Submitted to the Brazilian Journal of Probability and Statistics

A new log-linear bimodal Birnbaum-Saunders regression model with application to survival data

Francisco Cribari-Neto[a] **and Rodney V. Fonseca**[a]

[a]*Departamento de Estatística, Universidade Federal de Pernambuco, Recife, PE, Brazil*

Abstract. The log-linear Birnbaum-Saunders model has been widely used in empirical applications. We introduce an extension of this model based on a recently proposed version of the Birnbaum-Saunders distribution which is more flexible than the standard Birnbaum-Saunders law since its density may assume both unimodal and bimodal shapes. We show how to perform point estimation, interval estimation and hypothesis testing inferences on the parameters that index the regression model we propose. We also present a number of diagnostic tools, such as residual analysis, local influence, generalized leverage, generalized Cook's distance and model misspecification tests. We investigate the usefulness of model selection criteria and the accuracy of prediction intervals for the proposed model. Results of Monte Carlo simulations are presented. Finally, we also present and discuss an empirical application.

1 Introduction

The Birnbaum-Saunders distribution ($\mathcal{BS}$) was proposed by Birnbaum and Saunders (1969) to analyze fatigue lifetime data. It has been widely discussed in the literature and was used as the baseline for several related probability distributions. A concise review on the Birnbaum-Saunders distribution and its extensions can be found in Leiva (2015).

Regression models based on the Birnbaum-Saunders distribution were also discussed in the literature, the first model being introduced by Rieck and Nedelman (1991), who proposed the log-linear Birnbaum-Saunders regression model (BSRM). Their model was later analyzed by Galea, Leiva-Sánchez and Paula (2004), who developed some diagnostic tools for the model, and by Xie and Wei (2007), who provided additional tools for detecting atypical observations.

The BSRM was extended in several different directions. For instance, Barros, Paula and Leiva (2008) developed a model using the Student-t Birnbaum-Saunders distribution and Lemonte and Cordeiro (2009) proposed

MSC 2010 subject classifications: 62J99, 62F03, 62F40

Keywords and phrases. Birnbaum-Saunders distribution, diagnostic analysis, model selection criteria, prediction intervals, misspecification test

a nonlinear version of the BSRM. A mixed model for censored data based on the Birnbaum-Saunders distribution was introduced by Villegas, Paula and Leiva (2011) and Lemonte (2013) proposed a log-linear model based on an extended Birnbaum-Saunders distribution. The log-linear Birnbaum-Saunders power regression model was proposed by Martínez-Flórez, Bolfarine and Gómez (2017), who also introduced the nonlinear sinh-power-normal regression model. Model misspecification tests for the BSRM were proposed by Santos and Cribari-Neto (2017). Bayesian inference for the BSRM was developed by Tsionas (2001). More recently, Vilca, Azevedo and Balakrishnan (2017) introduced the nonlinear sinh-normal/independent regression model, which encompasses several other $\mathcal{BS}$ regression models and developed Bayesian inference for such a model.

The chief goal of our paper is to propose a log-linear regression model based on a bimodal version of the Birnbaum-Saunders distribution that has been recently introduced by Owen and Ng (2015). Such a distribution is more flexible than the original $\mathcal{BS}$ law and we use it as the basis for developing a regression model that is more general than the BSRM introduced by Rieck and Nedelman (1991). Parameter estimation and standard inferential strategies are presented. A second goal of our paper is to provide diagnostic tools for the proposed regression model, thus allowing practitioners to verify whether the model assumptions are satisfied and making it possible for them to detect atypical observations, much in the same spirit as done for the BSRM by Galea, Leiva-Sánchez and Paula (2004) and Xie and Wei (2007). We also present a RESET-like misspecification test which is similar to the one introduced by Santos and Cribari-Neto (2017) for the log-linear Birnbaum-Saunders model. The test can be used to check whether the model's functional form is correctly specified. In addition, we consider the issue of performing model selection based on model selection criteria and the construction of prediction intervals for the regression model we propose similarly to what was done by Bayer and Cribari-Neto (2015) and Espinheira, Ferrari and Cribari-Neto (2014), respectively, for beta regressions.

In summary, in this paper we introduce a new regression model that generalizes the standard log-linear Birnbaum-Saunders regression model. We thus offer applied statisticians an additional model that they can consider in their data analyses. We develop point estimation, interval estimation, hypothesis testing inference and diagnostic analyses for the proposed model. Model selection strategies are also presented. It is interesting to note that in the empirical application we present in Section 8 the regression model we propose is shown to be more adequate than the standard log-linear Birnbaum-Saunders model.

The remainder of the paper is organized as follows. In Section 2, the bimodal Birnbaum-Saunders distribution is presented as well as some of its key properties. The log-linear model for responses that follow such a bimodal Birnbaum-Saunders distribution is proposed in Section 3. A broad variety of diagnostic tools for the proposed model are presented in Section 4. In Section 5 we address the issue of constructing prediction intervals for non-observed response values and in Section 6 we consider different model selection strategies. Results from Monte Carlo simulations are reported in Section 7, and an empirical application is presented and discussed in Section 8. Finally, Section 9 offers some concluding remarks.

2 The bimodal Birnbaum-Saunders distribution

The generalized Birnbaum-Saunders distribution considered in this paper was introduced by Dıaz-Garcıa and Domınguez-Molina (2006). They obtained it by adding a second shape parameter to the $\mathcal{BS}$ distribution function. More recently, Owen and Ng (2015) analyzed the distribution, which the authors denoted by $\mathcal{GBS}_2$. In particular, they investigated the relationships among the inverse Gaussian, $\mathcal{BS}$ and $\mathcal{GBS}_2$ distributions.

A random variable T is said to follow the $\mathcal{GBS}_2(\alpha, \eta, \nu)$ law if its distribution function is given by

$$F_T(t|\alpha,\eta,\nu) = \Phi\left(\frac{1}{\alpha}\left[\left(\frac{t}{\eta}\right)^{\nu} - \left(\frac{\eta}{t}\right)^{\nu}\right]\right), \quad t > 0, \tag{1}$$

where $\Phi(\cdot)$ denotes the standard normal distribution function, $\alpha > 0$, $\eta > 0$ and $\nu > 0$. Here, η is the scale parameter whereas α and ν are shape parameters. The probability density function of T is

$$f_T(t|\alpha,\eta,\nu) = \frac{\nu}{t\alpha\sqrt{2\pi}}\left[\left(\frac{t}{\eta}\right)^{\nu} + \left(\frac{\eta}{t}\right)^{\nu}\right]\exp\left\{-\frac{1}{2\alpha^2}\left[\left(\frac{t}{\eta}\right)^{\nu} - \left(\frac{\eta}{t}\right)^{\nu}\right]^2\right\}, \quad t > 0.$$

The $\mathcal{GBS}_2$ distribution has a noteworthy advantage over the original $\mathcal{BS}$ distribution: the density of the former can be unimodal and bimodal whereas that of the latter does not allow for more than one mode. The $\mathcal{GBS}_2$ density is bimodal whenever $\alpha > 2$ and $\nu > 2$ simultaneously. Figure 1 contains $\mathcal{GBS}_2$ density plots for different values of α, η and ν. It is noteworthy (Figure 1a) that the $\mathcal{GBS}_2$ density becomes more symmetric as the value of ν increases and the values of the remaining parameters are held constant. The value of α also impacts the distribution asymmetry: the distribution becomes less asymmetric as the value of α decreases (Figure 1b). In Figure 1c we see an

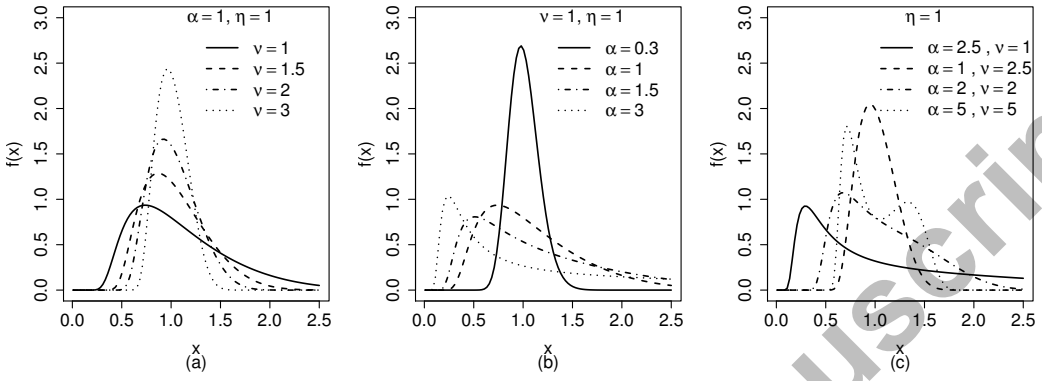

Figure 1 *$\mathcal{GBS}_2(\alpha, \eta, \nu)$ densities for some parameter values.*

example of a bimodal density when we take $\alpha > 2$ and $\nu > 2$. Notice that the $\mathcal{GBS}_2$ density is quite flexible, since it may assume a variety of different shapes.

Several useful properties of the $\mathcal{GBS}_2$ distribution were obtained by Owen and Ng (2015), some of them also holding for the $\mathcal{BS}$ distribution. For instance, it was shown that η is the distribution median, which can be easily verified from Equation (1). Additionally, the $\mathcal{GBS}_2$ distribution is closed under reciprocity and proportionality, i.e., $T^{-1} \sim \mathcal{GBS}_2(\alpha, \eta^{-1}, \nu)$ and $aT \sim \mathcal{GBS}_2(\alpha, a\eta, \nu)$, for $a > 0$. Pseudo-random number generation from $T \sim \mathcal{GBS}_2$ can be performed using the stochastic representation $T = \eta\left[\alpha Z/2 + \sqrt{(\alpha Z/2)^2 + 1}\right]^{1/\nu}$, where Z follows the standard normal distribution.

Another stochastic representation for the $\mathcal{GBS}_2$ distribution is as follows: if $T \sim \mathcal{GBS}_2(\alpha, \eta, \nu)$, then $Y = \log(T)$ is distributed as hyperbolic sine normal ($\mathcal{SHN}$), whose distribution function is $F_Y(y|\alpha, \mu, \sigma) = \Phi\{2\sinh[(y-\mu)/\sigma]/\alpha\}$, $y \in \mathbb{R}$, where $\mu = \log(\eta)$, $\sigma = 1/\nu$ and $\sinh(\cdot)$ denotes the hyperbolic sine function. We shall write $Y \sim \mathcal{SHN}(\alpha, \mu, \sigma)$, α being a shape parameter, σ being a scale parameter and μ being a location parameter and the mean of the distribution.

