Influence diagnostics in Birnbaum–Saunders nonlinear regression models

Artur J. Lemonte, Alexandre G. Patriota

Departamento de Estatística, Universidade de São Paulo, Rua do Matão, 1010, São Paulo/SP, 05508-090, Brazil

Abstract

We consider the issue of assessing influence of observations in the class of Birnbaum–Saunders nonlinear regression models, which is useful in lifetime data analysis. Our results generalize those in Galea et al. [2004, Influence diagnostics in log-Birnbaum–Saunders regression models. *Journal of Applied Statistics* **31**, 1049–1064] which are confined to Birnbaum–Saunders linear regression models. Some influence methods, such as the local influence, total local influence of an individual and generalized leverage are discussed. Additionally, the normal curvatures for studying local influence are derived under various perturbation schemes. We also give an application to a real fatigue data set.

Key words: Birnbaum–Saunders distribution; Fatigue life distribution; Influence diagnostic; Generalized leverage; Lifetime data; Local influence; Maximum likelihood estimation.

1 Introduction

The family of distributions proposed by Birnbaum and Saunders (1969), also known as the fatigue life distribution, has been widely applied for describing fatigue life, and lifetimes in general. This family of distributions was originally obtained from a model for which failure follows from the development and growth of a dominant crack. It was later derived by Desmond (1985) using a biological model which followed from relaxing some of the assumptions originally made by Birnbaum and Saunders (1969).

The random variable T is said to have a Birnbaum–Saunders distribution, say $\mathcal{B}\text{-}\mathcal{S}(\alpha, \eta)$, if its density function is given by

$$f_T(t; \alpha, \eta) = \frac{1}{2\alpha\eta\sqrt{2\pi}}\left[\left(\frac{\eta}{t}\right)^{1/2} + \left(\frac{\eta}{t}\right)^{3/2}\right]\exp\left\{-\frac{1}{2\alpha^2}\left(\frac{t}{\eta} + \frac{\eta}{t} - 2\right)\right\}, \quad t > 0,$$

where $\alpha > 0$ and $\eta > 0$ are shape and scale parameters, respectively. The density is right skewed, the skewness decreasing with α. For any $k > 0$, it follows that $kT \sim \mathcal{B}\text{-}\mathcal{S}(\alpha, k\eta)$. Some interesting results about improved statistical inference for the $\mathcal{B}\text{-}\mathcal{S}(\alpha, \eta)$ may be revised in Lemonte et al. (2007, 2008). Some generalizations and extensions of the Birnbaum–Saunders distribution are presented in Díaz–García and Leiva (2005) and Gómes et al. (2009).

Rieck and Nedelman (1991) proposed a log-linear regression model based on the Birnbaum–Saunders distribution. They showed that if $T \sim \mathcal{B}\text{-}\mathcal{S}(\alpha, \eta)$, then $Y = \log(T)$ is sinh-normal distributed, say $Y \sim \mathcal{SN}(\alpha, \mu, \sigma)$, with shape, location and scale parameters given by α, $\mu = \log(\eta)$ and $\sigma = 2$, respectively. Diagnostic tools for the Birnbaum–Saunders regression model were developed by Galea et al. (2004), Leiva et al. (2007) and Xi and Wei (2007). Small-sample adjustments for the likelihood ratio test can be found in Lemonte et al. (2009).

Recently, Lemonte and Cordeiro (2009) proposed a new class of Birnbaum–Saunders nonlinear regression models. This class generalizes the regression model described by Rieck and Nedelman (1991). Additionally, the authors discussed maximum likelihood estimation for the model parameters, and derive closed-form expressions for the second-order biases of these estimates.

Diagnostic analysis is an efficient way to detect influential observations. The first technique developed to assess the individual impact of cases on the estimation process is, perhaps, the case deletion which became a very popular tool. However, case deletion excludes all information from an observation and we can hardly say whether that observation has some influence on a specific aspect of the model. To overcome this problem, one can resort to local influence approach where one again investigates the model sensibility under small perturbations. In this context, Cook (1986) proposes a general framework to detect influential observations which give a measure of this sensibility under small perturbations on the data or in the model. Several authors have extended the local influence method to various regression models; see, for example, Lawrance (1988), Thomas and Cook (1990), Paula (1993), Lesaffre and Verbeke (1998) and, more recently, Osorio et al. (2007), Espinheira et al. (2008), Paula et al. (2009), among others.

In this article, we present diagnostic methods based on local influence and generalized leverage in the class of Birnbaum–Saunders nonlinear regression models.

Our results generalize those in Galea et al. (2004) which are confined to Birnbaum–Saunders linear regression models. In Section 2, we present the class of Birnbaum–Saunders nonlinear regression models. The score functions and observed Fisher information matrix are given as well as the process for estimating the regression coefficients and the shape parameter. Derivations of the normal curvature under different perturbation schemes together with generalized leverage are made in Section 3. An application to a real dataset are analyzed in Section 4. Finally, Section 5 concludes the paper.

2 Birnbaum–Saunders nonlinear regression model

Let $T \sim \mathcal{B}\text{-}\mathcal{S}(\alpha, \eta)$. The density function of $Y = \log(T)$ has the form

$$\pi(y; \alpha, \mu, \sigma) = \frac{2}{\alpha\sigma\sqrt{2\pi}} \cosh\left(\frac{y-\mu}{\sigma}\right) \exp\left\{-\frac{2}{\sigma^2}\sinh^2\left(\frac{y-\mu}{\sigma}\right)\right\}, \quad y \in \mathbb{R}.$$

This distribution has a number of interesting properties: (i) It is symmetric around the location parameter μ; (ii) It is unimodal for $\alpha \leq 2$ and bimodal for $\alpha > 2$; (iii) $\mathbb{E}(y) = \mu$ and its variance is a function of α only, and has no closed-form expression, but Rieck (1989) obtained asymptotic approximations for both small and large values of α; (iv) If $y_\alpha \sim \mathcal{SN}(\alpha, \mu, \sigma)$, then $Z_\alpha = 2(y_\alpha - \mu)/(\alpha\sigma)$ converges in distribution to the standard normal distribution when $\alpha \to 0$.

Lemonte and Cordeiro (2009) proposed the following regression model:

$$y_i = f_i(\boldsymbol{x}_i; \boldsymbol{\beta}) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where y_i is the logarithm of the ith observed lifetime, $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})^\top$ is an $m \times 1$ vector of known explanatory variables associated with the ith observable response y_i, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^\top$ is a vector of unknown nonlinear parameters, and $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$. We assume a nonlinear structure for the location parameter μ_i in model (1), say $\mu_i = f_i(\boldsymbol{x}_i; \boldsymbol{\beta})$, where f_i is assumed to be a known and twice continuously differentiable function with respect to $\boldsymbol{\beta}$.

