Chilean Journal of Statistics
Vol. 7, No. 1, April 2016, 17–30

Regression
Research Paper

Comparative inference and diagnostic in a reparametrized Birnbaum-Saunders regression model

Rodney Vasconcelos Fonseca[1,*], Juvêncio Santos Nobre[2] and Rafael Bráz Azevedo Farias[2]

[1]Departamento de Estatística, Universidade Federal de Pernambuco, Recife, Brazil
[2]Departamento de Estatística e Matemática Aplicada, Universidade Federal do Ceará, Fortaleza, Brazil

(Received: 29 October 2015 · Accepted in final form: 11 March 2016)

Abstract

The Birnbaum-Saunders distribution has received great attention in recent years, providing many useful techniques for analysis of positive response variables in several works. A regression model in a similar framework of a generalized linear models was proposed using a new parametrization of the Birnbaum-Saunders distribution. The purpose of this paper is to develop a Bayesian approach for this recent regression model. In addition, Bayesian influence diagnostic procedures will be discussed and compared with classical alternatives using Monte Carlo methods. Both approaches have similar results, but we propose a way to improve the Bayesian method.

Keywords: Birnbaum-Saunders distribution · Bayesian inference · Bayesian influence diagnostic.

Mathematics Subject Classification: Primary 62F15 · Secondary 62J20.

1. Introduction

The Birnbaum-Saunders distribution, $\mathcal{BS}$, was initially proposed by Birnbaum and Saunders (1969a) in order to analyze the lifetime of objects that suffer from a cyclic damage that causes a crack when a certain threshold of damage is reached. Afterwards, Birnbaum and Saunders (1969b) presented inference methods for this initial distribution. A regression model framework was introduced at first by Rieck and Nedelman (1991), who used a relation between the $\mathcal{BS}$ and Sinh-Normal distributions to create a log-Birnbaum-Saunders model. Thereafter, accelerated lifetime models were discussed by Owen and Padgett (2000). After these papers, many works have been done about $\mathcal{BS}$ regression models, as Díaz-García and Leiva (2002) who introduced an extension of early models with a generalized Birnbaum-Saunders distribution, allowing different kernels of elliptical distributions to be used in the $\mathcal{BS}$ model. Influence diagnostic for log-$\mathcal{BS}$ models was discussed, for example, by Galea et al. (2004) and generalization of their results for log-$\mathcal{BS}$ nonlinear models was given by Lemonte and Patriota (2011). Recently, Leiva (2016) presented an excellent

*Corresponding author. Email: rvf1@de.ufpe.br

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

description about inference, modeling and applications using the $\mathcal{BS}$ distribution.

On the other hand, Bayesian procedures for $\mathcal{BS}$ models present only a few papers in literature. We highlight Achcar (1993), who has developed Bayesian estimation procedures for the $\mathcal{BS}$ distribution by using various approximations to the marginal posteriors and compared with classical methods and Tsionas (2001), who developed a Bayesian approach for a $\mathcal{BS}$ regression model proposed by Rieck and Nedelman (1991). Influence diagnostic analysis in Bayesian framework was discussed for a non-linear Birnbaum-Saunders regression model by Farias and Lemonte (2011) and for a $\mathcal{BS}$ distribution with t-Student kernel by Cancho et al. (2010).

A $\mathcal{BS}$ distribution parametrized by its mean was discussed by Santos-Neto et al. (2012), based on generalized linear models (GLM), with the following probability density function

$$f(y;\mu,\delta) = \frac{\exp\{\delta/2\}\sqrt{1+\delta}}{4\sqrt{\pi\mu}} y^{-3/2}\left(y+\frac{\delta\mu}{(1+\delta)}\right)\exp\left\{-\frac{\delta}{4}\left(\frac{(1+\delta)y}{\delta\mu}+\frac{\delta\mu}{(1+\delta)y}\right)\right\},$$

with $\delta > 0$, $\mu > 0$ representing the mean of Y, whose variance is given by $\mathrm{Var}(Y) = \mu^2(2\delta+5)/(\delta+1)^2$. Properties from the original $\mathcal{BS}$ were retained, such as reciprocity and proportionality closure, i.e., $Y^{-1} \sim \mathcal{BS}\left(\frac{(\delta+1)^2}{\delta^2\mu}, \delta\right)$ and $aY \sim \mathcal{BS}(a\mu,\delta)$ whith $a > 0$, respectively.

A regression model based on this new reparametrization was proposed by Leiva et al. (2014), in the similar way done in GLM, where a function links the explanatory variables to the conditional expectation of the response variable, which is explained by a linear predictor

$$g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} = \eta_i, \tag{1}$$

where $\mathbf{x}_i = (x_{i1},\ldots,x_{ip})^\top$ is a vector of explanatory variables, $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)^\top$ is a regression coefficients vector and η_i is the linear predictor. The link function $g : \mathbb{R}^+ \to \mathbb{R}$ is monotone and at least twice differentiable. Estimation methods, residual analysis and influence diagnostic where done in the frequentist approach by Leiva et al. (2014) for this new model. One of the objectives of this paper is to develop and discuss estimation methods and diagnostic in the Bayesian approach.

The paper is organized as follows. In Section 2 we introduced the Bayesian approach for the reparametrized Birnbaum-Saunders regression model. The structure of the model and possible *prior* distributions for the parameters are discussed. In Section 3 we presented comparison model criteria currently used in the literature of Bayesian regression models and Section 4 presented results to perform Bayesian influence diagnostic under such models. In Section 5, we give a brief description of local influence diagnostic under the frequentist approach. A Monte Carlo simulation study is performed in section 6, analyzing prior choices for the Bayesian model and comparing some influence diagnostic methods for the $\mathcal{BS}$ model. An application to real data is made in section 7 under Bayesian and frequentist methodologies, with a comparison of both. Finally, section 8 concludes the paper with a discussion.

