

Birnbaum–Saunders nonlinear regression models

Artur J. Lemonte

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

Gauss M. Cordeiro

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

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 [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. We also give an application to a real fatigue data set.

Key words: Bias correction, Birnbaum–Saunders distribution, maximum likelihood estimation, nonlinear regression.

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 has the appealing

feature of providing satisfactory tail fitting. 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 It was later derived by Desmond (1985) using a biological model which followed from relaxing some of the assumptions originally made by Birnbaum and Saunders (1969).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 be Birnbaum–Saunders distributed with parameters $\alpha, \eta > 0$, say $\mathcal{B}\text{-}\mathcal{S}(\alpha, \eta)$, 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 $kT \sim \mathcal{B}\text{-}\mathcal{S}(\alpha, k\eta)$.

McCarter (1999) considered parameter estimation under type II data censoring for the $\mathcal{B}\text{-}\mathcal{S}(\alpha, \eta)$ 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).

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, § 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 bias of the MLE in these models which is used to define a bias corrected estimate.

Bias adjustment has been extensively studied in the statistical literature. In fact, Cook et al. (1986) proposed bias correction in normal nonlinear models. Young and Bakir (1987) obtained bias corrected estimates for a generalized

log-gamma regression model. Cordeiro and McCullagh (1991) gave general matrix formulae for bias correction in generalized linear models, whereas Paula (1992) derive the second-order biases in exponential family nonlinear models. Cordeiro et al. (2000) obtained bias correction for symmetric nonlinear regression models. More recently, Vasconcellos and Cribari-Neto (2005) calculate the biases of the MLEs in a new class of beta regression. Cordeiro and Demétrio (2008) propose formulae for the second-order biases of the maximum quasi-likelihood estimates, whereas Cordeiro and Toyama (2008) derive the second-order biases in generalized nonlinear models with dispersion covariates.

The rest of the paper is 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 the corresponding ones 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 $\text{Var}(Y) = \sigma^2 w(\alpha)$, 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 $\alpha \rightarrow 0$.

We define the nonlinear regression model

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

where y_i is the logarithm of the i th observed lifetime, \mathbf{x}_i is an $m \times 1$ vector

of known explanatory variables associated with the i th observable response y_i , $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ 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 $\mu_i = f_i(\mathbf{x}_i; \boldsymbol{\beta})$, where f_i is assumed to be a known and twice continuously differentiable function with respect to $\boldsymbol{\beta}$. For the linear regression $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$, 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 $\mathbf{y} = (y_1, \dots, y_n)^\top$ obtained from (1), except for constants, can be expressed as

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

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 $i = 1, \dots, n$. 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 $\mathbf{D} = \mathbf{D}(\boldsymbol{\beta}) = \partial \boldsymbol{\mu} / \partial \boldsymbol{\beta}$ of partial derivatives of $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^\top$ with respect to $\boldsymbol{\beta}$ is assumed to be of full rank, i.e., $\text{rank}(\mathbf{D}) = p$ for all $\boldsymbol{\beta}$. The nonlinear predictors $\mathbf{x}_1, \dots, \mathbf{x}_n$ are embedded in an infinite sequence of $m \times 1$ 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 $\boldsymbol{\beta}$.

It is easy to see by differentiating (2) that

$$\begin{aligned} U_r &= \frac{1}{2} \sum_{i=1}^n d_{ir} \left(\xi_{i1} \xi_{i2} - \frac{\xi_{i2}}{\xi_{i1}} \right), & U_\alpha &= -\frac{n}{\alpha} + \frac{1}{\alpha} \sum_{i=1}^n \xi_{i2}^2, \\ U_{rs} &= \frac{1}{2} \sum_{i=1}^n g_{irs} \left(\xi_{i1} \xi_{i2} - \frac{\xi_{i2}}{\xi_{i1}} \right) - \frac{1}{4} \sum_{i=1}^n d_{ir} d_{is} \left(2\xi_{i2}^2 + \frac{4}{\alpha^2} - 1 + \frac{\xi_{i2}^2}{\xi_{i1}^2} \right), \\ U_{r\alpha} &= -\frac{1}{\alpha} \sum_{i=1}^n d_{ir} \xi_{i1} \xi_{i2} & \text{and} & \quad U_{\alpha\alpha} = \frac{n}{\alpha^2} - \frac{3}{\alpha^2} \sum_{i=1}^n \xi_{i2}^2. \end{aligned}$$