The log-likelihood function for the vector parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha)^\top$ from a random sample $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$ obtained from (1), can be expressed as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{2}$$

where $\ell_i(\boldsymbol{\theta}) = -\log(8\pi)/2 + \log(\xi_{i1}) - \xi_{i2}^2/2$,

$$\xi_{i1} = \xi_{i1}(\boldsymbol{\theta}) = \frac{2}{\alpha}\cosh\left(\frac{y_i - \mu_i}{2}\right), \quad \xi_{i2} = \xi_{i2}(\boldsymbol{\theta}) = \frac{2}{\alpha}\sinh\left(\frac{y_i - \mu_i}{2}\right),$$

for $i = 1, \ldots, n$. The $n \times p$ local matrix $\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{\beta}) = \partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}^\top$ of partial derivatives of $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^\top$ with respect to $\boldsymbol{\beta}$ is assumed to be of full rank, i.e., $\mathrm{rank}(\boldsymbol{D}) = p$ for all $\boldsymbol{\beta}$.

The score functions for $\boldsymbol{\beta}$ and α can be expressed, respectively, as

$$\boldsymbol{U}_{\boldsymbol{\beta}} = \boldsymbol{D}^\top \boldsymbol{s} \quad \text{and} \quad U_\alpha = -\frac{n}{\alpha} + \frac{1}{\alpha}\sum_{i=1}^{n} \xi_{i2}^2,$$

where $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{\theta}) = (s_1, s_2, \ldots, s_n)^\top$ with $s_i = (\xi_{i1}\xi_{i2} - \xi_{i2}/\xi_{i1})/2$. The MLE $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\alpha})^\top$ satisfies $p+1$ equations: $\boldsymbol{U}_{\boldsymbol{\beta}} = \boldsymbol{0}$ and $U_\alpha = 0$. A joint iterative procedure to obtain the MLEs of $\boldsymbol{\beta}$ and α is given by (Lemonte and Cordeiro, 2009)

$$\boldsymbol{\beta}^{(m+1)} = (\boldsymbol{D}^{(m)\top}\boldsymbol{D}^{(m)})^{-1}\boldsymbol{D}^{(m)\top}\boldsymbol{\zeta}^{(m)}, \quad \alpha^{(m+1)} = \frac{1}{2}\alpha^{(m)}(1 + \bar{\xi}_2^{(m)}), \quad m = 0, 1, \ldots,$$

where $\boldsymbol{\zeta}^{(m)} = \boldsymbol{D}^{(m)}\boldsymbol{\beta}^{(m)} + \{4/\psi(\alpha^{(m)})\}\boldsymbol{s}^{(m)}$, $\bar{\xi}_2^{(m)} = \sum_{i=1}^{n} \xi_{i2}^{2(m)}/n$ and $\psi(\alpha) = 2 + 4/\alpha^2 - \alpha^{-1}\sqrt{2\pi}\{1 - \mathtt{erf}(\sqrt{2}/\alpha)\}\exp(2/\alpha^2)$. Also, $\mathtt{erf}(\cdot)$ is the *error function* (see, for example, Gradshteyn and Ryzhik, 2007). It can be shown that $\psi(\alpha) \approx 1 + 4/\alpha^2$ for α small and $\psi(\alpha) \approx 2$ for α large. The above equations show that any software with a weighted linear regression routine can be used to calculate the MLEs of $\boldsymbol{\beta}$ and α iteratively. Starting values $\boldsymbol{\beta}^{(0)}$ and $\alpha^{(0)}$ for the iterative algorithm are required.

The asymptotic inference for the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha)^\top$ can be based on the normal approximation of the MLE of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\alpha})^\top$. Let $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ the asymptotic variance-covariance matrix for $\widehat{\boldsymbol{\theta}}$. Then, for n large, $\widehat{\boldsymbol{\theta}} \overset{a}{\sim} \mathcal{N}_{p+1}(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$, where $\overset{a}{\sim}$ denotes approximately distributed. Additionally, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ may be approximated by $-\ddot{\boldsymbol{L}}_{\widehat{\boldsymbol{\theta}}\widehat{\boldsymbol{\theta}}}^{-1}$, where $\ddot{\boldsymbol{L}}_{\widehat{\boldsymbol{\theta}}\widehat{\boldsymbol{\theta}}}$ is the $(p+1) \times (p+1)$ observed information matrix evaluated at $\widehat{\boldsymbol{\theta}}$, obtained from

$$\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{bmatrix} \ddot{\boldsymbol{L}}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \ddot{\boldsymbol{L}}_{\boldsymbol{\beta}\alpha} \\ \ddot{\boldsymbol{L}}_{\alpha\boldsymbol{\beta}} & \ddot{L}_{\alpha\alpha} \end{bmatrix} = \begin{bmatrix} \boldsymbol{D}^\top\boldsymbol{V}\boldsymbol{D} + [\boldsymbol{s}^\top][\boldsymbol{G}] & \boldsymbol{D}^\top\boldsymbol{h} \\ \boldsymbol{h}^\top\boldsymbol{D} & \mathrm{tr}(\boldsymbol{K}) \end{bmatrix},$$

where $\boldsymbol{V} = \mathrm{diag}\{v_1, v_2, \ldots, v_n\}$, $v_i = v_i(\boldsymbol{\theta}) = -\{2\xi_{i2}^2 + 4/\alpha^2 - 1 + \xi_{i2}^2/\xi_{i1}^2\}/4$, $\boldsymbol{h} = (h_1, h_2, \ldots, h_n)^\top$, $h_i = h_i(\boldsymbol{\theta}) = -\xi_{i1}\xi_{i2}/\alpha$, $\boldsymbol{K} = \mathrm{diag}\{k_1, k_2, \ldots, k_n\}$, $k_i = k_i(\boldsymbol{\theta}) = 1/\alpha^2 - 3\xi_{i2}^2/\alpha^2$, $\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\beta}) = \partial^2\boldsymbol{\mu}/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top$ is an array of dimension $n \times p \times p$ and $\mathrm{tr}(\cdot)$ is the trace operator. Finally, $[\cdot][\cdot]$ represents the bracket product of a matrix by an array as defined by Wei (1998, p. 188).[1]

[1] If $\boldsymbol{B}$ is an $m \times n$ matrix and $\boldsymbol{A}$ is an $n \times p \times q$ array, then $\boldsymbol{C} = [\boldsymbol{B}][\boldsymbol{A}]$ is called the bracket product of $\boldsymbol{B}$ and $\boldsymbol{A}$, that is an $m \times p \times q$ array with elements $Y_{tij} = \sum_{k=1}^{n} B_{tk}A_{kij}$.

3 Diagnostic analysis

3.1 Local Influence

The local influence method is recommended when the concern is related to investigate the model sensibility under some minor perturbations in the model (or data). Let $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ be a k-dimensional vector of perturbations, where $\boldsymbol{\Omega} \subset \mathbb{R}^k$ is an open set. The perturbed log-likelihood function is denoted by $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$. We consider that exists a non perturbation vector, namely $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$, such that $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \ell(\boldsymbol{\theta})$. The influence of minor perturbations on the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ can be assessed by using the log-likelihood displacement $LD_{\boldsymbol{\omega}} = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}$, where $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denotes the maximum likelihood estimate under $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$.

