ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Birnbaum–Saunders nonlinear regression models

Artur J. Lemonte [a,*], Gauss M. Cordeiro [b]

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

[b] Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Recife/PE, 52171-900, Brazil

ARTICLE INFO

Article history:
Received 30 January 2009
Received in revised form 22 June 2009
Accepted 27 June 2009
Available online xxxx

ABSTRACT

We introduce, for the first time, a new class of Birnbaum–Saunders nonlinear regression models potentially useful in lifetime data analysis. The class generalizes the regression model described by Rieck and Nedelman [Rieck, J.R., Nedelman, J.R., 1991. A log-linear model for the Birnbaum–Saunders distribution. Technometrics 33, 51–60]. We discuss maximum-likelihood estimation for the parameters of the model, and derive closed-form expressions for the second-order biases of these estimates. Our formulae are easily computed as ordinary linear regressions and are then used to define bias corrected maximum-likelihood estimates. Some simulation results show that the bias correction scheme yields nearly unbiased estimates without increasing the mean squared errors. Two empirical applications are analysed and discussed.

1. Introduction

Different regression models have been proposed for lifetime data such as those based on the gamma, lognormal and Weibull distributions. These models typically provide a satisfactory fit in the middle portion of the data, but very often fail to deliver a good fit at the tails, where only a few observations are generally available. The family of distributions proposed by Birnbaum and Saunders (1969) can also be used to model lifetime data and it is widely applicable to model failure times of fatiguing materials. This family was originally obtained from a model for which failure follows from the development and growth of a dominant crack and has the appealing feature of providing satisfactory tail fitting. It was later derived by Desmond (1985) using a biological model which followed from relaxing some of the assumptions originally made by Birnbaum and Saunders (1969).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 be Birnbaum–Saunders distributed with parameters $\alpha, \eta > 0$, say $\mathcal{B}$-$\mathcal{S}(\alpha, \eta)$, if its cumulative distribution function (cdf) is given by

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

where $\Phi(\cdot)$ is the standard normal distribution function and α and η are shape and scale parameters, respectively. It is easy to show that η is the median of the distribution: $F_T(\eta) = \Phi(0) = 1/2$. For any $k > 0$, then $kT \sim \mathcal{B}$-$\mathcal{S}(\alpha, k\eta)$.

McCarter (1999) considered parameter estimation under type II data censoring for the $\mathcal{B}$-$\mathcal{S}(\alpha, \eta)$ distribution. Lemonte et al. (2007) derived the second-order biases of the maximum-likelihood estimates (MLEs) of α and η, and obtained a corrected likelihood ratio statistic for testing the parameter α. Lemonte et al. (2008) proposed several bootstrap bias corrected estimates of α and η. Further details on the Birnbaum–Saunders distribution can be found in Johnson et al. (1995).

* Corresponding author.
E-mail address: arturlemonte@gmail.com (A.J. Lemonte).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2009.06.015

Please cite this article in press as: Lemonte, A.J., Cordeiro, G.M., Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.06.015

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. Their model has been widely used as an alternative model to the gamma, lognormal and Weibull regression models; see Rieck and Nedelman (1991, Section 7). Diagnostic tools for the Birnbaum–Saunders regression model were developed by Galea et al. (2004), Leiva et al. (2007) and Xie and Wei (2007), and the Bayesian inference was considered by Tisionas (2001).

In this paper we propose a class of Birnbaum–Saunders nonlinear regression models which generalizes the regression model introduced by Rieck and Nedelman (1991). We discuss maximum-likelihood estimation of the regression parameters and obtain the Fisher information matrix. As is well known, however, the MLEs, although consistent, are typically biased in finite samples. In order to overcome this shortcoming, we derive a closed-form expression for the biases of the MLEs in these models which are used to define bias corrected estimates.

Bias adjustment has been extensively studied in the statistical literature. For example, Pázman and Denis (1999) derived expressions for the asymptotic approximation of the bias of the least-squares estimators in nonlinear regression models with parameters which are subject to nonlinear equality constraints. Cordeiro et al. (2000) obtained bias correction for symmetric nonlinear regression models. Vasconcellos and Silva (2005) discussed analytical bias corrections for MLEs in a regression model where the errors are Student-t distributed with unknown degrees of freedom. Ospina et al. (2006) derived closed-form expressions for the second-order biases of MLEs for beta regression models. Yang and Lin (2007) discussed improved maximum-likelihood estimation for the common shape parameter of several Weibull populations. Cordeiro and Demétrio (2008) proposed formulae for the second-order biases of the maximum quasi-likelihood estimates, whereas Cordeiro and Toyama-Udo (2008) derived the second-order biases in generalized nonlinear models with dispersion covariates. More recently, Patriota and Lemonte (2009) derive the second-order biases of the MLEs in a multivariate normal model where the mean vector and the covariance matrix have parameters in common, whereas Patriota et al. (2009) develops a bias correction scheme for a multivariate heteroskedastic errors-in-variables model.

The rest of the paper is organized as follows. Section 2 introduces the class of Birnbaum-Saunders nonlinear regression models and discusses maximum-likelihood estimation. Using general results from Cox and Snell (1968), we derive in Section 3 the second-order biases of the MLEs of the nonlinear parameters in our class of models and define bias corrected estimates. Some special models are considered in Section 4. Simulation results are presented and discussed in Section 5 for two nonlinear regression models. We show that the bias corrected estimates are nearly unbiased with mean squared errors very close to those of the uncorrected estimates. Section 6 gives an application of the proposed regression model to a real fatigue data set, which provides a better fit at the tail of the data. Finally, Section 7 concludes the paper.

2. Model specification

Let $T \sim \mathcal{B}\text{-}\mathcal{S}(\alpha, \eta)$. The density function of $Y = \log(T) \sim \mathcal{SN}(\alpha, \mu, \sigma)$ has the form (Rieck and Nedelman, 1991)

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

This distribution has a number of interesting properties (Rieck, 1989): (i) It is symmetric around the location parameter μ; (ii) It is unimodal for $\alpha \leq 2$ and bimodal for $\alpha > 2$; (iii) The mean and variance of Y are $\mathbb{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \sigma^2 w(\alpha)$, respectively. There is no closed-form expression for $w(\alpha)$, 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 $S_\alpha = 2(Y_\alpha - \mu)/(\alpha\sigma)$ converges in distribution to the standard normal distribution when $\alpha \to 0$.

We define the nonlinear 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, $\boldsymbol{x}_i$ is an $m \times 1$ vector of known explanatory variables associated with the ith observable response y_i, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ 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 $\boldsymbol{\beta}$. For the linear regression $\mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, the model (1) reduces to Rieck and Nedelman's (1991) model.

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

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log(\xi_{i1}) - \frac{1}{2}\sum_{i=1}^{n} \xi_{i2}^2, \tag{2}$$

where $\xi_{i1} = \xi_{i1}(\boldsymbol{\theta}) = 2\alpha^{-1}\cosh([y_i - \mu_i]/2)$, $\xi_{i2} = \xi_{i2}(\boldsymbol{\theta}) = 2\alpha^{-1}\sinh([y_i - \mu_i]/2)$ for $i = 1, \ldots, n$. The function $\ell(\boldsymbol{\theta})$ is assumed to be regular (Cox and Hinkley, 1974, Ch. 9) with respect to all $\boldsymbol{\beta}$ and α derivatives up to third order. Further, the $n \times p$ local matrix $\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{\beta}) = \partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}^\top$ of partial derivatives of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$ with respect to $\boldsymbol{\beta}$ is assumed to be of full rank, i.e., $\mathrm{rank}(\boldsymbol{D}) = p$ for all $\boldsymbol{\beta}$. The nonlinear predictors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are embedded in an infinite sequence of

Please cite this article in press as: Lemonte, A.J., Cordeiro, G.M., Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.06.015

$m \times 1$ vectors that must satisfy these regularity conditions for the asymptotics to be valid. Under these assumptions, the MLEs have good asymptotic properties such as consistency, sufficiency and normality.

