{1i;\sigma} \exp\left\{-\frac{1}{2}Q(\boldsymbol{\beta})\right\}, \\ \phi(\gamma|\alpha, \boldsymbol{\beta}, \mathbf{y}) &\propto \prod_{i=1}^n \phi(\xi_{2i;\mu,\sigma})\Phi(\lambda\xi_{2i;\mu,\sigma})\xi_{1i;\sigma},\end{aligned}$$

and

$$\begin{aligned}\pi(h|\alpha, \beta, \gamma, \mathbf{y}_c) &\propto \prod_{i=1}^n \phi(\xi_{2i}^*)\xi_{1i}^* \exp\left\{-\frac{1}{2}h_i^2\right\}, \\ \pi(\alpha|\beta, \gamma, \mathbf{y}_c) &\propto \prod_{i=1}^n \phi(\xi_{2i}^*)\xi_{1i}^* \exp\left\{-\frac{1}{2}(h_i^2 + 2\alpha\lambda_\alpha)\right\} \alpha^{r_\alpha-1}, \\ \pi(\beta|\alpha, \gamma, \mathbf{y}_c) &\propto \prod_{i=1}^n \phi(\xi_{2i}^*)\xi_{1i}^* \exp\left\{-\frac{1}{2}(h_i^2 + Q(\boldsymbol{\beta}))\right\}, \\ \pi(\gamma|\alpha, \beta, \mathbf{y}_c) &\propto \prod_{i=1}^n \phi(\xi_{2i}^*)\xi_{1i}^* \exp\left\{-\frac{1}{2}h_i^2\right\}.\end{aligned}$$

The MCMC algorithms iteratively simulates from the full conditionals above, depending on the adopted likelihood.

3.1 Residual analysis

The residual analysis is an important tool for model fit assessment. It is possible, through the residual analysis, to check for 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 $Y_i|\boldsymbol{\theta} \sim \text{SNHSCP}(\alpha, \mathbf{x}_i^\top \boldsymbol{\beta}, \sigma = 2, \gamma)$ be a random variable with a conditional cumulative distribution function (cdf) given by $F_{Y_i|\boldsymbol{\theta}}(y_i) = \Phi_\gamma(\xi_{2i})$, where $\Phi_\gamma(\cdot)$ is the cdf of the SNCP(0, 1, γ), see Chaves et al. (2016a). Therefore we can define the quantile residual as

$$R_{q,i} = \Phi^{-1}\left\{\Phi_\gamma\left[\xi_2\left(y_i; \hat{\alpha}, \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}\right)\right]\right\}, \quad (8)$$

where $(\hat{\cdot})$ is the respective Bayesian estimator (i.e., posterior expectation, posterior median or posterior mode). Therefore, with $\hat{\alpha}$, $\hat{\gamma}$ and $\hat{\boldsymbol{\beta}}$ being consistent estimators (in the frequentist sense) of α , γ and $\boldsymbol{\beta}$, respectively, we have that $R_{q,i}$ converges

in distribution to the standard normal distribution. The methodology used for the construction of simulated envelopes follows the usual one, that is, since the observed residuals are expected to follow a standard normal distribution, under the well fit of the model, the envelopes are simulated from a standard normal distributions, as described in Atkinson (1985) (see also Vilca et al. (2016)).

3.2 Statistics for Model comparison

When MCMC algorithms are used to obtain the posterior distributions, some statistics for model comparison can be easily calculated, see Spiegelhalter et al. (2002). To introduce these statistics, we first define $D(\boldsymbol{\theta}) = -2\ell(\boldsymbol{\theta}|\mathbf{y})$, where $\ell(\boldsymbol{\theta}|\mathbf{y})$ is the logarithm of the original likelihood, that is

$$\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \left\{ \log [\phi(\xi_{2i;\mu,\sigma})] + \log \Phi(\lambda\xi_{2i;\mu,\sigma}) + \log(\xi_{1i;\sigma}) \right\}.$$

Also, let $\boldsymbol{\theta}^{(m)}$, $m = 1, \dots, M$, be the m -th value of the valid simulated MCMC sample, that is, the MCMC sample obtained after discarding the burn-in and a proper spacing (lag) between the values. Finally, let $\bar{\boldsymbol{\theta}}$ be the vector with the posterior expectation, of each parameter, based on the valid MCMC sample, and $\bar{D}(\bar{\boldsymbol{\theta}}) = \frac{1}{M} \sum_{m=1}^M D(\boldsymbol{\theta}^{(m)})$. Denote also the deviance by $D(\bar{\boldsymbol{\theta}}) = -2\ell(\bar{\boldsymbol{\theta}}|\mathbf{y})$, and the deviance information criterion (DIC) by $\text{DIC} = D(\bar{\boldsymbol{\theta}}) + 2p_D$, where $p_D = \bar{D}(\bar{\boldsymbol{\theta}}) - D(\bar{\boldsymbol{\theta}})$.

The EAIC (posterior expectation of AIC) and EBIC (posterior expectation of BIC) are given, respectively, by $\text{EAIC} = D(\bar{\boldsymbol{\theta}}) + 2p_{\boldsymbol{\theta}}$ and $\text{EBIC} = D(\bar{\boldsymbol{\theta}}) + p_{\boldsymbol{\theta}} \log(n)$, where $p_{\boldsymbol{\theta}}$ is the total number of parameters of the model. Finally, let $L(\boldsymbol{\theta}|y_i)$, $i = 1, \dots, n$, be the original likelihood (see Equation (3)) related to the i th observation. Then, the LPML (logarithm of the pseudo-marginal likelihood) is calculated as $\text{LPML} = \sum_{i=1}^n \ln(\widehat{\text{CPO}}_i)$ and

$$\widehat{\text{CPO}}_i = \left\{ \frac{1}{M} \sum_{m=1}^M \frac{1}{L(\boldsymbol{\theta}^{(m)}|y_i)} \right\}^{-1}, \quad (9)$$

represents the conditional predictive ordinate, see Ibrahim et al. (2001) and Gelfand et al. (1992). The smaller the values of DIC, EAIC, EBIC and deviance, the better the model fit, but the opposite occurs with the LPML.