The score function for $\boldsymbol{\beta}$ is $\mathbf{U}_\beta = \frac{1}{2} \mathbf{D}^\top \mathbf{s}$, where $\mathbf{s} = \mathbf{s}(\boldsymbol{\theta})$ is an n -vector whose i th 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 $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\alpha})^\top$ be the MLE of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha)^\top$. We can write $\hat{\boldsymbol{\theta}} \stackrel{a}{\sim} \mathcal{N}_{p+1}(\boldsymbol{\theta}, \mathbf{K}_\theta^{-1})$ for n large, where $\stackrel{a}{\sim}$ denotes approximately distributed, \mathbf{K}_θ is the block-diagonal Fisher information matrix given by $\mathbf{K}_\theta = \text{diag}\{\mathbf{K}_\beta, \kappa_{\alpha,\alpha}\}$, \mathbf{K}_θ^{-1} is its inverse, $\mathbf{K}_\beta = \psi_1(\alpha)(\mathbf{D}^\top \mathbf{D})/4$ is the information matrix for $\boldsymbol{\beta}$ and $\kappa_{\alpha,\alpha} = 2n/\alpha^2$ is the information for α . Also,

$$\psi_1(\alpha) = 2 + \frac{4}{\alpha^2} - \frac{\sqrt{2\pi}}{\alpha} \left\{ 1 - \text{erf}\left(\frac{\sqrt{2}}{\alpha}\right) \right\} \exp\left(\frac{2}{\alpha^2}\right),$$

where $\text{erf}(\cdot)$ is the error function given by

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

Details on $\text{erf}(\cdot)$ can be found in Gradshteyn and Ryzhik (2007). Since \mathbf{K}_θ is block-diagonal, the vector $\boldsymbol{\beta}$ and the scalar α are globally orthogonal (Cox and Reid, 1987) and $\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}$ are asymptotically independent. It can be shown (Rieck, 1989) that $\psi_1(\alpha) \approx 1 + 4/\alpha^2$ for α small and $\psi_1(\alpha) \approx 2$ for α large.

The MLE $\hat{\boldsymbol{\theta}}$ 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

$$\begin{aligned} \boldsymbol{\beta}^{(m+1)} &= \boldsymbol{\beta}^{(m)} + (\mathbf{D}^{(m)\top} \mathbf{D}^{(m)})^{-1} \mathbf{D}^{(m)\top} \boldsymbol{\zeta}^{(m)}, \\ \alpha^{(m+1)} &= \frac{1}{2} \alpha^{(m)} (1 + \bar{\xi}_2^{(m)}), \end{aligned}$$

where $\boldsymbol{\zeta}^{(m)} = 2\mathbf{s}^{(m)}/\psi_1(\alpha^{(m)})$ and $\bar{\xi}_2^{(m)} = \sum_{i=1}^n \xi_{i2}^{2(m)}/n$ for $m = 0, 1, 2, \dots$

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 $\mathbf{D}^{(0)}$, $\boldsymbol{\zeta}^{(0)}$ 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 \mathbf{D} , $\boldsymbol{\zeta}$ 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:

$$\kappa_{rs} = -\frac{\psi_1(\alpha)}{4} \sum_{i=1}^n d_{ir} d_{is}, \quad \kappa_{r\alpha} = \kappa_{r\alpha\alpha} = 0, \quad \kappa_{\alpha\alpha} = -\frac{2n}{\alpha^2}, \quad \kappa_{\alpha\alpha\alpha} = \frac{10n}{\alpha^3},$$

$$\begin{aligned}\kappa_{rst} &= -\frac{\psi_1(\alpha)}{4} \sum_{i=1}^n (g_{irs}d_{it} + g_{irt}d_{is} + g_{ist}d_{ir}), & \kappa_{rs\alpha} &= \frac{(2 + \alpha^2)}{\alpha^3} \sum_{i=1}^n d_{ir}d_{is}, \\ \kappa_{rs}^{(t)} &= -\frac{\psi_1(\alpha)}{4} \sum_{i=1}^n (g_{irt}d_{is} + g_{ist}d_{ir}), & \kappa_{r\alpha}^{(\alpha)} &= \kappa_{r\alpha}^{(s)} = 0 \quad \text{and} \quad \kappa_{\alpha\alpha}^{(\alpha)} = \frac{4n}{\alpha^3}.\end{aligned}$$