2. Section Two

The Bayesian approach for the model presented in equation (1) may be done similarly to Bayesian generalized linear models (see, for example, Dey et al. (2000)), assigning the Birnbaum-Saunders distribution for the response variable $Y_1,\ldots,Y_n$, and suitable *prior* distributions for the coefficients regression vector $\boldsymbol{\beta}$ and the parameter related to disper-

sion, δ. Consider the following hierarchical structure for the Bayesian Birnbaum-Saunders model

$$\begin{aligned} y_i|\boldsymbol{\beta},\delta &\sim \mathcal{BS}(g^{-1}\{\mathbf{x}_i^\top\boldsymbol{\beta}\},\delta), \quad i=1,\cdots,n \\ (\boldsymbol{\beta},\delta)^\top &\sim \pi(\boldsymbol{\beta},\delta). \end{aligned} \tag{2}$$

The likelihood function for the $\mathcal{BS}$ regression model is given by

$$L(\boldsymbol{\beta},\delta|\mathbf{y}) \propto \left\{\prod_{i=1}^{n} y_i^{-3/2}\left(y_i+\frac{\delta\mu_i}{(1+\delta)}\right)\left(\frac{\delta+1}{\mu_i}\right)^{\frac{1}{2}}\right\}\exp\left\{\frac{n\delta}{2}-\frac{(1+\delta)}{4}\sum_{i=1}^{n}\frac{y_i}{\mu_i}-\frac{\delta^2}{4(1+\delta)}\sum_{i=1}^{n}\frac{\mu_i}{y_i}\right\},$$

with $\mu_i = g^{-1}(\mathbf{x}_i^\top\boldsymbol{\beta})$. Nevertheless, since we have no prior information about dependence between $\boldsymbol{\beta}$ and δ, we consider the conditional independence $\pi(\boldsymbol{\beta},\delta)=\pi(\boldsymbol{\beta})\pi(\delta)$ in equation (2). Furthermore, it may be used a distribution with support in $\mathbb{R}^p$ for $\pi(\boldsymbol{\beta})$ and a positive support distribution for δ. As a first choice, we propose a multivariate Normal distribution for the regression coefficients, and since δ assume positive values only, a Gamma distribution is a natural choice for its prior distribution. In this way, the first hierarchical structure for the Bayesian Birnbaum-Saunders regression model, denoted by M_1, is of the form

$$y_i|\boldsymbol{\beta},\delta \sim \mathcal{BS}(g^{-1}\{\mathbf{x}_i^\top\boldsymbol{\beta}\},\delta), \quad i=1,\cdots,n \tag{3}$$

$$\boldsymbol{\beta} \sim \mathcal{N}_p(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0), \tag{4}$$

$$\delta \sim \text{Gamma}(a,b), \tag{5}$$

with $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$, a and b denoting the hyperparameters of the model and for δ, we assume the following Gamma distribution $\pi(\delta|a,b)\propto\delta^{a-1}\exp\{-b\delta\}$, with $a,b>0$.

In the model M_1 presented in equations (3)-(5) it is given just one way in many ways to choose prior distributions for the parameters of interest. Different prior distributions can be and should be tested to modeling the data. One of the objective of doing the test is to check if different prior distributions leads to different inferential results.

The second model considered here, denoted by M_2, differ of M_1 on the covariance matrix of the prior distribution of the coefficients regression $\boldsymbol{\Sigma}_0$ given in equation (4). It is assumed in M_2 that $\boldsymbol{\Sigma}_0$ is of the form $\tau^{-2}\boldsymbol{\Sigma}_0$, where τ^2 is a hyperparameter of precision associated to a Gamma prior distribution, $\tau^2\sim\text{Gamma}(c,d)$, with $c,d>0$. So we can hierarchically have

$$y_i|\boldsymbol{\beta},\delta \sim \mathcal{BS}(g^{-1}\{\mathbf{x}_i^\top\boldsymbol{\beta}\},\delta), \quad i=1,\cdots,n \tag{6}$$

$$\boldsymbol{\beta}|\tau^2 \sim \mathcal{N}_p(\boldsymbol{\mu}_0,\tau^{-2}\boldsymbol{\Sigma}_0), \tag{7}$$

$$\tau^2 \sim \text{Gamma}(c/2,d/2), \tag{8}$$

$$\delta \sim \text{Gamma}(a,b). \tag{9}$$

The representation of the equations (7)-(8) in model M_2 allows us to have a good range of flexible distributions belonging to the class of scale normal distributions for the prior distributions for $\boldsymbol{\beta}$. This class present distributions with heavier or lighter tails than the normal distribution, besides having normal distribution as a particular case. We highlighted Normal distribution, Cauchy distribution, Student t-distribution and Person Type VII family as particular cases of model M_2 (see for example, Choy and Chan (2008)).

The normal distribution is obtained assuming that τ^2 follows a degenerate distribution on one (model M_1) and a Cauchy distribution is obtained assuming $c=d=1$ in the prior distribution in equation (8). There exist quite a few forms of the multivariate Student

t-distribution (see Kotz and Nadarajah (2004)) with many of the cited variations focusing on introducing non-centrality. Among all the possible multivariate representations, the most common form considered is obtained when $c = d = \nu$, where ν denotes the degrees of freedom. This one will be considered in this paper. Finally, Person Type VII family is obtained for any $c, d > 0$ in equation (8).

Is is important to observe in model M_2 that the hyperparameters c and d controls shape and tails of the prior distribution of $\boldsymbol{\beta}$ is considered fixed and known. That it is not always a good alternative to modeling data. Therefore, the last model considered in this paper, denoted by M_3, consider that $c = d = \nu$ is unknown and a prior distribution is chosen for the degrees of freedom ν. The prior distribution resulting is a multivariate Student-t distribution with ν degrees of freedom for $\boldsymbol{\beta}$. It is also assumed that $\nu \sim \text{Unif}(c, d)$, with $c, d > 0$. The hierarchical representation of model M_3 is

$$y_i|\boldsymbol{\beta}, \delta \sim \mathcal{BS}(g^{-1}\{\mathbf{x}_i^\top \boldsymbol{\beta}\}, \delta), \quad i = 1, \cdots, n \tag{10}$$

$$\boldsymbol{\beta}|\tau^2 \sim \mathcal{N}_p(\boldsymbol{\mu}_0, \tau^{-2}\boldsymbol{\Sigma}_0), \tag{11}$$

$$\tau^2 \sim \text{Gamma}(\nu/2, \nu/2), \tag{12}$$

$$\nu \sim \text{Unif}(c, d), \tag{13}$$

$$\delta \sim \text{Gamma}(a, b). \tag{14}$$

Since the posterior distributions obtained for models M_1, M_2 and M_3 do not present a known distribution and do not have closed form, Monte Carlo Markov Chain (MCMC) methods are usually good to generate samples from the joint posterior distributions, allowing us to make inferences.