The Cook's idea for assessing local influence is essentially to analyse the local behavior of $LD_{\boldsymbol{\omega}}$ around $\boldsymbol{\omega}_0$ by evaluating the curvature of the plot of $LD_{\boldsymbol{\omega}_0+a\boldsymbol{d}}$ against a, where $a \in \mathbb{R}$ and $\boldsymbol{d}$ is a unit norm direction. One of the measures of particular interest is the direction $\boldsymbol{d}_{\max}$ corresponding to the largest curvature $C_{\boldsymbol{d}_{\max}}$. The index plot of $\boldsymbol{d}_{\max}$ may evidence those observations that have considerable influence on $LD_{\boldsymbol{\omega}}$ under minor perturbations. Also, plots of $\boldsymbol{d}_{\max}$ against covariate values may be helpful for identifying atypical patterns. Cook (1986) shows that the normal curvature at the direction $\boldsymbol{d}$ is given by

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

where $\boldsymbol{\Delta} = \partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top$, both $\boldsymbol{\Delta}$ and $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}$ are evaluated at $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega}_0$. Hence, $C_{\boldsymbol{d}_{\max}}/2$ is the largest eigenvalue of $\boldsymbol{B} = -\boldsymbol{\Delta}^\top \ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} \boldsymbol{\Delta}$ and $\boldsymbol{d}_{\max}$ is the corresponding unit norm eigenvector. The index plot of $\boldsymbol{d}_{\max}$ for the matrix $\boldsymbol{B}$ may show how to perturb the model (or data) to obtain large changes in the estimate of $\boldsymbol{\theta}$.

However, if the interest lies in computing the local influence for $\boldsymbol{\beta}$, the normal curvature in the direction of the vector $\boldsymbol{d}$ is $C_{\boldsymbol{d};\boldsymbol{\beta}}(\boldsymbol{\theta}) = 2|\boldsymbol{d}^\top \boldsymbol{\Delta}^\top (\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} - \ddot{\boldsymbol{L}}_{22}) \boldsymbol{\Delta} \boldsymbol{d}|$, where

$$\ddot{\boldsymbol{L}}_{22} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{L}_{\alpha\alpha}^{-1} \end{bmatrix},$$

and $\boldsymbol{d}_{\max;\boldsymbol{\beta}}$ here is the unit norm eigenvector corresponding to the largest eigenvalue of $\boldsymbol{B}_1 = -\boldsymbol{\Delta}^\top (\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} - \ddot{\boldsymbol{L}}_{22}) \boldsymbol{\Delta}$ (see Cook, 1986, Eq. (26)). The index plot of the largest eigenvector of $\boldsymbol{B}_1$ may reveal those influential observations on $\widehat{\boldsymbol{\beta}}$.

Another procedure is the total local curvature corresponding to the ith element, which follows by taking $\boldsymbol{d}_i$ or an $n \times 1$ vector of zeros with one at the ith position. Thus, the curvature at the direction $\boldsymbol{d}_i$ assumes the form $C_i(\boldsymbol{\theta}) = 2|\boldsymbol{\Delta}_i^\top \ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} \boldsymbol{\Delta}_i|$, where $\boldsymbol{\Delta}_i^\top$ denotes the ith row of $\boldsymbol{\Delta}$. This is named total local influence (see, for instance, Lesaffre and Verbeke, 1998). It is also possible to compute the total local influence of the ith individual when estimating a subset of the elements of $\boldsymbol{\theta}$. For instance, if the interest lies in $\boldsymbol{\beta}$, we have that $C_{i;\boldsymbol{\beta}}(\boldsymbol{\theta}) = 2|\boldsymbol{\Delta}_i^\top(\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} - \ddot{\boldsymbol{L}}_{22})\boldsymbol{\Delta}_i|$. Verbeke and Molembergs (2000, § 11.3) propose considering as point out those cases such that $C_i \geq 2\bar{C}$, where $\bar{C} = \sum_{i=1}^n C_i/n$.

3.2 Curvature calculations

Next, we calculate, for three different perturbation scheme, the matrix

$$\boldsymbol{\Delta} = \{\Delta_{ri}\}_{(p+1)\times n} = \left\{\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\theta_r \partial\omega_i}\right\}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}, \quad r = 1,\ldots,p+1 \quad \text{and} \quad i = 1,\ldots,n,$$

considering the model defined in (1) and its log-likelihood function given by (2).

3.2.1 Case-weights perturbation

The perturbation of cases is done by defining some weights for each observation in the log-likelihood function as follows:

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

where $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)^\top$ is the total vector of weights, with $0 \leq \omega_i \leq 1$, for $i = 1,\ldots,n$, and $\boldsymbol{\omega}_0 = (1,1,\ldots,1)^\top$ is the vector of no perturbations. The matrix $\boldsymbol{\Delta}$ is given by

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_{\boldsymbol{\beta}} \\ \boldsymbol{\Delta}_\alpha \end{pmatrix},$$

where $\boldsymbol{\Delta}_{\boldsymbol{\beta}} = \widehat{\boldsymbol{D}}^\top\widehat{\boldsymbol{S}}$ with $\widehat{\boldsymbol{S}} = \mathrm{diag}\{\widehat{s}_1, \widehat{s}_2, \ldots, \widehat{s}_n\}$, and $\boldsymbol{\Delta}_\alpha = (\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n)$ with $\widehat{b}_i = -1/\widehat{\alpha} + \widehat{\xi}_{i2}^2/\widehat{\alpha}$. Note that, for linear models, the matrix $\boldsymbol{\Delta}$ reduces to the one given in Galea et al. (2004).

3.2.2 Response perturbation

We will consider here that each y_i is perturbed as $y_{iw} = y_i + \omega_i S_y$, where S_y is a scale factor that may be estimated standard deviation of $\boldsymbol{y}$. In this case, the perturbed log-likelihood function is given by

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\frac{n}{2}\log(8\pi) + \sum_{i=1}^n \log(\xi_{i1w_1}) - \frac{1}{2}\sum_{i=1}^n \xi_{i2w_1}^2,$$

where $\xi_{i1w_1} = \xi_{i1w_1}(\boldsymbol{\theta}) = 2\alpha^{-1}\cosh([y_{iw} - \mu_i]/2)$, $\xi_{i2w_1} = \xi_{i2w_1}(\boldsymbol{\theta}) = 2\alpha^{-1}\sinh([y_{iw} - \mu_i]/2)$ and $\boldsymbol{\omega}_0 = (0,0,\ldots,0)^\top$ is the vector of no perturbations. The matrix $\boldsymbol{\Delta}$ assumes the form

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_{\boldsymbol{\beta}} \\ \boldsymbol{\Delta}_\alpha \end{pmatrix},$$

where $\boldsymbol{\Delta}_{\boldsymbol{\beta}} = -S_y \widehat{\boldsymbol{D}}^\top \widehat{\boldsymbol{V}}$ and $\boldsymbol{\Delta}_\alpha = -S_y \widehat{\boldsymbol{h}}$. It is noteworthy that the matrix $\boldsymbol{\Delta}$ reduces to the one given in Galea et al. (2004) for linear models.