Let $B(\widehat{\beta}_a)$ and $B(\widehat{\alpha})$ be the n^{-1} biases of $\widehat{\beta}_a$ ($a = 1, \dots, p$) 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 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

$$B(\widehat{\beta}_a) = \sum'_{s,t,u} \kappa^{a,s} \kappa^{t,u} \left(\kappa_{st}^{(u)} - \frac{1}{2} \kappa_{stu} \right) + \kappa^{\alpha,\alpha} \sum'_s \kappa^{a,s} \left(\kappa_{s\alpha}^{(\alpha)} - \frac{1}{2} \kappa_{s\alpha\alpha} \right) \quad (3)$$

and

$$B(\widehat{\alpha}) = (\kappa^{\alpha,\alpha})^2 \left(\kappa_{\alpha\alpha}^{(\alpha)} - \frac{1}{2} \kappa_{\alpha\alpha\alpha} \right) + \kappa^{\alpha,\alpha} \sum'_{t,u} \kappa^{t,u} \left(\kappa_{\alpha t}^{(u)} - \frac{1}{2} \kappa_{\alpha tu} \right), \quad (4)$$

where $\kappa^{r,s}$ is the (r, s) th element of the inverse $\mathbf{K}_{\boldsymbol{\beta}}^{-1}$ of the information matrix for $\boldsymbol{\beta}$, $\kappa^{\alpha,\alpha} = \kappa_{\alpha,\alpha}^{-1}$ and \sum' denotes the summation over all combinations of parameters $\beta_1, \beta_2, \dots, \beta_p$.

First, we consider equation (3) from which we readily have that the second sum is zero since $\kappa_{s\alpha\alpha} = \kappa_{s\alpha}^{(\alpha)} = 0$. It follows that

$$B(\widehat{\beta}_a) = -\frac{\psi_1(\alpha)}{8} \sum'_{s,t,u} \kappa^{a,s} \kappa^{t,u} \sum_{i=1}^n (g_{isu}d_{it} - g_{ist}d_{iu} + g_{itu}d_{is}).$$

By rearranging the summation terms we obtain

$$B(\widehat{\beta}_a) = -\frac{\psi_1(\alpha)}{8} \sum_{i=1}^n \sum'_s \kappa^{a,s} d_{is} \sum'_{t,u} \kappa^{t,u} g_{itu}.$$

Let \mathbf{d}_i^\top ($1 \times p$) and \mathbf{g}_i^\top ($1 \times p^2$) be vectors containing the first and second partial derivatives of the mean μ_i with respect to the β 's. We can write the above equation in matrix notation as

$$B(\widehat{\beta}_a) = -\frac{\psi_1(\alpha)}{8} \boldsymbol{\rho}_a^\top \mathbf{K}_{\boldsymbol{\beta}}^{-1} \sum_{i=1}^n \{ \mathbf{d}_i \mathbf{g}_i^\top \} \text{vec}(\mathbf{K}_{\boldsymbol{\beta}}^{-1}),$$

where $\boldsymbol{\rho}_a^\top$ is the a th row of the $p \times p$ identity matrix and $\text{vec}(\cdot)$ is the operator which transforms a matrix into a vector by stacking the columns of the matrix

one underneath the other. It is straightforward to check that

$$B(\hat{\beta}_a) = -\frac{\psi_1(\alpha)}{8} \boldsymbol{\rho}_a^\top \mathbf{K}_\beta^{-1} \mathbf{D}^\top \mathbf{G} \text{vec}(\mathbf{K}_\beta^{-1}),$$

where $\mathbf{D} = \partial \boldsymbol{\mu} / \partial \boldsymbol{\beta} = (\mathbf{d}_1, \dots, \mathbf{d}_n)^\top$ and $\mathbf{G} = \partial^2 \boldsymbol{\mu} / \partial \boldsymbol{\beta}^\top \partial \boldsymbol{\beta} = (\mathbf{g}_1, \dots, \mathbf{g}_n)^\top$ are $n \times p$ and $n \times p^2$ matrices of the first and second partial derivatives of the mean vector $\boldsymbol{\mu}$ with respect to $\boldsymbol{\beta}$, respectively. The n^{-1} bias vector $\mathbf{B}(\hat{\boldsymbol{\beta}})$ of $\hat{\boldsymbol{\beta}}$ can then be written as

$$\mathbf{B}(\hat{\boldsymbol{\beta}}) = (\mathbf{D}^\top \mathbf{D})^{-1} \mathbf{D}^\top \mathbf{d}, \quad (5)$$

where \mathbf{d} is an $n \times 1$ vector defined as $\mathbf{d} = -\{2/\psi_1(\alpha)\} \mathbf{G} \text{vec}\{(\mathbf{D}^\top \mathbf{D})^{-1}\}$.