3. Section Three

Since a prior distribution has to be chosen to fit the model and, if there is no additional knowledge about the distribution of the parameters, an alternative is to choose the better prior distribution by using some criteria to compare how different models are by the fitting the data. The method used in this paper is based in predictive functions, verifying how good model predicts values from the data, using the Conditional Predictive Ordinate (CPO), proposed by Geisser and Eddy (1979). Firstly, consider D as the full data set and $D^{[i]}$ the data with the ith observation removed. The CPO_i measures the probability of the model to predict the value of the ith observation given the data without the predicted observation, i.e., $\text{CPO}_i = \pi(y_i|D^{[i]}) = \left\{\int \pi(\boldsymbol{\theta}|D)/\pi(y_i|\boldsymbol{\theta})d\boldsymbol{\theta}\right\}^{-1}$, which according to Christensen et al. (2011) may be approximated from the MCMC samples of the parameters $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(s)}$ by

$$\widehat{\text{CPO}_i} = \left\{ \frac{1}{s} \sum_{i=1}^{s} \frac{1}{f_i(y_i|\boldsymbol{\theta}^{(k)}, D)} \right\}^{-1},$$

where y_i is the observed value and $f_i(y_i|\boldsymbol{\theta}^{(k)}, D)$ is the probability density of y_i for the current value of $\boldsymbol{\theta}$ and the full data. Higher is the CPO_i value, more credible is the observation to be explained by the fitted model. Given the CPO values for each observation, two different models M_1 and M_2, for example, can be compared by using the log pseudo-marginal likelihood (LPML), given by $\text{LPML} = \sum_{i=1}^{n} \log(\text{CPO}_i)$, with n denoting the sample size of the data. By this criteria, the best model would be the one with largest LPML. Another possibility would be to use the pseudo-Bayes factor (PBF), also discussed by Geisser and

Eddy (1979), given by

$$\text{PBF}(\text{M}_1, \text{M}_2) = \frac{\pi(y_i|\hat{D^{[i]}}, \text{M}_1)}{\pi(y_i|\hat{D^{[i]}}, \text{M}_2)} = \frac{\prod_{i=1}^{n} \widehat{\text{CPO}}_i(\text{M}_1)}{\prod_{i=1}^{n} \widehat{\text{CPO}}_i(\text{M}_2)},$$

where $\widehat{\text{CPO}}_i(\text{M}_j)$ is the estimated value of CPO of the ith observation under model j. The model M_2 (M_1) is preferable than model M_1 (M_2) if $PBF[M_1, M_2] < 1 (> 1)$.

4. Section Four

In order to verify the presence of influential observations in regression models under Bayesian framework, Cho et al. (2009) proposed a method that uses the Kullback-Leibler (K-L) divergence measure to identify those observations. The K-L divergence is given by

$$K(\mathbb{P}, \mathbb{P}^{[i]}) = \int \pi(\boldsymbol{\theta}|D) \log \left\{ \frac{\pi(\boldsymbol{\theta}|D)}{\pi(\boldsymbol{\theta}|D^{[i]})} \right\} d\boldsymbol{\theta},$$

where $\boldsymbol{\theta}$ is the parameters vector of the model, $\mathbb{P}$ is the posterior distribution of $\boldsymbol{\theta}$ for the full data and $\mathbb{P}^{[i]}$ is the posterior distribution for the data with the ith observation removed. According to Cho et al. (2009), these measures can also be obtained through the expression

$$K(\mathbb{P}, \mathbb{P}^{[i]}) = -\log(\text{CPO}_i) + \mathbb{E}_{\boldsymbol{\theta}} \left[\log \left\{ f(y_i|\boldsymbol{\theta}) \right\} |D \right], \tag{15}$$

where $\mathbb{E}_{\boldsymbol{\theta}} [\cdot|D]$ is the posterior expectation of the distribution of $\boldsymbol{\theta}$ given D. The equation (15) allows to calculate an approximation of the K-L divergence using MCMC samples of the posterior distribution of $\boldsymbol{\theta}$ for the full data. Also, comparing values that might define a coin as biased or unbiased, Vidal and Castro (2010) derived a cut-off point for the K-L divergence, where an observation with $K(\mathbb{P}, \mathbb{P}^{[i]}) > 0.14$ could be considered as influential. Besides having a intuitive justification, this cut-off point might be too rigorous, thus we propose using twice the mean value of $K(\mathbb{P}, \mathbb{P}^{[1]}), \ldots, K(\mathbb{P}, \mathbb{P}^{[n]})$, similarly the way done in classical approach for total local influence (see next section), allowing more flexibility for the cut-off point to vary according to the sample analyzed. These rules of thumb are very useful in applications and provide criteria of assessing influence rather than visual inspection.

Another quantity used with the K-L divergence is the calibration, whose definition is given by McCulloch (1989). Given that $K(\mathbb{Q}_1, \mathbb{Q}_2) = k$ for two posterior distributions $\mathbb{Q}_1$ and $\mathbb{Q}_2$, the objective consists to find a value for $q(k)$ in order to $K\left[B\{1/2\}, B\{q(k)\}\right] = k$, where $B\{p\}$ represents a Bernoulli distribution whith parameter p. According to McCulloch (1989) the value of $q(k)$ is the calibration of k, and allows to observe similarities between the K-L divergences in the situations $\mathbb{Q}_1$ and $\mathbb{Q}_2$, and $B\{1/2\}$ and $B\{q(k)\}$, where $q(k) \approx 1$ means that models $\mathbb{Q}_1$ and $\mathbb{Q}_2$ are notoriously different, while if $q(k) \approx 0.5$, the models are quite similar. Taking $q(k)$ as p_i, according to Cho et al. (2009), $K(\mathbb{P}, \mathbb{P}^{[i]})$ may be measured by

$$K(\mathbb{P}, \mathbb{P}^{[i]}) = K(B\{1/2\}, B\{p_i\}) = \frac{-\log\left\{4p_i(1-p_i)\right\}}{2},$$

where using the divergence measure to calculate p_i, we have

$$p_i = \frac{1 + \sqrt{1 - \exp\{-2K(\mathbb{P}, \mathbb{P}^{[i]})\}}}{2},$$

where $p_i \in [0.5, 1]$. The values p_i can be used to verify how well the current model fitted the ith observation, where values further from 0.5 might be an indicative of influence.