3.2.3 Explanatory variable perturbation

Consider now an additive perturbation on a particular continuous explanatory variable, namely $\boldsymbol{x}_j$, by making $x_{ijw} = x_{ij} + \omega_i S_x$, where S_x is a scale factor that may be estimated standard deviation of $\boldsymbol{x}_j$. This perturbation scheme leads to the following expression for the log-likelihood function:

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\frac{n}{2}\log(8\pi) + \sum_{i=1}^{n}\log(\xi_{i1w_2}) - \frac{1}{2}\sum_{i=1}^{n}\xi_{i2w_2}^2,$$

where $\xi_{i1w_2} = \xi_{i1w_2}(\boldsymbol{\theta}) = 2\alpha^{-1}\cosh([y_i - \mu_{iw}]/2)$, $\xi_{i2w_2} = \xi_{i2w_2}(\boldsymbol{\theta}) = 2\alpha^{-1}\sinh([y_i - \mu_{iw}]/2)$ and $\mu_{iw} = f_i(\boldsymbol{x}_{iw}, \boldsymbol{\beta})$, with $\boldsymbol{x}_{iw} = (x_{i1}, \ldots, x_{ijw}, \ldots, x_{im})^\top$. Here, $\boldsymbol{\omega}_0 = (0, 0, \ldots, 0)^\top$ is the vector of no perturbations. The matrix $\boldsymbol{\Delta}$ is given by

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_{\boldsymbol{\beta}} \\ \boldsymbol{\Delta}_\alpha \end{pmatrix},$$

where $\boldsymbol{\Delta}_{\boldsymbol{\beta}}$ is a $p \times n$ matrix with Δ_{ri} elements that assume the form (for $r = 1, \ldots, p$ and $i = 1, \ldots, n$)

$$\Delta_{ri} = \ddot{\mu}_{irw}\widehat{s}_i + \dot{\mu}_{iw}\dot{\mu}_{irw}\widehat{v}_i,$$

with

$$\ddot{\mu}_{irw} = \left.\frac{\partial^2 \mu_{iw}}{\partial\beta_r \partial\omega_i}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}, \quad \dot{\mu}_{iw} = \left.\frac{\partial \mu_{iw}}{\partial\omega_i}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} \quad \text{and} \quad \dot{\mu}_{irw} = \left.\frac{\partial \mu_{iw}}{\partial\beta_r}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}.$$

Additionally, $\boldsymbol{\Delta}_\alpha = (\widehat{e}_1, \widehat{e}_2, \ldots, \widehat{e}_n)$ with $\widehat{e}_i = \dot{\mu}_{iw}\widehat{h}_i$.

For linear models, i.e. $\mu_i = \boldsymbol{x}_i^\top\boldsymbol{\beta}$, the matrix $\boldsymbol{\Delta}_{\boldsymbol{\beta}}$ reduces to the one given in Galea et al. (2004). Note that $\mu_{iw} = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \beta_j w_i S_x$. Thus, $\ddot{\mu}_{irw} = 0$ $(r \neq j)$ and $\ddot{\mu}_{irw} = S_x$ $(r = j)$, $\dot{\mu}_{irw} = x_{ir}$ and $\dot{\mu}_{iw} = S_x\widehat{\beta}_j$. Clearly, $\boldsymbol{\Delta}_\alpha$ also reduces to the one given in Galea et al. (2004) for linear models.

3.3 Generalized leverage

In what follows we shall use the generalized leverage proposed by Wei et al. (1998), which is defined as $\boldsymbol{GL}(\widetilde{\boldsymbol{\theta}}) = \partial\widetilde{\boldsymbol{y}}/\partial\boldsymbol{y}^\top$, where $\boldsymbol{\theta}$ is an s-vector such that $\mathbb{E}(\boldsymbol{y}) = \boldsymbol{\mu}(\boldsymbol{\theta})$ and $\widetilde{\boldsymbol{\theta}}$ is an estimator of $\boldsymbol{\theta}$, with $\widetilde{\boldsymbol{y}} = \boldsymbol{\mu}(\widetilde{\boldsymbol{\theta}})$. Here, the (i, l) element of $\boldsymbol{GL}(\widetilde{\boldsymbol{\theta}})$, i.e. the generalized leverage of the estimator $\widetilde{\boldsymbol{\theta}}$ at (i, l), is the instantaneous rate of change in ith predicted value with respect to the lth response value. As noted by the authors, the generalized leverage is invariant under reparameterization and observations with large GL_{ij} are leverage points. Wei et al. (1998) have shown that the generalized leverage is obtained by evaluating

$$\boldsymbol{GL}(\boldsymbol{\theta}) = \boldsymbol{D}_{\boldsymbol{\theta}}(-\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}})^{-1}\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{y}},$$

at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, where $\boldsymbol{D}_{\boldsymbol{\theta}} = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}^\top$ and $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{y}} = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{y}^\top$.

Under model defined in (1), we have that

$$\boldsymbol{D}_{\boldsymbol{\theta}} = \begin{bmatrix}\boldsymbol{D} & \boldsymbol{0}\end{bmatrix} \quad \text{and} \quad \ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{y}} = -\begin{bmatrix}\boldsymbol{D}^\top\boldsymbol{V} \\ \boldsymbol{h}\end{bmatrix}.$$

It is noteworthy that $\boldsymbol{GL}(\boldsymbol{\theta})$ reduces to the one given in Galea et al. (2004) for linear models.

4 Application

The fatigue processes are by excellence ideally modeled by the Birnbaum–Saunders distribution due to its genesis. We analyze an application to a biaxial fatigue data set reported by Rieck and Nedelman (1991) on the life of a metal piece in cycles to failure. The response N is the number of cycles to failure and the explanatory variable w is the work per cycle (mJ/m^3). The data of forty six observations were taken from Table 1 of Galea et al. (2004).

Rieck and Nedelman (1991) proposed the following model for the biaxial fatigue data:

$$y_i = \beta_1 + \beta_2 \log w_i + \varepsilon_i, \quad i = 1, \ldots, 46 \tag{3}$$

where $y_i = \log N_i$ and $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$. Rieck and Nedelman (1991) takes the logarithm of w to ensure a linear relationship between the response variable (y) and the covariate (w) in (3). However, this model does not consider the real scale of the covariate. Galea et al. (2004) apply diagnostics methods in this dataset considering model (3).