The EAIC and EBIC statistics tend to select the model with the smallest number of parameters ($p_{\boldsymbol{\theta}}$) since it gives more penalties to models with more parameters. On the other hand, the DIC tends to select the most complex (or the most general) model, that is, it tends to select the overfitted model, see Ando (2007). Finally, the LPML and the Deviance statistics tend to select the model that presents the largest likelihood. This corresponds to the most general model when the competing models are nested.

3.3 Bayesian case influence diagnostics

Since regression models are sensitive to the underlying model assumptions, it is important to perform sensitivity analysis. Here, we consider the measure of divergence within the Bayesian context as in Cho et al. (2009), who developed case deletion influence diagnostics for both joint and marginal posterior distributions based on the Kullback-Leibler (K-L) divergence, and presented a simple way of calculating such influence measure by using MCMC outputs. Let $K(P, P_{(-i)})$ be the K-L divergence between P and $P_{(-i)}$, where P stands for the posterior distribution of $\boldsymbol{\theta}$ for the full data and $P_{(-i)}$ stands for the posterior distribution of $\boldsymbol{\beta}$ without the i th observation. Then, we have

$$K(P, P_{(-i)}) = \int \pi(\boldsymbol{\theta}|\mathbf{y}) \ln \left\{ \frac{\pi(\boldsymbol{\theta}|\mathbf{y})}{\pi(\boldsymbol{\theta}|\mathbf{y}_{(-i)})} \right\} d\boldsymbol{\theta},$$

where $\mathbf{y}_{(-i)}$ corresponds to \mathbf{y} without the i th observation. Also, using the notation introduced earlier in Section 3.2, the MCMC estimate of $K(P, P_{(-i)})$ is $\widehat{K}(P, P_{(-i)}) = -\ln(\widehat{CPO}_i) + \frac{1}{M} \times \sum_{m=1}^M \ln[L(\boldsymbol{\theta}^{(m)}|\mathbf{y}_i)]$, where \widehat{CPO}_i is as in (9).

4 Simulation studies

In this section, we present four simulation studies: parameter recovery of the two MCMC algorithms with a comparison with the ML estimates (PRC), the behavior of the proposed residuals (R), the behavior of the K-L divergence measure (D) and the performance of the statistics of model comparison (SMC). From the results related to a convergence study (not presented for the sake of simplicity) we observed that to set a burn-in of 4,000, with a spacing of 100, generating a total of 104,000 values was enough to have valid MCMC samples of 1,000 values for each parameter.

Also, 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 (0.5, 1.5), that is, low and moderate variability, and value of the parameter γ (-0.67, -0.45, 0, 0.45, 0.67), 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 two others, only one replica and only one scenario were used. More specific details concerning each study are presented in the following subsections.

The general model used was

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i | \boldsymbol{\theta} \stackrel{iid}{\sim} \text{SNHSCP}(\alpha, 0, \sigma = 2, \gamma), i = 1, \dots, n, \quad (10)$$

where $\beta = (1, 2)^\top$. The prior distributions considered in all simulation studies were: $\beta_k \stackrel{iid}{\sim} \mathcal{N}(0, 1000)$, $k = 0, 1$, $\alpha \sim \text{gamma}(1, 0.8)$ and $\gamma \sim U(-a, a)$, where $a = 0.99527$. The first prior is quite flat, the second is reasonably concentrated in the interval $(0, 3)$ (90% of the mass), and was based on the literature which indicates that, in general, the estimates usually lie in this interval and the last prior is non-informative.

4.1 Parameter recovery and comparison of the estimation methods

We will refer to the maximum likelihood estimates by ML and the Bayesian estimates as augmented, when the “augmented” likelihood was used and “original”, when the original likelihood was considered. For the sake of simplicity, we present only the results for three values of γ ($-0.67, 0, 0.45$) and one value of α (0.5), since for the other scenarios the patterns were similar. We calculated the usual statistics to measure the accuracy of the estimates: bias, variance (Var), root mean squared error (RMSE) and absolute value of the relative bias (AVRB). Let θ be the parameter of interest and let $\hat{\theta}_r$ be some estimate related to the replica r , and $\bar{\hat{\theta}} = (1/R) \sum_{r=1}^R \hat{\theta}_r$. The adopted statistics are defined as: $\text{BIAS} = \bar{\hat{\theta}} - \theta$, $\text{Var} = (1/R) \sum_{r=1}^R (\hat{\theta}_r - \bar{\hat{\theta}})^2$, $\text{RMSE} = \sqrt{(1/R) \sum_{r=1}^R (\theta - \hat{\theta}_r)^2}$, $\text{AVRB} = |\bar{\hat{\theta}} - \theta|/|\theta|$.

Tables 1, 2 and 3 present some results. We can notice that the original and augmented approach performed quite similarly. We can also notice that under the smallest sample size ($n = 10$) the ML estimates for α and γ are more accurate than the Bayesian ones. On the other hand, for this same sample size, the augmented approach provides the best estimates for β_0 and β_1 . For the other two sample sizes, the Bayesian ones and ML estimates are equivalent, in the sense that, for some scenarios and parameters, the former is better, while for others, the latter is better. Additionally, the Bayesian estimates presented smaller variance and RMSE whereas the ML presented smaller bias and AVRB. In conclusion, we can notice certain equivalence between the Bayesian and ML approach. However, the Bayesian approach is easier to implement and for developing tools for case influence diagnostics and it allows to incorporating prior information. Also, depending on the optimization method adopted for the ML approach, it is possible to obtain undesired estimates, whereas the Bayesian estimates always lie in the parameter space. Since the results of the two Bayesian set of estimates were quite similar, we decide to use the original approach, once it is easier to implement and faster.

4.2 Behavior of the residuals

For this study, we considered the scenario where $\alpha = 0.5$, $\gamma = 0.67$ and $n = 200$. Here we simulated only one set of observations for four different models: SNCP log-BS, log-BS, Student-t log-BS and Skew t log-BS. The first one is the model given by (10) while

Table 1: Results of simulation study (PRC) - $\gamma = -0.67$.

