Computational Statistics and Data Analysis 64 (2013) 34–50

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

A new extended Birnbaum–Saunders regression model for lifetime modeling

Artur J. Lemonte *

Departamento de Estatística, Universidade de São Paulo, São Paulo/SP, Brazil

ARTICLE INFO

Article history:
Received 6 August 2012
Received in revised form 15 December 2012
Accepted 19 February 2013
Available online 4 March 2013

Keywords:
Birnbaum–Saunders distribution
Censored data
Fatigue life distribution
Lifetime data
Local influence analysis

ABSTRACT

A new class of extended Birnbaum–Saunders regression models is introduced. It can be applied to censored data and be used more effectively in survival analysis and fatigue life studies. Maximum likelihood estimation of the model parameters with censored data as well as influence diagnostics for the new regression model are investigated. The normal curvatures for studying local influence are derived under various perturbation schemes and a martingale-type residual is considered to assess departures from the extended Birnbaum–Saunders error assumption as well as to detect outlying observations. Further, a test of homogeneity of the shape parameters of the new regression model is proposed. Two real data sets are analyzed for illustrative purposes.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The two-parameter Birnbaum–Saunders (BS) distribution, also known as the fatigue life distribution, was introduced by Birnbaum and Saunders (1969) and has received considerable attention in recent years. It was originally derived from a model for a physical fatigue process where dominant crack growth causes failure. It was later derived by Desmond (1985) using a biological model which followed from relaxing some of the assumptions originally made by Birnbaum and Saunders (1969).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 relationship between the BS distribution and the inverse Gaussian distribution was investigated by Desmond (1986) who demonstrated that the BS distribution is an equal-weight mixture of an inverse Gaussian distribution and its complementary reciprocal. For book treatments of inverse Gaussian and BS distributions and their relationships, see Marshall and Olkin (2007, Chapter 13) and especially Saunders (2007, Chapter 10). More recently, Jones (2012) also discussed the relationship between the BS and the inverse Gaussian distributions.

The cumulative distribution function of a random variable T with BS distribution, say $T \sim \mathrm{BS}(\alpha, \eta)$, is $G(t) = \Phi(v)$, with $t > 0$, where $\Phi(\cdot)$ is the standard normal cumulative function, $v = v(t) = \rho(t/\eta)/\alpha$, $\rho(z) = z^{1/2} - z^{-1/2}$, and $\alpha > 0$ and $\eta > 0$ are the shape and scale parameters, respectively. The shape of the hazard function of the BS distribution is discussed in Kundu et al. (2008). The authors showed that the hazard function is not monotone and is unimodal for all ranges of the parameter values. Some interesting results on improved statistical inference as well as interval estimation for the BS distribution may be revised in Wu and Wong (2004), Lemonte et al. (2007, 2008) and Wang (2012). The BS distribution has been applied in a wide variety of fields. For the applications of the BS distributions, read, for example, Balakrishnan et al. (2007) in reliability and Leiva et al. (2008, 2009) in other fields. It is worthwhile to mention that there has been a great deal of progress recently in developing statistical methodology for the BS model and its generalizations. Notable contributions include Professor Narayanaswamy Balakrishnan (http://www.math.mcmaster.ca/bala/bala.html) and co-workers, and Professor Victor Leiva (http://staff.deuv.cl/leiva/) and co-workers.

* Tel.: +55 11 38526071.
E-mail address: arturlemonte@gmail.com.

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.02.025

On the basis of the scheme proposed by Marshall and Olkin (1997), Lemonte (2013) introduced a quite flexible distribution which can be used to model failure times for materials subject to fatigue and lifetime data. The new distribution was called by the author as the Marshall–Olkin extended Birnbaum–Saunders (MOEBS) distribution. Hereafter, the random variable T is said to have a MOEBS distribution with shape parameters $\alpha > 0$ and $\lambda > 0$, and scale parameter $\eta > 0$, say $T \sim \text{MOEBS}(\lambda, \alpha, \eta)$, if its cumulative function is given by

$$G(t) = \frac{\Phi(v)}{1 - \bar{\lambda}\,\Phi(-v)}, \quad t > 0, \tag{1}$$

where $\bar{\lambda} = 1 - \lambda$. The survival function is $S(t) = \lambda\,\Phi(-v)/[1 - \bar{\lambda}\,\Phi(-v)]$, whereas the probability density function corresponding to (1) takes the form $g(t) = \lambda\,\kappa(\alpha, \eta)\,t^{-3/2}\,(t+\eta)\,[1-\bar{\lambda}\,\Phi(-v)]^{-2}\exp\left\{-\tau(t/\eta)/(2\alpha^2)\right\}$, where $\kappa(\alpha, \eta) = \exp(\alpha^{-2})/(2\alpha\sqrt{2\pi\eta})$ and $\tau(z) = z + z^{-1}$. It can be shown that if $T \sim \text{MOEBS}(\lambda, \alpha, \eta)$, then $kT \sim \text{MOEBS}(\lambda, \alpha, k\eta)$, for $k > 0$, i.e. the class of MOEBS distributions is closed under scale transformations. The two-parameter BS distribution arises from (1) when $\lambda = 1$, that is, $T \sim \text{BS}(\alpha, \eta) = \text{MOEBS}(1, \alpha, \eta)$.