Lemonte and Cordeiro (2009) proposed the nonlinear regression model

$$y_i = \beta_1 + \beta_2 \exp(\beta_3/w_i) + \varepsilon_i, \quad i = 1, \ldots, 46, \tag{4}$$

for the biaxial fatigue data, where $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$. The authors showed that the nonlinear model (4) fits satisfactorily to the fatigue data. The maximum likelihood estimates (the standard errors in parentheses) for the parameters of model (4) are: $\widehat{\beta}_1 = 8.9876\,(0.7454)$, $\widehat{\beta}_2 = -5.1802\,(0.5075)$, $\widehat{\beta}_3 = -22.5196\,(7.3778)$ and $\widehat{\alpha} = 0.40\,(0.0417)$.

Here, we proposed a new nonlinear model for the biaxial fatigue data, that is, we consider

$$y_i = \beta_1 w_i^{\beta_2} + \varepsilon_i, \quad i = 1, \ldots, 46, \tag{5}$$

where also $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$. The maximum likelihood estimates (the standard errors in parentheses) for the parameters of model (5) are: $\widehat{\beta}_1 = 15.9166\,(0.9271)$, $\widehat{\beta}_2 = -0.2618\,(0.0169)$, and $\widehat{\alpha} = 0.41\,(0.0422)$. Figure 1 gives the scatter-plot of the data, together with the fitted curves of models (4) and (5). As can be seen, both fitted models fits satisfactorily to the fatigue data.

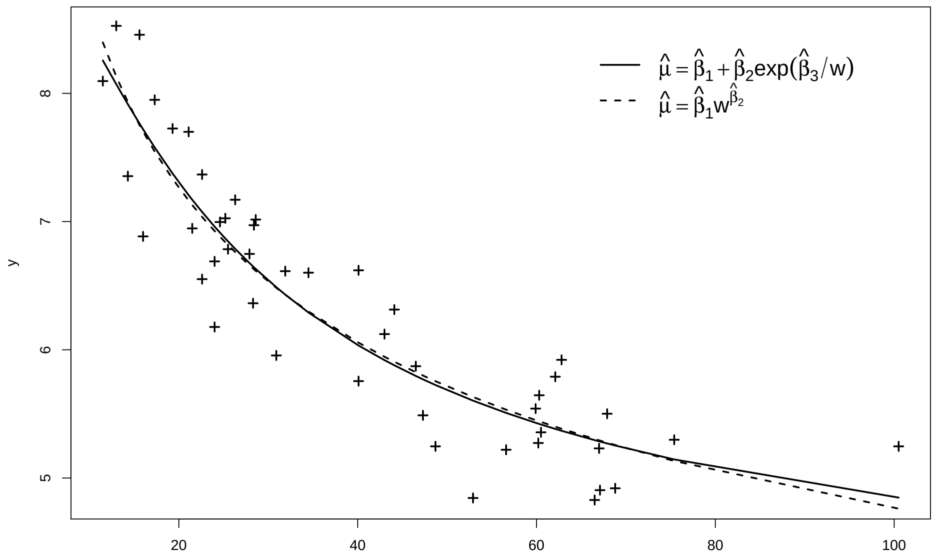

Figure 1: Scatter-plot and the fitted models.

Following Xie and Wei (2007), we obtain the residuals $\widehat{\varepsilon}_i = y_i - \widehat{\mu}_i$ and $\widehat{R}_i = 2\widehat{\alpha}^{-1}\sinh(\widehat{\varepsilon}_i/2)$. Figure 2 gives the scatter-plot of $\widehat{R}_i$ versus the predicted values $\widehat{\mu}_i$ for models (4) and (5). Note that the distribution of $\widehat{R}_i$ is approximately normal and have the same behaviour for both models. Based upon the fact that $U \sim \mathcal{SN}(\alpha, \mu, \sigma)$ if $2\alpha^{-1}\sinh\{(U - \mu)/\sigma\} \sim \mathcal{N}(0, 1)$, then the residual $\widehat{\varepsilon}_i$ should follow approximately a sinh-normal distribution.

For the purpose of verifying which nonlinear model better represent the true relationship between y and w, we also conduct a hypothesis testing proposed by Vuong (1989). Consider choosing between two nonnested models: model F_{μ_1} with density function $\pi(y_i|\mu_{1i})$ and model F_{μ_2} with density function $\pi(y_i|\mu_{2i})$, where $\mu_{1i} =$

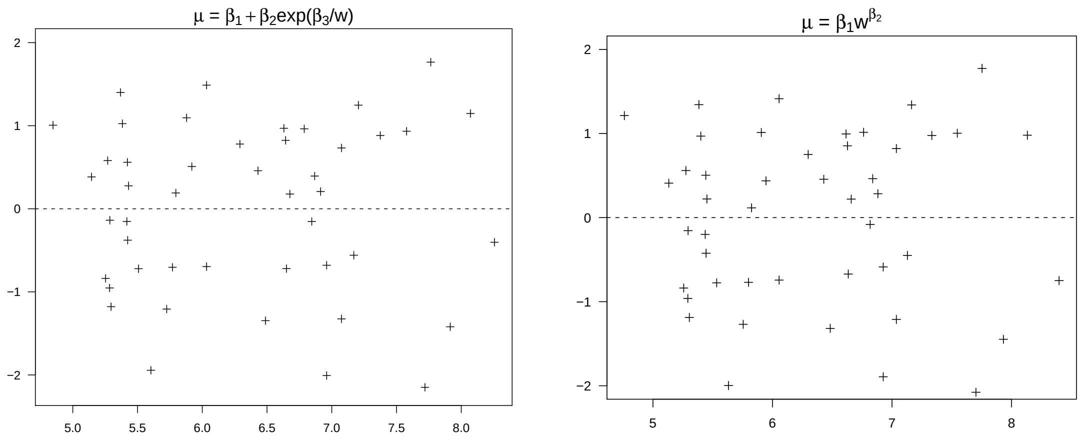

Figure 2: Scatter-plot of $\widehat{R}_i$ versus $\widehat{\mu}_i$.