The relationship between the $\mathcal{BS}$ and $\mathcal{SHN}$ distributions was established by Rieck and Nedelman (1991), who noted that the log-Birnbaum-Saunders distribution is a particular case of the $\mathcal{SHN}$ distribution: the latter reduces to the former when $\sigma = 2$. The authors have also presented several properties of the $\mathcal{SHN}$ distribution and used the relationship between the two

distributions to propose the log-linear Birnbaum-Saunders regression model.

3 Log-linear $\mathcal{GBS}_2$ regression model

We use the relationship between the $\mathcal{GBS}_2$ and $\mathcal{SHN}$ distributions to propose a log-linear $\mathcal{GBS}_2$ regression model, the response variable being $Y = \log(T)$, where T follows the $\mathcal{GBS}_2$ law. Consider n independent random variables $T_1, \ldots, T_n$, where $T_i \sim \mathcal{GBS}_2(\alpha, \eta_i, \nu)$, $i = 1, \ldots, n$. The log-linear $\mathcal{GBS}_2$ regression model (GBS2RM) we propose is given by

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n,$$

where $y_i = \log(t_i)$, with $t_1, \ldots, t_n$ representing observations on the random variables $T_1, \ldots, T_n$. Here, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ is a vector of explanatory variables associated with the response variable, $\boldsymbol{\beta}=(\beta_1, \ldots, \beta_p)^\top$ is a p-vector of unknown parameters and $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{SHN}(\alpha, 0, \nu^{-1})$, iid indicating that the random variables are independent and identically distributed. Note that $t_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) e^{\epsilon_i}$, where $e^{\epsilon_i} \sim \mathcal{GBS}_2(\alpha, 1, \nu)$. Since the $\mathcal{GBS}_2$ distribution is closed under proportionality, it follows that T_i is $\mathcal{GBS}_2(\alpha, \exp(\mathbf{x}_i^\top \boldsymbol{\beta}), \nu)$ distributed.

Estimation of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \alpha, \nu)^\top$, the GBS2RM parameter vector, can be carried out by maximum likelihood. The log-likelihood function is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ \log\left(\frac{2}{\sqrt{2\pi}}\right) + \log(\xi_{i1}) - \frac{1}{2}\xi_{i2}^2 \right\},$$

where $\xi_{i1} = \nu\alpha^{-1}\cosh[\nu(y_i - \mu_i)]$ and $\xi_{i2} = 2\alpha^{-1}\sinh[\nu(y_i - \mu_i)]$, $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ being the linear predictor, $i = 1, \ldots, n$. The first derivatives of $\ell(\boldsymbol{\theta})$ with respect to the model's parameters are

$$\ell_{\boldsymbol{\beta}} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = X^\top \mathbf{a}, \quad \ell_\alpha = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = \sum_{i=1}^{n} \left\{ \frac{1}{\alpha}\left(\xi_{i2}^2 - 1\right) \right\},$$

$$\ell_\nu = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \nu} = \sum_{i=1}^{n} \left\{ \frac{1}{\nu} + \frac{\xi_{i2}}{\xi_{i1}}\frac{\nu}{2}(y_i - \mu_i) - \xi_{i1}\xi_{i2}\frac{2}{\nu}(y_i - \mu_i) \right\},$$

where $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$ is an $n \times p$ full column rank matrix and $\mathbf{a} = (a_1, \ldots, a_n)^\top$, with $a_i = 2\xi_{i1}\xi_{i2} - \nu^2\xi_{i2}/(2\xi_{i1})$, $i = 1, \ldots, n$.

The maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\alpha}, \hat{\nu})^\top$ of the parameters that index the model are obtained by solving $\ell_{\boldsymbol{\theta}} = 0$, where $\ell_{\boldsymbol{\theta}} =$

$\partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. They cannot be expressed in closed-form. Estimates can be obtained by numerically maximizing $\ell(\boldsymbol{\theta})$ using a Newton (e.g., Newton-Raphson) or quasi-Newton (e.g., BFGS) nonlinear optimization algorithm.

In order to perform maximum likelihood estimation it is necessary to specify a starting point for the parameter vector. Our proposal is to use the least squares estimate $\boldsymbol{\beta}_0 = (X^\top X)^{-1}X^\top \mathbf{y}$ as a starting value for $\hat{\boldsymbol{\beta}}$, where $\mathbf{y} = (y_1, \ldots, y_n)^\top$, along with $\nu_0 = 0.5$ as a starting point for ν, which is the value of ν that corresponds to the $\mathcal{BS}$ distribution. Finally, we recommend using as starting value $\alpha_0 = (2/\sqrt{n})\left(\sum_{i=1}^n \sinh^2[\nu_0(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_0)]\right)^{1/2}$ for α, which corresponds to the solution of $\ell_\alpha|_{(\boldsymbol{\beta},\nu)=(\boldsymbol{\beta}_0,\nu_0)} = 0$.

It is possible to show, after some algebra, that the model Hessian matrix is given by

$$\ell_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{bmatrix} X^\top VX & X^\top\mathbf{d} & X^\top\mathbf{g} \\ \mathbf{d}^\top X & \mathrm{tr}[D(\mathbf{b})] & \mathrm{tr}[D(\mathbf{c})] \\ \mathbf{g}^\top X & \mathrm{tr}[D(\mathbf{c})] & \mathrm{tr}[D(\mathbf{f})] \end{bmatrix},$$

where $V = \mathrm{diag}\{v_1, \ldots, v_n\}$, $D(\mathbf{b}) = \mathrm{diag}\{b_1, \ldots, b_n\}$, $D(\mathbf{c}) = \mathrm{diag}\{c_1, \ldots, c_n\}$, $D(\mathbf{f}) = \mathrm{diag}\{f_1, \ldots, f_n\}$, $\mathbf{d} = (d_1, \ldots, d_n)^\top$ and $\mathbf{g} = (g_1, \ldots, g_n)^\top$, with diag denoting a diagonal matrix and tr denoting the trace operator. The components of these vectors and matrices are $v_i = \nu^2\cosh^{-2}[\nu(y_i - \mu_i)] - 4\nu^2\cosh[2\nu(y_i - \mu_i)]/\alpha^2$, $b_i = -3\xi_{i2}^2/\alpha^2 + 1/\alpha^2$, $c_i = 4\xi_{i1}\xi_{i2}(y_i - \mu_i)/(\alpha\nu)$, $d_i = -4\xi_{i1}\xi_{i2}/\alpha$, $f_i = -\nu^{-2} + [(y_i - \mu_i)\nu/\alpha\xi_{i1}]^2 - 4\alpha^{-2}(y_i - \mu_i)^2\cosh[2\nu(y_i - \mu_i)]$ and $g_i = -\nu(y_i - \mu_i)\cosh^{-2}[\nu(y_i - \mu_i)] - \tanh[\nu(y_i - \mu_i)] + 4\nu\alpha^{-2}\cosh[2\nu(y_i - \mu_i)](y_i - \mu_i) - 2\alpha^{-2}\sinh[2\nu(y_i - \mu_i)]$.

Under mild regularity conditions (Severini, 2000), it can be shown that $\hat{\boldsymbol{\theta}}$ is asymptotically distributed as $\mathcal{N}_{p+2}(\boldsymbol{\theta}, \Sigma_{\boldsymbol{\theta}})$. The asymptotic covariance of $\hat{\boldsymbol{\theta}}$, $\Sigma_{\boldsymbol{\theta}}$, can be approximated by $-\hat{\ell}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}$, where $\hat{\ell}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ denotes $\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}$ evaluated at $\hat{\boldsymbol{\theta}}$. Hence, based on the asymptotic normality of $\hat{\boldsymbol{\theta}}$, it is possible to obtain an approximated $100\times(1-\gamma)\%$ confidence region for $\boldsymbol{\theta}$, $0 < \gamma < 1$, which is given by the set of values of $\boldsymbol{\theta}$ such that $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top(-\hat{\ell}_{\boldsymbol{\theta}\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq \chi^2_{p+2}(\gamma)$, where $\chi^2_{p+2}(\gamma)$ denotes the $1-\gamma$ quantile of the chi-square distribution with $p+2$ degrees of freedom.

We shall now turn to testing inferences. Consider the following partition of the parameter vector: $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\lambda})^\top$, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_r)^\top$ is the vector of parameters of interest and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_s)^\top$ is the vector of nuisance parameters, with $r+s = p+2$. We shall focus on the test of $\mathcal{H}_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0$ against $\mathcal{H}_1 : \boldsymbol{\psi} \neq \boldsymbol{\psi}_0$ in the context the GBS2RM model. The likelihood ratio statistic is given by $W = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})\}$, where $\tilde{\boldsymbol{\theta}}$ denotes the restricted maximum likelihood estimator of $\boldsymbol{\theta}$, which is obtained by maximizing $\ell(\boldsymbol{\theta})$ subject to

$\boldsymbol{\psi} = \boldsymbol{\psi}_0$, i.e., imposing the null hypothesis. Under standard regularity conditions, the asymptotic distribution of W under $\mathcal{H}_0$ is χ^2_r. Therefore, the null hypothesis is rejected at significance level γ $(0 < \gamma < 1)$ if $W > \chi^2_r(\gamma)$.

It is possible to define a global goodness-of-fit measure. Following Nagelkerke (1991), we define the following pseudo-R^2:

$$R^2_N = \frac{1 - \{L(0)/L(\hat{\boldsymbol{\theta}})\}^{2/n}}{1 - L(0)^{2/n}}.$$

Here, $L(\hat{\boldsymbol{\theta}})$ is the full model likelihood function and $L(0)$ is the likelihood function obtained only using the intercept in the linear predictor. Such a measure assumes values in $[0, 1]$. The closer R^2_N is to one, the better the model fit.

We note that the proposed regression model belongs to the class of Generalized Additive Models for Location, Scale and Shape (GAMLSS) Models. For details on GAMLSS modeling, see Stasinopoulos et al. (2017) and `http://www.gamlss.org`. The *gamlss* package developed for the R statistical computing environment (R Core Team, 2016) can be used to fit the model.

In the next section we shall develop diagnostic analysis tools for the proposed regression model.

4 Diagnostic methods

Diagnostic analysis tools allow practitioners to verify whether a fitted regression model represents well the data at hand. In particular, such tools can be used to verify whether the model assumptions are satisfied and also whether parameter estimation is considerably affected by a few atypical observations. In what follows we shall develop some diagnostic analysis tools for the GBS2RM model.

4.1 Residual analysis

We propose two different residuals for the GBS2RM model. They can both be easily computed. The first residual we introduce is based on the stochastic relationship between the normal and $\mathcal{SHN}$ distributions, namely: if $Y \sim \mathcal{SHN}(\alpha, \mu, \nu^{-1})$, then $Z = 2\alpha^{-1}\sinh[\nu(Y - \mu)]$ follows the standard normal distribution. Let $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_n)^\top$ be the estimated linear predictor. The first residual is given by

$$r_{SHN_i} = 2\sinh[\hat{\nu}(y_i - \hat{\mu}_i)]/\hat{\alpha} = \hat{\xi}_{i2}, \quad i = 1, \ldots, n.$$

It is standard normally distributed if the model's distributional assumptions are correct.