5. Section Five

Local influence method was firstly proposed by Cook (1986) to analyze differences between an adopted model and a model with perturbed observations. Considering $\boldsymbol{\theta}$ a vector $p \times 1$ of the model parameter and $\boldsymbol{\theta}_{\boldsymbol{w}}$ the parameters vector of the model under perturbation, a influence measure caused by the perturbation vector $\boldsymbol{w} \subset \mathbb{R}^n$, with n being the sample size, can be obtained by

$$LD(\boldsymbol{w}) = 2[l(\boldsymbol{\theta}) - l_{\boldsymbol{w}}(\boldsymbol{\theta})],$$

where $l(\boldsymbol{\theta})$ is the log-likelihood function for the non-perturbed data, $l_{\boldsymbol{w}}(\boldsymbol{\theta})$ is the log-likelihood function for the model under perturbation and $LD(\boldsymbol{w})$ is a deviance measure of the likelihood functions caused by the perturbation vector $\boldsymbol{w}$.

Considering $\boldsymbol{w}_0$ a non-perturbation vector where $l_{\boldsymbol{w}_0}(\boldsymbol{\theta}) = l(\boldsymbol{\theta})$, the method consists in verify the behavior of $LD(\boldsymbol{w})$ around $\boldsymbol{w}_0$. The influence analysis is done using a unit direction $||\boldsymbol{d}|| = 1$ and evaluating the plot of $LD(\boldsymbol{w}_0 + a\boldsymbol{d})$ versus a, $a \in \mathbb{R}$. The usual method consists in consider the direction $\boldsymbol{d}_{max}$ which corresponds to the largest normal curvature $C_{\boldsymbol{d}}(\boldsymbol{\theta})$ around $a = 0$, and Cook (1986) gives its expression which is

$$C_{\boldsymbol{d}}(\boldsymbol{\theta}) = 2|\boldsymbol{d}^{\mathsf{T}}\boldsymbol{\Delta}^{\mathsf{T}}\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}\boldsymbol{\Delta}\boldsymbol{d}|,$$

where $-\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ is the observed Fisher's matrix and $\boldsymbol{\Delta}$ is a perturbation matrix with elements

$$\boldsymbol{\Delta}_{ij} = \frac{\partial^2 l_{\boldsymbol{w}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{w}_j},$$

evaluated in $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, with $i = 1, \ldots, p$ and $j = 1, \ldots, n$. According to Cook (1986), we have that C_{max} is taken as the largest eigenvalue of $\ddot{\mathbf{F}} = -\boldsymbol{\Delta}^{\mathsf{T}}\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}\boldsymbol{\Delta}$ and $\boldsymbol{d}_{max}$ and its associated eigenvector. The graphical analysis of influence consists in evaluate the plot of $\boldsymbol{d}_{max}$ versus the observations indexes.

The local influence analysis for the reparametrized $\mathcal{BS}$ regression model was discussed by Leiva et al. (2014) considering many cases of perturbation. Another useful way to verify the influence is called total local influence, proposed by Lesaffre and Verbeke (1998), evaluating $C_i = 2|\ddot{\mathbf{F}}_{ii}|$, where $\ddot{\mathbf{F}}_{ii}$ is the ith diagonal element of $\ddot{\mathbf{F}}$. The authors also proposed a cut-off point, where observations with C_i greater than $2\bar{C}$ should be highlighted as influential, where $\bar{C} = n^{-1}\sum_{i=1}^{n} C_i$. This technique can be very useful in applications and could be used in comparison with other methods of influence diagnostic, like the K-L divergence in Bayesian models.

6. Section Six

6.1 Different priors comparison

A Monte Carlo simulation study was performed to compare the different prior distributions presented in the three models M_1, M_2 and M_3 from Section 2. Observations from $Y_i \sim \mathcal{BS}(\mu_i, \delta)$ and $X_i \sim \text{Unif}(0,1)$ were simulated, with

$$\mu_i = \exp\{\beta_0 + \beta_1 x_i\}, \quad i = 1, \ldots, 50.$$

In the Bayesian models was considered $\boldsymbol{\Sigma}_0 = \tau^{-2}\mathbf{I}_p$ in equation (4), with $\mathbf{I}_p$ denoting the identity matrix of dimension p and $\tau^2 > 0$ a fixed hyperparameter for the precision for the prior distribution of $\boldsymbol{\beta}$. Furthermore, it is assumed $\boldsymbol{\mu}_0 = \mathbf{0}_p$ in (4), where $\mathbf{0}_p$ is a vector of p zeros. The three considered models are presented below:

The first model, denoted by M_1, is the $\mathcal{BS}$-Normal prior model given in equations (3)-(5) with $\boldsymbol{\mu}_0 = \mathbf{0}_2$, $\boldsymbol{\Sigma}_0 = 1000\mathbf{I}_2$ and $a = b = 0.01$. The second model, denoted by M_2, is the $\mathcal{BS}$-Person Type VII prior model given in equations (6)-(9) with $\boldsymbol{\mu}_0 = \mathbf{0}_3$, $\boldsymbol{\Sigma}_0 = 1000\mathbf{I}_2$, $a = 0.01$, $b = 0.01$, $c = 2$ and $d = 4.002$. Finally, the last model, denoted by M_3, is the $\mathcal{BS}$-Student-t prior model given in equations (10)-(14) with $\boldsymbol{\mu}_0 = \mathbf{0}_2$, $\boldsymbol{\Sigma}_0 = 1000\mathbf{I}_2$, $a_0 = 0$, $b_0 = 100$. These hyperparameters were chosen in order to guarantee a large variance (low precision) for the prior distributions, i.e., they consist in non informative prior distributions.

The models presented previously were fitted considering four different scenarios of parameters. They are $\boldsymbol{\theta}_j = (\beta_0, \beta_1, \delta)^\top, j = 1, 2, 3, 4$: $\boldsymbol{\theta}_1 = (2, -1.5, 3)^\top$; $\boldsymbol{\theta}_2 = (-2.5, 2, 3)^\top$; $\boldsymbol{\theta}_3 = (-.5, -1.7, 5)^\top$ and $\boldsymbol{\theta}_4 = (1.3, 2.2, 1)^\top$. The simulations were done using the softwares `R` (R Core Team (2014)) and `OpenBUGS` (Thomas et al. (2006)), through the package `R2OpenBUGS`. The initial values for the MCMC chains were obtained using the classical approach for the regression model, with MCMC samples of size 10000 with the first 1000 values being discarded in the burn-in period. To check the convergence of the MCMC chains, ACF plots and traceplots were analyzed using the `CODA` package (see Plummer et al. (2006)) and the convergence was confirmed in all cases.