$\mu_{1i}(\boldsymbol{x}_i;\boldsymbol{\theta})$ and $\mu_{2i} = \mu_{2i}(\boldsymbol{x}_i;\boldsymbol{\gamma})$. The test statistic can be written as

$$\begin{aligned} T_{LR,NN} = &\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\log\frac{\pi(y_i|\widehat{\mu}_{1i})}{\pi(y_i|\widehat{\mu}_{2i})}\right\}\times \\ &\times\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{\pi(y_i|\widehat{\mu}_{1i})}{\pi(y_i|\widehat{\mu}_{2i})}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\log\frac{\pi(y_i|\widehat{\mu}_{1i})}{\pi(y_i|\widehat{\mu}_{2i})}\right)^2\right\}^{-1/2}, \end{aligned} \tag{6}$$

where $\widehat{\mu}_{1i} = \widehat{\mu}_{1i}(\boldsymbol{x}_i;\widehat{\boldsymbol{\theta}})$ and $\widehat{\mu}_{2i} = \widehat{\mu}_{2i}(\boldsymbol{x}_i;\widehat{\boldsymbol{\gamma}})$, $i = 1,\ldots,n$. For strictly nonnested models, the statistic (6) converges in distribution to a standard normal distribution under the null hypothesis of equivalence of the models (Vuong, 1989). Thus, the null hypothesis is not rejected if $|T_{LR,NN}| \leq z_{\rho/2}$. On the other hand, we reject at significance level ρ the null hypothesis of equivalence of the models in favor of model F_{μ_1} being better (or worse) than model F_{μ_2} if $T_{LR,NN} > z_\rho$ (or $T_{LR,NN} < -z_\rho$). Let $\widehat{\mu}_{1i} = \widehat{\beta}_1 + \widehat{\beta}_2\exp(\widehat{\beta}_3/w_i)$ and $\widehat{\mu}_{2i} = \widehat{\beta}_1 w^{\widehat{\beta}_2}$. The test statistic ($T_{LR,NN}$) equals 0.9123 and the corresponding p-value is 0.17. Therefore, the test indicates that both nonlinear models are equivalent.

Next, we will apply the generalized leverage and local influence methods developed in the previous sections considering both nonlinear models (4) and (5).

Figures 3 and 4 give the $|\boldsymbol{l}_{\max}|$ (local influence) and C_i (total local influence), respectively, corresponding to $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\alpha}$ for different perturbation schemes considering model (4). Based on these figures, we observed that cases 1, 2, 3, 4, 5, 12, 32 and 46, have more pronounced influence than the other observations. Additionally, Figures 5 and 6 give the $|\boldsymbol{l}_{\max}|$ and C_i, respectively, corresponding to $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\alpha}$ for different perturbation schemes for the nonlinear model (5). From Figures 5 and 6, the same observations detected in Figures 3 and 4 are detected.

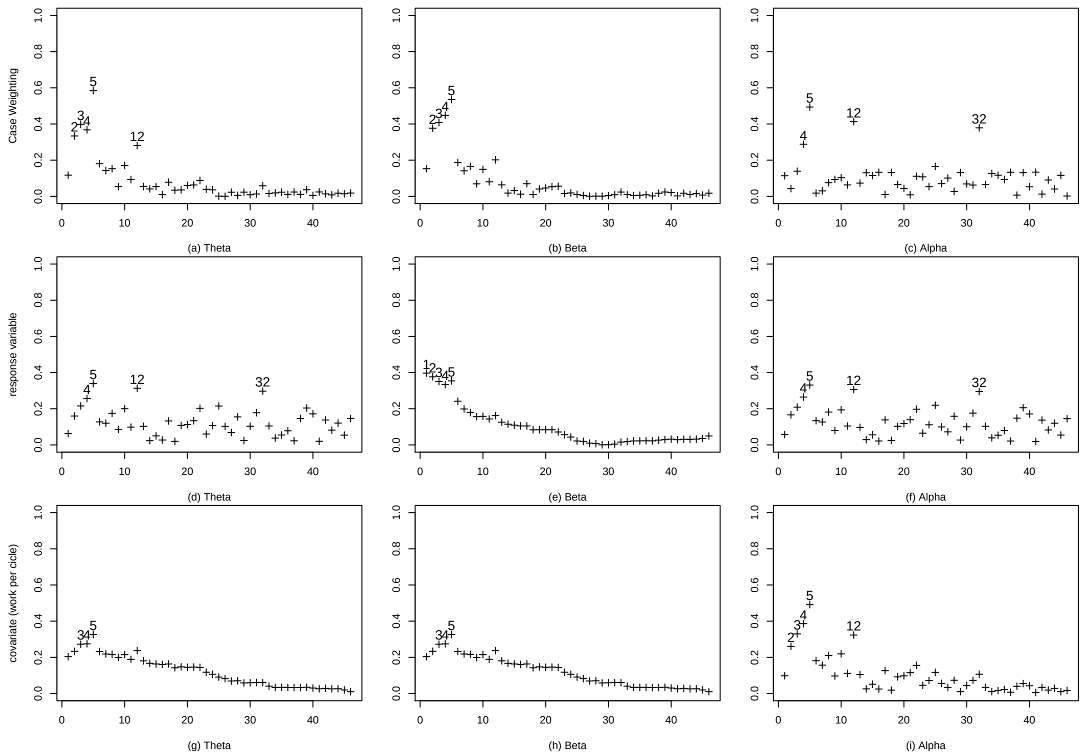

Figure 3: Index plots of $|\boldsymbol{l}_{\max}|$ for $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\alpha}$ under case weighting, response and covariate perturbation schemes; $\mu = \beta_1 + \beta_2 \exp(\beta_3/w)$.

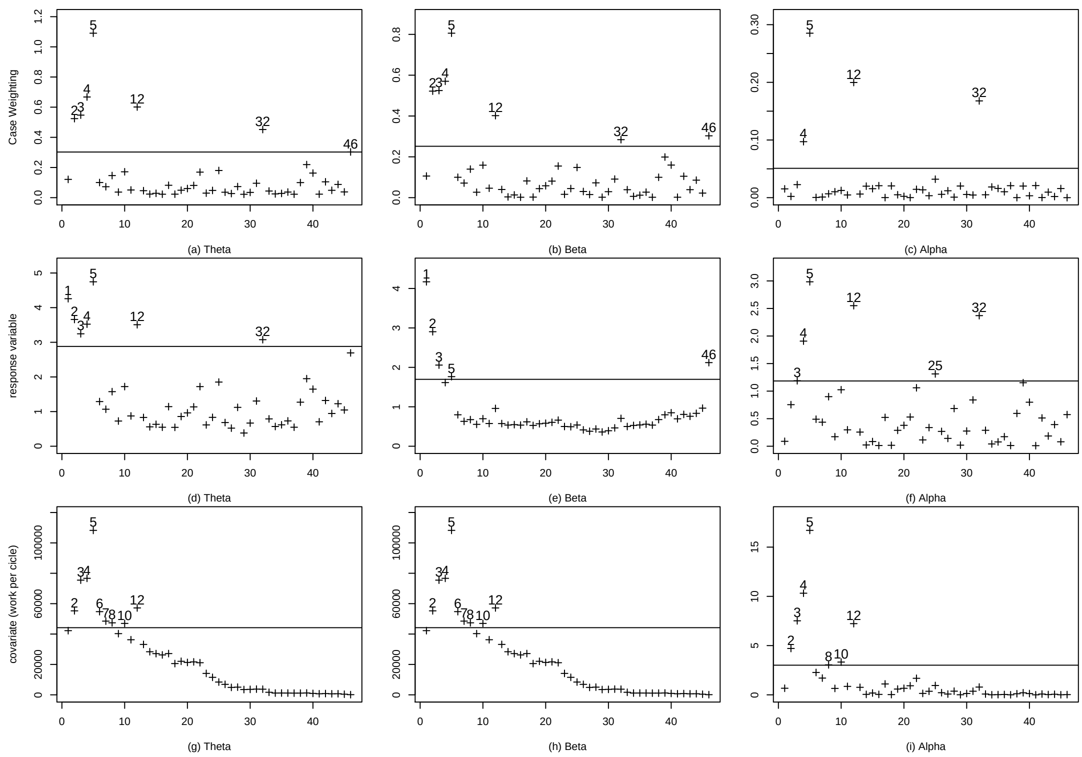

Figure 4: Index plots of C_i for $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\alpha}$ under case weighting, response and covariate perturbation schemes; $\mu = \beta_1 + \beta_2 \exp(\beta_3/w)$.