The second proposed residual is the generalized Cox-Snell residual, which in the GBS2RM is given by

$$r_{CSG_i} = -\log\left(1 - F_Y(y_i|\hat{\boldsymbol{\theta}})\right) = -\log\left(1 - \Phi(\hat{\xi}_{i2})\right).$$

This residual is exponentially distributed with unit mean if the model's distributional assumptions hold true.

A common practice is to use the idea outlined by Atkinson (1985) when performing a residual analysis. He suggested constructing confidence bands for quantile-quantile (QQ) plots of the residuals. That can be easily done in the context of the GBS2RM model, as we explain in the supplementary material. One can then compute confidence bands for the residuals r_{SHN} and r_{CSG} and use them to check whether the fitted model represents well the data.

4.2 Local influence

Practitioners are oftentimes interested in measuring the impact of different observations on the resulting parameter estimates. That can be accomplished using the local influence method proposed by Cook (1986), which is based on the likelihood displacement $LD(\boldsymbol{\omega}) = 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega})]$ as a function of a given perturbation vector $\boldsymbol{\omega} \in \Omega$. Here, $\boldsymbol{\omega}$ is a $q \times 1$ vector, Ω is an open subset of $\mathbb{R}^q$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ is the model parameter vector. The no perturbation vector $\boldsymbol{\omega}_0$ yields the minimal likelihood displacement and is such that $\ell(\hat{\boldsymbol{\theta}}) = \ell(\hat{\boldsymbol{\theta}}|\boldsymbol{\omega}_0)$. The interest lies in evaluating the behavior of $LD(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$ by analyzing the normal curvature of the plot of $LD(\boldsymbol{\omega}_0 + a\mathbf{l})$ against a, where $a \in \mathbb{R}$ and $\mathbf{l}$ is the unit norm direction. Such a curvature is given by $C_{\mathbf{l}}(\hat{\boldsymbol{\theta}}) = 2|\mathbf{l}^\top \Delta^\top \ell_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} \Delta \mathbf{l}|$, where Δ is the perturbation matrix, whose (i, j) element is

$$\Delta_{ij} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\theta_i \partial\omega_j}, \quad i = 1, \ldots, p \text{ and } j = 1, \ldots, q.$$

This matrix is evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

The main interest lies in the maximal curvature, C_{max}, which is given by the largest eigenvalue of $B = \Delta^\top \ell_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} \Delta$. The direction $\mathbf{l}_{max}$ is the eigenvector of B corresponding to C_{max}. The index plot of $\mathbf{l}_{max}$ may reveal which data points lead to the largest changes in $LD(\boldsymbol{\omega})$. It can thus be used to detect influential observations.

Consider the partition of the parameter vector as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^\top$ and suppose the interest lies in evaluating the influence on $\boldsymbol{\theta}_1$. The normal curvature is $C_{\mathbf{l}}(\hat{\boldsymbol{\theta}}_1) = 2|\mathbf{l}^\top \Delta^\top (\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} - B_1)\Delta \mathbf{l}|$, where

$$B_1 = \begin{bmatrix} 0 & 0 \\ 0 & \ell_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}^{-1} \end{bmatrix},$$

with $\ell_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2} = \partial^2 \ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_2\boldsymbol{\theta}_2^\top$. Hence, an analysis of influence can be based on the index plot of the eigenvector corresponding to the largest eigenvalue of $\Delta^\top(\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} - B_1)\Delta$. Similarly, when the interest lies in $\boldsymbol{\theta}_2$, the normal curvature is given by $C_{\mathbf{l}}(\hat{\boldsymbol{\theta}}_2) = 2|\mathbf{l}^\top \Delta^\top (\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} - B_2)\Delta \mathbf{l}|$, where

$$B_2 = \begin{bmatrix} \ell_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Here, $\ell_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} = \partial^2 \ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_1\boldsymbol{\theta}_1^\top$. Again, the analysis is based on the index plot of the eigenvector corresponding to the largest eigenvalue of $\Delta^\top(\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} - B_2)\Delta$.

We shall consider three different perturbation schemes for local influence analysis in the GBS2RM, namely: case-weights perturbation, response variable perturbation, and explanatory variables perturbation. In the following subsections, we provide closed-form expressions for the perturbation matrix in such perturbation schemes.

4.2.1 *Case-weights perturbation* In this scheme, the weight ω_i represents the contribution of y_i to the log-likelihood function, $i = 1, \ldots, n$. The perturbed log-likelihood function is thus given by

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i \ell_i(\boldsymbol{\theta}|\boldsymbol{\omega}),$$

where $\ell_i(\boldsymbol{\theta}|\boldsymbol{\omega}) = \log(2/\sqrt{2\pi}) + \log(\xi_{i1}) - \xi_{i2}^2/2$. The no perturbation vector is $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. After some algebra, we obtained the components of the perturbation matrix, which are $\Delta_{\boldsymbol{\beta}} = X^\top D(\mathbf{a})$, $\Delta_{\alpha i} = (\xi_{i2}^2 - 1)/\alpha$ and $\Delta_{\nu i} = \nu^{-1} + \xi_{i2}\nu(y_i - \mu_i)/(2\xi_{i1}) - 2\xi_{i2}\xi_{i1}(y_i - \mu_i)/\nu$, where $\mathbf{a}$ is as before. Let $\Delta_\alpha = (\Delta_{\alpha 1} \cdots \Delta_{\alpha n})$ and $\Delta_\nu = (\Delta_{\nu 1} \cdots \Delta_{\nu n})$ be row vectors. The perturbation matrix can be conveniently expressed as $\Delta = (\Delta_{\boldsymbol{\beta}}^\top, \Delta_\alpha^\top, \Delta_\nu^\top)^\top$, a matrix of dimension $(p+2) \times n$.

4.2.2 *Response variable perturbation* In this perturbation scheme, a modified response variable of the form $y_{i\omega} = y_i + \omega_i S_y$ is considered, $i = 1, \ldots, n$, where ω_i is the ith component of the perturbation vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$

and S_y is a scaling factor, usually taken to be the standard deviation of $\mathbf{y} = (y_1, \ldots, y_n)^\top$. The no perturbation vector is $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$. After some algebra, we obtained $\Delta_{\boldsymbol{\beta}} = S_y X^\top D(\mathbf{m})$, $\Delta_{\alpha i} = 4\xi_{i1}\xi_{i2}S_y/\alpha$ and

$$\Delta_{\nu i} = \frac{\nu S_y(y_i - \mu_i)}{\cosh^2[\nu(y_i - \mu_i)]} + S_y \tanh[\nu(y_i - \mu_i)] - \frac{4\nu S_y}{\alpha^2}(y_i - \mu_i)\cosh[2\nu(y_i - \mu_i)] - \frac{4S_y}{\alpha^2}\sinh[\nu(y_i - \mu_i)]\cosh[\nu(y_i - \mu_i)],$$

where $D(\mathbf{m}) = \mathrm{diag}\{m_1, \ldots, m_n\}$, with $m_i = \nu^2\{4\alpha^{-2}\cosh[2\nu(y_i - \mu_i)] - \cosh^{-2}[\nu(y_i - \mu_i)]\}$. The perturbation matrix is given by $\Delta = (\Delta_{\boldsymbol{\beta}}^\top, \Delta_\alpha^\top, \Delta_\nu^\top)^\top$.

4.2.3 *Explanatory variables perturbation* This scheme is considered when we are interested in analyzing the impact of a perturbation on a specific explanatory variable, of index j say. The perturbation here is of the form $x_{ij\omega} = x_{ij} + \omega_i S_x$, where $j \in \{1, \ldots, p\}$, $i = 1, \ldots, n$ and S_x is a scaling factor, usually equal to the standard deviation of $(x_{1j}, \ldots, x_{nj})^\top$. The no perturbation vector is $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$. It is possible to show that the perturbation matrix is composed by $\Delta_{\boldsymbol{\beta}} = S_x\beta_j X^\top D(\mathbf{o}_1) + S_x\mathbf{q}^{(j)}\mathbf{o}_2^\top$, $\Delta_{\alpha i} = -4S_x\beta_j\xi_{i1}\xi_{i2}/\alpha$ and

$$\Delta_{\nu i} = -S_x\beta_j\left\{\frac{(y_i - \mu_i)\nu^3}{\alpha^2\xi_{i1}^2} + \frac{\nu\xi_{i2}}{2\xi_{i1}} - \frac{4\nu}{\alpha^2}(y_i - \mu_i)\cosh[2\nu(y_i - \mu_i)] - \frac{2}{\nu}\xi_{i2}\xi_{i1}\right\},$$

where $\mathbf{o}_1 = (o_{11}, \ldots, o_{1n})^\top$ and $\mathbf{o}_2 = (o_{21}, \ldots, o_{2n})^\top$, with $o_{1i} = \nu^2\cosh^{-2}[\nu(y_i - \mu_i)] - 4\nu^2\cosh[2\nu(y_i - \mu_i)]/\alpha^2$, $o_{2i} = 2\nu\sinh[2\nu(y_i - \mu_i)]/\alpha^2 - \nu\tanh[\nu(y_i - \mu_i)]$ and $\mathbf{q}^{(j)}$ is a vector with one in the jth position and zero elsewhere. The perturbation matrix is once again given by $\Delta = (\Delta_{\boldsymbol{\beta}}^\top, \Delta_\alpha^\top, \Delta_\nu^\top)^\top$.

4.3 Generalized leverage

The generalized leverage method was proposed by Wei, Hu and Fung (1998) and aims at measuring the influence of observed values on predicted values. Let $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_n)^\top$ be the vector of predicted values. The generalized leverage is given by $\partial\tilde{y}_i/\partial y_j$, i.e., it measures the change in the ith predicted value induced by the jth response value. The leverage matrix proposed by the authors is given by $GL(\theta) = D_{\boldsymbol{\theta}}(-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}})^{-1}(\ell_{\boldsymbol{\theta}\mathbf{y}})$, where $\ell_{\boldsymbol{\theta}\mathbf{y}} = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\mathbf{y}^\top$ and $D_{\boldsymbol{\theta}} = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}^\top$, $\boldsymbol{\mu}$ denoting the expected value of $\mathbf{y}$. Such a matrix is evaluated at the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ and the leverage points are those observations with large values of GL_{ii}, the ith diagonal element of $GL(\theta)$, $i = 1, \ldots, n$.