The derivatives with respect to the components of $\boldsymbol{\beta}$ and α are denoted by: $U_r = \partial\ell(\boldsymbol{\theta})/\partial\beta_r$, $U_\alpha = \partial\ell(\boldsymbol{\theta})/\partial\alpha$, $U_{rs} = \partial^2\ell(\boldsymbol{\theta})/\partial\beta_r\partial\beta_s$, $U_{r\alpha} = \partial^2\ell(\boldsymbol{\theta})/\partial\beta_r\partial\alpha$, $U_{rst} = \partial^3\ell(\boldsymbol{\theta})/\partial\beta_r\partial\beta_s\partial\beta_t$, $U_{rs\alpha} = \partial^3\ell(\boldsymbol{\theta})/\partial\beta_r\partial\beta_s\partial\alpha$, etc. Further, we use the following notation for joint cumulants of log-likelihood derivatives: $\kappa_{rs} = \mathbb{E}(U_{rs})$, $\kappa_{r,\alpha} = \mathbb{E}(U_r U_\alpha)$, $\kappa_{rst} = \mathbb{E}(U_{rst})$, etc. Let $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t$, etc. All κ's and their derivatives are assumed to be of order $\mathcal{O}(n)$. Also, we adopt the notation $d_{ir} = \partial\mu_i/\partial\beta_r$ and $g_{irs} = \partial^2\mu_i/\partial\beta_r\partial\beta_s$ for the first and second partial derivatives of μ_i with respect to the elements of $\boldsymbol{\beta}$.

By differentiating (2) we have

$$U_r = \frac{1}{2}\sum_{i=1}^{n} d_{ir}\left(\xi_{i1}\xi_{i2} - \frac{\xi_{i2}}{\xi_{i1}}\right), \qquad U_\alpha = -\frac{n}{\alpha} + \frac{1}{\alpha}\sum_{i=1}^{n}\xi_{i2}^2,$$

$$U_{rs} = \frac{1}{2}\sum_{i=1}^{n} g_{irs}\left(\xi_{i1}\xi_{i2} - \frac{\xi_{i2}}{\xi_{i1}}\right) - \frac{1}{4}\sum_{i=1}^{n} d_{ir}d_{is}\left(2\xi_{i2}^2 + \frac{4}{\alpha^2} - 1 + \frac{\xi_{i2}^2}{\xi_{i1}^2}\right),$$

$$U_{r\alpha} = -\frac{1}{\alpha}\sum_{i=1}^{n} d_{ir}\xi_{i1}\xi_{i2} \quad \text{and} \quad U_{\alpha\alpha} = \frac{n}{\alpha^2} - \frac{3}{\alpha^2}\sum_{i=1}^{n}\xi_{i2}^2.$$

The score function for $\boldsymbol{\beta}$ is $\boldsymbol{U}_{\boldsymbol{\beta}} = \frac{1}{2}\boldsymbol{D}^\top\boldsymbol{s}$, where $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{\theta})$ is an n-vector whose ith element is equal to $\xi_{i1}\xi_{i2} - \xi_{i2}/\xi_{i1}$.

It is well known that, under general regularity conditions (Cox and Hinkley, 1974, Ch. 9), the MLEs are consistent, asymptotically efficient and asymptotically normal. Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\alpha})^\top$ be the MLE of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha)^\top$. We can write $\widehat{\boldsymbol{\theta}} \overset{a}{\sim} \mathcal{N}_{p+1}(\boldsymbol{\theta}, \boldsymbol{K}_{\boldsymbol{\theta}}^{-1})$ for n large, where $\overset{a}{\sim}$ denotes approximately distributed, $\boldsymbol{K}_{\boldsymbol{\theta}}$ is the block-diagonal Fisher information matrix given by $\boldsymbol{K}_{\boldsymbol{\theta}} = \mathrm{diag}\{\boldsymbol{K}_{\boldsymbol{\beta}}, \kappa_{\alpha,\alpha}\}$, $\boldsymbol{K}_{\boldsymbol{\theta}}^{-1}$ is its inverse, $\boldsymbol{K}_{\boldsymbol{\beta}} = \psi(\alpha)(\boldsymbol{D}^\top\boldsymbol{D})/4$ is the information matrix for $\boldsymbol{\beta}$ and $\kappa_{\alpha,\alpha} = 2n/\alpha^2$ is the information for α. Also,

$$\psi(\alpha) = 2 + \frac{4}{\alpha^2} - \frac{\sqrt{2\pi}}{\alpha}\left\{1 - \mathtt{erf}\left(\frac{\sqrt{2}}{\alpha}\right)\right\}\exp\left(\frac{2}{\alpha^2}\right),$$

where $\mathtt{erf}(\cdot)$ is the error function given by $\mathtt{erf}(x) = (2/\sqrt{\pi})\int_0^x \mathrm{e}^{-t^2}\,\mathrm{d}t$. Details on $\mathtt{erf}(\cdot)$ can be found in Gradshteyn and Ryzhik (2007). Since $\boldsymbol{K}_{\boldsymbol{\theta}}$ is block-diagonal, the vector $\boldsymbol{\beta}$ and the scalar α are globally orthogonal (Cox and Reid, 1987) and $\widehat{\boldsymbol{\beta}}$ and $\widehat{\alpha}$ are asymptotically independent. It can be shown (Rieck, 1989) that $\psi(\alpha) \approx 1 + 4/\alpha^2$ for α small and $\psi(\alpha) \approx 2$ for α large.

The MLE $\widehat{\boldsymbol{\theta}}$ satisfies $p+1$ equations $U_r = U_\alpha = 0$ for the components of $\boldsymbol{\beta}$ and α. The Fisher scoring method can be used to estimate $\boldsymbol{\beta}$ and α simultaneously by iteratively solving the equations

$$\boldsymbol{\beta}^{(m+1)} = (\boldsymbol{D}^{(m)\top}\boldsymbol{D}^{(m)})^{-1}\boldsymbol{D}^{(m)\top}\boldsymbol{\delta}^{(m)} \quad \text{and} \quad \alpha^{(m+1)} = \frac{\alpha^{(m)}}{2}\left(1 + \bar{\xi}_2^{(m)}\right),$$

where $\boldsymbol{\delta}^{(m)} = \boldsymbol{D}^{(m)}\boldsymbol{\beta}^{(m)} + 2\boldsymbol{s}^{(m)}/\psi(\alpha^{(m)})$ and $\bar{\xi}_2^{(m)} = \sum_{i=1}^{n}\xi_{i2}^{2(m)}/n$ for $m = 0, 1, 2, \ldots$

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. Initial approximations $\boldsymbol{\beta}^{(0)}$ and $\alpha^{(0)}$ for the iterative algorithm are used to evaluate $\boldsymbol{D}^{(0)}$, $\boldsymbol{\delta}^{(0)}$ and $\bar{\xi}_2^{(0)}$ from which these equations can be used to obtain the next estimates $\boldsymbol{\beta}^{(1)}$ and $\alpha^{(1)}$. These new values can update $\boldsymbol{D}$, $\boldsymbol{\delta}$ and $\bar{\xi}_2$ and so the iterations continue until convergence is achieved.