We now calculate the n^{-1} bias of $\hat{\alpha}$. Using (4), we obtain

$$\begin{aligned} B(\hat{\alpha}) &= -\frac{\alpha}{4n} - \frac{(2 + \alpha^2)}{4\alpha n} \sum_{i=1}^n \sum'_{t,u} \kappa^{t,u} d_{it} d_{iu} = -\frac{\alpha}{4n} - \frac{(2 + \alpha^2)}{4\alpha n} \sum_{i=1}^n \mathbf{d}_i^\top \mathbf{K}_\beta^{-1} \mathbf{d}_i \\ &= -\frac{\alpha}{4n} - \frac{(2 + \alpha^2)}{4\alpha n} \text{tr}(\mathbf{D} \mathbf{K}_\beta^{-1} \mathbf{D}^\top), \end{aligned}$$

where $\text{tr}(\cdot)$ denotes the trace operator. Now, making use of the fact that $\text{tr}(\mathbf{D} \mathbf{K}_\beta^{-1} \mathbf{D}^\top) = 4p/\psi_1(\alpha)$, we can rewrite the n^{-1} bias of $\hat{\alpha}$ as

$$B(\hat{\alpha}) = -\frac{1}{n} \left\{ p \left(\frac{2 + \alpha^2}{\alpha \psi_1(\alpha)} \right) + \frac{\alpha}{4} \right\}. \quad (6)$$

Equations (5) and (6) represent the main results of the paper. The bias vector $\mathbf{B}(\hat{\boldsymbol{\beta}})$ can be obtained from a simple ordinary least-squares regression of \mathbf{d} on the columns of \mathbf{D} . It depends on the nonlinearity of the regression function f and the parameter α . The bias vector $\mathbf{B}(\hat{\boldsymbol{\beta}})$ will be small when \mathbf{d} is orthogonal to the columns of \mathbf{D} . Also, it can be large when $\psi_1(\alpha)$ and n are both small. Equation (5) 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 $\mathbf{B}(\hat{\boldsymbol{\beta}})$. For linear models, the matrix \mathbf{G} and the vector \mathbf{d} vanish and hence $\mathbf{B}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$, 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 (6) depends directly on the nonlinear structure of the regression model only through the rank p of \mathbf{D} . It shows that the bias is always a linear function of the dimension p of $\boldsymbol{\beta}$.

In the right-hand sides of expressions (5) and (6), which are both of order n^{-1} , consistent estimates of the parameters $\boldsymbol{\beta}$ and α can be inserted to define bias corrected estimates $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \mathbf{B}(\hat{\boldsymbol{\beta}})$ and $\tilde{\alpha} = \hat{\alpha} - B(\hat{\alpha})$, where $\mathbf{B}(\hat{\boldsymbol{\beta}})$ and $B(\hat{\alpha})$ are the values of $\mathbf{B}(\hat{\boldsymbol{\beta}})$ and $B(\hat{\alpha})$, respectively, at $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\alpha})^\top$. The bias

corrected estimates $\tilde{\boldsymbol{\beta}}$ and $\tilde{\alpha}$ are expected to have better sampling properties than the classical MLEs $\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}$. In fact, we present some simulations in Section 5 to show that $\tilde{\boldsymbol{\beta}}$ and $\tilde{\alpha}$ 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 can not 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 bias 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 $B(\hat{\mu}_i)$ of the MLE $\hat{\mu}_i$ of the i th mean $\mu_i = f_i(\mathbf{x}_i; \boldsymbol{\beta})$. We can easily show by Taylor series expansion that

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

where \mathbf{M}_i is a $p \times p$ matrix of second partial derivatives $\partial^2 \mu_i / \partial \beta_r \partial \beta_s$ (for $r, s = 1, \dots, p$), $\text{Cov}(\hat{\boldsymbol{\beta}}) = \mathbf{K}_\beta^{-1}$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ and the vectors \mathbf{d}_i and $\mathbf{B}(\hat{\boldsymbol{\beta}})$ were mentioned previously. All quantities in the above equation should be evaluated at $\hat{\boldsymbol{\beta}}$.

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

$$\text{Var}(\hat{\mu}_i) = \text{tr}\{(\mathbf{d}_i \mathbf{d}_i^\top) \text{Cov}(\hat{\boldsymbol{\beta}})\}.$$

4 Special models

Equation (5) 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 $\mathbf{B}(\hat{\boldsymbol{\beta}})$ algebraically with minimal effort. In particular, (5) may simplify considerably if the number of nonlinear parameters is small. Moreover, for any nonlinear special model, we can calculate the bias $\mathbf{B}(\hat{\boldsymbol{\beta}})$ numerically via a software with numerical linear algebra facilities such as Ox (Doornik, 2001) and R (R Development Core Team, 2008).