In the GBS2RM, we have that $\boldsymbol{\mu} = X\boldsymbol{\beta}$, and thus $D_{\boldsymbol{\theta}} = [X\ \mathbf{0}\ \mathbf{0}]$, an $n \times (p+2)$ matrix, where $\mathbf{0}$ denotes an n-vector of zeros. Additionally,

$$\ell_{\boldsymbol{\theta}\mathbf{y}} = (-1)\begin{bmatrix} X^\top V \\ \mathbf{d}^\top \\ \mathbf{g}^\top \end{bmatrix},$$

where the expressions of V, $\mathbf{d}$ and $\mathbf{g}$ are given in Section 3.

When we only focus on the vector $\boldsymbol{\beta}$, we obtain $D_{\boldsymbol{\beta}} = X$, $\ell_{\boldsymbol{\beta}\mathbf{y}} = -X^\top V$ and $\ell_{\boldsymbol{\beta}\boldsymbol{\beta}} = X^\top V X$. Consequently, the leverage matrix for the regression parameters is given by $GL(\boldsymbol{\beta}) = X\left(X^\top V X\right)^{-1} X^\top V$.

4.4 Generalized Cook's distance

According to Xie and Wei (2007) and Cook and Weisberg (1982), the generalized Cook's distance is given by $GD_i = (\hat{\boldsymbol{\theta}}_{(i)} - \boldsymbol{\theta})^\top M(\hat{\boldsymbol{\theta}}_{(i)} - \boldsymbol{\theta})$, where $\hat{\boldsymbol{\theta}}_{(i)}$ denotes the estimate of $\boldsymbol{\theta}$ obtained after excluding the ith observation from the sample and M is a nonnegative definite matrix, usually taken to be $M = -\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}$, the observed information matrix.

The computation of $\hat{\boldsymbol{\theta}}_{(i)}$ may be cumbersome when the sample size is large. An alternative is to use the one-step approximation to $\hat{\boldsymbol{\theta}}_{(i)}$ as proposed by Xie and Wei (2007) for the log-linear Birnbaum-Saunders model. The one-step approximation is given by $\hat{\boldsymbol{\theta}}^1_{(i)} = \hat{\boldsymbol{\theta}} + \{-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}\}^{-1}\ell_{\boldsymbol{\theta}(i)}$, where $\ell_{\boldsymbol{\theta}(i)} = \partial\ell_{(i)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, with $\ell_{(i)}$ denoting the log-likelihood function of model without the ith observation. The terms on the right hand side of the equality are evaluated at the maximum likelihood estimates. The index 1 in $\hat{\boldsymbol{\theta}}^1_{(i)}$ indicates that a one-step approximation is used.

Using the fact that $\ell_{\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0$, we obtained, after some algebra, the following components of $\ell_{\boldsymbol{\theta}(i)}$:

$$\begin{aligned}
\frac{\partial\ell_{(i)}(\boldsymbol{\theta})}{\partial\beta_l}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= x_{il}\left\{\frac{\nu^2}{2}\frac{\xi_{i2}}{\xi_{i1}} - 2\xi_{i2}\xi_{i1}\right\}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad l = 1,\ldots,p,\\
\frac{\partial\ell_{(i)}(\boldsymbol{\theta})}{\partial\alpha}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \left\{\frac{1-\xi_{i2}^2}{\alpha}\right\}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},\\
\frac{\partial\ell_{(i)}(\boldsymbol{\theta})}{\partial\nu}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \left\{\xi_{i2}\xi_{i1}\frac{2(y_i-\mu_i)}{\nu} - \frac{1}{\nu} - \frac{\xi_{i2}}{\xi_{i1}}\frac{\nu}{2}(y_i-\mu_i)\right\}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.
\end{aligned}$$

Since $\hat{\boldsymbol{\theta}}^1_{(i)} - \hat{\boldsymbol{\theta}} = \{-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}\}^{-1}\ell_{\boldsymbol{\theta}(i)}$, when $M = \{-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}\}$ it follows that the generalized Cook's distance can be approximated by

$$GD^1_i = \ell^\top_{\boldsymbol{\theta}(i)}\{-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}\}^{-1}\ell_{\boldsymbol{\theta}(i)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

In order to evaluate the impact of the ith observation on $\boldsymbol{\beta}$, α or ν, we approximate GD_i in each of these cases in the following manner, respectively:

$$GD_i^1(\boldsymbol{\beta}) = \ell_{\boldsymbol{\beta}(i)}^\top \{[-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}]^{-1}\}^{\boldsymbol{\beta}\boldsymbol{\beta}} \ell_{\boldsymbol{\beta}(i)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$
$$GD_i^1(\alpha) = \ell_{\alpha(i)}^\top \{[-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}]^{-1}\}^{\alpha\alpha} \ell_{\alpha(i)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$
$$GD_i^1(\nu) = \ell_{\nu(i)}^\top \{[-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}]^{-1}\}^{\nu\nu} \ell_{\nu(i)}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

where $\{\cdot\}^{\boldsymbol{\theta}_j\boldsymbol{\theta}_j}$ represents the diagonal block corresponding to $\boldsymbol{\theta}_j$ in the matrix.

4.5 A misspecification test

A key assumption of the GBS2RM is that the variable $\mathbf{y}$ is linearly related to the vector of regression parameters $\boldsymbol{\beta}$, which may not hold true in some applications. Other misspecifications may take place, such as the omission of an important covariate or of interactions between covariates. Therefore, it is important to test whether the functional form of a fitted GBS2RM is adequate. In short, we wish to test whether the GBS2RM is misspecified.

The effect of model misspecification on the residuals of classic linear regression models was investigated by Ramsey (1969), who proposed the RESET test (Regression Specification Error Test) that can be used to determine whether a given classic linear model is correctly specified; see also Ramsey and Gilbert (1972). A RESET-type test for the log-linear Birnbaum-Saunders model was considered by Santos and Cribari-Neto (2017), who investigated the test size distortions in small samples and its power under different types of misspecification. We now outline the RESET misspecification test for the GBS2RM:

1. Estimate the parameters of the GBS2RM
$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \epsilon_i, \quad i = 1, \ldots, n,$$
and obtain the predicted values $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_n)^\top$.
2. Estimate the parameters of the augmented GBS2RM given by
$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \gamma_1\hat{\mu}_i^2 + \cdots + \gamma_{k-1}\hat{\mu}_i^k + \varepsilon_i, \quad i = 1, \ldots, n,$$
where $k \geq 2$.
3. Test $\mathcal{H}_0 : \gamma_1 = \cdots = \gamma_{k-1} = 0$ (correct model specification) against $\mathcal{H}_1$ that $\gamma_j \neq 0$ for at least one $j \in \{1, \ldots, k-1\}$ (model misspecification).
4. If the null hypothesis is rejected, reject the model under evaluation, i.e., conclude that there is evidence of model misspecification.

5 Prediction intervals

In this section we address the issue of obtaining prediction intervals for a non-observed response value. We shall use an approach similar to that of Stine (1985) for linear models. Stine's proposal involves the use bootstrap resampling to estimate the prediction error distribution, which is then used to obtain the prediction intervals. In similar fashion, Davison and Hinkley (1997) provide an algorithm to compute bootstrap prediction intervals for generalized linear models, which was recently extended by Espinheira, Ferrari and Cribari-Neto (2014) for beta regression models. In fact, building upon the work of Mojirsheibani and Tibshirani (1996) on confidence intervals for parameters based on future samples, Espinheira, Ferrari and Cribari-Neto (2014) proposed a method that can be used to compute BC_a (bias-corrected and accelerated) prediction intervals, which we shall now apply to the GBS2RM.

Consider a sample $y_1, \ldots, y_n$ of the response variable and let X be the corresponding matrix of covariates. We wish to obtain a prediction interval for a non-observed response value y_+ based on a new observation of the covariates, denoted by $\mathbf{x}_+$. We consider a prediction error function $\mathcal{R}(y;\mu)$, which is a monotonic function of y, has constant variance and whose ϱth quantile is denoted by δ_ϱ. Here, μ denotes the mean of y. The lower and upper prediction limits of a $1-\varrho$ prediction interval for y_+ are, respectively, $y_{+,\varrho/2}$ and $y_{+,1-\varrho/2}$, such that $\mathcal{R}(y_{+,\varrho/2};\mu) = \delta_{\varrho/2}$ and $\mathcal{R}(y_{+,1-\varrho/2};\mu) = \delta_{1-\varrho/2}$. Since the distribution of $\mathcal{R}(y_+;\mu)$ is usually unknown, we make use of resampling methods to estimate it. We follow Espinheira, Ferrari and Cribari-Neto (2014) and use $\tilde{\varrho} = \Phi\left(\hat{z}_0 + [\hat{z}_0 + z_\varrho]/[1-\hat{a}(\hat{z}_0 + z_\varrho)]\right)$, where $\hat{z}_0$ is a bias correction constant, $\hat{a}$ is a factor known as acceleration constant and z_ϱ is the ϱth standard normal quantile. The following estimates of z_0 and a are used for constructing BC_a prediction intervals for new response observations:

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\mathcal{R}_+^* < \mathcal{R}_m\}}{B}\right) \quad \text{and} \quad a = \frac{\sqrt{n}\,\mathbb{E}(\dot{\ell}_+^3)}{6\mathrm{Var}(\dot{\ell}_+)^{3/2}},$$

where $\mathcal{R}_m$ is the median of the prediction errors $\mathcal{R}_1, \ldots, \mathcal{R}_n$ of the fitted model, $\mathcal{R}_+^*$ denotes the bootstrap estimate of the prediction error for the non-observed response value, B is the number of bootstrap replications used and $\dot{\ell} = \partial \log f_Y(y_+|\boldsymbol{\theta})/\partial\mu$.

In the GBS2RM, $a = 0$. To see that notice that $\dot{\ell} = 2\xi_1\xi_2 - \nu^2\xi_2/(2\xi_1)$, where $\xi_1 = \nu\cosh[\nu(y-\mu)]/\alpha$ and $\xi_2 = 2\sinh\nu(y-\mu)]/\alpha$. Therefore, the

numerator of a is proportional to

$$\mathbb{E}(\dot{\ell}^3) = -\frac{\nu^6}{8}\mathbb{E}\left(\frac{\xi_2^3}{\xi_1^3}\right) + \frac{3\nu^4}{2}\mathbb{E}\left(\frac{\xi_2^3}{\xi_1}\right) - 6\nu^2\mathbb{E}\left(\xi_1\xi_2^3\right) + 8\nu^2\mathbb{E}\left(\xi_1^3\xi_2^3\right).$$

Using the fact that $\xi_1 = \nu\sqrt{1+(\alpha\xi_2/2)^2}/2$ and that expected values of odd functions of a standard normal random variable equal zero, it is possible to show that $\mathbb{E}(\dot{\ell}^3) = 0$ and, as a consequence, $a = 0$. Hence, in the GBS2RM the BC_a method (Efron, 1987) reduces to the BC (bias-corrected) method proposed by Efron (1981), where the estimate of ϱ turns out to be $\hat{\varrho} = \Phi\left(2\hat{z}_0 + z_\varrho\right)$.