From the results in the Table 1, we can observe that for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ the best model was M_1, but with a LPML close to M_2 and the worst model being M_3 in both scenarios. For θ_3 the three models had quite similar performance, with M_2 being the best. On the other hand, for $\boldsymbol{\theta}_4$ the best model was M_3, and the closest LPML from M_3 was the one from M_2. This results suggest that the prior choice in the $\mathcal{BS}$ regression model has a slight impact in its quality, but before choosing the model it would be useful compare the possible prior distributions.

6.2 Influence diagnostic comparison

A second Monte Carlo simulation was performed to evaluate this time the influence diagnostic of classical and Bayesian approaches in the $\mathcal{BS}$ regression model. We used the expressions given by Leiva et al. (2014) of generalized leverage (GL), local influence considering the case-weight perturbation (LI_{cw}), response perturbation (LI_{resp}) and covariate perturbation (LI_{cov}) schemes. A number of 1000 replicas of samples of the vectors $\mathbf{y}$ and $\mathbf{x}$ were generated, with $n = 50$ observations, where $y_i \sim \mathcal{BS}(\mu_i, \delta)$, $x_i \sim \text{Unif}(0,1)$, $\mu_i = \exp\{-0.5 - 1.7x_i\}$, $\delta = 5$ and $i = 1, \ldots, 50$. The covariate generated was kept fixed, being the same for each replica. To simulate influential observations, some values y_j randomly chosen from $\mathbf{y}$ were replaced by $y_j^* = y_j + 5\sigma_y$, where σ_y denotes the standard deviation from the sample $\mathbf{y}$. The effect of modifications in the covariate was also studied, changing some values of $\mathbf{x}$ in the same way, $x_j^* = x_j + 5\sigma_x$, where σ_x denotes the

Table 1. Comparison of the models M_1, M_2 and M_3 within different scenarios. The criteria LPML was calculated and the PBF was obtained between the models in the rows with the ones in the columns.

Scenario	Model	LPML	PBF	
			M_2	M_3
$\boldsymbol{\theta}_1$	M_1	-118.7370	1.062801	1.415612
	M_2	-118.7979	-	1.331963
	M_3	-119.0845	-	-
$\boldsymbol{\theta}_2$	M_1	24.5828	1.041390	1.283002
	M_2	24.5422	-	1.232009
	M_3	24.3336	-	-
$\boldsymbol{\theta}_3$	M_1	17.6647	0.983168	0.979655
	M_2	17.6816	-	0.996428
	M_3	17.6852	-	-
$\boldsymbol{\theta}_4$	M_1	-173.8033	0.853929	0.792765
	M_2	-173.6454	-	0.928373
	M_3	-173.5711	-	-

standard deviation from the sample $\mathbf{x}$. The percentage of modified values from the sample size was denoted by r, being 2%, 6% or 10%. For the Bayesian model an option for model M_2 was made and the initial estimates from the classical model were used for the MCMC chains, the later with sample size 3000, burn-in period of 300 observations and spacing of 2 observations. Before performing the complete simulation, the Bayesian model was analyzed for some replicas in order to verify the convergence of the MCMC chains, which was adequate considering traceplots and ACF plots. The influence measures were obtained for the complete parameter vector $\boldsymbol{\theta} = (\beta_0, \beta_1)^\top$. Observations were considered as influential in each approach through the cut-off points $2\bar{C}$ proposed by Lesaffre and Verbeke (1998) for total local influence in the classical model, and $K_{cut1} = 0.14$ for the K-L divergence in the Bayesian model discussed by Vidal and Castro (2010). We also proposed using $K_{cut2} = 2\overline{\text{K-L}}$ as cut-off point for the K-L divergence, similarly in the total local influence, where $\overline{\text{K-L}}$ is the mean of K-L values. We denoted the results using K_{cut2} as K-Ld. For generalized leverage we used a cut-off point of $2p/n$, where p is the number of parameters and n the sample size. Percentages of correct detection of the modified values were obtained for each diagnostic method, evaluating the cut-off points.

In the Table 2, we can notice LI_{cw} and K-Ld had the best performance in general, i.e., the cut-off point $2\bar{C}$ of total local influence under case-weight perturbation and K_{cut2} were the most efficient in the simulation study detecting the perturbed observations. We can observe that LI_{cov} had the best results detecting perturbations like x^*, but for perturbations like y^* the results were worse, as expected. On the other hand, LI_{resp} performed poorly in both schemes of perturbation compared with the other methods, thus its results were not presented. The GL had a better performance detecting perturbation of the form x^* (which is similar to generating leverage points) than y^* (similar to generating outliers). The influence diagnostic in the Bayesian model using K_{cut1} detected the perturbed observations less efficiently than LI_{cw} and K-Ld, which might have occurred because its cut-off point is fixed for all samples, making this Bayesian criteria more rigorous in taking an observation as influential.

Therefore, we see that LI_{cw} and K-Ld were the most efficient methods to detect influent points in the simulation study, with a slight advantage for the latter. We also noticed that using $K_{cut2} = 2\overline{\text{K-L}}$ improved the performance of the Bayesian diagnostic, allowing more flexibility for the K-L cut-off point.

Table 2. Rates of correct detection of influence points by the total local influence under different schemes of perturbation, generalized leverage and K-L divergence (for two cut-off points). The proportion of modified variables to simulate influent observations is r and the terms y^* and x^* denote if the modification was on the response or in the covariate.

Modification	r	LI_{cw}	LI_{cov}	GL	K-L	K-Ld
x^*	%2	0.985	0.995	0.771	0.978	0.993
	%6	0.618	0.718	0.503	0.484	0.625
	%10	0.420	0.539	0.371	0.266	0.427
y^*	%2	1.000	0.551	0.000	0.961	1.000
	%6	0.941	0.351	0.000	0.724	0.936
	%10	0.731	0.248	0.000	0.443	0.767

7. Application

The data set analyzed is the biaxial fatigue data, first studied by Brown and Miller (1978), which consists in 46 observations about the lifetime until its failure, in cycles, of a piece of metal. The variables are Y, the number of cycles until failure, and $Work$, measuring the work per cycle. Was noticed by Rieck and Nedelman (1991) that using $\log(Work)$ (which we denote by x) is a more appropriate choice. The models were adjusted using the softwares `R` and `OpenBUGS`.

The dataset was also analyzed by other authors with different models using the Birnbaum-Saunders distribution, like the already cited Galea et al. (2004), Lemonte and Patriota (2011), Farias and Lemonte (2011), for example. Our analysis differ from the previous works using a reparametrized $\mathcal{BS}$ and comparing the classical and Bayesian approaches for the model.