3. Biases of estimates of $\boldsymbol{\beta}$ and α

We now obtain some joint cumulants of log-likelihood derivatives and their derivatives:

$$\kappa_{rs} = -\frac{\psi(\alpha)}{4}\sum_{i=1}^{n} d_{ir}d_{is}, \qquad \kappa_{r\alpha} = \kappa_{r\alpha\alpha} = 0, \qquad \kappa_{\alpha\alpha} = -\frac{2n}{\alpha^2}, \qquad \kappa_{\alpha\alpha\alpha} = \frac{10n}{\alpha^3},$$

$$\kappa_{rst} = -\frac{\psi(\alpha)}{4}\sum_{i=1}^{n}(g_{irs}d_{it} + g_{irt}d_{is} + g_{ist}d_{ir}), \qquad \kappa_{rs\alpha} = \frac{(2+\alpha^2)}{\alpha^3}\sum_{i=1}^{n} d_{ir}d_{is},$$

$$\kappa_{rs}^{(t)} = -\frac{\psi(\alpha)}{4}\sum_{i=1}^{n}(g_{irt}d_{is} + g_{ist}d_{ir}), \qquad \kappa_{r\alpha}^{(\alpha)} = \kappa_{r\alpha}^{(s)} = 0 \quad \text{and} \quad \kappa_{\alpha\alpha}^{(\alpha)} = \frac{4n}{\alpha^3}.$$

Let $B(\widehat{\beta}_a)$ and $B(\widehat{\alpha})$ be the n^{-1} biases of $\widehat{\beta}_a$ $(a = 1, \ldots, p)$ and $\widehat{\alpha}$, respectively. The use of Cox and Snell's (1968) formula to obtain these biases is greatly simplified, since $\boldsymbol{\beta}$ and α are globally orthogonal and the cumulants corresponding to the

Please cite this article in press as: Lemonte, A.J., Cordeiro, G.M., Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.06.015

parameters in $\boldsymbol{\beta}$ are invariant under permutation of these parameters. From now on, we use Einstein summation convention with the indices varying over the corresponding parameters. We have

$$B(\widehat{\beta}_a) = \sum_{s,t,u}{}' \kappa^{a,s}\kappa^{t,u}\left(\kappa_{st}^{(u)} - \frac{1}{2}\kappa_{stu}\right) + \kappa^{\alpha,\alpha}\sum_{s}{}' \kappa^{a,s}\left(\kappa_{s\alpha}^{(\alpha)} - \frac{1}{2}\kappa_{s\alpha\alpha}\right)$$

and

$$B(\widehat{\alpha}) = (\kappa^{\alpha,\alpha})^2\left(\kappa_{\alpha\alpha}^{(\alpha)} - \frac{1}{2}\kappa_{\alpha\alpha\alpha}\right) + \kappa^{\alpha,\alpha}\sum_{t,u}{}' \kappa^{t,u}\left(\kappa_{\alpha t}^{(u)} - \frac{1}{2}\kappa_{\alpha tu}\right),$$

where $\kappa^{r,s}$ is the (r,s)th element of the inverse $\boldsymbol{K}_{\boldsymbol{\beta}}^{-1}$ of the information matrix for $\boldsymbol{\beta}$, $\kappa^{\alpha,\alpha} = \kappa_{\alpha,\alpha}^{-1}$ and $\sum'$ denotes the summation over all combinations of parameters $\beta_1, \ldots, \beta_p$. Plugging the cumulants given before into these two expressions and after some algebra we can obtain the n^{-1} bias of $\widehat{\boldsymbol{\beta}}$, say $\boldsymbol{B}(\widehat{\boldsymbol{\beta}})$, in matrix form, and $B(\widehat{\alpha})$.

We now define the matrix $\boldsymbol{G}$ $(n \times p^2)$ from the elements of $\boldsymbol{\beta}$ by $\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\beta}) = \{\partial^2\mu_i/\partial\beta_r\partial\beta_s\}$ (for $r, s = 1, 2, \ldots, p$). Thus, the ith row of $\boldsymbol{G}$ is given by $\partial^2\mu_i/\partial\beta_1^2$, $\partial^2\mu_i/\partial\beta_1\partial\beta_2, \ldots, \partial^2\mu_i/\partial\beta_p^2$. We can show after some algebra that the $p \times 1$ bias vector $\boldsymbol{B}(\widehat{\boldsymbol{\beta}})$ reduces to

$$\boldsymbol{B}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{D}^\top\boldsymbol{D})^{-1}\boldsymbol{D}^\top\boldsymbol{d}, \tag{3}$$

where $\boldsymbol{d}$ is an $n \times 1$ vector defined as $\boldsymbol{d} = -\{2/\psi(\alpha)\}\boldsymbol{G}\mathrm{vec}\{(\boldsymbol{D}^\top\boldsymbol{D})^{-1}\}$ and "vec" is the operator which transforms a matrix into a vector by stacking the columns of the matrix one underneath the other. The n^{-1} bias of $\widehat{\alpha}$ is given by simple formula

$$B(\widehat{\alpha}) = -\frac{1}{n}\left\{p\left(\frac{2+\alpha^2}{\alpha\psi(\alpha)}\right) + \frac{\alpha}{4}\right\}. \tag{4}$$

A number of remarks are worth making with respect to formulae (3) and (4). The bias vector $\boldsymbol{B}(\widehat{\boldsymbol{\beta}})$ can be obtained from a simple ordinary least-squares regression of $\boldsymbol{d}$ on the columns of $\boldsymbol{D}$. It depends on the nonlinearity of the regression function f and the parameter α. The bias vector $\boldsymbol{B}(\widehat{\boldsymbol{\beta}})$ will be small when $\boldsymbol{d}$ is orthogonal to the columns of $\boldsymbol{D}$. Also, it can be large when $\psi(\alpha)$ and n are both small. Eq. (3) is easily handled algebraically for any type of nonlinear regression, since it involves simple operations on matrices and vectors. For special models with closed-form information matrix for $\boldsymbol{\beta}$, it is possible to obtain closed-form expressions for $\boldsymbol{B}(\widehat{\boldsymbol{\beta}})$. For linear models, the matrix $\boldsymbol{G}$ and the vector $\boldsymbol{d}$ vanish and hence $\boldsymbol{B}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{0}$, which is in agreement with the result due to Rieck and Nedelman (1991, p. 54) that the MLEs are unbiased to order n^{-1}. Expression (4) depends directly on the nonlinear structure of the regression model only through the rank p of $\boldsymbol{D}$. It shows that the bias is always a linear function of the dimension p of $\boldsymbol{\beta}$.

On the right-hand sides of expressions (3) and (4), which are both of order n^{-1}, consistent estimates of the parameters $\boldsymbol{\beta}$ and α can be inserted to define bias corrected estimates $\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{B}}(\widehat{\boldsymbol{\beta}})$ and $\widetilde{\alpha} = \widehat{\alpha} - \widehat{B}(\widehat{\alpha})$, where $\widehat{\boldsymbol{B}}(\widehat{\boldsymbol{\beta}})$ and $\widehat{B}(\widehat{\alpha})$ are the values of $\boldsymbol{B}(\widehat{\boldsymbol{\beta}})$ and $B(\widehat{\alpha})$, respectively, at $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\alpha})^\top$. The bias corrected estimates $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\alpha}$ are expected to have better sampling properties than the classical MLEs $\widehat{\boldsymbol{\beta}}$ and $\widehat{\alpha}$. In fact, we present some simulations in Section 5 to show that $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\alpha}$ have smaller biases than their corresponding uncorrected estimates, thus suggesting that these bias corrections have the effect of shrinking the adjusted estimates toward to the true parameter values. However, we cannot say that the bias corrected estimates offer always some improvement over the MLEs, since they can have mean squared errors larger.