First, we consider a nonlinear regression model which depends on a single nonlinear parameter. Equation (5) gives

$$B(\hat{\beta}) = -\frac{2}{\psi_1(\alpha)} \frac{\kappa_2}{\kappa_1^2},$$

where $\kappa_1 = \sum_{i=1}^n (df_i/d\beta)^2$ and $\kappa_2 = \sum_{i=1}^n (df_i/d\beta)(d^2f_i/d\beta^2)$. The constants κ_1 and κ_2 are evaluated at $\hat{\beta}$ and $\hat{\alpha}$ to yield $\hat{B}(\hat{\beta})$ and the corrected estimate $\tilde{\beta} = \hat{\beta} - \hat{B}(\hat{\beta})$. 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

$$\boldsymbol{\mu} = \mathbf{Z}\boldsymbol{\lambda} + \eta\mathbf{g}(\gamma), \quad (7)$$

where \mathbf{Z} is a known $n \times (p-2)$ matrix of full rank, $\mathbf{g}(\gamma)$ is an $n \times 1$ vector, $\boldsymbol{\beta} = (\boldsymbol{\lambda}^\top, \eta, \gamma)^\top$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{p-2})^\top$ and η and γ are scalar parameters. This class of models occurs very often in statistical modeling; see Cook et al. (1986) and Cordeiro et al. (2000). 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 (7) 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 \mathbf{D} takes the form $\mathbf{D} = [\mathbf{Z}, \mathbf{g}(\gamma), \eta(d\mathbf{g}(\gamma)/d\gamma)]$ and, after some algebra, we can obtain from (5)

$$\mathbf{B}(\hat{\boldsymbol{\beta}}) = -\left[\frac{1}{\eta} \text{Cov}(\hat{\eta}, \hat{\gamma}) \boldsymbol{\tau}_p + \frac{\eta}{2} \text{Var}(\hat{\gamma}) \boldsymbol{\delta}_p \right], \quad (8)$$

where $\boldsymbol{\tau}_p$ is a $p \times 1$ vector with a one in the last position and zeros elsewhere, $\boldsymbol{\delta}_p = (\mathbf{D}^\top \mathbf{D})^{-1} \mathbf{D}^\top (d^2 \mathbf{g}(\gamma)/d\gamma^2)$ is simply the set of coefficients from the ordinary regression of the vector $d^2 \mathbf{g}(\gamma)/d\gamma^2$ on the matrix \mathbf{D} , and $\text{Var}(\hat{\gamma})$ and $\text{Cov}(\hat{\eta}, \hat{\gamma})$ are the large-sample second moments obtained from the appropriate elements of the asymptotic covariance matrix $\text{Cov}(\hat{\boldsymbol{\beta}}) = \mathbf{K}_{\boldsymbol{\beta}}^{-1} = (4/\psi_1(\alpha))(\mathbf{D}^\top \mathbf{D})^{-1}$. It is clear from (8) that $\mathbf{B}(\hat{\boldsymbol{\beta}})$ does not depend explicitly on the linear parameters in $\boldsymbol{\lambda}$ and it is proportional to $4/\psi_1(\alpha)$. Further, the covariance term $\text{Cov}(\hat{\eta}, \hat{\gamma})$ contributes only to the bias of $\hat{\gamma}$.

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 MLEs 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. Also, the number of Monte Carlo replications was 10,000. All simulations were performed using the Ox matrix programming language (Doornik, 2001).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 $\hat{\theta}$, defined as $\{\mathbb{E}(\hat{\theta}) - \theta\}/\theta$, is obtained by estimating $\mathbb{E}(\hat{\theta})$ by Monte Carlo) and the root mean square error ($\sqrt{\text{MSE}}$), 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 $i = 1, \dots, n$. 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 times greater than the average bias of the corrected estimates. This fact suggests that the second-order bias of the MLEs should not be ignored in samples of small to moderate size since they can be non-negligible. The figures in Table 2 show 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

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

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.

deteriorates (see Tables 1 and 2). For instance, when $n = 15$, the relative biases of $\hat{\gamma}$ (MLE) and $\tilde{\gamma}$ (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 known Michaelis–Menton model, which is very useful 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

$$\mu_i = \frac{\eta x_i}{\gamma + x_i}, \quad i = 1, 2, \dots, n,$$

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 regression 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.

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.