7.1 Classical approach

Using the results presented in Leiva et al. (2014), the classical approach for the Birnbaum-Saunders regression model was implemented for the following model

$$\log(\mu_i) = \beta_0 + \beta_1 x_i, \quad i = 1, \ldots, 46, \tag{16}$$

where $Y_i \sim \mathcal{BS}(\mu_i, \delta)$ and the logarithmic link function is used. Maximum likelihood estimates and its standard errors (SE) are presented in Table 3. We can observe that the covariate x has a negative influence in the expected value of the lifetime, and the regressions coefficients are significant since the intervals of two standard errors do not contain zero.

In Figure 1, the residuals plot show an apparent random behavior, with one observation outside the the limits of $(-2, 2)$, which indicates this point might be possibly outlier. A simulated envelope (see Atkinson (1985)) was done using the standardized residual proposed by Leiva et al. (2014), which has the form $r_i = \{\widehat{\mathrm{Var}}(Y_i)\}^{-1/2}(y_i - \hat{\mu}_i)$, where $\widehat{\mathrm{Var}}(Y_i) = \hat{\mu}_i^2(2\hat{\delta} + 5)/(\hat{\delta} + 1)^2$, indicates that the fitted model seems adequate.

Table 3. Estimates for the model (16) in classical approach and its respective standard errors.

Parameter	Estimate	SE
β_0	12.3606	0.4001
β_1	-1.6708	0.1111
δ	11.8771	2.4765

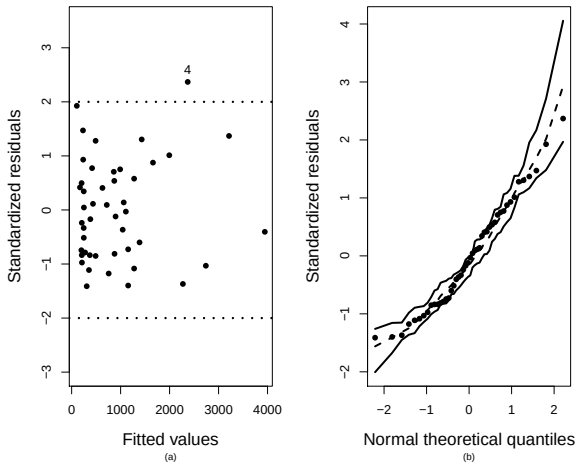

Figure 1. Plot of residuals (a) and simulated envelope (b)

7.2 Bayesian approach

The Birnbaum-Saunders regression model given in equation (16) was also considered under the Bayesian approach. However, it will be used three different prior distributions for the regression coefficients and it will be verified the sensibility of the model (16) for each prior distribution considered in this paper. It is important to emphasize since we have no prior information about the dependence structure among the regression coefficients, it is a fair choice to assume the non informative prior distributions used in the simulation study in the section 6.1. A MCMC sample of size 10000 was generated with a burn-in period of 1000. The convergence of the chains generated by MCMC algorithms were verified using procedures included in `CODA` package and the software `R`, being the latter also used to calculate the quality measures of fit.

Table 4. Estimates in the Bayesian model.

Model	Parameter	Mean	SE	$P_{2,5}$	Median	$P_{97,5}$
	β_0	12.3924	0.4138	11.5800	12.390	13.200
M_1	β_1	-1.6774	0.1148	-1.9000	-1.677	-1.451
	δ	11.1991	2.3376	7.1858	11.030	16.360
	β_0	12.3197	0.4001	11.530	12.3100	13.1200
M_2	β_1	-1.6572	0.1108	-1.877	-1.6560	-1.4400
	δ	11.3063	2.3613	7.114	11.1700	16.3800
	τ^2	0.0255	0.0183	0.003	0.0212	0.0719
	β_0	12.3307	0.4155	11.5097	12.3300	13.1300
M_3	β_1	-1.6597	0.1168	-1.8800	-1.6610	-1.4280
	δ	11.2473	2.3883	7.0839	11.1200	16.2300
	τ^2	0.0196	0.0169	0.0010	0.0150	0.0635
	ν	1.0813	0.8760	0.1153	0.8419	3.3532

We can observe in Table 4 that the inferential results for β_0, β_1 and δ in the three models are quite similar to the equivalent one obtained in classical approach presented in Table 3. Since the zero is not contained in the credible intervals of the regression coefficients, we can assume that the model presents an intercept and the variable x has significant information about the response Y. In Figure 2 we can see that the results for calibration were very similar in the three models (the correlations were greater than 0.99), meaning that the choice of the prior distribution would not have great influence on diagnostic analysis. Also, most points have a calibration value close to 0.5 and 0.6, with some exceptions that

are the points #2, #4, #5, #12, #32 and #46. These observations might be diagnosed as influential on further analysis.

Calculating the log pseudo-marginal likelihood (LPML) on these models, is obtained $\text{LPML}(\text{M}_1) = -317.9599$, $\text{LPML}(\text{M}_2) = -317.9022$ and $\text{LPML}(\text{M}_3) = -318.0876$, which are quite close, with M_2 a little larger. Using the pseudo-Bayes factor we obtained $\text{PBF}(\text{M}_1, \text{M}_2) = 0.9439$, $\text{PBF}(\text{M}_1, \text{M}_3) = 1.1362$ and $\text{PBF}(\text{M}_2, \text{M}_3) = 1.2036$, therefore, again we concluded that model M_2 is slightly better among those three for the data set in study. However, there are no strong evidence in favor of one model. This result shows that the model it is quite robust on the choice of the prior distribution of the coefficients regression.

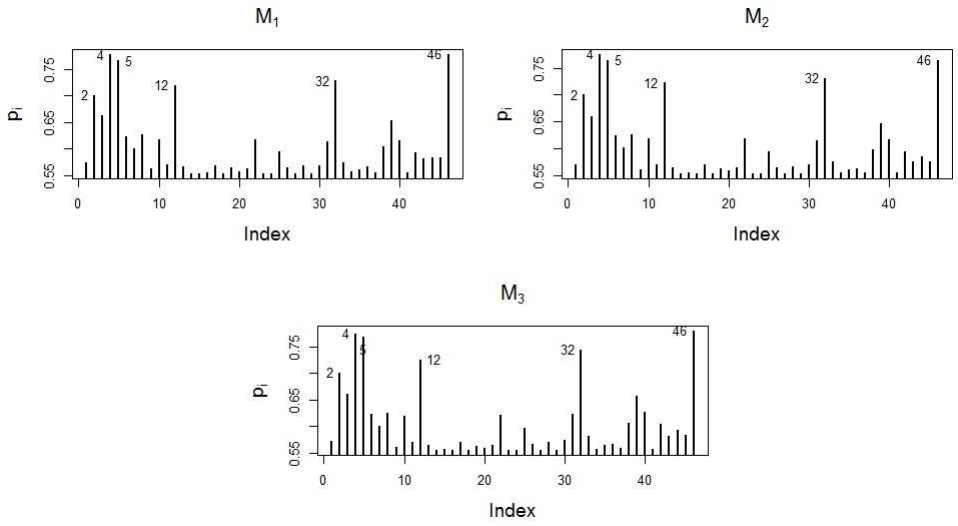

Figure 2. Calibrations of the Bayesian models.