We consider the r_{SHN} residual as the prediction error function for the GBS2RM, i.e., $\mathcal{R}_i = r_{SHN_i}$, $i = 1, \ldots, n$. Moreover, we notice that such a prediction error function is monotonic in y and has constant variance, since the reference distribution of the r_{SHN} residual is standard normal. Hence, an algorithm to compute prediction intervals similar to that proposed by Espinheira, Ferrari and Cribari-Neto (2014) for beta regressions can be used with the GBS2RM. Such an algorithm is presented in the paper supplementary material.

6 Model selection criteria

We shall now turn to model selection, which is of paramount importance in regression analysis. Model selection is oftentimes based on model selection criteria, i.e., on criteria that seek to identity the best fitting model. The most commonly used model selection criteria are the Akaike Information Criterion (AIC), proposed by Akaike (1973), and the Schwarz Information Criterion (SIC), introduced by Schwarz (1978). For a detailed account of the different model selection criteria and their use in regression modeling, readers are referred to McQuarrie and Tsai (1998). Our goal in what follows is to present some model selection criteria that can be used with the GBS2RM. We follow Bayer and Cribari-Neto (2015), who considered model selection in the class of varying dispersion beta regressions.

Several model selection criteria were developed as extensions of previously existing criteria, such as the AIC. Akaike (1973) derived the AIC as an estimator of $\Delta(\theta_0, k) = \mathbb{E}_0\left[-2\log f(\mathbf{y}|\boldsymbol{\theta})\right]|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, a measure of the discrepancy between the true model $f(\mathbf{y}|\boldsymbol{\theta}_0)$ and a fitted candidate model $f(\mathbf{y}|\hat{\boldsymbol{\theta}})$, where $\mathbb{E}_0$ indicates that the expectation is computed with respect to the true model and k is the dimension of $\boldsymbol{\theta}$. The AIC is given by $-2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}) + 2k$, the term $-2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}})$ being Akaike's estimator of $\Delta(\theta_0, k)$ and $2k$ being an asymptotic bias correction. The AIC is the most commonly used model

selection criterion. Nonetheless, it may perform poorly in small samples, as pointed out by Hurvich and Tsai (1989). This occurs because the AIC becomes progressively more negatively biased as k becomes larger relative to n, which leads the AIC to often select over-specified models. As a consequence, bias-adjusted variants of the AIC have been proposed and investigated in the literature. They are typically obtained by deriving bias corrections that are superior to $2k$, such as in the AIC_c of Sugiura (1978) and Hurvich and Tsai (1989).

Some authors suggested using bootstrap resampling to estimate the bias of $-2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}})$; see, e.g., Ishiguro and Sakamoto (1991), Cavanaugh and Shumway (1997) and Shibata (1997). In what follows, we shall use the notation of Bayer and Cribari-Neto (2015) to denote some of these model selection criteria that are bootstrap-based. The criteria EIC_1 and EIC_2 were proposed by Ishiguro and Sakamoto (1991) and by Cavanaugh and Shumway (1997), respectively. Shibata (1997) showed that the EIC_1 and EIC_2 are asymptotically equivalent and proposed the EIC_3, EIC_4 and EIC_5 criteria.

The five bootstrap-based information criteria mentioned above differ in the way they estimate the bias term, the goodness of fit factor $-2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}})$ being the same for all of them. Following Pan (1999), Bayer and Cribari-Neto (2015) proposed a bootstrap model selection criterion that focuses on the goodness of fit term rather than on the bias, which the authors called BQCV (bootstrap quasi cross-validation) criterion. The BQCV criterion estimates the discrepancy between the true model and a candidate model directly, but can overestimate $\Delta(\theta_0, k)$. Hence, the authors proposed using the criterion $\mathrm{632QCV} = 0.368 \times [-2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}})] + 0.632 \times \mathrm{BQCV}$, which follows from the results in Pan (1999) and from the fact that the terms $-2\log f(\mathbf{y}|\hat{\boldsymbol{\theta}})$ and BQCV underestimate and overestimate $\Delta(\theta_0, k)$, respectively. Thus, 632QCV aims at balancing these two terms.

Several other (not AIC-based) criteria were proposed in the literature, such as the aforementioned SIC, the SIC_c derived by McQuarrie (1999), the HQ proposed by Hannan and Quinn (1979) and its corrected version HQ_c presented by McQuarrie and Tsai (1998) for the normal linear model. The finite sample performances of the criteria listed in this section were investigated in the literature for a wide variety of models, such as regression, time series and nonparametric models; for details, see McQuarrie and Tsai (1998). In the next section we shall numerically evaluate how well model selection schemes based on model selection criteria work in the log-linear $\mathcal{GBS}_2$ model.

7 Numerical evaluation

In this section we shall present results from Monte Carlo simulation studies that were conducted to numerically evaluate the finite performances of the maximum likelihood estimators of the parameters that index the GBS2RM model and the quality of the approximation of the r_{SHN} and r_{CSG} residuals distributions by the respective reference distributions. Additionally, we shall numerically evaluate the finite sample performance of the RESET-type test for misspecification of the GBS2RM model, prediction intervals for non-observed response values and different model selection schemes. The simulations were performed using the Ox matrix programming language (Doornik, 2009). Log-likelihood maximizations were carried out using the BFGS quasi-Newton nonlinear optimization algorithm with analytical first derivatives. Using such an algorithm convergence when performing log-likelihood maximizations was reached in most samples, the convergence failure rate being around 7.5%.

7.1 Maximum likelihood estimation

The parameter vector is $\boldsymbol{\theta} = (\beta_1, \beta_2, \alpha, \nu)^\top$ and the regression model is

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i, \quad i = 1, \ldots, n, \tag{2}$$

where $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{SHN}(\alpha, 0, \nu^{-1})$. The parameter values are $\beta_1 = 1$, $\beta_2 = 0.5$, $\alpha = 1$ and $\nu = 1$ and the starting points for the log-likelihood maximization were obtained using the proposal described in Section 3.

The covariates values were randomly generated from three different distributions: uniform $\mathcal{U}(0, 1)$, exponential with unit mean, and standard normal. We shall denote such data generation schemes by $E1$, $E2$ and $E3$, respectively. The values of all covariates were kept constant during the simulations.

The results we report are based on 10,000 Monte Carlo replications, the sample sizes being $n \in \{30, 60, 90\}$. We computed the mean squared error (MSE) of the MLE of each parameter and also the corresponding relative biases: $[\mathbb{E}(\hat{\boldsymbol{\theta}}_j) - \boldsymbol{\theta}_j]/\boldsymbol{\theta}_j$, $j = 1, \ldots, 4$. The results for the different covariates generation schemes are presented in Table 1. We note that the MSEs tend to decrease as larger sample sizes are used, as expected. The estimates of the regression parameters β_1 and β_2 are more accurate than those of α and ν. It is noteworthy that point estimates for the shape parameters have considerable bias, especially in small samples. Nevertheless, it is possible to obtain less biased point estimates by using a bootstrap bias correction (Davison and Hinkley, 1997, p. 103). We carried out a separate simulation experiment to evaluate the effectiveness of such a correction. The results can be found

in the paper supplementary material. They show that the bootstrap bias correction can be quite effective. We also note that the results for the three covariates generation schemes are similar.

Table 1 *Relative biases (RB) and mean squared errors (MSE), Model (2).*

Generation	n	Measure	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}$	$\hat{\nu}$
$E1$	30	RB	0.0003	−0.0052	1.5198	0.9057
		MSE	0.1711	0.3156	2.3377	1.2659
	60	RB	0.0019	−0.0077	0.6301	0.4362
		MSE	0.1093	0.2078	1.0530	0.7104
	90	RB	0.0017	−0.0061	0.3922	0.2829
		MSE	0.0934	0.1650	0.7314	0.5285
$E2$	30	RB	−0.0016	0.0020	1.4943	0.8875
		MSE	0.1354	0.0999	2.3341	1.2546
	60	RB	−0.0006	0.0015	0.6219	0.4318
		MSE	0.0900	0.0567	1.0415	0.7050
	90	RB	−0.0003	0.0014	0.3859	0.2789
		MSE	0.0705	0.0498	0.7232	0.5242
$E3$	30	RB	−0.0004	0.0004	1.5218	0.8988
		MSE	0.0875	0.0926	2.3787	1.2661
	60	RB	0.0002	−0.0009	0.6344	0.4387
		MSE	0.0606	0.0586	1.0576	0.7101
	90	RB	0.0002	−0.0002	0.3919	0.2836
		MSE	0.0484	0.0458	0.7236	0.5234

7.2 Empirical distribution of the residuals

The second simulation study was performed to evaluate how well the distributions of the residuals r_{SHN} and r_{CSG} are approximated by the corresponding reference distributions. The results are for Model (2) and the sample size is $n = 60$. The covariates values are obtained as in the previous simulation. Based on all 10,000 replications, we computed the mean, standard deviation, asymmetry and kurtosis of each residual. For r_{SHN}, whose distribution is expected to be approximately standard normal, we expect such statistics to be close to 0, 1, 0 and 3, respectively. For r_{CSG}, whose reference distribution is exponential with unit mean, such statistics are expected to be close to 1, 1, 2 and 6, respectively. The results are presented in Table 2. They show that the distribution of r_{SHN} is better approximated by its reference distribution than that of r_{CSG}. Nonetheless, the results for the residual r_{CSG} were also satisfactory, the means and standard deviations being quite close to one. It is then possible to conclude that the distributions of the proposed residuals are well approximated by the respective reference

distributions. Practitioners can then compare the quantiles of residuals obtained from a fitted GBS2RM with those of the corresponding reference distributions.

Table 2 *Means, standard deviations (SD), asymmetries and kurtosis of the residuals r_{SHN} and r_{CSG}, Model (2).*

Generation	Residual	Mean	SD	Asymmetry	Kurtosis
$E1$	r_{SHN}	-2.11×10^{-5}	0.9989	0.0003	2.9700
	r_{CSG}	1.0000	0.9928	1.7866	6.7611
$E2$	r_{SHN}	-8.85×10^{-5}	0.9989	0.0007	2.9720
	r_{CSG}	0.9999	0.9931	1.7905	6.7936
$E3$	r_{SHN}	0.0003	0.9988	0.0001	2.9712
	r_{CSG}	1.0003	0.9929	1.7871	6.7705

7.3 RESET-type misspecification test

Next, we performed a set of Monte Carlo simulations to evaluate the finite sample performance of the RESET-type misspecification test. Since the previous results were similar for the different schemes of covariates values generation, we shall only report results obtained using standard uniform draws. The number of Monte Carlo replications is $10,000$ and the sample sizes are $n \in \{30, 60, 90\}$. The testing variable is the vector of squared predicted values and the test was performed using the likelihood ratio test criterion. The significance levels are 10%, 5% and 1%.