Parameter	n	Estimation method	Mean	Variance	Bias	RMSE	AVRB
α	10	Augmented	0.601	0.003	0.101	0.115	0.203
		Original	0.607	0.003	0.107	0.118	0.214
		ML	0.447	0.002	-0.053	0.068	0.105
	50	Augmented	0.515	<0.001	0.015	0.019	0.029
		Original	0.513	<0.001	0.013	0.020	0.026
		ML	0.504	<0.001	0.004	0.021	0.008
	200	Augmented	0.503	<0.001	0.003	0.006	0.006
		Original	0.503	<0.001	0.003	0.005	0.005
		ML	0.515	<0.001	0.015	0.025	0.030
β_0	10	Augmented	1.011	0.129	0.011	0.359	0.011
		Original	1.026	0.071	0.026	0.267	0.026
		ML	0.966	0.080	-0.034	0.285	0.034
	50	Augmented	0.984	0.011	-0.016	0.105	0.016
		Original	1.001	0.012	0.001	0.108	0.001
		ML	0.998	0.013	-0.002	0.114	0.002
	200	Augmented	0.987	0.002	-0.013	0.052	0.013
		Original	0.993	0.003	-0.007	0.052	0.007
		ML	0.996	0.003	-0.004	0.054	0.004
β_1	10	Augmented	1.976	0.542	-0.023	0.737	0.012
		Original	1.902	0.302	-0.098	0.558	0.049
		ML	2.038	0.328	0.038	0.574	0.019
	50	Augmented	2.023	0.040	0.022	0.202	0.011
		Original	1.982	0.046	-0.018	0.215	0.009
		ML	2.013	0.053	0.013	0.230	0.006
	200	Augmented	2.021	0.011	0.022	0.100	0.011
		Original	2.009	0.011	0.009	0.104	0.004
		ML	2.019	0.011	0.019	0.104	0.009
γ	10	Augmented	-0.094	0.044	0.576	0.613	0.859
		Original	-0.147	0.051	0.523	0.571	0.781
		ML	-0.575	0.377	0.095	0.622	0.142
	50	Augmented	-0.481	0.061	0.189	0.311	0.282
		Original	-0.504	0.061	0.166	0.297	0.248
		ML	-0.718	0.080	-0.048	0.287	0.072
	200	Augmented	-0.632	0.015	0.038	0.127	0.057
		Original	-0.607	0.015	0.063	0.137	0.094
		ML	-0.778	0.017	-0.108	0.169	0.161

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

Parameter	n	Estimation Method	Mean	Variance	Bias	RMSE	AVRB
α	10	Augmented	0.621	0.002	0.121	0.129	0.241
		Original	0.612	0.002	0.112	0.123	0.225
		ML	0.445	0.001	-0.055	0.064	0.110
	50	Augmented	0.519	<0.001	0.019	0.024	0.039
		Original	0.521	<0.001	0.021	0.023	0.041
		ML	0.488	<0.001	-0.012	0.016	0.024
	200	Augmented	0.506	<0.001	0.006	0.007	0.012
		Original	0.506	<0.001	0.006	0.007	0.013
		ML	0.498	<0.001	-0.002	0.004	0.005
β_0	10	Augmented	1.017	0.070	0.017	0.265	0.017
		Original	0.966	0.088	-0.034	0.299	0.034
		ML	1.017	0.082	0.017	0.286	0.017
	50	Augmented	1.017	0.022	0.017	0.149	0.017
		Original	1.000	0.016	> -0.001	0.127	<0.001
		ML	0.998	0.018	-0.002	0.136	0.002
	200	Augmented	1.000	0.004	<0.001	0.062	<0.001
		Original	0.992	0.003	-0.008	0.052	0.008
		ML	1.001	0.004	0.001	0.061	0.001
β_1	10	Augmented	1.966	0.295	-0.034	0.544	0.017
		Original	2.057	0.386	0.057	0.624	0.029
		ML	1.983	0.315	-0.017	0.562	0.008
	50	Augmented	1.972	0.082	-0.028	0.288	0.014
		Original	2.003	0.067	0.003	0.259	0.001
		ML	2.002	0.072	0.002	0.268	0.001
	200	Augmented	1.999	0.015	> -0.001	0.123	<0.001
		Original	2.016	0.011	0.016	0.106	0.008
		ML	1.996	0.015	-0.004	0.122	0.002
γ	10	Augmented	-0.220	0.050	-0.220	0.314	-
		Original	-0.007	0.038	-0.007	0.195	-
		ML	-0.016	0.011	-0.016	0.108	-
	50	Augmented	0.042	0.086	0.042	0.297	-
		Original	-0.002	0.098	-0.002	0.314	-
		ML	-0.039	0.026	-0.039	0.166	-
	200	Augmented	-0.012	0.034	-0.012	0.185	-
		Original	0.044	0.032	0.044	0.183	-
		ML	0.006	0.012	0.006	0.109	-

Table 3: Results of simulation study (PRC) - $\gamma = 0.45$.