7.3 Influence diagnostic comparison

The quantities for influence diagnostic and generalized leverage were calculated and normalized (divided by their norms) for comparison. The total local influence was used considering case-weight perturbation and covariate perturbation schemes, which had better results in the simulation study than the response perturbation scheme. For the Bayesian analysis the model M_2 was considered and the influence measure was the Kullback-Leibler divergence using two different cut-off points, $K_{cut1} = 0.14$ and $K_{cut2} = 2\overline{\text{K-L}}$. We could observe in the Figure 3 that LI_{cw} and K-L divergence appeared to be equivalent, having almost the same results for each observation, with correlation 0.9986 for their values. This result is interesting because C_i is obtained under classical approach while K-L is a Bayesian methodology, being criteria of influence of different methodologies, but identifying the influence points almost in the same way. The plots of LI_{cov} and GL were different from the others, but with common observations detected as leverage or influential points. The cut-off points of each method were also divided by their respective values of norm and plotted as the horizontal dotted line. We can observe that K_{cut1} from K-L is clearly more rigorous than LI_{cw} and K_{cut2} in defining an observation as influential, which agrees with the results obtained in the simulations in section 6.2. Observations #4 and #5 were the most detected by all methods.

To verify the influence of the observations detected by the total local influence and K-L divergence, the model (16) was refitted excluding individually the observations #2,

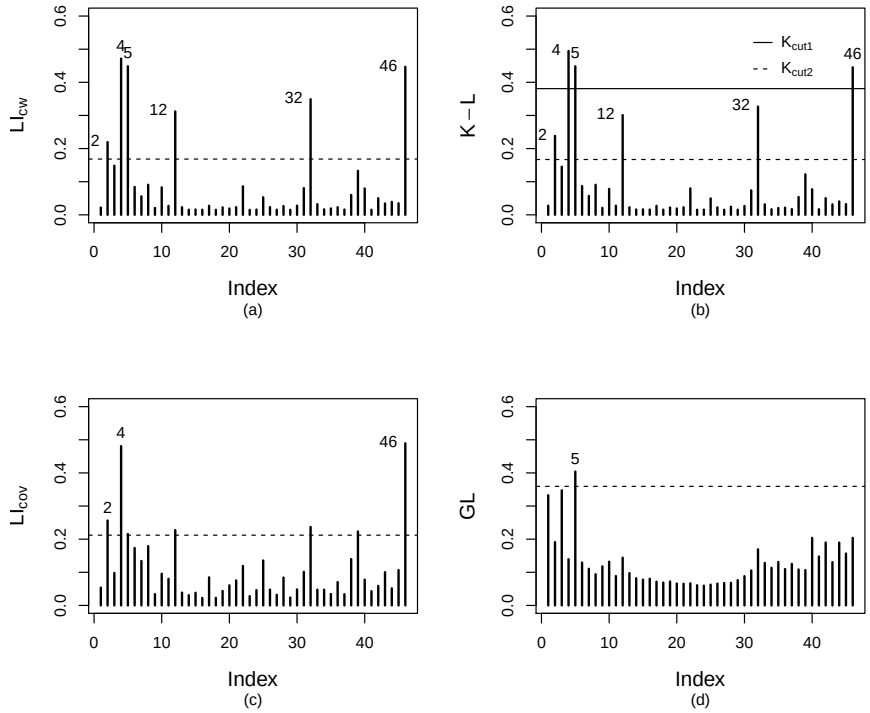

Figure 3. Influence diagnostic plots of local influence under case-weight (a) and covariate (c) perturbations, generalized leverage (d) and Kulback-Leibler divergence (b). All measures in this plot were normalized for comparison reasons.

#4, #5, #12, #32 and #46, with results given in Table 5. There, estimates from the regression models under classical and Bayesian approach were obtained and the relative changes $\left|(\hat{\alpha} - \hat{\alpha}^{[i]})/\hat{\alpha}\right|$ were calculated, where $\hat{\alpha}$ represents a estimate obtained with the full data and $\hat{\alpha}^{[i]}$ the same estimate without the ith observation in the data. We can observe that for both methodologies the greatest changes are on the estimates about δ, with difference of 7.7% of observation #12 in the classical model. The observations #2, #12 and #32 were not detected as influential in the Bayesian model using K_{cut1}, and we can see that the change in the regression coefficients caused by the elimination of these observations were the least in the Bayesian approach. Observation #5, was the only one detected in the generalized leverage plot and caused a change quite similar to that made by observation #4, suggesting that #5 could be viewed as influent and leverage point. For the estimates of parameters β_0 and β_1, in both methodologies, we can observe that observations #4, #5 and #46 were detected as the most influential, which could also be observed previously in the analysis of Figure 2, where these observations had the largest values of calibrations.

Table 5. Estimates under classical and Bayesian approach eliminating each observation identified as influential in Figure 3. The relative changes of the estimates are in parenthesis.

Eliminated	Classical			Bayesian		
case	β_0	β_1	δ	β_0	β_1	δ
None	12.361(0.000)	-1.671(0.000)	11.877(0.000)	12.318(0.000)	-1.657(0.000)	11.355(0.000)
2	12.202(0.013)	-1.630(0.024)	12.118(0.020)	12.161(0.013)	-1.616(0.025)	11.502(0.013)
4	12.160(0.016)	-1.621(0.030)	12.713(0.07)	12.091(0.018)	-1.600(0.034)	12.084(0.064)
5	12.543(0.015)	-1.718(0.029)	12.697(0.069)	12.491(0.014)	-1.702(0.027)	12.058(0.062)
12	12.455(0.008)	-1.694(0.014)	12.745(0.073)	12.403(0.007)	-1.677(0.012)	12.069(0.063)
32	12.280(0.007)	-1.644(0.016)	12.793(0.077)	12.232(0.007)	-1.629(0.017)	12.162(0.071)
46	12.488(0.014)	-1.710(0.032)	11.776(0.037)	12.548(0.015)	-1.729(0.035)	12.459(0.049)

Through the latter result, we might conclude that two models fitted under different

approaches detected quite the same results in identifying possible influent points, which is a indicator of the equivalence of classical and Bayesian methodologies for the $\mathcal{BS}$ regression models. This approach consists in a more intensive way to carry out influence detection in $\mathcal{BS}$ regression models, specially using K_{cut2} for the K-L divergence, which arrived at the same results of the classical approach and is better than the previous cut-off point proposed, according to the simulation study for the $\mathcal{BS}$ model.