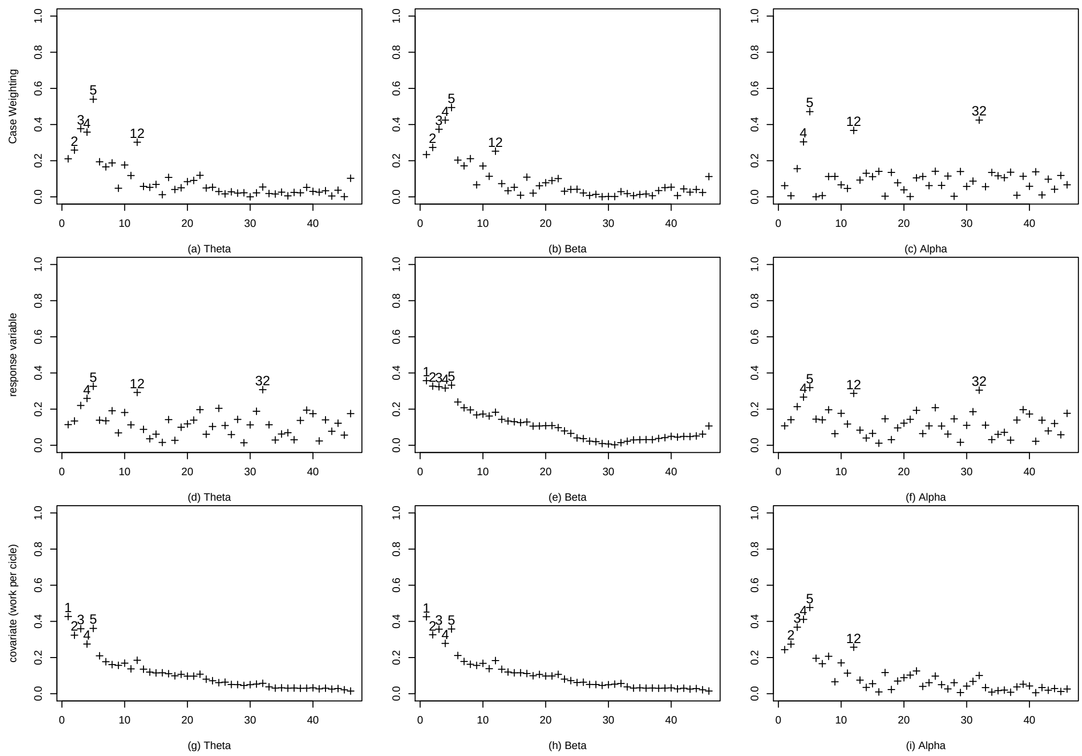

Figure 5: Index plots of $|\boldsymbol{l}_{\max}|$ for $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\alpha}$ under case weighting, response and covariate perturbation schemes; $\mu = \beta_1 w^{\beta_2}$.

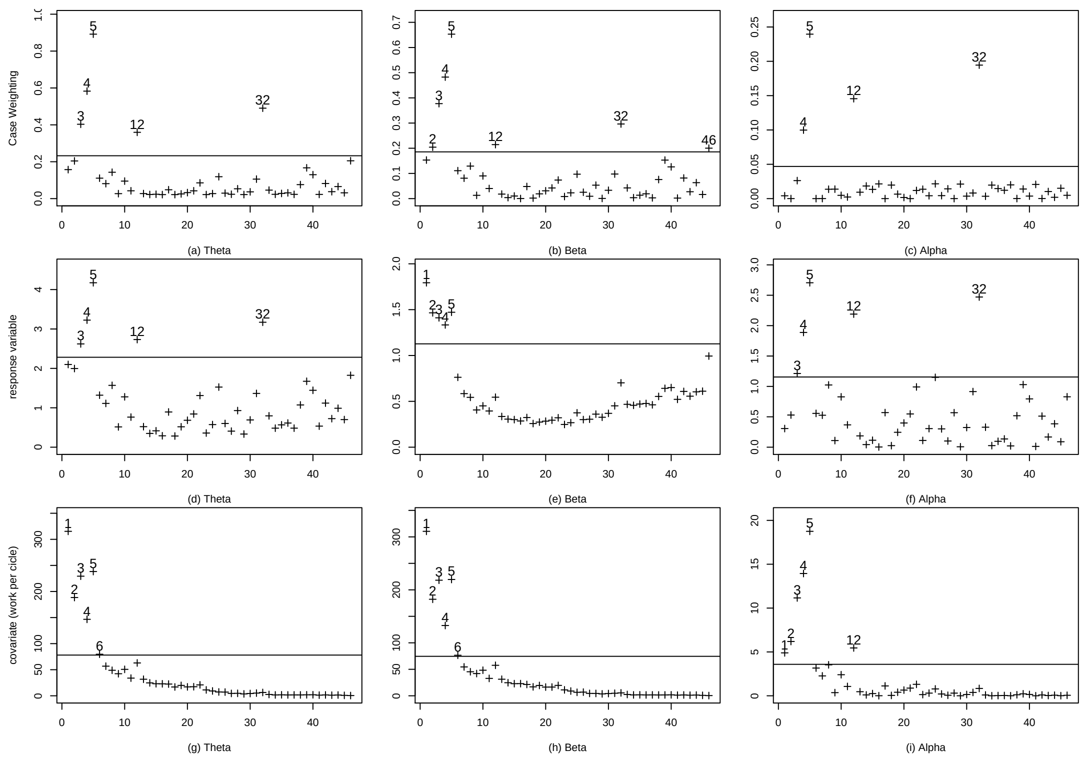

Figure 6: Index plots of C_i for $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\alpha}$ under case weighting, response and covariate perturbation schemes; $\mu = \beta_1 w^{\beta_2}$.