Parameter	n	Estimation Method	Mean	Variance	Bias	RMSE	AVRB
α	10	Augmented	0.611	0.003	0.111	0.123	0.222
		Original	0.619	0.002	0.119	0.129	0.238
		ML	0.439	0.003	-0.061	0.081	0.123
	50	Augmented	0.518	< 0.001	0.018	0.021	0.037
		Original	0.518	< 0.001	0.018	0.021	0.036
		ML	0.506	0.001	0.006	0.028	0.012
	200	Augmented	0.505	<0.001	0.005	0.006	0.010
		Original	0.504	< 0.001	0.004	0.005	0.009
		ML	0.513	<0.001	0.013	0.025	0.026
β_0	10	Augmented	1.001	0.106	0.001	0.325	0.001
		Original	1.040	0.087	0.040	0.298	0.040
		ML	1.051	0.128	0.051	0.361	0.051
	50	Augmented	1.011	0.013	0.011	0.112	0.011
		Original	1.008	0.015	0.008	0.121	0.008
		ML	0.992	0.016	-0.008	0.128	0.008
	200	Augmented	1.003	0.003	0.003	0.055	0.003
		Original	0.999	0.004	-0.001	0.060	0.001
		ML	1.001	0.003	0.001	0.056	0.001
β_1	10	Augmented	2.003	0.414	0.003	0.643	0.002
		Original	1.912	0.359	-0.088	0.606	0.044
		ML	1.907	0.426	-0.093	0.659	0.047
	50	Augmented	1.985	0.051	-0.015	0.225	0.008
		Original	1.992	0.057	-0.008	0.239	0.004
		ML	2.001	0.067	0.001	0.258	0.001
	200	Augmented	1.996	0.012	-0.004	0.108	0.002
		Original	2.004	0.014	0.004	0.120	0.002
		ML	1.992	0.012	-0.008	0.112	0.004
γ	10	Augmented	0.050	0.045	-0.399	0.452	0.889
		Original	0.042	0.040	-0.408	0.455	0.907
		ML	0.432	0.461	-0.018	0.679	0.040
	50	Augmented	0.324	0.085	-0.126	0.317	0.279
		Original	0.322	0.076	-0.128	0.304	0.285
		ML	0.525	0.156	0.075	0.401	0.166
	200	Augmented	0.431	0.022	-0.019	0.148	0.043
		Original	0.414	0.026	-0.036	0.164	0.081
		ML	0.614	0.033	0.164	0.245	0.364

the second corresponds to its particular case when $\gamma = 0$. The third and the fourth models correspond to the model (10) using in (2) instead of a SNCP distribution, an Student-t and a skew-t distribution (according to Azzalini and Capitanio (2003)), with $\nu = 4$ degrees of freedom and asymmetry parameter $\gamma = 0.67$, respectively. For each simulated data set, we fit a SNCP log-BS model and calculate the residuals presented in (8). Four plots were built for each situation, including an envelope for the residuals, and they are presented in Figures (1), (2), (3) and (4). To simulate from the skew-t distribution we used the function *rst* from the R package *sn*.

We can notice that, when the SNCP log-BS model (or its particular case, the log-BS model) is the underlying one, the residuals present a symmetric behavior, resembling a standard normal distribution, with all of them within the envelope within the interval $(-2, 2)$, with no systematic behavior. On the other hand, when the underlying model is the Student t log-BS, we observe several residuals with large absolute values (close to 3), i.e., possibly outliers, with many of them falling outside the envelope, which, in its turn, presents a behavior compatible with a heavy tails distribution. Finally, when the underlying model is the Skew t log-BS, we observe some outliers, a skewed behavior of the residuals, with many observations lying outside the envelope, which, in its turn, presents a behavior compatible with a skewed heavy tails distribution. 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 SNCP (the generating distribution).

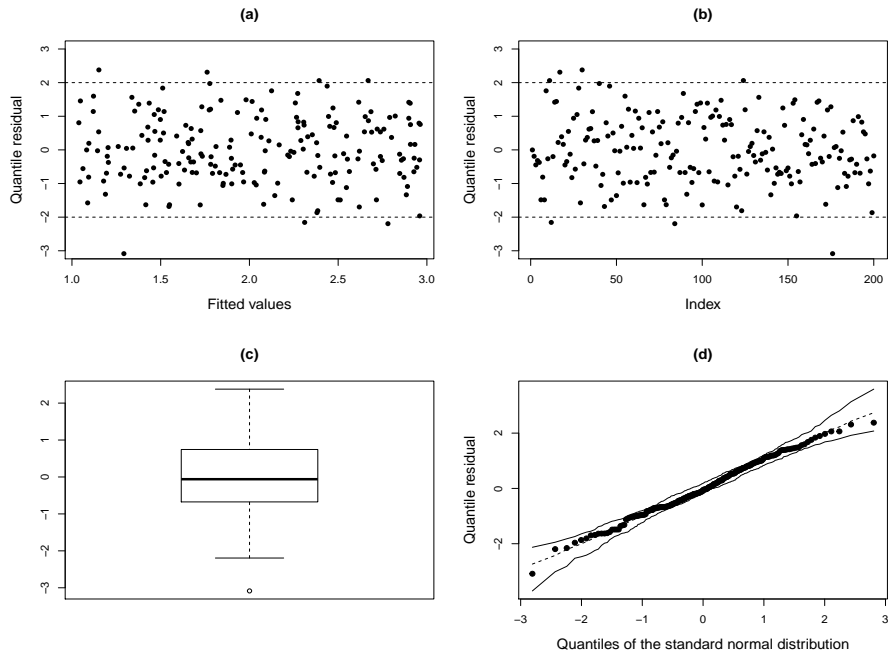


Figure 1: Residual plots for the observations generated from a SNCP log-BS model.

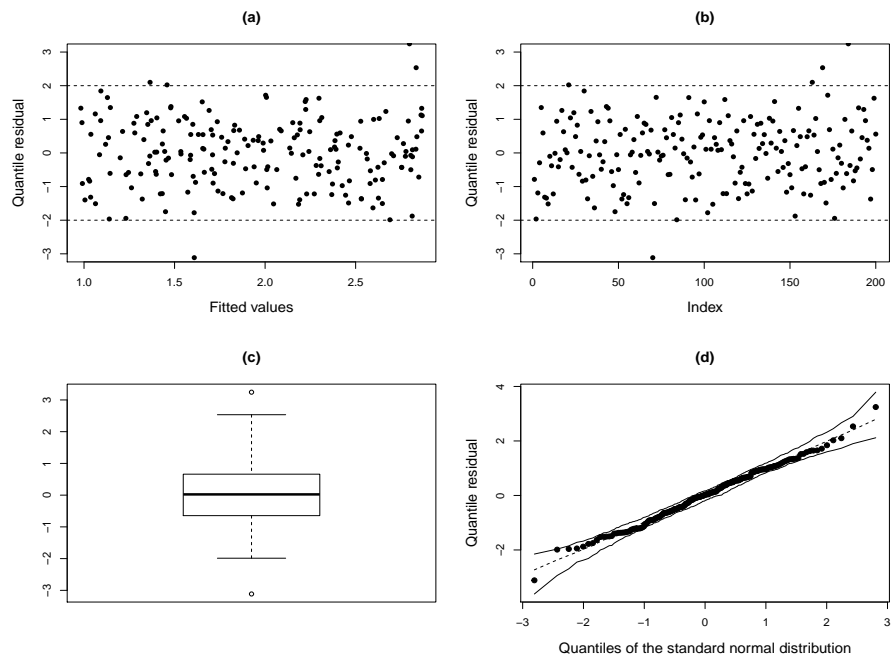


Figure 2: Residual plots for the observations generated from a log-BS model.