The first simulation study was carried out to compute the RESET-type test null rejection rates. In each Monte Carlo replication we generated $y_i = 1 - 0.5x_{2i} + 1.3x_{3i} + \epsilon_i$, where $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{SHN}(1, 0, 1)$, $i = 1, \ldots, n$, fitted the GBS2RM and obtained the predicted values. We then fitted the model $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \gamma \hat{\mu}_i^2 + \epsilon_i$ and tested the null hypothesis $\mathcal{H}_0 : \gamma = 0$ against a two-sided alternative hypothesis. The test null rejection rates are presented in Table 3. We note that the test is considerably size-distorted when the sample size is small ($n = 30$); such distortions become much smaller when the sample size increases to $n = 60$ and $n = 90$. With 90 observations, the test null rejection rate at the 5% nominal level is 6.9%.

Table 3 *Null rejection rates of the RESET-type test for the GBS2RM.*

n	Significance level		
	10%	5%	1%
30	0.2107	0.1367	0.0437
60	0.1388	0.0808	0.0218
90	0.1268	0.0690	0.0170

We also performed simulations aimed at evaluating the power of the RESET-type test under different sources of model misspecification. We consider four different data generation processes (schemes), denoted by $P1$, $P2$, $P3$ and $P4$. In scheme $P1$, $y_i = 1 + 0.5x_{2i} + 1.8x_{3i} + 1.8x_{2i} \times x_{3i} + \epsilon_i$; notice the interaction between the two covariates. In scheme $P2$, $y_i = 1 + 0.5x_{2i} + 1.8x_{3i} + 1.8x_{i2}^2 + \epsilon_i$; notice that the linear predictor includes a squared regressor. In scheme $P3$, $y_i = 1 + 0.5x_{2i} + 1.8x_{3i} + 1.5x_{4i} + \epsilon_i$; notice the covariate $\mathbf{x}_4$. In scheme $P4$, the data generating process is nonlinear: $y_i = (1 + 0.5x_{2i} + 1.8x_{3i})^{\varphi} + \epsilon_i$, $\varphi \in \mathbb{R}$. In the simulations, we used $\varphi = 1.7$. In all four schemes, $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{SHN}(1, 0, 1)$, $i = 1, \ldots, n$, and the following (misspecified) model was estimated: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$. Since the test is liberal, testing inference was based on exact critical values which were estimated in the previous simulations (size simulations). We thus compute the power of a size-corrected test. The test nonnull rejection rates are presented in Table 4. We note that the RESET-type test for the GBS2RM displays good power, with nonnull rejection rates increasing when larger sample sizes are considered. The test was more powerful under scheme $P4$, where linearity is incorrectly assumed.

Table 4 *Nonnull rejection rates of the RESET-type test for the GBS2RM model under different schemes of model misspecification.*

Scheme	n	Significance level		
		10%	5%	1%
$P1$	30	0.2456	0.1573	0.0496
	60	0.4778	0.3405	0.1407
	90	0.6471	0.5183	0.2856
$P2$	30	0.1845	0.1103	0.0316
	60	0.3226	0.2083	0.0718
	90	0.4428	0.3234	0.1522
$P3$	30	0.2090	0.1243	0.0354
	60	0.5286	0.3635	0.1276
	90	0.7724	0.6314	0.3428
$P4$	30	0.3601	0.2470	0.0959
	60	0.6571	0.5212	0.2631
	90	0.8352	0.7320	0.5104

7.4 Prediction intervals

A simulation study was performed to evaluate the performances of the percentile and BC_a prediction intervals for a non-observed occurrence y_+ of the response variable given a new observation of the covariate in a GBS2RM model. The data generating process is $y_i = 1 + 0.5x_{2i} + \epsilon_i$, where $\epsilon_i \overset{\text{iid}}{\sim}$

$\mathcal{SHN}(\alpha, 0, 1/\nu)$, $i = 1, \ldots, n$. The results are presented in Table 5. The sample sizes considered were as in the previous simulation studies, the number of Monte Carlo replications used was $5,000$ and the covariate values were obtained as standard uniform random draws. We used $B = 1,000$ bootstrap replications for constructing 95% confidence prediction intervals. In Table 5 we present the empirical coverages of the percentile and BC_a prediction intervals and the proportions of the replications for which y_+ was smaller (larger) than the lower (upper) prediction intervals. The figures in Table 5 show that both prediction intervals display empirical coverages close to 95% and that the non-coverages are well distributed in both sides. The percentile intervals perform slightly better than the BC_a intervals. Additionally, the performances of both intervals improve as the sample size increases, as expected. It is thus noteworthy that bootstrap prediction intervals perform well when used with the GBS2RM, especially the percentile method.

Table 5 *Empirical coverages and left and right non-coverages of* 95% *prediction intervals in a GBS2RM for different values of* α *and* ν.

n	Percentile			BC_a		
	Left	Coverage	Right	Left	Coverage	Right
			$\boldsymbol{\theta} = (1, 0.5, 1, 1)^\top$			
30	0.027	0.942	0.031	0.029	0.939	0.032
60	0.026	0.946	0.028	0.025	0.945	0.030
90	0.026	0.947	0.027	0.027	0.945	0.028
			$\boldsymbol{\theta} = (1, 0.5, 0.5, 1.5)^\top$			
30	0.028	0.941	0.031	0.029	0.937	0.034
60	0.025	0.950	0.025	0.025	0.948	0.027
90	0.024	0.950	0.026	0.025	0.950	0.025
			$\boldsymbol{\theta} = (1, 0.5, 1.5, 0.5)^\top$			
30	0.029	0.945	0.026	0.029	0.941	0.030
60	0.026	0.949	0.025	0.026	0.949	0.025
90	0.028	0.949	0.023	0.028	0.949	0.023

7.5 Model selection criteria

We performed a simulation study to evaluate the finite sample performances of the model selection criteria presented in Section 6. The simulation is similar to that conducted by Hurvich and Tsai (1989) for linear models. The true model is $y_i = -1 + x_{2i} + x_{3i} + x_{4i} + \epsilon_i$, where $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{SHN}(1, 0, 1)$, $i = 1, \ldots, n$, and the true parameter vector $\boldsymbol{\theta}_0$ has dimension $k_0 = 6$. We consider the $n \times 6$ matrix of candidate regressors $\tilde{X}$, whose first column is a vector of ones. The first four columns of such a matrix contain the regressors that are present in the true data generating process. The models were fitted in nested

fashion using as covariates the first $2, 3, \ldots, 6$ columns of $\tilde{X}$. The number of Monte Carlo and bootstrap replications are as before, as well as the sample sizes considered. All covariate values in $\tilde{X}$ were obtained as random standard uniform draws. The bootstrap extensions of the AIC criteria as well as BQCV and 632QCV were computed using parametric resampling. Nonparametric bootstrap resampling leads to very similar results which are not presented for brevity. We computed the proportion of model under-specification, of model over-specification and of correct specification for each criterion. The results are presented in Table 6. They show that the AIC tends to select over-specified models, especially for small sample sizes. The bootstrap extensions of the AIC performed well when $n = 60$ and $n = 90$. For instance, for $n = 60$ the largest frequency of correct model selection was achieved by the EIC_1. The performances of the EIC_2 and EIC_4 were also very good for $n = 90$, with proportions of correct specification in excess of 97%. We note that BQCV and 632QCV performed well when the sample contained 60 or 90 observations, but performed poorly when $n = 30$, often selecting under-specified models. Overall and on balance, the SIC_c was the best performer. It outperformed all competing criteria when $n = 30$ and was very competitive with the bootstrap criteria when $n = 60$ and $n = 90$. Additionally, SIC_c does not entail the computational burden of performing data resampling. We thus recommend that model selection be based on such a criterion when performing GBS2RM modeling.

Table 6 *Proportions of model under-specification ($k < k_0$), correct specification ($k = k_0$) and over-specification ($k > k_0$) of a GBS2RM using the selection criteria discussed in Section 6.*

	$n = 30$			$n = 60$			$n = 90$		
	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$
AIC	0.018	0.454	0.527	0.000	0.676	0.324	0.000	0.723	0.277
AIC_c	0.052	0.658	0.290	0.000	0.762	0.238	0.000	0.775	0.225
SIC	0.052	0.625	0.323	0.001	0.887	0.112	0.000	0.929	0.071
SIC_c	0.151	0.741	0.107	0.004	0.936	0.060	0.000	0.960	0.039
HQ	0.031	0.517	0.451	0.001	0.790	0.209	0.000	0.844	0.156
HQ_c	0.083	0.705	0.212	0.001	0.866	0.133	0.000	0.890	0.110
EIC_1	0.909	0.088	0.003	0.021	0.945	0.034	0.000	0.941	0.059
EIC_2	0.953	0.045	0.002	0.099	0.887	0.014	0.000	0.978	0.022
EIC_3	0.059	0.576	0.365	0.001	0.781	0.219	0.000	0.790	0.210
EIC_4	0.954	0.044	0.002	0.099	0.888	0.012	0.000	0.976	0.024
EIC_5	0.059	0.607	0.334	0.001	0.778	0.221	0.000	0.766	0.234
BQCV	0.898	0.098	0.004	0.013	0.930	0.057	0.000	0.886	0.114
632QCV	0.830	0.161	0.009	0.002	0.890	0.107	0.000	0.776	0.224

8 Empirical application

We shall now use the GBS2RM to analyze real (not simulated) data. The analysis was carried out using the Ox matrix programming language and the R statistical computing environment (R Core Team, 2016). The computer codes we used are available upon request. The data contain information on patients who suffered from an acute type of leukemia and are provided by Feigl and Zelen (1965). The variable of interest is the patient's lifetime (in weeks) since the disease diagnosis. One of the covariates is the patient's white blood cells count upon leukemia diagnosis. The other covariate marks the presence or absence of a specific factor in the patient white cells, which can be classified as AG positive or AG negative, respectively.

Let $y_i = \log(t_i)$ be the logarithm of the ith patient lifetime, $x_{i2} = \log(W_i)$ be the logarithm of the patient's white blood cells count and x_{i3} be a dummy variable that equals one if the ith patient is AG positive and zero otherwise, $i = 1, \ldots, 33$. Figure 2 contains boxplots of $\mathbf{y}$ and $\mathbf{x}_2$ for different levels of $\mathbf{x}_3$ and a scatterplot of $\mathbf{y}$ vs. $\mathbf{x}_2$. We note from Figure 2a that lifetime tends to decrease as the white cells count increases and that AG positive patients ($x_3 = 1$) tend to live longer than AG negative patients ($x_3 = 0$), which can also be observed in Figure 2b. Figure 2c shows that there is no clear relationship between $\mathbf{x}_2$ and $\mathbf{x}_3$.