8. Discussion

In this paper we developed a Bayesian approach for the regression model proposed by Leiva et al. (2014), where we directly analyze the expected value of variable with the reparametrized distribution. Different model structures for Bayesian modeling were given as well as comparison criteria from the literature to decide on which model to use. Local influence diagnostic for classical and Bayesian approaches was discussed. In simulation studies we could observe that the model is quite robust for different prior distributions choices and that the cut-off points of the influence measures in the classical models performed better than the Bayesian model using cut-off points suggested in the literature, especially the total local influence under case-weight perturbation (LI_{cw}). On the other hand, we proposed the use of a cut-off point similar from the classical method in the Bayesian model, which improved the results from the latter. An application to real data was presented to illustrate the usage of the model under both methodologies and we could conclude that besides the differences, both handled equivalent results, giving special attention for the similarities between the influence diagnostic plots of LI_{cw} and K-L divergence. The compared model consists in a more intensive approach to analyze possible influent observations, which can be used by practioners that wish to pay more attention for such data points that usually affect the estimates significantly. We could observe that under both methodologies the results were equivalent using the best methods from the simulation study, but the comparison is a good way to confirm the results from one way or another.

Acknowledgements

We are grateful to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq projects 445294/2014-0 and 306770/2014-8) and Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) for partial financial support.

References

Achcar, J.A., 1993. Inferences for the Birnbaum-Saunders fatigue life model using Bayesian methods. Comput. Statist. Data Anal., 15, 367–380.

Atkinson, A.C., 1985. Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. Oxford University Press, New York.

Birnbaum, Z.W., Saunders, S.C., 1969a. A new family of life distributions. J. Appl. Probab., 6, 319–327.

Birnbaum, Z.W., Saunders, S.C., 1969b. Estimation for a Family of Life Distributions with Applications to Fatigue. J. Appl. Probab., 6, 328–347.

Brown, M.W., Miller, K.J., 1978. Biaxial Fatigue Data, Report CEMR1/78, Department of Mechanical Engineering. University of Sheffield, Sheffield, UK.

Cancho, V.G., Ortega, E.M.M., Paula, G.A., 2010. On estimation and influence diagnostics for log-Birnbaum-Saunders Student-t regression models: Full Bayesian analysis. J. Statist. Plann. Inference, 140, 2486–2496.

Cho, H., Ibrahim, J.G., Sinha, D., Zhu, H., 2009. Bayesian Case Influence Diagnostics for Survival Models. Biometrics, 65, 116–124.

Choy, B., Chan, J., 2008. Scale Mixtures distributions in Statistical Modelling. Australian and New Zealand Journal of Statistics, 50, 135–146.

Christensen, R., Johnson, W.O., Branscum, A.J., Timothy, E., 2011. Bayesian Ideas and Data Analysis:An Introduction for Scientists and Statisticians. Chapman and Hall, New York.

Cook, R.D., 1986. Assessment of local influence. J. R. Stat. Soc. Ser. B. Stat. Methodol., 48, 133–169.

Dey, D., Ghosh, S., Mallick, B., 2000. Generalized Linear Models: A Bayesian Perspective. CRC Press, New York.

Díaz-García, J.A., Leiva, V., 2002. A new family of life distributions based on birnbaum-saunders distribution. Tech. Report I-02-17(PE/CIMAT), Mexico.

Farias, R.B.A., Lemonte, A.J., 2011. Bayesian inference for the Birnbaum-Saunders non-linear regression model. Stat. Methods Appl., 20, 423–438.

Galea, M., Leiva, V., Paula, G.A., 2004. Influence diagnostic in log-Birnbaum-Sunders regression models. J. Appl. Stat., 31, 1049–1064.

Geisser, S., Eddy, W., 1979. A Predictive Approach to Model Selection. J. Amer. Statist. Assoc., 74, 153–160.

Kotz, S., Nadarajah, S., 2004. Multivariate t Distributions and their Applications. Cambridge University Press, Cambridge.

Leiva, V., 2016. The Birnbaum-Saunders distribution. Academic Press, London.

Leiva, V., Santos-Neto, M., Cysneiros, F.J.A., Barros, M.E., 2014. Birnbaum-Saunders statistical modelling: a new approach. Stat. Model., 14, 21–48.

Lemonte, A.J., Patriota, A.G., 2011. Influence diagnostics in Birnbaum-Saunders regression models. J. Appl. Stat., 38, 871–884.

Lesaffre, E., Verbeke, G., 1998. Local Influence in Linear Mixed Models. Biometrics, 54, 570–582.

McCulloch, R.E., 1989. Local Model Influence. J. Amer. Statist. Assoc., 84, 473–478.

Owen, W.J., Padgett, W.J., 2000. A Birnbaum-Saunders accelerated life model. IEEE Transactions on Reliability, 49, 224–229.

Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: Convergence diagnosis and output analysis for MCMC. R News, 6, 7–11.

R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna; Software available at <http://www.R-project.org/>.

Rieck, J.R., Nedelman, J.R., 1991. A Log-Linear Model for the Birnbaum-Saunders Distribution. Technometrics, 33, 51–60.

Santos-Neto, M., Cysneiros, F.J.A., Leiva, V., Ahmed, S.E., 2012. On New Parametrizations Of The Birnbaum-Saunders Distribution. Pak. J. Statist., 28, 1–26.

Thomas, A., O'Hara, B., Ligges, U., Sturtz, S., 2006. Making BUGS Open. R News, 6, 12–17. Available at: <http://www.openbugs.net/w/Downloads>.

Tsionas, E.G., 2001. Bayesian Inference in Birnbaum-Saunders Regression. Communication in Statistics - Theory and Methods, 30, 179–193.

Vidal, I., Castro, L.M., 2010. Influential observations in the independent Student-t measurement error model with weak nondifferential error. Chil. J. Stat., 1, 17–34.