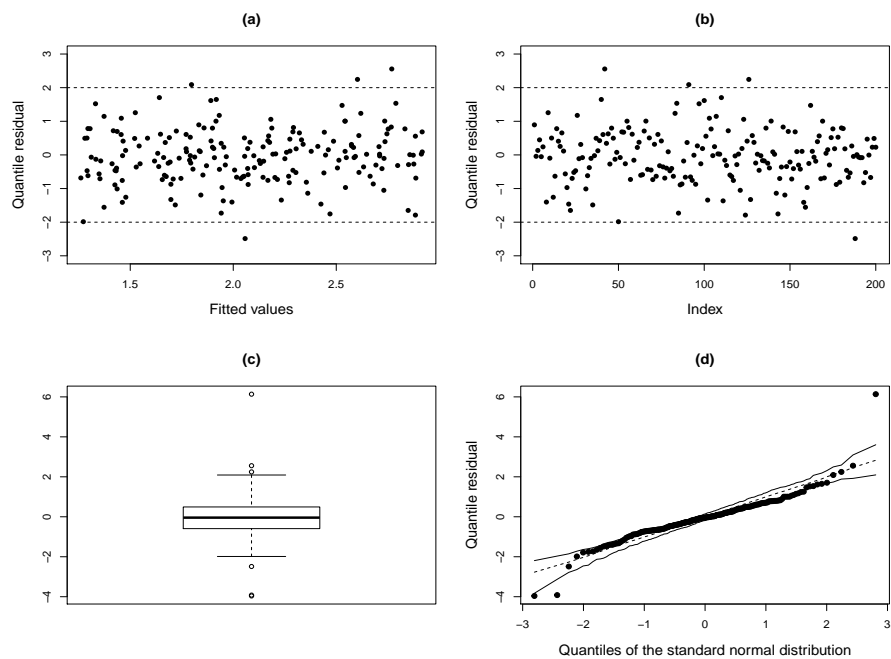


Figure 3: Residual plots for the observations generated from a Student-t log-BS model.

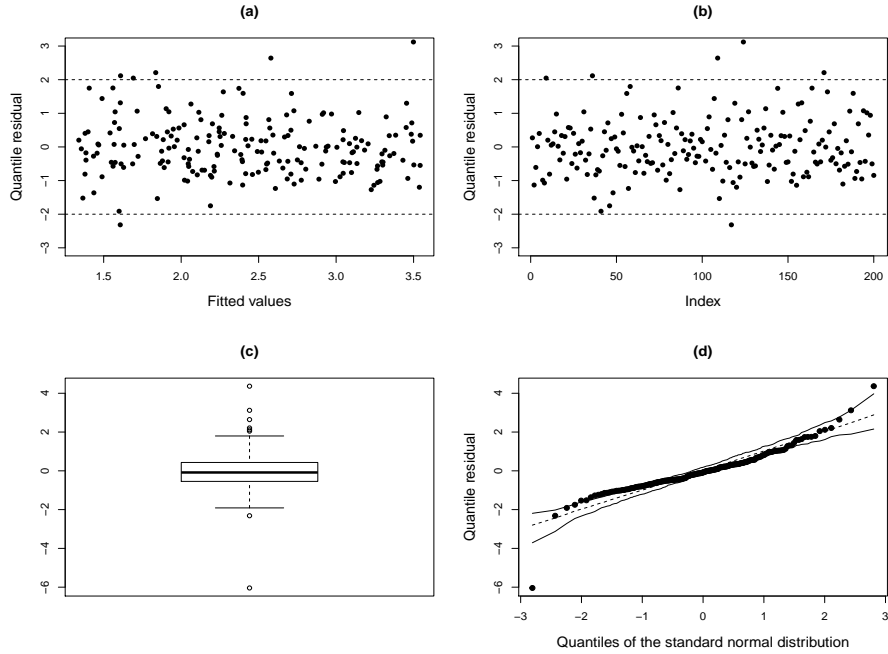


Figure 4: Residual plots for the observations generated from a Skew t log-BS model.

4.3 Behavior of the K-L divergence

The scenario and the framework (four underlying models) considered here are exactly those presented in Subsection 4.2. That is, we fitted our SNCP log-BS model to the each one of four data sets, generating observations according to the four models presented in the previous Subsection. It can be notice that a larger number of large values for the K-L divergence are observed when the data come from the Student-t log-BS or the Skew-t log-BS models, compared with the data generated from the other two models, as expected. The values that stand out when the data is generated from a SNCP log-BS or the log-BS model (only one for each model), can be considered as sampling fluctuations, which is also expected. Therefore, we can conclude that this measure is suitable to detect potential influential observations.

4.4 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 SNCP log-BS regression model with $\alpha = 0.5$, $\beta = (1, 2)^\top$, $\gamma = 0.67$, considering two samples sizes ($n = 50$, $n = 200$) and we fit two competing models, the SNCP log-BS and log-BS. The last two scenarios are equivalent to the two first, but the replicas were simulated from the log-BS model.