It is worth emphasizing that there are other methods to obtain bias corrected estimates. In regular parametric problems, Firth (1993) developed the so-called "preventive" method, which also allows for the removal of the second-order bias. His method consists of modifying the original score function to remove the first-order term from the asymptotic biases of these estimates. In exponential families with canonical parameterizations, his correction scheme consists in penalizing the likelihood by the Jeffreys invariant priors. This is a preventive approach to bias adjustment which has its merits, but the connections between our results and his work are not pursued in this paper since they could be developed in future research. Additionally, it should be mentioned that it is possible to avoid cumbersome and tedious algebra on cumulant calculations by using Efron's bootstrap (Efron and Tibshirani, 1993). We use the analytical approach here since this leads to a nice formula. Moreover, the application of the analytical bias approximation seems to generally be the most feasible procedure to use and it continues to receive attention in the literature.

We now calculate the second-order bias $B(\widehat{\mu}_i)$ of the MLE $\widehat{\mu}_i$ of the ith mean $\mu_i = f_i(\boldsymbol{x}_i; \boldsymbol{\beta})$. We can easily show by Taylor series expansion that

$$B(\widehat{\mu}_i) = \boldsymbol{d}_i^\top\boldsymbol{B}(\widehat{\boldsymbol{\beta}}) + \frac{1}{2}\mathrm{tr}\{\boldsymbol{M}_i\mathrm{Cov}(\widehat{\boldsymbol{\beta}})\},$$

where $\boldsymbol{d}_i^\top$ is a $1 \times p$ vector of first partial derivatives $\partial\mu_i/\partial\beta_r$ (for $r = 1, \ldots, p$), $\boldsymbol{M}_i$ is a $p \times p$ matrix of second partial derivatives $\partial^2\mu_i/\partial\beta_r\partial\beta_s$ (for $r, s = 1, \ldots, p$), $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{K}_{\boldsymbol{\beta}}^{-1}$ is the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{B}(\widehat{\boldsymbol{\beta}})$ was given previously. All quantities in the last equation should be evaluated at $\widehat{\boldsymbol{\beta}}$.

The asymptotic variance of $\widehat{\mu}_i$ can also be expressed explicitly in terms of the covariance of $\widehat{\boldsymbol{\beta}}$ by

$$\mathrm{Var}(\widehat{\mu}_i) = \mathrm{tr}\{(\boldsymbol{d}_i\boldsymbol{d}_i^\top)\mathrm{Cov}(\widehat{\boldsymbol{\beta}})\}.$$

Please cite this article in press as: Lemonte, A.J., Cordeiro, G.M., Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.06.015

4. Special models

Eq. (3) is easily handled algebraically for any type of nonlinear model, since it involves simple operations on matrices and vectors. This equation, in conjunction with a computer algebra system such as MAPLE (Abell and Braselton, 1994) will compute $\boldsymbol{B}(\widehat{\boldsymbol{\beta}})$ algebraically with minimal effort. In particular, Eq. (3) may simplify considerably if the number of nonlinear parameters is small. Moreover, for any special nonlinear model, we can calculate the bias $\boldsymbol{B}(\widehat{\boldsymbol{\beta}})$ numerically via a software with numerical linear algebra facilities such as Ox (Doornik, 2006) and R (R Development Core Team, 2008).

First, we consider a nonlinear regression model which depends on a single nonlinear parameter. Eq. (3) gives

$$B(\widehat{\beta}) = -\frac{2}{\psi(\alpha)}\frac{\kappa_2}{\kappa_1^2},$$

where $\kappa_1 = \sum_{i=1}^{n}(\mathrm{d}f_i/\mathrm{d}\beta)^2$ and $\kappa_2 = \sum_{i=1}^{n}(\mathrm{d}f_i/\mathrm{d}\beta)(\mathrm{d}^2 f_i/\mathrm{d}\beta^2)$. The constants κ_1 and κ_2 are evaluated at $\widehat{\beta}$ and $\widehat{\alpha}$ to yield $\widehat{B}(\widehat{\beta})$ and the corrected estimate $\widetilde{\beta} = \widehat{\beta} - \widehat{B}(\widehat{\beta})$. For example, the simple exponential model $f_i = \exp(\beta x_i)$ leads to $\kappa_1 = \sum_{i=1}^{n} x_i^2 \exp(2\beta x_i)$ and $\kappa_2 = \sum_{i=1}^{n} x_i^3 \exp(2\beta x_i)$.

As a second example, we consider a partially nonlinear regression model defined by

$$\boldsymbol{\mu} = \boldsymbol{Z}\boldsymbol{\lambda} + \eta \boldsymbol{g}(\gamma), \tag{5}$$

where $\boldsymbol{Z}$ is a known $n \times (p-2)$ matrix of full rank, $\boldsymbol{g}(\gamma)$ is an $n \times 1$ vector, $\boldsymbol{\beta} = (\boldsymbol{\lambda}^\top, \eta, \gamma)^\top$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{p-2})^\top$ and η and γ are scalar parameters. This class of models occurs very often in statistical modeling. For example, $\mu = \lambda_1 z_1 + \lambda_2 z_2 + \eta \exp(\gamma x)$ (Gallant, 1975), $\mu = \lambda - \eta \log(x_1 + \gamma x_2)$ (Darby and Ellis, 1976) and $\mu = \lambda + \eta \log(x_1/(\gamma + x_2))$ (Stone, 1980). Ratkowsky (1983, Ch. 5) discusses several models of the form (5) which include the asymptotic regression and Weibull-type models given by $\mu = \lambda - \eta\gamma^x$ and $\mu = \lambda - \eta \exp(-\gamma x)$, respectively.

The $n \times p$ local model matrix $\boldsymbol{D}$ takes the form $\boldsymbol{D} = [\boldsymbol{Z}, \boldsymbol{g}(\gamma), \eta(\mathrm{d}\boldsymbol{g}(\gamma)/\mathrm{d}\gamma)]$ and, after some algebra, we can obtain from Eq. (3)

$$\boldsymbol{B}(\widehat{\boldsymbol{\beta}}) = -\left[\frac{1}{\eta}\mathrm{Cov}(\widehat{\eta}, \widehat{\gamma})\boldsymbol{\tau}_p + \frac{\eta}{2}\mathrm{Var}(\widehat{\gamma})\boldsymbol{\delta}_p\right], \tag{6}$$

where $\boldsymbol{\tau}_p$ is a $p \times 1$ vector with a one in the last position and zeros elsewhere, $\boldsymbol{\delta}_p = (\boldsymbol{D}^\top\boldsymbol{D})^{-1}\boldsymbol{D}^\top(\mathrm{d}^2\boldsymbol{g}(\gamma)/\mathrm{d}\gamma^2)$ is simply the set of coefficients from the ordinary regression of the vector $\mathrm{d}^2\boldsymbol{g}(\gamma)/\mathrm{d}\gamma^2$ on the matrix $\boldsymbol{D}$, and $\mathrm{Var}(\widehat{\gamma})$ and $\mathrm{Cov}(\widehat{\eta}, \widehat{\gamma})$ are the large-sample second moments obtained from the appropriate elements of the asymptotic covariance matrix $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{K}_{\boldsymbol{\beta}}^{-1} = (4/\psi(\alpha))(\boldsymbol{D}^\top\boldsymbol{D})^{-1}$. It is clear from formula (6) that $\boldsymbol{B}(\widehat{\boldsymbol{\beta}})$ does not depend explicitly on the linear parameters in $\boldsymbol{\lambda}$ and it is proportional to $4/\psi(\alpha)$. Further, the covariance term $\mathrm{Cov}(\widehat{\eta}, \widehat{\gamma})$ contributes only to the bias of $\widehat{\gamma}$.