We estimated the following log-linear model $\mathcal{GBS}_2$:

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i, \tag{3}$$

where $\epsilon_i \overset{\text{iid}}{\sim} \mathcal{SHN}(\alpha, 0, \nu^{-1})$. The maximum likelihood estimates of the parameters (standard errors in parentheses) are $\hat{\beta}_1 = 6.159$ (0.8280), $\hat{\beta}_2 = -0.360$ (0.0828), $\hat{\beta}_3 = 0.055$ (0.2786), $\hat{\alpha} = 6.914$ (3.8980) and $\hat{\nu} = 1.272$ (0.2794). It is noteworthy that the standard error of β_3 is large, which might be indicative that the variable $\mathbf{x}_3$ is not relevant for the analysis at hand. Indeed, for the model in Equation (3) $\text{SIC}_c = 120.97$ whereas when the model is fitted without the covariate $\mathbf{x}_3$ $\text{SIC}_c = 115.92$. Simultaneous removal of $\mathbf{x}_2$ and of the intercept yielded a larger value of the model selection criterion. The other model selection criteria led to the same conclusions. We shall address the significance of the factor AG later in this section. The likelihood ratio test statistic for testing $\mathcal{H}_0 : \nu = 0.5$ against $\mathcal{H}_1 : \nu \neq 0.5$ equals 4.65, the corresponding p-value being 0.0309. That is, there is evidence (at the 5% significance level) that the GBS2RM is superior to the BSRM for analyzing these data.

The value of the pseudo-R^2 for the GBS2RM model is $R^2_N = 0.350$. Figure 3 contains residual (r_{SHN} and r_{CSG}) plots. In Figures 3a and 3b we

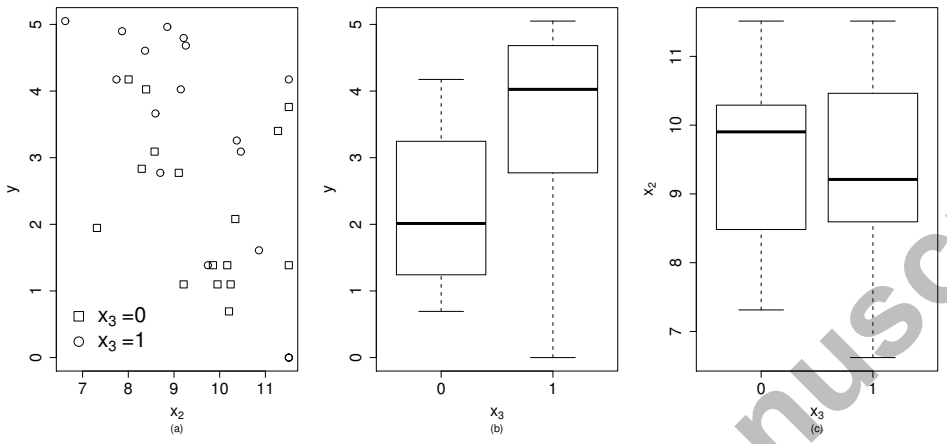

Figure 2 *Scatterplot of the response variable vs.* $\boldsymbol{x}_2$ *(a) and boxplots for different levels of* $\boldsymbol{x}_3$ *(b and c).*

see plots of residuals against predicted values $\hat{\boldsymbol{\mu}}$ with 95% confidence bands. Notice that most residuals lie inside the intervals and that there is no noticeable pattern in the residuals. Visual inspection of Figures 3c and 3d reveals that all residuals are inside the confidence regions, thus indicating that the distributional assumptions hold. We also performed the RESET-type test for model misspecification using the square of the predicted values ($\hat{\boldsymbol{\mu}}$) as testing variable. The test statistic equals 0.73, the corresponding p-value being 0.3916. Hence, there is no evidence of model misspecification at the usual significance levels.

Diagnostic plots can be found in the paper supplementary material. Local influence analysis for the GBS2RM model should be able to identify data points that might be largely influencing the parameters estimates. Figure 1 in the supplementary material contains local influence plots relative to the regression and shape parameters using the three perturbation schemes discussed earlier. It is noteworthy that the most influential data points are observations 14, 15 and 17. We computed the generalized Cook's distance for each observation. The GD measures for the model parameters are presented in Figure 2 of the supplementary material. Notice that such results agree with those obtained using local influence analysis, i.e., observations 14, 15 and 17 are singled out as atypical. Generalized leverage measures are presented in Figure 3 of the supplementary material. Observations 2 and 21 stand out. The former corresponds to the patient with the lowest white cells count among all AG positive patients, whereas the latter corresponds to the

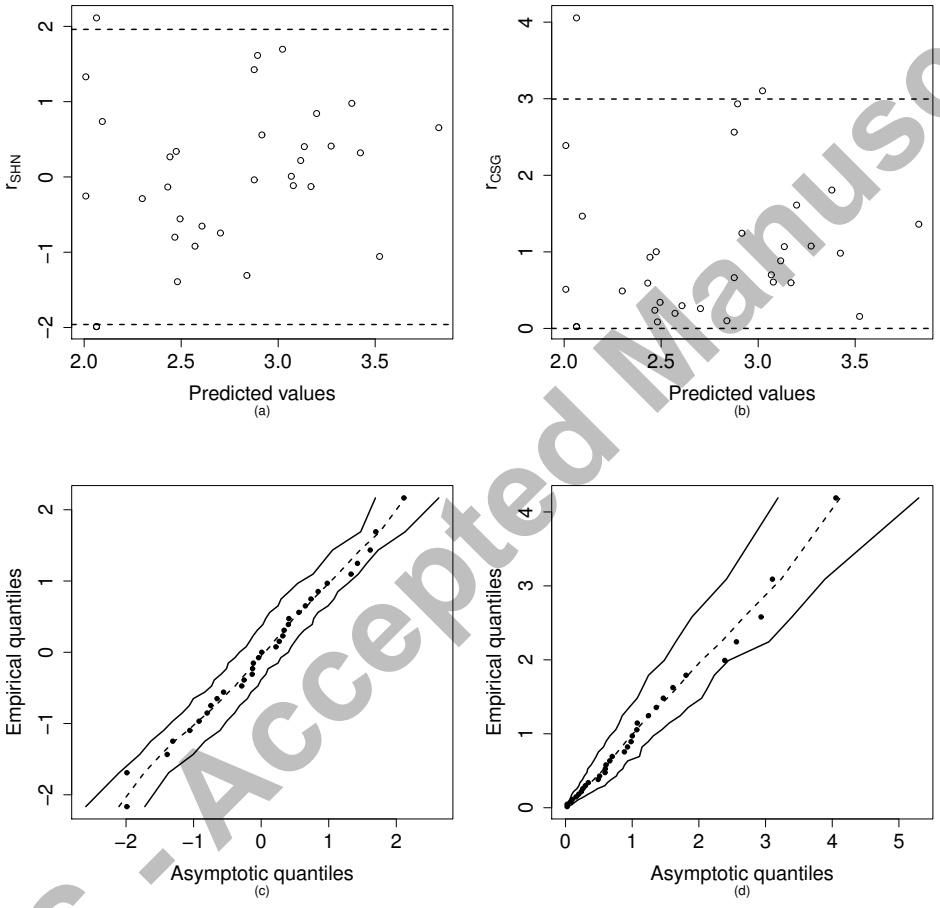

Figure 3 *Predicted values $\hat{\boldsymbol{\mu}}$ against the residuals r_{SHN} (a) and r_{CSG} (b), and simulated envelopes with bands of 95% of confidence for the residuals r_{SHN} (c) and r_{CSG} (d). The dashed lines in the panels (a) and (b) indicate approximate confidence regions (95% confidence).*

patient with the lowest white cells count of all AG negative patients.

We sequentially removed each atypical observation from the data and fitted the model after each data point removal. In each case, we computed the absolute relative change in the estimates, i.e., we computed $|(\hat{\boldsymbol{\theta}}_{j(i)} - \hat{\boldsymbol{\theta}}_j)/\hat{\boldsymbol{\theta}}_j|$, where $\hat{\boldsymbol{\theta}}_j$ represents the jth parameter estimate obtained using the complete data and $\hat{\boldsymbol{\theta}}_{j(i)}$ is the corresponding estimate obtained after the ith observation removal. Additionally, we tested the significance of each regressor when the reduced data were used. We also tested whether $\nu = 0.5$ in order to distinguish between GBS2RM and the log-linear Birnbaum-Saunders model. The main goal is to determine whether any relevant inferential decision was reversed after the atypical observations were removed from the data. The relative changes in the parameter estimates and the tests p-values are presented in Table 7.

The figures in Table 7 show that the intercept is statistically different from zero and the regressor that accounts for white cells count are statistically significantly in all scenarios at the usual significance levels. We also note that the $\mathcal{GBS}_2$-based model remains superior to that based on the $\mathcal{BS}$ law at the 10% significance level, thus strengthening the evidence in favor of the GBS2RM. The most intriguing result relates to $\mathbf{x}_3$. Such a regressor is not statistically significant at the usual significance levels when all observations are used in the model fit, as noted earlier. However, after observation 14 or observation 15 is removed from the data the estimate of β_2 changes considerably. Such observations were detected as influential data points and correspond to AG positive patients with short lifetimes but that present high counts of white blood cells, contrary to what is implied by the fitted model. When the model is fitted without these two observations in the data the covariate $\mathbf{x}_3$ becomes statistically significant. Additionally, the SIC_c for the model without $\mathbf{x}_3$ equals 104.68 whereas the model fitted with such a regressor yields $\mathrm{SIC}_c = 101.01$, i.e., the SIC_c also provides indication that the AG factor impacts the patients' lifetime. Most of the other selection criteria also lead to the same conclusion. Simultaneous removal of cases 14, 15 and 17 leads to the same testing inference, i.e., $\mathbf{x}_3$ is found to be statistically significant. Hence, observations 14 and 15 are the cases responsible for the reversal in the inference decision regarding the statistical significance of $\mathbf{x}_3$.

We decided to consider the model fitted without the atypical cases (observations 14 and 15) in the data. The parameter estimates (standard errors in parentheses) are $\hat{\beta}_1 = 4.219$ (0.7047), $\hat{\beta}_2 = -0.179$ (0.0691), $\hat{\beta}_3 = 0.643$ (0.1701), $\hat{\alpha} = 11.135$ (6.2724) and $\hat{\nu} = 1.807$ (0.3499). Also, $R^2_N = 0.4176$, which indicates that the new model fit is superior to that obtained using the complete data. By applying the bootstrap bias correction, we obtain

Table 7 *Absolute relative changes in the parameter estimates of Model (3) with tests p-values in parentheses after removal of the indicated data point(s). The null hypotheses are* $\mathcal{H}_0 : \beta_j = 0$, $j = 1, 2, 3$, *and* $\mathcal{H}_0 : \nu = 0.5$.