In the first two scenarios, for $n = 50$, the criteria AIC, BIC, EAIC, EBIC, DIC

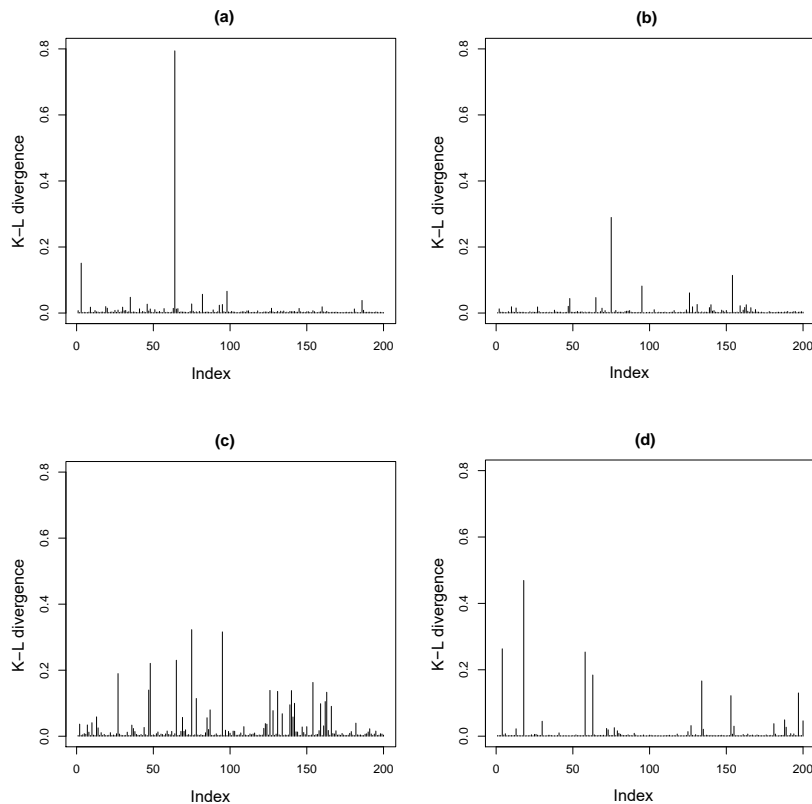


Figure 5: K-L divergence for the observations generated from the models: (a) SNCP log-BS, (b) log-BS, (c) Student-t log-BS, (d) Skew t log-BS.

and LMPL chose the SNPC log-BS model (the underlying model) in 97%, 95%, 99%, 97%, 100% and 99% of the replicas, respectively, whereas, for $n = 200$, all statistics chosen the true model in 100% of the scenarios. On the other hand, for the two last scenarios, under $n = 50$, the criteria AIC, BIC, EAIC, EBIC, DIC e LMPL chose the log-BS model (the underlying model) in 97%, 99%, 95%, 96%, 100% and 91% of the scenarios, respectively, whereas, under $n = 200$, these percentages were 100%, 100%, 97%, 99%, 100% and 95%, 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.

5 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 matches where (i) there was at least one goal scored by the

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

True underlying model: SNPC log-BS							
Model	n	AIC	BIC	EAIC	EBIC	DIC	LPML
SNPC log-BS	50	69.481	77.129	80.825	88.473	215.656	-38.630
	200	270.105	283.299	297.969	311.162	867.065	-146.664
log-BS	50	73.389	79.125	85.424	91.160	235.421	-41.268
	200	282.964	292.859	329.219	339.114	967.708	-162.727
True underlying model: log-BS							
Model	n	AIC	BIC	EAIC	EBIC	DIC	LPML
SNPC log-BS	50	74.930	82.578	84.529	92.177	225.907	-40.521
	200	287.320	300.514	303.617	316.810	877.804	-152.385
log-BS	50	73.945	79.681	85.513	91.249	234.210	-41.534
	200	284.139	294.034	303.751	313.646	883.921	-152.953

home team, and (ii) there was at least one goal scored by either team from the penalty spot, lack of a kick, or any other meaning unclear. 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 to be scored by the home team based on the time in minutes the first goal is scored by either team. From Figure 6, it can be seen that a linear model can be suitable to link the natural logarithm of these two variables.

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

The two proposed models are (the SNCP log-BS model):

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, \dots, 37,$$

where $Y_i = \log(T_{2i})$, $x_i = \log(T_{1i})$, $T_{ji}, j = 1, 2$, is the value of the variable j for the team i , $\varepsilon_i | \boldsymbol{\theta} \stackrel{i.i.d.}{\sim} \text{SNHSCP}(\alpha, 0, \sigma = 2, \gamma)$ and the log-BS model (i.e., considering $\gamma = 0$). We assume the same prior distributions considered in the Section 4 and the Bayesian approach based on the original likelihood. Figures 7 and 8 present the residual analysis for the two models. It can be seen that the SNPC log-BS model fitted to the data properly and much better than the log-BS model. Figure 9 indicates that one observation highlights more under the SNCP log-BS than under the log-BS (and other three present, approximately, the same values for the KL divergence measure). Therefore, we can conclude that the log-BS presents a slight advantage, under this criterion. Table 5 presents the statistics of model comparison. We can see that all selected the SNCP log-BS as the most appropriate. In conclusion, we can see that the SNCP log-BS model is more appropriate than the log-BS model and it fitted to the data quite well.

Table 6 presents the Bayesian estimates, that is, the posterior expectations (PE), posterior standard deviations (PSD) and the 95% equi-tailed credibility intervals for the two models. We can see that the asymmetry parameter was statistically significant and that the estimates are similar between the two models. Also the larger the time is for the first goal to be scored by either team, the higher the time is for a goal of any sort to be scored.