5. Numerical results

We now use Monte Carlo simulation to evaluate the finite-sample performance of the MLEs of the parameters and of their corrected versions in two nonlinear regression models. The estimates of the parameters were obtained by maximizing the log-likelihood function using the BFGS quasi-Newton method with analytical derivatives. This method is generally regarded as the best-performing nonlinear optimization method (Mittelhammer et al., 2000, p. 199). The covariate values were selected as random draws from the uniform $\mathcal{U}(0, 1)$ distribution and for fixed n those values were kept constant throughout the experiment. The number of Monte Carlo replications was 10,000 and all simulations were performed using the Ox matrix programming language (Doornik, 2006).[1]

In order to analyze the performance of the estimates, we computed, for each sample size and for each estimate: the relative bias (the relative bias of an estimate $\widehat{\theta}$, defined as $\{\mathbb{E}(\widehat{\theta}) - \theta\}/\theta$, is obtained by estimating $\mathbb{E}(\widehat{\theta})$ by Monte Carlo) and the root mean square error ($\sqrt{\mathrm{MSE}}$), where MSE is the estimated mean square error from the 10,000 Monte Carlo replications.

First, we consider the nonlinear regression model

$$\mu_i = \lambda_1 z_{i1} + \lambda_2 z_{i2} + \eta \exp(\gamma x_i),$$

where $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$ for $i = 1, \ldots, n$. The sample sizes were $n = 15$, 30 and 45. Without loss of generality, the true values of the regression parameters were taken as $\lambda_1 = 4$, $\lambda_2 = 5$, $\eta = 3$, $\gamma = 1.5$ and $\alpha = 0.5$ and 1.5.

Table 1 gives the relative biases of both uncorrected and corrected estimates to show that the bias corrected estimates are much closer to the true parameters than the unadjusted estimates. For instance, when $n = 15$ and $\alpha = 1.5$, the average of the estimated relative biases for the estimates of the model parameters is -0.03244, whereas the average of the estimated relative biases for the corrected estimates is -0.0083. Hence, the average bias (in absolute value) of the MLEs is almost four

[1] Ox is freely distributed for academic purposes and available at http://www.doornik.com.

Please cite this article in press as: Lemonte, A.J., Cordeiro, G.M., Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.06.015

Table 1
Relative biases of the uncorrected and corrected estimates.

α	n		λ_1	λ_2	η	γ	α
0.5	15	MLE	0.0006	−0.0013	0.0011	0.0020	−0.1691
		BCE	0.0007	−0.0011	0.0001	0.0008	−0.0395
	30	MLE	0.0001	−0.0013	0.0013	0.0009	−0.0811
		BCE	0.0002	−0.0012	0.0007	−0.0001	−0.0092
	45	MLE	0.0003	−0.0012	0.0007	0.0008	−0.0537
		BCE	0.0003	−0.0011	0.0003	0.0001	−0.0042
1.5	15	MLE	−0.0068	−0.0083	0.0248	0.0197	−0.1916
		BCE	−0.0055	−0.0046	0.0113	0.0056	−0.0481
	30	MLE	−0.0016	−0.0034	0.0079	0.0078	−0.0933
		BCE	−0.0011	−0.0018	0.0027	0.0012	−0.0116
	45	MLE	−0.0028	−0.0027	0.0052	0.0026	−0.0614
		BCE	−0.0023	−0.0018	0.0023	−0.0005	−0.0048

BCE: bias corrected estimate.

Table 2
Root mean squared errors of the uncorrected and corrected estimates.

α	n		λ_1	λ_2	η	γ	α
0.5	15	MLE	0.4093	0.4920	0.2707	0.0924	0.1234
		BCE	0.4093	0.4921	0.2709	0.0922	0.1067
	30	MLE	0.3006	0.3806	0.2113	0.0688	0.0763
		BCE	0.3006	0.3806	0.2114	0.0686	0.0702
	45	MLE	0.2434	0.2874	0.1768	0.0567	0.0590
		BCE	0.2434	0.2874	0.1769	0.0566	0.0555
1.5	15	MLE	1.6302	1.1230	0.9756	0.3235	0.3938
		BCE	1.6333	1.1274	0.9819	0.3152	0.3315
	30	MLE	0.9684	0.7003	0.5785	0.1931	0.2399
		BCE	0.9693	0.7011	0.5807	0.1908	0.2155
	45	MLE	0.6505	0.5575	0.3895	0.1318	0.1837
		BCE	0.6507	0.5577	0.3901	0.1311	0.1700

BCE: bias corrected estimate.

times greater than the average bias of the corrected estimates. This fact suggests that the second-order biases of the MLEs should not be ignored in samples of small to moderate size since they can be non-negligible. Table 2 shows that the root mean squared errors of the uncorrected and corrected estimates are very close. Hence, the figures in both tables suggest that the corrected estimates have good properties.

When the parameter α increases, the finite-sample performance of the MLEs deteriorates (see Tables 1 and 2). For instance, when $n = 15$, the relative biases of $\widehat{\gamma}$ (MLE) and $\widetilde{\gamma}$ (BCE) were 0.0020 and 0.0008 (for $\alpha = 0.5$) and 0.0197 and 0.0056 (for $\alpha = 1.5$), which indicate an increase in the relative biases of nearly 10 and 7 times, respectively. Also, the root mean squared errors in the same order were 0.0924 and 0.0922 (for $\alpha = 0.5$) and 0.3235 and 0.3152 (for $\alpha = 1.5$).

Next, we consider the very useful Michaelis–Menton model for estimating growth curves, where it is common for the response to approach an asymptote as the stimulus increases. The Michaelis–Menton model (Ratkowsky, 1983) provides an hyperbolic form for μ_i against x_i given by

$$\mu_i = \frac{\eta x_i}{\gamma + x_i}, \quad i = 1, \ldots, n,$$

where the curve has an asymptote at $\mu = \eta$. Here, the sample sizes were $n = 20$, 30, 40 and 50. Also, the true values of the nonlinear parameters were taken as $\eta = 3$ and $\gamma = 0.5$, with $\alpha = 0.5$.

Table 3 gives the relative biases and root mean squared errors of the uncorrected and corrected estimates. The figures in this table reveal that the MLEs of the parameters can be substantially biased, even when $n = 50$, and that the bias correction is very effective. In terms of MSE, the adjusted estimates are slightly better than the ordinary MLEs.

Following Paolino (2001), we compute an efficiency measure for the corrected estimator (BCE) relative to the MLE by comparing the average values of the first differences for the mean response (for the Michaelis–Menton model) from the estimates of the parameters with and without bias correction with the *true* values of the first differences. This measure is related to the estimation of the impact of a covariate change on the mean. Measuring such impact is oftentimes of interest. The first differences are computed by varying the covariate from one standard deviation above its average value to one

Please cite this article in press as: Lemonte, A.J., Cordeiro, G.M., Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.06.015

Table 3
Relative biases and root mean squared errors of the uncorrected and corrected estimates; $\alpha = 0.5$ and different sample sizes.

n		Relative Bias			$\sqrt{\text{MSE}}$		
		η	γ	α	η	γ	α
20	MLE	0.0476	0.1718	−0.0669	0.6984	0.3947	0.0859
	BCE	−0.0016	−0.0081	−0.0061	0.5264	0.2783	0.0847
30	MLE	0.0313	0.1077	−0.0439	0.5245	0.2750	0.0684
	BCE	0.0004	0.0012	−0.0024	0.4478	0.2252	0.0678
40	MLE	0.0215	0.0754	−0.0330	0.4222	0.2207	0.0582
	BCE	−0.0001	−0.0003	−0.0015	0.3835	0.1954	0.0578
50	MLE	0.0160	0.0558	−0.0259	0.3609	0.1862	0.0516
	BCE	0.0000	−0.0001	−0.0005	0.3380	0.1710	0.0514

BCE: bias corrected estimate.