Deleted	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\alpha}$	$\hat{\nu}$
None	— (< 0.001)	— (< 0.001)	— (0.8428)	—	— (0.0309)
2	0.025 (< 0.001)	0.044 (< 0.001)	0.417 (0.7808)	0.053	0.023 (0.0316)
14	0.041 (< 0.001)	0.058 (0.0006)	2.503 (0.4781)	0.119	0.018 (0.0554)
15	0.041 (< 0.001)	0.058 (< 0.001)	2.503 (0.4781)	0.119	0.018 (0.0554)
17	0.100 (< 0.001)	0.178 (< 0.001)	0.842 (0.9695)	0.035	0.056 (0.0245)
21	0.022 (< 0.001)	0.037 (< 0.001)	0.187 (0.8202)	0.117	0.048 (0.0819)
14 and 15	0.315 (< 0.001)	0.504 (0.0175)	10.669 (0.0029)	0.610	0.419 (0.0014)
14, 15, 17	0.304 (< 0.001)	0.487 (0.0288)	10.587 (0.0067)	0.444	0.367 (0.0033)

the following corrected estimates: $\hat{\beta}_1^{bc} = 4.199$, $\hat{\beta}_2^{bc} = -0.177$, $\hat{\beta}_3^{bc} = 0.650$, $\hat{\alpha}_1^{bc} = 2.257$ and $\hat{\nu}_1^{bc} = 1.487$. The only estimate that changed substantially when we applied the bias correction was that of α. We note that uncorrected and corrected regression parameters estimates are similar, which is in line with our Monte Carlo simulation evidence. The correct model specification is not rejected by the RESET-type test at the usual nominal levels. Figure 4 presents prediction intervals obtained with this model for values of x_2 ranging between 6 and 12, for each level of x_3. We note that patients for whom the presence of the AG factor was detected tend to live longer than AG negative patients, thus corroborating the evidence in Figure 2. Moreover, we note that higher white blood cells counts in the patients are significantly associated with shorter lifetimes, as can be observed in Figure 2. Furthermore, nearly all points in Figure 4 lie inside the respective prediction intervals, but the observations 14 and 15 are clearly atypical data points.

9 Concluding remarks

Log-linear Birnbaum-Saunders regression models have been frequently used in the literature. The model is based on the standard Birnbaum-Saunders distribution. In this paper, we propose a log-linear model based on a bimodal version of the Birnbaum-Saunders law. The log-linear Birnbaum-Saunders regression model is a particular case of our model. Parameter estimation is carried out by maximum likelihood. We provide an expression for the observed information matrix, discuss hypothesis testing inference and present a pseudo-R^2 measure. Several different diagnostic tools for the proposed model were discussed and two different residuals were introduced. We showed how to perform local influence analysis under three different perturbation schemes (case-weights perturbation, response variable perturbation and ex-

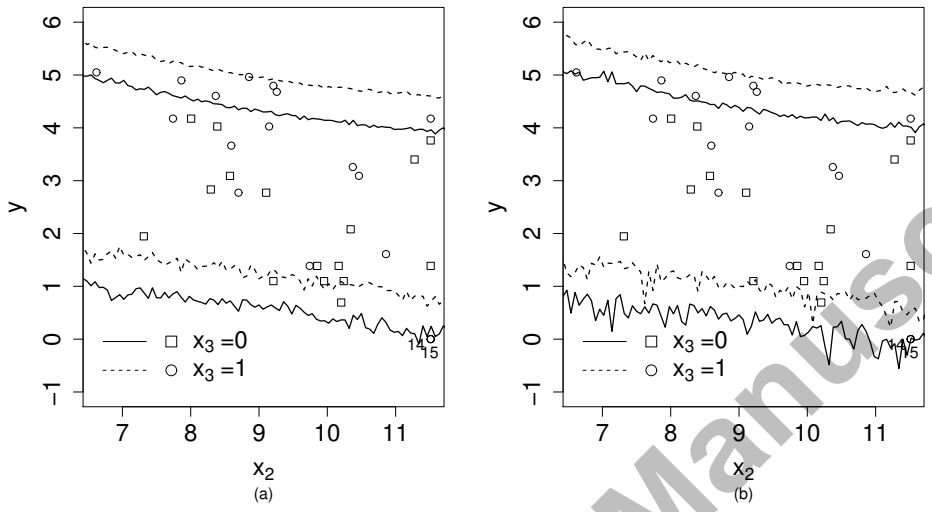

Figure 4 *95% percentile (a) and BC_a (b) prediction intervals for $\boldsymbol{y}$. Solid lines indicate the intervals for AG negative patients ($x_3 = 0$) and dashed lines indicate intervals for AG positive patients ($x_3 = 1$).*

planatory variable perturbation), derived generalized leverage measures, obtained the generalized Cook's distance, and presented a model misspecification test. We also provided an algorithm that can be used to construct prediction intervals for out of sample response values. In addition, we investigated the finite sample performances of different model selection criteria for the proposed model. Simulation results and a empirical application were presented and discussed. In the empirical application, the regression model we proposed was shown to be more adequate than the standard log-linear Birnbaum-Saunders model.

Finally, we note that since the proposed regression model is useful for lifetime data analyses, in future research we shall focus on extending it to deal with censored observations, such as, for instance, data subject to type I and type II censoring schemes.

Acknowledgements

We gratefully acknowledge partial financial support from CNPq (grant numbers 300845/2010-3 and 132540/2015-0). We also thank two anonymous referees and an associate editor for comments and suggestions.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the second international symposium on information theory* (B. N. Petrov and F. Csáki, eds.) 267–281.

Atkinson, A. C. (1985). *Plots, Transformations, and Regression*. Clarendon Press, Oxford.

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** 316–332.

Bayer, F. M. and Cribari-Neto, F. (2015). Bootstrap-based model selection criteria for beta regressions. *TEST* **24** 776–795.

Birnbaum, Z. W. and Saunders, S. C. (1969). A new family of life distributions. *Journal of Applied Probability* **6** 319–327.

Cavanaugh, J. E. and Shumway, R. H. (1997). A bootstrap variant of AIC for state-space model selection. *Statistica Sinica* **7** 473–496.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society B* **48** 133–169.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall, New York.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, New York.

Diaz-Garcia, J. A. and Dominguez-Molina, J. R. (2006). Some generalisations of Birnbaum-Saunders and sinh-normal distributions. *International Mathematical Forum* **1** 1709–1727.

Doornik, J. A. (2009). *Ox: An Object-Oriented Matrix Programming Language*. Timberlake Consultants Press, London.

Efron, B. (1981). Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics* **9** 139–158.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association* **82** 171–185.

Espinheira, P. L., Ferrari, S. L. and Cribari-Neto, F. (2014). Bootstrap prediction intervals in beta regressions. *Computational Statistics* **29** 1263–1277.

Feigl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics* **21** 826–838.

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

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society B* **41** 190–195.

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* **76** 297–307.

Ishiguro, M. and Sakamoto, Y. (1991). WIC: An estimation-free information criterion. *Research Memorandum of the Institute of Statistical Mathematics* **410**.

Leiva, V. (2015). *The Birnbaum-Saunders Distribution*. Academic Press, London.

Lemonte, A. J. (2013). A new extended Birnbaum-Saunders regression model for lifetime modeling. *Computational Statistics & Data Analysis* **64** 34–50.

Lemonte, A. J. and Cordeiro, G. M. (2009). Birnbaum-Saunders nonlinear regression models. *Computational Statistics & Data Analysis* **53** 4441–4452.

Martínez-Flórez, G., Bolfarine, H. and Gómez, H. W. (2017). The log-linear Birnbaum-Saunders power model. *Methodology and Computing in Applied Probability*

19 913–933.

McQuarrie, A. D. (1999). A small-sample correction for the Schwarz SIC model selection criterion. *Statistics & Probability Letters* **44** 79–86.

McQuarrie, A. D. and Tsai, C.-L. (1998). *Regression and Time Series Model Selection*. World Scientific, Singapore.

Mojirsheibani, M. and Tibshirani, R. (1996). Some results on bootstrap prediction intervals. *Canadian Journal of Statistics* **24** 549–568.

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika* **78** 691–692.

Owen, W. J. and Ng, H. K. T. (2015). Revisit of relationships and models for the Birnbaum-Saunders and inverse-Gaussian distributions. *Journal of Statistical Distributions and Applications* **2** 1–23.

Pan, W. (1999). Bootstrapping likelihood for model selection with small samples. *Journal of Computational and Graphical Statistics* **8** 687–698.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society B* **31** 350–371.

Ramsey, J. and Gilbert, R. (1972). A Monte Carlo study of some small sample properties of tests for specification error. *Journal of the American Statistical Association* **67** 180–186.

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

Santos, J. and Cribari-Neto, F. (2017). Hypothesis Testing in Log-Birnbaum–Saunders Regressions. *Communications in Statistics-Simulation and Computation* **46** 3990–4003.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6** 461–464.

Severini, T. A. (2000). *Likelihood Methods in Statistics*. Oxford University Press, New York.

Shibata, R. (1997). Bootstrap estimate of Kullback-Leibler information for model selection. *Statistica Sinica* **7** 375–394.

Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V. and De Bastiani, F. (2017). *Flexible Regression and Smoothing: Using GAMLSS in R*. Chapman and Hall/CRC, New York.

Stine, R. A. (1985). Bootstrap prediction intervals for regression. *Journal of the American Statistical Association* **80** 1026–1031.

Sugiura, N. (1978). Further analysts of the data by Akaike's information criterion and the finite corrections: Further analysts of the data by Akaike. *Communications in Statistics-Theory and Methods* **7** 13–26.

R Core Team (2016). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.

Tsionas, E. G. (2001). Bayesian inference in Birnbaum-Saunders regression. *Communications in Statistics – Theory and Methods* **30** 179–193.

Vilca, F., Azevedo, C. L. and Balakrishnan, N. (2017). Bayesian inference for sinh-normal/independent nonlinear regression models. *Journal of Applied Statistics* **44** 2052–2074.

Villegas, C., Paula, G. A. and Leiva, V. (2011). Birnbaum-Saunders mixed models for censored reliability data analysis. *IEEE Transactions on Reliability* **60** 748–758.

Wei, B. C., Hu, Y. Q. and Fung, W. K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics* **25** 25–37.

Xie, F. C. and Wei, B. C. (2007). Diagnostics analysis for log-Birnbaum-Saunders regression models. *Computational Statistics & Data Analysis* **51** 4692–4706.