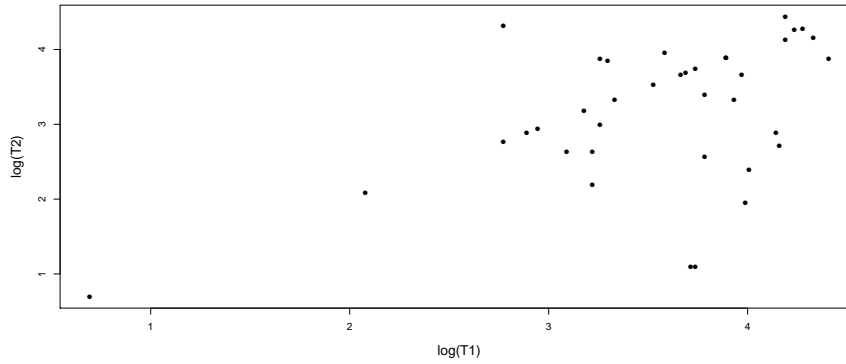


Figure 6: Scatter plot between the natural logarithm of the T1 and T2

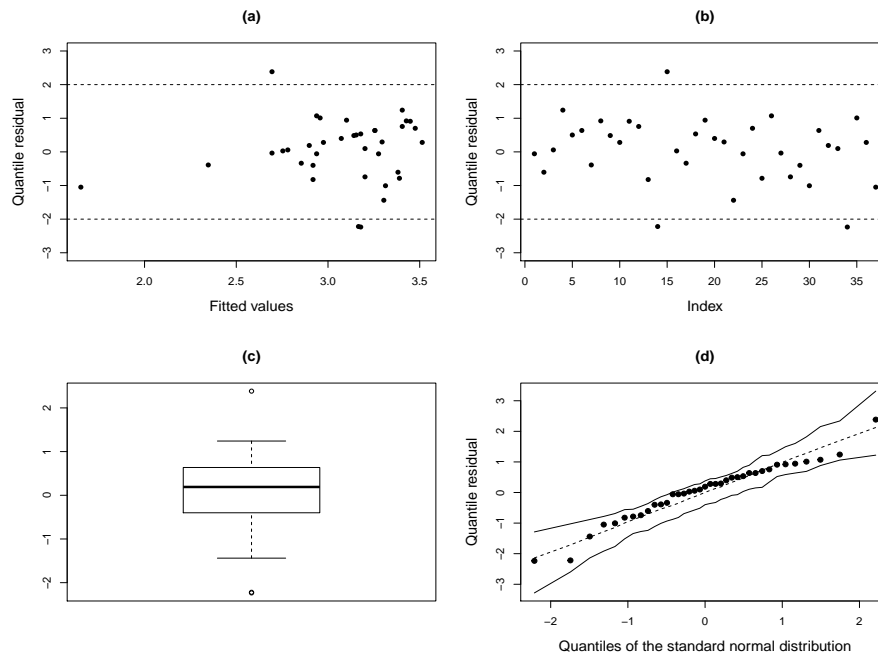


Figure 7: Residual analysis for the SNCP log-BS model.

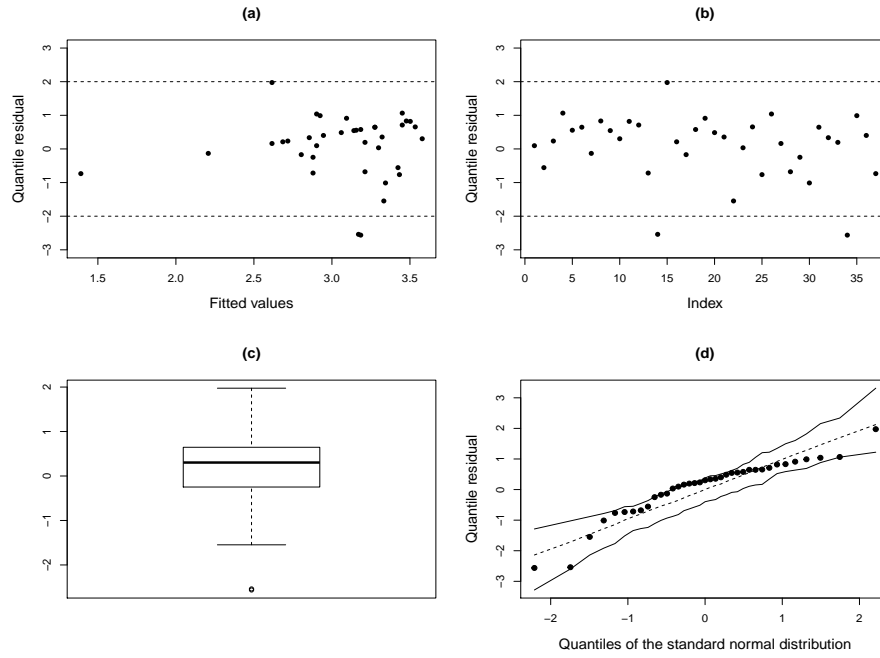


Figure 8: Residual analysis for the log-BS model.

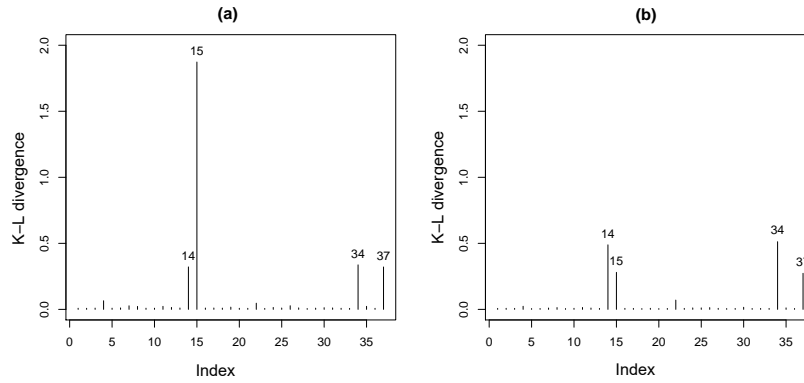


Figure 9: K-L divergence measure for the models: (a) SNCP log-BS and (b) log-BS.

Table 5: Statistics of model comparison

	Statistics					
	AIC	BIC	EAIC	EBIC	DIC	LPML
SNPC log-BS	92.865	99.308	97.053	103.496	263.933	-47.880
log-BS	97.468	102.300	100.692	105.524	281.259	-49.268

Table 6: Bayesian estimates for the SNCP log-BS and log-BS models.