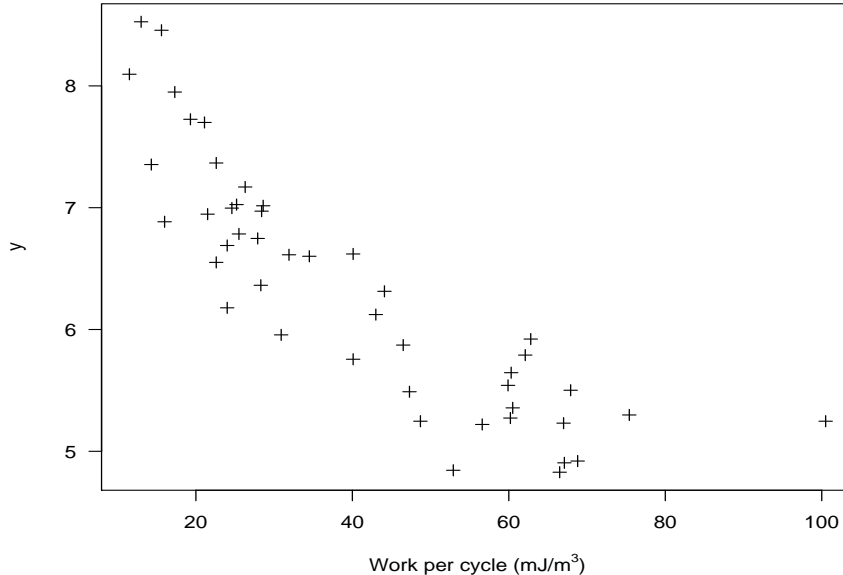


Fig. 1. Scatter-plot of the data set.

6 Application

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

$$y_i = \beta_1 + \beta_2 \log w_i + \varepsilon_i, \quad (9)$$

where $y_i = \log N_i$ and $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$, for $i = 1, \dots, 46$. The MLEs (the corresponding standard errors in parentheses) are: $\hat{\beta}_1 = 12.2797 (0.3942)$, $\hat{\beta}_2 = -1.6708 (0.1096)$ and $\hat{\alpha} = 0.4104 (0.0428)$. We take the logarithm of w to ensure a linear relationship between the response variable (y) and the covariate in (9); see Galea et al. (2004, Figure 1). However, Figure 1 suggests a nonlinear relationship between the response variable and the covariate w .

Here, we proposed the nonlinear regression model

$$y_i = \beta_1 + \beta_2 \exp(\beta_3/w_i) + \varepsilon_i, \quad i = 1, \dots, 46, \quad (10)$$

where $\varepsilon_i \sim \mathcal{SN}(\alpha, 0, 2)$. The MLEs (the standard errors in parentheses) are:

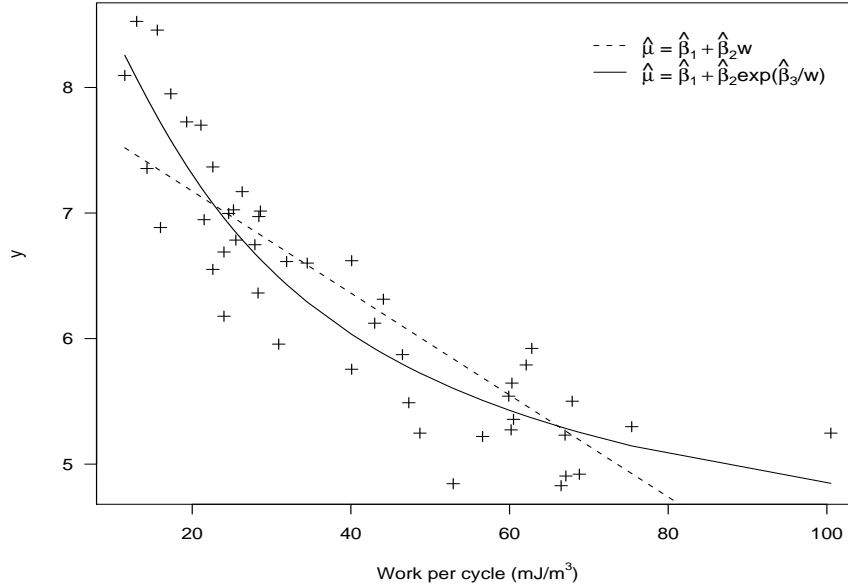


Fig. 2. Scatter-plot and the fitted models.