Table 4
Efficiency with respect to the MLE: mean change.

n	20	25	30	35	40	50	80	100
EFFIC	110.50	109.00	107.46	105.37	105.21	104.30	102.91	102.12

standard deviation below it. That is, we compute

$$\text{EFFIC} = 100 \times \frac{\sqrt{\sum_{l=1}^{R}(\Delta_{l,\text{MLE}} - \Delta_{\text{true}})^2}}{\sqrt{\sum_{l=1}^{R}(\Delta_{l,\text{BCE}} - \Delta_{\text{true}})^2}},$$

where Δ_{true} is the true value of the first difference, i.e., $\Delta_{\text{true}} = \Delta\mu$, Δ_{MLE} is the estimate of the first difference evaluated at the MLEs, Δ_{BCE} is the estimate using the bias corrected estimates and R is the total number of Monte Carlo replications. Eight different sample sizes were considered: $n = 20, 25, 30, 35, 40, 50, 80$ and 100. The values of the parameters (η, γ and α) are as those considered above. Numerical values greater than 100 indicate that the bias corrected estimator is more efficient, in the sense discussed above, than the classical estimator. The figures in Table 4 show that the adjusted bias estimator is more efficient than the MLE. As expected, when the sample size increases the statistic EFFIC approaches 100.

6. Applications

Obviously, due to the genesis of the Birnbaum–Saunders distribution, the fatigue processes are by excellence ideally modeled by this model. First, we consider 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:

$$y_i = \beta_1 + \beta_2 \log(w_i) + \varepsilon_i, \tag{7}$$

where $y_i = \log(N_i)$ and $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$ for $i = 1, \ldots, 46$. The MLEs (the corresponding standard errors in parentheses) are: $\widehat{\beta}_1 = 12.2797$ (0.3942), $\widehat{\beta}_2 = -1.6708$ (0.1096) and $\widehat{\alpha} = 0.4104$ (0.0428). We take the logarithm of w to ensure a linear relationship between the response variable (y) and the covariate in (7); see Galea et al. (2004, Fig. 1). However, Fig. 1 suggests a nonlinear relationship between the response variable and the covariate w.

Hence, we propose the nonlinear regression model

$$y_i = \beta_1 + \beta_2 \exp(\beta_3/w_i) + \varepsilon_i, \quad i = 1, \ldots, 46, \tag{8}$$

where $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$. The MLEs (the standard errors in parentheses) 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). The bias corrected estimates are: $\widetilde{\beta}_1 = 8.7806$ (0.7734), $\widetilde{\beta}_2 = -4.9362$ (0.5266), $\widetilde{\beta}_3 = -22.1713$ (7.6548) and $\widetilde{\alpha} = 0.4157$ (0.0433). The uncorrected estimates are slightly different from the bias corrected estimates even for large samples ($n = 46$ observations).

Fig. 2 gives the scatter-plot of the data, the fitted model (8) and the fitted straight line, say $y_i = \beta_1 + \beta_2 w_i + \varepsilon_i$, where the MLEs are: $\widehat{\beta}_1 = 7.9864$ (0.1622), $\widehat{\beta}_2 = -0.0406$ (0.0036) and $\widehat{\alpha} = 0.52$ (0.0542). Fig. 2 shows that the nonlinear model (8) (unlike the linear model) fits satisfactorily to the fatigue data.

Following Xie and Wei (2007), we obtain the residuals $\widehat{\varepsilon}_i = y_i - \widehat{\mu}_i$ and $\widehat{R}_i = 2\widehat{\alpha}^{-1}\sinh(\widehat{\varepsilon}_i/2)$. Fig. 3 gives the scatter-plot of $\widehat{R}_i$ versus the predicted values $\widehat{\mu}_i$ for both fitted models: (i) $y_i = \beta_1 + \beta_2 w_i + \varepsilon_i$; and (ii) $y_i = \beta_1 + \beta_2 \exp(\beta_3/w_i) + \varepsilon_i$.

Please cite this article in press as: Lemonte, A.J., Cordeiro, G.M., Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.06.015

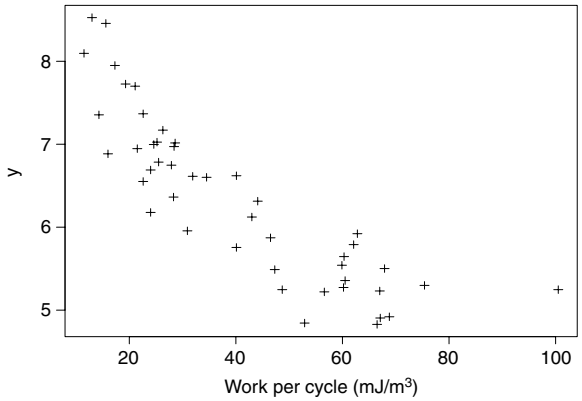

Fig. 1. Scatter-plot of the data set.

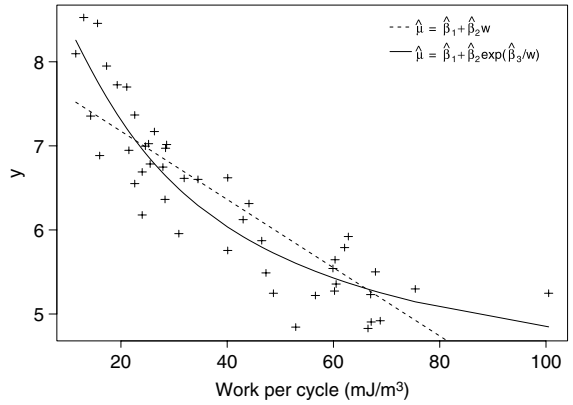

Fig. 2. Scatter-plot and the fitted models.

Table 5
MLEs of the parameters, relative changes (%) on $\boldsymbol{\beta}$ dropping 46th observation and the corresponding p-values.

Parameter	Linear model			Nonlinear model		
	Estimate	RC	p-value	Estimate	RC	p-value
β_1	8.1906	−2.56	0.00	8.8005	2.08	0.00
β_2	−0.0469	15.5	0.00	−5.1695	0.21	0.00
β_3	–	–	–	−25.0287	−11.2	0.00

Fig. 3 shows that the distribution of $\widehat{R}_i$ is approximately normal for model (ii) but this is not true for model (i). 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.

The 46th case (the one with work per cycle near 100) can be an influential observation. We deleted this observation and the estimates for the models $y_i = \beta_1 + \beta_2 w_i + \varepsilon_i$ and $y_i = \beta_1 + \beta_2 \exp(\beta_3/w_i) + \varepsilon_i$ are presented in Table 5, together with the relative changes (in percentage) of each parameter estimate, defined by $\mathrm{RC}_{\theta_r} = \{(\widehat{\theta}_r - \widehat{\theta}_{r[i]})/\widehat{\theta}_r\} \times 100\%$, where $\widehat{\theta}_{r[i]}$ denotes the MLE of θ_r after the observation ith being removed. We note that the 46th observation changes the estimate of β_3 more than 11% but the significance of β_3 is not modified.