Parameter	SNCP log-BS			log-BS		
	PE	PSD	CI _{95%}	PE	PSD	CI _{95%}
α	.943	.124	[.739; 1.239]	.969	.121	[.766; 1.244]
β_0	1.305	.789	[-.270; 2.735]	.982	.778	[-.561; 2.430]
β_1	.501	.218	[.098; .932]	.589	.215	[.168; 1.010]
γ	-.611	.219	[-.936; -.116]	-	-	-

6 Concluding Remarks

In this paper, we developed Bayesian inference for the SNCP log-BS regression model proposed by Chaves et al. (2016b). Parameter estimation, model fit assessment, model comparison and influence diagnostics were developed through MCMC algorithms. The results from the simulation studies indicated that the estimation methods (including the maximum likelihood one) recovered all parameters properly and that the Bayesian paradigm is equivalent, in terms of the accuracy of the estimates, to the ML method. However, the computational implementation of the MCMC algorithms and of the influence diagnostics are easier to be implemented than the EM algorithm based approach developed by Chaves et al. (2016b). Besides, the tools for case influence diagnostics are easier implemented through the Bayesian approach. Also, we can incorporate prior information in the Bayesian analysis. Therefore, we can conclude that our approach is as good as that developed by Chaves et al. (2016b). Also, the Bayesian tools for model comparison and model fit assessment indicated that the SNCP log-BS regression model fitted to the data well and better than the log-BS model.

As future developments, we suggest the use of Jeffreys-rule prior and independent Jeffreys prior. Other auxiliary algorithms such as the Hamiltonian Monte Carlo (see Homan and Gelman (2014) and Carpenter et al. (2016)), adaptive reject sampling and slice sampling (see Gamerman and Lopes (2006)) could be used and compared among them and with the Metropolis-Hastings algorithm. Finally, other numerical methods to obtain approximation for the marginal posterior distributions, such as INLA algorithm, can be considered, see Rue and Martino (2009).

Acknowledgements

The authors would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the financial support through a Masters Scholarship granted to the first author under the guidance of the second.

References

- Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika*, 94, 2, 443-458.
- Atkinson, A. C. (1985). *Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis*. Oxford: Clarendon Press.
- Azevedo, C.L.N., Bolfarine H. and Andrade, D.F. (2011). Bayesian inference for a skew-normal IRT model under the centred parameterization. *Comput. Stat. Data Anal.* 55, 1, 353-365.
- Azevedo, C.L.N., Bolfarine H. and Andrade, D.F. (2012). Parameter recovery for a skew-normal IRT model under a Bayesian approach: hierarchical framework, prior and kernel sensitivity and sample size. *Journal of Statistical Computation and Simulation*. 55, 353-365.
- Azzalini, A. (1985). A class of distribution which includes the normal ones. *Scandinavian journal of statistics*, 12, 2, 171-178.
- Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *J.Roy. Statist. Soc. B*, 65, 367-389.
- Balakrishnan, N., Leiva, V., Sanhueza, A. and Vilca-Labra, F. (2009). Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm. *SORT*, 33, 2, 171-192.
- Barros, M., Paula, G.A. and Leiva, V. (2008). A New Class of Survival Regression Models with Heavy-Tailed Errors: Robustness and Diagnostics. *Lifetime Data Analysis*, 14, 3, 1-17.
- Birnbaum, Z. W. and Saunders, S. C. (1969). A new family of life distributions. *Journal of Applied Probability*, 6, 2, 637-652.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P. and Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software (in press)*.
- Chaves, N. L., Azevedo, C. L. N., Vilca-Labra, F. and Nobre, J. S. (2016 a). A new class of Birnbaum-Saunders distribution: frequentist and Bayesian inference. *manuscript under preparation*.
- Chaves, N. L.; Azevedo, C. L. N.; Vilca-Labra, F. Nobre, J. S. N. (2016 b). A new log-Birnbaum-Saunders regression model based on the skew normal distribution under the centered parameterization. *manuscript under preparation*.

- Cho, H., Ibrahim, J. G., Sinha, D. and Zhu, H. (2009). Bayesian case influence diagnostics for survival models. *Biometrics*, 65, 1, 116-124.
- Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5, 3, 236-244.
- Gamerman, D. and Lopes, H. (2006). *Stochastic simulation for bayesian inference, second edition*. Chapman & Hall/CRC, New York-NY.
- Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In: Bernardo J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds) Bayesian statistics. 4. Oxford University Press, Oxford pp 147-169.
- Homan, M. D. and Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, 15, 1, 1593-1623 .
- Ibrahim, J. G., Chen, M-H. and Sinha, D. (2001). *Bayesian Survival Anal.* Spring, New York.
- Leiva, V., Barros, M., Paula, G. A. and Galea, M. (2009). Generalized Birnbaum-Saunders Models using R. *Brazilian Statistical Association*. http://staff.deuv.cl/leiva/archivos/books/leiva_barros_paula_2009.pdf
- Meintanis, S. G. (2007). Test of fit for marshallolkin distributions with applications. *Journal of Statistical Planning and inference*, 137, 12, 3954-3963.
- Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of applied statistics*, 27, 7, 859-870.
- 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>
- Rieck, J. R. (1989). *Statistical analysis for the Birnbaum-Saunders fatigue life distribution*, Ph.D dissertation, Clemson University, South Carolina, USA.
- Rieck, J. R. and Nedelman, J. R. (1991). A log-linear model for the Birnbaum-Saunders distribution. *Technometrics*, 33, 1, 51-60.
- Rue, H. and Martino, S. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B*, 71, 2, 319-392.

- Santana, L., Vilca-Labra, F. and Leiva, V. (2011). Influence analysis in skew-Birnbaum-Saunders regression models and applications. *Journal of Applied Statistics*, 38, 8, 1633-1649.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Der Linde, A. V. (2002). Bayesian measures of model complexity and fit. *J R Stat Soc Ser B* , 64, 4, 583-639.
- Spiegelhalter, D. J., Thomas, A., Best, N. and Lunn, D. (2014). *OpenBUGS User Manual*. <http://www.openbugs.net/Manuals/Manual.html>.
- Sturtz, S., Ligges, U., and Gelman, A. E. (2005) R2winbugs: a package for running winbugs from R. *Journal of Statistical software* 12, 3, 1-16.
- Vilca, F. and 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. and Leiva, V. (2011). Estimation of extreme percentiles in Birnbaum-Saunders distributions. *Computational Statistics & Data Analysis*, 55, 5, 1665-1678.
- Vilca, F., Zeller, C. B. and 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.
- Vilca, F., Azevedo, C. L. N. and Balakrishnan, N. (2016). Bayesian Inference for Sinh-normal/Independent Nonlinear Regression Models. *Journal of Applied Statistics*, 1-23.