Rieck and Nedelman (1991) proposed a log-linear regression model based on the BS distribution. They showed that if $T \sim \text{BS}(\alpha, \eta)$, then $Y = \log(T)$ is sinh-normal (SN) distributed with shape, location and scale parameters given by α, $\mu = \log(\eta)$ and $\sigma = 2$, respectively; that is, the log-BS (LBS) distribution is a special case of the SN distribution introduced by them and, in this case, the notation $Y \sim \text{LBS}(\alpha, \mu)$ is considered. The SN distribution is symmetrical, presents greater and smaller degrees of kurtosis than the normal model and also has bi-modality. Their regression model has received significant attention over the last few years by many researchers. For some recent references about the BS regression model, the reader is refereed to Desmond et al. (2008), Xiao et al. (2010), Lemonte et al. (2010), Lemonte (2011), Lemonte and Ferrari (2011a,b,c), Qu and Xie (2011) and Li et al. (2012), among others.

Some generalizations of the log-linear BS regression model have been proposed in the statistical literature. For example, some efforts can be found in the works by Barros et al. (2008), Lemonte and Cordeiro (2009), Santana et al. (2011), Lemonte (2012), Desmond et al. (2012) and Villegas et al. (2011). Barros et al. (2008) introduced the generalized BS regression model based on the BS-t_ν distribution (that is, based on the BS Student-t model with ν degrees of freedom), Lemonte and Cordeiro (2009) proposed a non-linear BS regression model, Santana et al. (2011) and Lemonte (2012) introduced the skewed BS regression model, whereas Villegas et al. (2011) and Desmond et al. (2012) studied a mixed log-linear model based on the BS distribution.

In this paper, in addition to the existing generalizations of the BS regression model, we shall propose the extended BS regression model based on the MOEBS distribution; that is, we will introduce a new class of lifetime regression models in which the errors follow the log-MOEBS distribution. The main motivation for introducing this new class of regression models relies on the fact that the practitioners will have a new BS regression model to use in practical applications. Moreover, the formulas related with the new regression model are manageable and with the use of modern computer resources and its numerical capabilities, the proposed model may prove to be an useful addition to the arsenal of applied statisticians. Additionally, the new model is quite flexible and can be widely applied in analyzing lifetime data. Further, we provide two applications to real data sets which show that the new regression model yields a better fit than the usual BS regression model. Furthermore, the new extended BS regression model can be used for modeling censored data as well as data without censoring. It should be mentioned that censored data is very common in lifetime data because of time limits and other restrictions on data collection. In a engineering life test experiment, for example, it is usually not feasible to continue experimentation until all items under study have failed. In a survival study, patients follow-up may be lost and also data analysis is usually done before all patients have reached the event of interest. The partial information contained in the censored observations is just a lower bound on the lifetime distribution. Reliability studies usually finish before all units have failed, even making use of accelerated tests. This is a special source of difficulty in the analysis of reliability data. Such data are said to be censored at right and they arise when some units are still running at the time of the data analysis, removed from test before they fail or because they failed from an extraneous cause. We refer the reader to Gijbels (2010) for a recent overview on censored data.

It is nowadays a well spread practice, after modeling, to check the model assumptions and conduct diagnostic studies in order to detect possible influential observations that may distort the results of the analysis. Diagnostic analysis is an efficient way to detect influential observations. The first technique developed to assess the individual impact of cases on the estimation process is, perhaps, the case deletion which became a very popular tool. Cook (1977) presents a great development of case deletion diagnostics for a general statistical model. Case deletion is an example of a global influence analysis, that is, the effect of an observation is assessed by completely removing it. However, case deletion excludes all information from an observation and we can hardly say whether this observation has some influence on a specific aspect of the model. To overcome this problem, one can resort to local influence approach where one investigates the model sensitivity under small perturbations. In this context, Cook (1986) proposed a general framework to detect influential observations which gives a measure of this sensitivity under small perturbations on the data or in the model. Many applications of the local influence method may be found in the statistical literature for various models and under different perturbation schemes. For instance, Espinheira et al. (2008), Vasconcellos and Fernandez (2009), Patriota et al. (2010), Lemonte and Patriota (2011), Zevallos et al. (2012) and Matos et al. (2013), among others. In this paper, we also propose a similar methodology to detect influential subjects in the new extended BS regression model. In particular, we obtain explicit formulas for Cook's (1986) normal curvature measure under three perturbation schemes.