Thus, based on the Figures 3 to 6, we eliminated those most influential observations and refitted the models. Tables 1 and 2 depict the relative changes of the MLE of the location parameters (RC_1) and their asymptotic standard deviations estimates (RC_2) considering models (4) and (5), respectively. These relative chances are defined as follows: $\mathrm{RC}_1 = (\widehat{\beta}_j - \widehat{\beta}_{j(i)})/\widehat{\beta}_j$ and $\mathrm{RC}_2 = (\widehat{\sigma}_{\widehat{\beta}_j} - \widehat{\sigma}_{\widehat{\beta}_{j(i)}})/\widehat{\sigma}_{\widehat{\beta}_j}$, where $\widehat{\beta}_{j(i)}$ and $\widehat{\sigma}_{\widehat{\beta}_{j(i)}}$ denotes the maximum likelihood estimate for β_j and $\sigma_{\widehat{\beta}_j}$, respectively, after observation ith is removed. Note that RC_2 can also be interpreted as the relative change of the confidence interval amplitude by removing the ith observation. It is another measure to verify the influence of one observation in the maximum likelihood estimation, since when the hypothesis is far away from the estimates the p-value will not indicate a change in the conclusions even if great changes occur in the estimates of the asymptotic standard deviations of the MLEs.

The results are grouped in Tables 1 and 2. As can be seen, figures in Table 1 shows that the relative changes in the location estimates of β_3 are large and also the relative changes in the confidence interval amplitude are very pronounced. On the other hand, Table 2 shows no large chances for either location and confidence interval amplitude relative changes. Therefore, we conclude that model (5) has lesser influence of such observations than model (4).

Table 1: Relatives changes (%) dropping the cases indicated – $\mu = \beta_1 + \beta_2 \exp(\beta_3/w)$.

	β_1		β_2		β_3	
Eliminated	RC_1	RC_2	RC_1	RC_2	RC_1	RC_2
1	−3.0066	−48.9443	−3.7817	−60.8956	9.5434	−16.6047
2	4.1151	13.2806	4.9367	18.7542	−15.8260	−14.5069
3	−4.0912	−25.5935	−5.8549	−36.7778	11.2002	4.3046
4	3.0072	13.4428	4.4898	16.0104	−9.1079	−4.0841
5	−3.4389	−11.2195	−5.6328	−18.0936	7.9576	9.1715
12	1.4290	15.7545	0.3269	19.6946	−8.9273	1.5680
32	0.2551	3.0730	1.5735	2.7544	0.0585	1.6377
42	2.0782	13.1594	0.2066	20.1666	−11.1417	−6.3048

In all cases p-value < 0.01 ($H_0 : \beta_j = 0$).

1) Graficos dos leverages

2) Paragrafo final falando das vantagens do novo modelo em relacao ao modelo inicial: interpretecao, numero de parametros, mudanca das estimativas (erros padrao), etc.

Table 2: Relatives changes (%) dropping the cases indicated – $\mu = \beta_1 w^{\beta_2}$.

	β_1		β_2	
Eliminated	RC_1	RC_2	RC_1	RC_2
1	−1.7690	−9.9580	-1.8071	−7.1242
2	1.8828	−3.6739	1.9408	−4.7617
3	−2.5050	−4.9348	−2.4887	−1.7996
4	2.6404	1.8452	2.6597	−0.1303
5	−3.1094	−1.3577	−3.0219	2.1018
12	−0.9929	1.8466	−0.7859	2.9390
32	1.4066	4.3495	1.8816	2.7667
46	−1.6605	−3.8179	−1.9948	−2.7323

In all cases p-value < 0.01 ($H_0 : \beta_j = 0$).

5 Concluding remarks

The Birnbaum–Saunders distribution is widely used to model times to failure for materials subject to fatigue. In this paper, we developed influence diagnostics for the class of Birnbaum–Saunders nonlinear regression models which can be useful for modeling lifetime or reliability data. Appropriate matrices for assessing local influence on the parameter estimates under different perturbation schemes are obtained. Our results are very general and can be applied to any nonlinear regression model defined by (1). In particular, our results generalize those in Galea et al. (2004) which are confined to Birnbaum–Saunders linear regression models. Additionally, we also present an application to real data.

Acknowledgments

We gratefully acknowledge grants from FAPESP (Brazil).

References

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

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

Desmond, A.F. (1985). Stochastic models of failure in random environments. *Canadian Journal of Statistics* **13**, 171–183.

Díaz–García, J.A., Leiva, V. (2005). A new family of life distributions based on the elliptically contoured distributions. *Journal of Statistical Planning and Inference* **128**, 445–457.

Espinheira, P.L., Ferrari, S.L.P., Cribari–Neto, F. (2008). Influence diagnostics in beta regression. *Computational Statistics and Data Analysis* **52**, 4417–4431.

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

Gómes, H.W., Olivares-Pacheco, J.F., Bolfarine, H. (2009). An extension of the generalized Birnbaum–Saunders distribution. *Statistics and Probability Letters* **79**, 331–338.

Gradshteyn, I.S., Ryzhik, I.M. (2007). *Table of Integrals, Series, and Products*. Academic Press, New York.

Lawrance, A.J. (1988). Regression transformation diagnostic using local influence. *Journal of the American Statistical Association* **83**, 1067–1072.

Leiva, V., Barros, M., Paula, G.A., Galea, M. (2007). Influence diagnostics in log-Birnbaum–Saunders regression models with censored data. *Computational Statistics and Data Analysis* **51**, 5694–5707.

Lemonte, A.J., Cordeiro, G.M. (2009). Birnbaum–Saunders nonlinear regression models. Working paper `arXiv:0901.4881v2`.

Lemonte, A.J., Cribari–Neto, F., Vasconcellos, K.L.P. (2007). Improved statistical inference for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics and Data Analysis* **51**, 4656–4681.

Lemonte, A.J., Ferrari, S.L.P., Cribari–Neto, F. (2009). Improved likelihood inference in Birnbaum–Saunders regressions. Working paper `arXiv:0806.2208v2`.

Lemonte, A.J., Simas, A.B., Cribari–Neto, F. (2008). Bootstrap-based improved estimators for the two-parameter Birnbaum–Saunders distribution. *Journal of Statistical Computation and Simulation* **78**, 37–49.

Lesaffre, E., Verbeke, G. (1998). Local influence in linear mixed model. *Biometrics* **54**, 570–583.

Osorio, F., Paula, G.A., Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis* **51**, 4354–4368.

Paula, G.A. (1993). Assessing local influence in restricted regression models. *Computational Statistics and Data Analysis* **16**, 63–79.

Paula, G.A., Medeiros, M., Vilca-Labra, F.E. (2009). Influence diagnostics for linear models with first-order autoregressive elliptical errors. *Statistics and Probability Letters* **79**, 339–346.

Rieck, J.R. (1989). *Statistical Analysis for the Birnbaum–Saunders Fatigue Life Distribution*. Ph.D. thesis, Clemson University, Dept. of Mathematical Sciences.

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

Thomas, W., Cook, R.D. (1990). Assessing influence on predictions from generalized linear models. *Technometrics* **32**, 59–65.

Verbeke, G., Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.

Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**, 307–333.

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

Wei, B.C. (1998). *Exponential Family Nonlinear Models*. Singapore: Springer.

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