A generalized LR statistic can be used for discriminating among nonnested models as discussed in the book of Camerom and Trived (1998, p. 184). 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 $\mu_{1i} = \mu_{1i}(\boldsymbol{x}_i; \boldsymbol{\theta})$ and $\mu_{2i} = \mu_{2i}(\boldsymbol{x}_i; \boldsymbol{\gamma})$. This statistic is a distance between the two models measured in terms of the Kullback–Liebler information criterion. The generalized LR statistic can be written as

$$T_{LR,NN} = \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\log\frac{\pi(y_i|\widehat{\mu}_{1i})}{\pi(y_i|\widehat{\mu}_{2i})}\right\} \times \left\{\frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{\pi(y_i|\widehat{\mu}_{1i})}{\pi(y_i|\widehat{\mu}_{2i})}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\log\frac{\pi(y_i|\widehat{\mu}_{1i})}{\pi(y_i|\widehat{\mu}_{2i})}\right)^2\right\}^{-1/2}, \tag{9}$$

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 (9) converges in distribution to a standard normal distribution under the null hypothesis of equivalence of the models (Vuong, 1989). Thus,

Please cite this article in press as: Lemonte, A.J., Cordeiro, G.M., Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.06.015

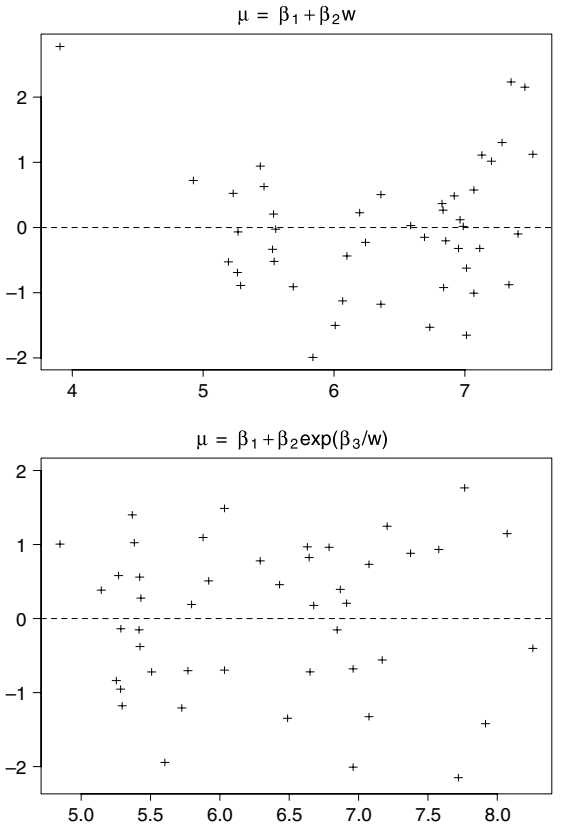

Fig. 3. Index plot of $\widehat{R}_i$ versus $\widehat{\mu}_i$.

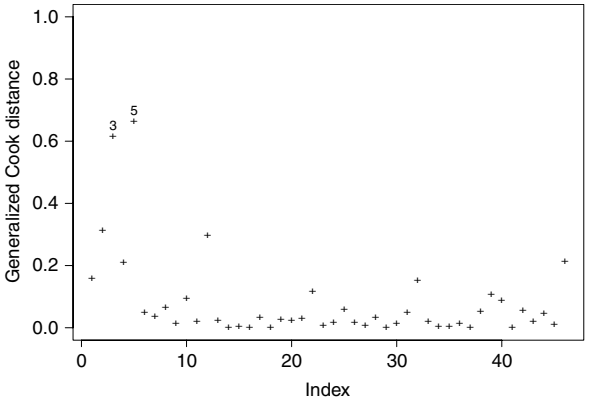

Fig. 4. Index plot of generalized Cook distance for the estimate $\widehat{\boldsymbol{\beta}}$.

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 $T_{LR,NN} < -z_\rho$).

We now use (9) for comparing the linear and nonlinear models fitted to the data. Let $\widehat{\mu}_{1i} = \widehat{\beta}_1 + \widehat{\beta}_2 \exp(\widehat{\beta}_3/w_i)$ (nonlinear model) and $\widehat{\mu}_{2i} = \widehat{\beta}_1 + \widehat{\beta}_2 w_i$ (linear model). The generalized LR test statistic ($T_{LR,NN}$) equals 2.4064 (p-value < 0.01). Therefore, the nonlinear model is significantly better than the linear model according to the generalized LR statistic.

Fig. 4 gives the generalized Cook distance for the nonlinear model (8) and shows that the 46th observation is not influential. The Cook distance identifies the cases 3 and 5 as possible influential observations on $\widehat{\boldsymbol{\beta}}$. Again, we eliminate the most influential observations (3 and 5) and refitted the model. Relative changes on the estimates of $\boldsymbol{\beta}$ are given in Table 6. We note that the observation 3 changes more than 11% the estimate of β_3 but the significance of β_3 is not modified.

Please cite this article in press as: Lemonte, A.J., Cordeiro, G.M., Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.06.015

Table 6
Relatives changes (%) on $\boldsymbol{\beta}$ dropping the cases indicated.

Eliminated	$\widehat{\beta}_1$		$\widehat{\beta}_2$		$\widehat{\beta}_3$	
	RC	p-value	RC	p-value	RC	p-value
None	0.00	0.00	0.00	0.00	0.00	0.00
3	−4.09	0.00	−5.84	0.00	11.2	0.00
5	−3.44	0.00	−5.63	0.00	7.96	0.00

Table 7
Fatigue data for SAE 1137 carbon steel.

	Strain amplitude	Fatigue life (cycles)
1	0.00188	336546
2	0.00200	218749
3	0.00225	92467
4	0.00250	66838
5	0.00300	38552
6	0.00350	20884
7	0.00400	15264
8	0.00450	10035
9	0.00500	7384
10	0.00600	5502
11	0.00700	3699
12	0.00800	2351
13	0.00900	2117

Table 8
MLEs of the parameters, relative changes (%) on the estimates of β_r by dropping the first observation and the corresponding p-values.

Parameter	Estimate	RC	p-value
β_1	1.7373	−6.52	0.00
β_2	−0.3109	3.87	0.00

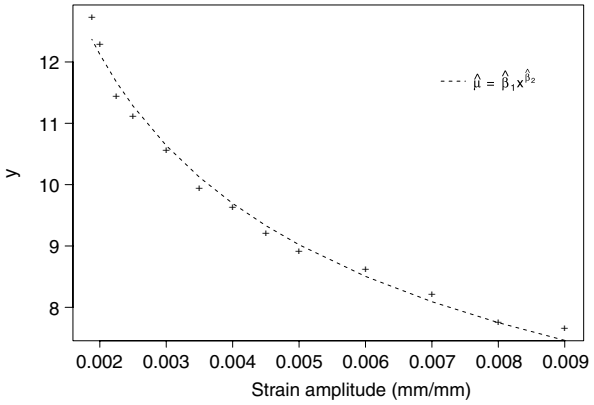

Fig. 5. Scatter-plot and the fitted model.

Next, as a second application, we consider a fatigue data set reported by Williams et al. (2003) on the life of a SAE 1137 carbon steel in cycles to failure. The response N is the number of cycles to failure and the explanatory variable x is the strain amplitude (mm/mm). The data are given in Table 7.

Here, we proposed the nonlinear regression model

$$y_i = \beta_1 x^{\beta_2} + \varepsilon_i, \quad i = 1, \ldots, 13, \tag{10}$$

where $y_i = \log(N_i)$ and $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$. The MLEs (the standard errors in parentheses) are: $\widehat{\beta}_1 = 1.6309$ (0.0899), $\widehat{\beta}_2 = -0.3228$ (0.0096) and $\widehat{\alpha} = 0.1704$ (0.0334). The bias corrected estimates are: $\widetilde{\beta}_1 = 1.6285$ (0.0984), $\widetilde{\beta}_2 = -0.3228$ (0.0106) and $\widetilde{\alpha} = 0.1869$ (0.0367). Fig. 5 gives the scatter-plot of the data and the fitted model (10).

Fig. 6 gives the generalized Cook distance for the nonlinear model (10). The Cook distance shows that the first case is a possible influential observation on $\widehat{\boldsymbol{\beta}}$. We eliminate this observation and refitted the model. Relative changes on the estimates of $\boldsymbol{\beta}$ are given in Table 8. The significance of β_2 is not modified dropping the first observation.