$\hat{\beta}_1 = 8.9876$ (0.7454), $\hat{\beta}_2 = -5.1802$ (0.5075), $\hat{\beta}_3 = -22.5196$ (7.3778) and $\hat{\alpha} = 0.40$ (0.0417). The bias corrected estimates are: $\tilde{\beta}_1 = 8.7806$ (0.7734), $\tilde{\beta}_2 = -4.9362$ (0.5266), $\tilde{\beta}_3 = -22.1713$ (7.6548) and $\tilde{\alpha} = 0.4157$ (0.0433). Hence, the uncorrected estimates are slightly different from the bias corrected estimates even for large samples ($n = 46$ observations).

Figure 2 gives the scatter-plot of the data, the fitted model (10) and the fitted straight line, say $y_i = \beta_1 + \beta_2 w_i + \varepsilon_i$, where the MLEs are: $\hat{\beta}_1 = 7.9864$ (0.1622), $\hat{\beta}_2 = -0.0406$ (0.0036) and $\hat{\alpha} = 0.52$ (0.0542). Figure 2 shows that the non-linear model (10) (unlike the linear model) fits satisfactorily to the fatigue data. The 46th observation (the one with work per cycle near 100) can be an influential data. However, it is not possible to say whether this observation is influential or not without using an efficient way to detect influential observations in the new class of models. Influence diagnostic analysis for this class of models will be developed in future research.

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

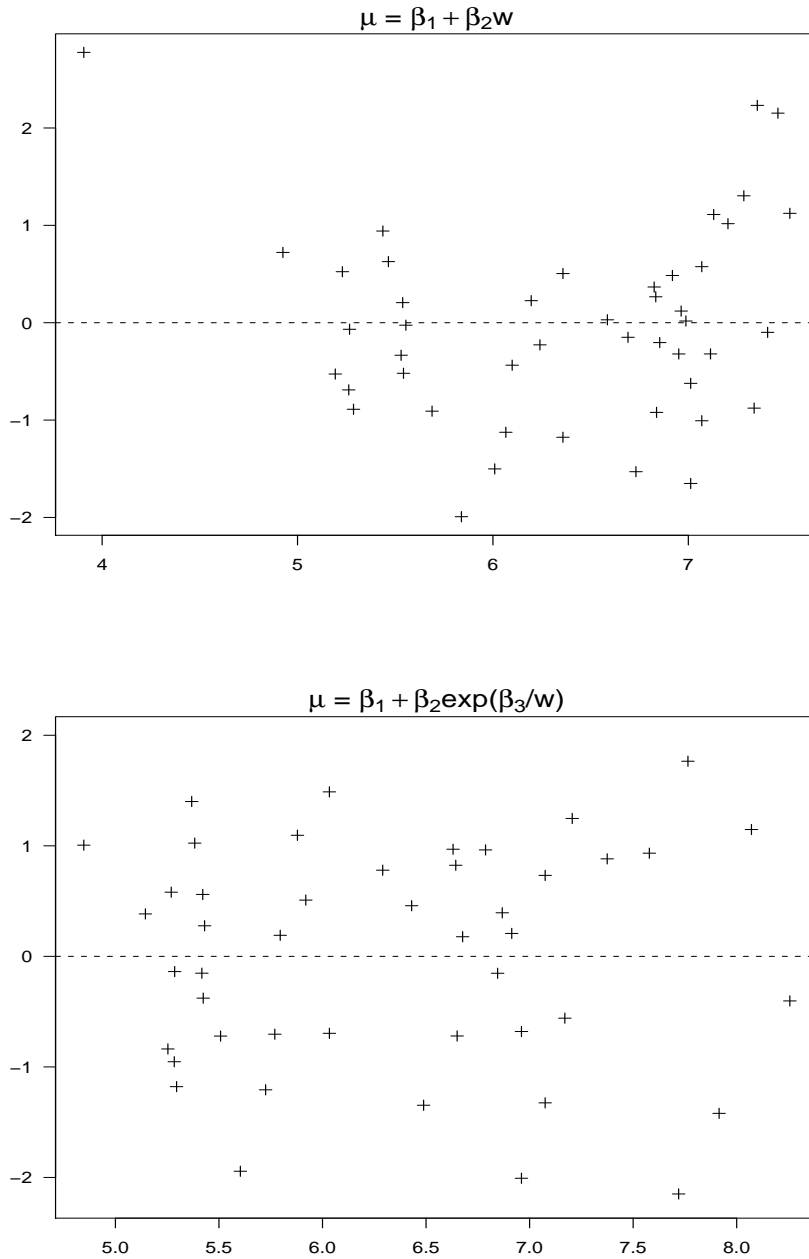


Fig. 3. Index plot of \widehat{R}_i versus $\widehat{\mu}_i$.

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, we propose a new class of Birnbaum–Saunders nonlinear regression models which generalizes the regression model described in Rieck and Nedelman

(1991). Second, we give simple formulae for calculating bias corrected maximum likelihood estimates of the parameters of these models. The simulation results presented show 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 an application to a real fatigue data set that illustrates the usefulness of the proposed model. Future research will be devoted to a study of diagnostics and influence analysis 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

- [1] Abell, M.L., Braselton, J.P. (1994). *The Maple V Handbook*. AP Professional, New York.
- [2] Birnbaum, Z.W., Saunders, S.C. (1969). A new family of life distributions. *Journal of Applied Probability*, **6**, 319–327.
- [3] Cook, R.D., Tsai, C.L., Wei, B.C. (1986). Bias in nonlinear regression. *Biometrika*, **73**, 615–623.
- [4] Cordeiro, G.M., Demétrio, C.G.B. (2008). Corrected estimators in extended quasi-likelihood models. *Communications in Statistics, Theory and Methods*, **37**, 873–880.
- [5] 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.
- [6] Cordeiro, G.M., McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society B*, **53**, 629–643.
- [7] Cordeiro, G.M., Toyama, M.C. (2008). Bias correction in generalized nonlinear models with dispersion covariates. *Communications in Statistics, Theory and Methods*, **37**, 2219–2225.
- [8] Cox, D.R., Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

- [9] Cox, D.R., Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society B*, **40**, 1–39.
- [10] Cox, D.R., Snell, E.J. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society B*, **30**, 248–275.
- [11] Darby, S.C., Ellis, M.J. (1976). A test for synergism between two drugs. *Journal of Applied Statistics*, **25**, 296–299.
- [12] Desmond, A.F. (1985). Stochastic models of failure in random environments. *Canadian Journal of Statistics*, **13**, 171–183.
- [13] Doornik, J.A. (2001). *An Object-Oriented Matrix Language – Ox 4*. Timberlake Consultants Press, London and Oxford. 5th ed. URL <http://www.doornik.com>.
- [14] Efron, B., Tibshirani, R.J. (1993). *An introduction to the bootstrap*. Chapman and Hall, New York.
- [15] Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27–38.
- [16] Galea, M., Leiva, V., Paula, G.A. (2004). Influence diagnostics in log-Birnbaum–Saunders regression models. *Journal of Applied Statistics*, **31**, 1049–1064.
- [17] Gallant, A.R. (1975). *Nonlinear Statistical Models*. Chapman and Hall, London.
- [18] Gradshteyn, I.S., Ryzhik, I.M. (2007). *Table of Integrals, Series, and Products*. Academic Press, New York.
- [19] Johnson, N., Kotz, S., Balakrishnan, N. (1995). *Continuous Univariate Distributions, Volume 2*, 2nd ed. Wiley, New York.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] Mittelhammer, R.C., Judge, G.G., Miller, D.J. (2000). *Econometric Foundations*. New York: Cambridge University Press.
- [25] Paula, G.A. (1992). Bias correction for exponential family nonlinear models. *Journal of Statistical Computation and Simulation*, **40**, 43–54.

- [26] 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.
- [27] Ratkowsky, D.A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach*. Marcel Dekker, New York.
- [28] Rieck, J.R. (1989). *Statistical Analysis for the Birnbaum–Saunders Fatigue Life Distribution*. Ph.D. dissertation, Clemson University.
- [29] Rieck, J.R., Nedelman, J.R. (1991). A log-linear model for the Birnbaum–Saunders distribution. *Technometrics*, **33**, 51–60.
- [30] Stone, M. (1980). Discussion of paper by D.M. Bate and D.G. Watts. *Journal of the Royal Statistical Society B*, **42**, 17–19.
- [31] Tisonas, E.G. (2001). Bayesian inference in Birnbaum–Saunders regression. *Communications in Statistics, Theory and Methods*, **30**, 179–193.
- [32] Vasconcellos, K.L.P., Cribari–Neto, F. (2005). Improved maximum likelihood estimation in a new class of beta regression. *Brazilian Journal of Probability and Statistics*, **19**, 13–31.
- [33] Xie, F.C., Wei, B.C. (2007). Diagnostics analysis for log-Birnbaum–Saunders regression models. *Computational Statistics and Data Analysis*, **51**, 4692–4706.
- [34] Young, D. H., Bakir, S. T. (1987). Bias correction for a generalized log-gamma regression model. *Technometrics*, **29**, 183–191.