Please cite this article in press as: Lemonte, A.J., Cordeiro, G.M., Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.06.015

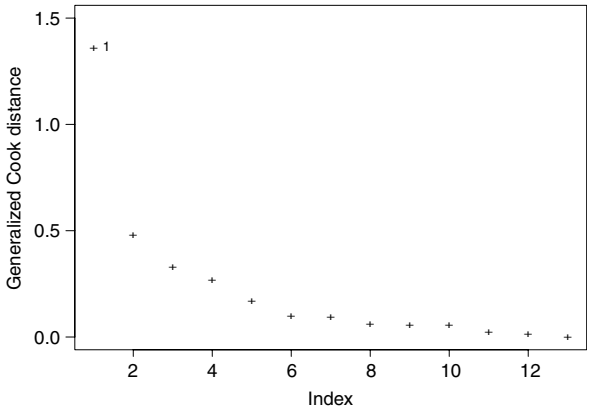

Fig. 6. Index plot of generalized Cook distance for the estimate $\widehat{\boldsymbol{\beta}}$.

7. Conclusions

The Birnbaum–Saunders distribution is widely used to model times to failure for materials subject to fatigue. The purpose of the paper was two fold. First, and most important, we propose a new class of Birnbaum–Saunders nonlinear regression models which is the natural extension of the regression model described in Rieck and Nedelman (1991). An important milestone in the development of more general Birnbaum–Saunders models was the emergence of exponential family nonlinear models (Cordeiro and Paula, 1989), which now have a wide range of applications in statistics. Second, we give simple formulae for calculating bias corrected maximum-likelihood estimates of the parameters of these models. We show by simulation that the bias correction derived is very effective, even when the sample size is large. Indeed, the bias correction mechanism adopted yields adjusted maximum-likelihood estimates which are nearly unbiased. We also present two applications to real data which illustrate the usefulness of the proposed model. Future research will be devoted to study local influence analysis (Cook, 1986) in the new class of nonlinear models.

Acknowledgments

We gratefully acknowledge grants from FAPESP and CNPq (Brazil). The authors are also grateful to an associate editor and two referees for helpful comments and suggestions.

References

Abell, M.L., Braselton, J.P., 1994. The Maple V Handbook. AP Professional, New York.
Birnbaum, Z.W., Saunders, S.C., 1969. A new family of life distributions. Journal of Applied Probability 6, 319–327.
Camerom, A.C., Trived, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press.
Cook, R.D., 1986. Assessment of local influence (with discussion). Journal of the Royal Statistical Society B 48, 133–169.
Cordeiro, G.M., Demétrio, C.G.B., 2008. Corrected estimators in extended quasi-likelihood models. Communications in Statistics, Theory and Methods 37, 873–880.
Cordeiro, G.M., Ferrari, S.L.P., Uribe-Opazo, M.A, Vasconcellos, K.L.P., 2000. Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. Statistics and Probability Letters 46, 317–328.
Cordeiro, G.M., Paula, G.A., 1989. Improved likelihood ratio statistics for exponential family nonlinear models. Biometrika 76, 93–100.
Cordeiro, G.M., Toyama–Udo, M.C., 2008. Bias correction in generalized nonlinear models with dispersion covariates. Communications in Statistics, Theory and Methods 37, 2219–2225.
Cox, D.R., Hinkley, D.V., 1974. Theoretical Statistics. Chapman and Hall, London.
Cox, D.R., Reid, N., 1987. Parameter orthogonality and approximate conditional inference (with discussion). Journal of the Royal Statistical Society B 40, 1–39.
Cox, D.R., Snell, E.J., 1968. A general definition of residuals (with discussion). Journal of the Royal Statistical Society B 30, 248–275.
Darby, S.C., Ellis, M.J., 1976. A test for synergism between two drugs. Journal of Applied Statistics 25, 296–299.
Desmond, A.F., 1985. Stochastic models of failure in random environments. Canadian Journal of Statistics 13, 171–183.
Doornik, J.A., 2006. An Object-Oriented Matrix Language – Ox 4, 5th ed. Timberlake Consultants Press, London, Oxford.
Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.
Firth, D., 1993. Bias reduction of maximum likelihood estimates. Biometrika 80, 27–38.
Galea, M., Leiva, V., Paula, G.A., 2004. Influence diagnostics in log-Birnbaum–Saunders regression models. Journal of Applied Statistics 31, 1049–1064.
Gallant, A.R., 1975. Nonlinear Statistical Models. Chapman and Hall, London.
Gradshteyn, I.S., Ryzhik, I.M., 2007. Table of Integrals, Series, and Products. Academic Press, New York.
Johnson, N., Kotz, S., Balakrishnan, N., 1995. Continuous Univariate Distributions, Volume 2, 2nd ed. Wiley, New York.
Leiva, V., Barros, M.K., 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., 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., 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.

Please cite this article in press as: Lemonte, A.J., Cordeiro, G.M., Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.06.015

McCarter, K.S., 1999. Estimation and Prediction for the Birnbaum–Saunders Distribution Using Type-II Censored Samples, With a Comparison to the Inverse Gaussian Distribution. Ph.D. dissertation, Kansas State University.
Mittelhammer, R.C., Judge, G.G., Miller, D.J., 2000. Econometric Foundations. Cambridge University Press, New York.
Ospina, R., Cribari–Neto, F., Vasconcellos, K.L.P., 2006. Improved point and interval estimation for a beta regression model. Computational Statistics and Data Analysis 51, 960–981.
Paolino, P., 2001. Maximum likelihood estimation of models with beta-distributed dependent variables. Political Analysis 9, 325–346.
Patriota, A.G., Lemonte, A.J., 2009. Bias correction in a multivariate regression model with general parameterization. Statistics and Probability Letters 79, 1655–1662.
Patriota, A.G., Lemonte, A.J., Bolfarine, H., 2009. Improved maximum likelihood estimators in a heteroskedastic errors-in-variables model. Statistical Papers doi:10.1007/s00362-009-0243-7.
Pázman, B.A., Denis, J.B., 1999. Bias of LS estimators in nonlinear regression models with constraints. Part I: General case. Applied Mathematics 44, 359–374.
R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN:3-900051-07-0.
Ratkowsky, D.A., 1983. Nonlinear Regression Modeling: A Unified Practical Approach. Marcel Dekker, New York.
Rieck, J.R., 1989. Statistical Analysis for the Birnbaum–Saunders Fatigue Life Distribution. Ph.D. dissertation, Clemson University.
Rieck, J.R., Nedelman, J.R., 1991. A log-linear model for the Birnbaum–Saunders distribution. Technometrics 33, 51–60.
Stone, M., 1980. Discussion of paper by D.M. Bater and D.G. Watts. Journal of the Royal Statistical Society B 42, 17–19.
Tisionas, E.G., 2001. Bayesian inference in Birnbaum–Saunders regression. Communications in Statistics, Theory and Methods 30, 179–193.
Vasconcellos, K.L.P., Silva, S.G., 2005. Corrected estimates for student t regression models with unknown degrees of freedom. Journal of statistical computation and simulation 75, 409–423.
Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.
Xie, F.C., Wei, B.C., 2007. Diagnostics analysis for log-Birnbaum–Saunders regression models. Computational Statistics and Data Analysis 51, 4692–4706.
Williams, C.R., Lee, Y.L., Rilly, J.T., 2003. A practical method for statistical analysis of strain-life fatigue data. International Journal of Fatigue 25, 427–436.
Yang, Z., Lin, D.K.J., 2007. Improved maximum-likelihood estimation for the common shape parameter of several Weibull populations. Applied Stochastic Models in Business and Industry 23, 373–453.

Please cite this article in press as: Lemonte, A.J., Cordeiro, G.M., Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.06.015