The paper unfolds as follows. The log-MOEBS distribution is proposed in Section 2. In Section 3, we introduce the extended BS regression model and discuss estimation of the model parameters. Specifically, we compute the maximum likelihood estimating equations by assuming random censoring. In Section 4, the normal curvatures of local influence are derived under various perturbation schemes and a kind of deviance residual is proposed to assess departures from the underlying log-MOEBS distribution as well as to detect outlying observations. In Section 5, we propose a likelihood ratio statistic for testing the homogeneity of the shape parameters. Two real data illustrations are considered in Section 6. The paper ends up with some concluding remarks in Section 7.

2. The log-MOEBS distribution

Let T be a random variable having the MOEBS cumulative function (1). The random variable $Y = \log(T)$ has a log-MOEBS (LMOEBS) distribution. After some algebra, the survival function, the cumulative function and the density function of Y, parameterized in terms of $\mu = \log(\eta)$, can be expressed, respectively, as

$$S(y) = \frac{\lambda\,\Phi(-\xi_2)}{1-\bar{\lambda}\,\Phi(-\xi_2)}, \qquad F(y) = \frac{\Phi(\xi_2)}{1-\bar{\lambda}\,\Phi(-\xi_2)}, \quad y \in \mathbb{R},$$

$$f(y) = \frac{\lambda\,\xi_1\,\phi(\xi_2)}{2\,[1-\bar{\lambda}\,\Phi(-\xi_2)]^2}, \quad y \in \mathbb{R}, \tag{2}$$

where $\phi(\cdot)$ is the standard normal density function,

$$\xi_1 = \frac{2}{\alpha}\cosh\left(\frac{y-\mu}{2}\right), \qquad \xi_2 = \frac{2}{\alpha}\sinh\left(\frac{y-\mu}{2}\right).$$

Evidently, the density function (2) does not involve any complicated function and hence can be easily computed numerically. If Y is a random variable having density function (2), we write $Y \sim \text{LMOEBS}(\lambda, \alpha, \mu)$. Thus, if $T \sim \text{MOEBS}(\lambda, \alpha, \eta)$, then $Y = \log(T) \sim \text{LMOEBS}(\lambda, \alpha, \mu)$; that is, if $Y \sim \text{LMOEBS}(\lambda, \alpha, \mu)$, then $T = \exp(Y)$ follows the MOEBS distribution with shape parameters λ and α, and scale parameter $\eta = \exp(\mu)$. We have that the LMOEBS and MOEBS models correspond to a logarithmic distribution and its associated distribution, respectively (Marshall and Olkin, 2007, Chapter 12). The special case $\lambda = 1$ corresponds to the LBS distribution introduced by Rieck and Nedelman (1991).

The LMOEBS density function can take various forms depending on the parameter values. Plots of the density (2) for selected parameter values are given in Fig. 1. These plots show great flexibility of the new distribution for different values of the shape parameters λ and α. So, the LMOEBS density function (2) allows for great flexibility and hence it can be very useful in many more practical situations. In fact, it can be symmetric, asymmetric and it can also exhibit bi-modality. Unfortunately, it is very difficult (or even impossible) to determine analytically (theoretically) for what values of α and λ the LMOEBS density function (2) is bi-modal. The plots in Fig. 1 indicate that the LMOEBS distribution is very versatile and that the shape parameter λ has a substantial effect on its skewness and kurtosis.

The sth moment of $Y \sim \text{LMOEBS}(\lambda, \alpha, \mu)$ is given by

$$E(Y^s) = \lambda \sum_{k=0}^{s} \binom{s}{k}\, 2^k\, \mu^{s-k}\, I_k(\lambda, \alpha), \tag{3}$$

where

$$I_k(\lambda, \alpha) = \int_0^1 \text{arcsinh}\left(-\frac{\alpha\,\Phi^{-1}(u)}{2}\right)^k (1-\bar{\lambda}\,u)^{-2}\,du.$$

The function $\Phi^{-1}(\cdot)$ denotes the standard normal quantile function. It is not known how $I_k(\lambda, \alpha)$ can be reduced to a closed-form expression. However, this integral can be easily computed numerically in software such as Ox (Doornik, 2009) and R (R Development Core Team, 2012). The skewness and kurtosis measures can be calculated from the ordinary moments given in (3) using well-known relationships. In particular, we have $E(Y) = \mu + 2\,\lambda\, I_1(\lambda, \alpha)$. The function $I_1(\lambda, \alpha)$ is plotted as function of λ and α in Fig. 2. This figure reveals that for $\lambda < 1$, $I_1(\lambda, \alpha)$ is a decreasing function of α, whereas for $\lambda > 1$, $I_1(\lambda, \alpha)$ is an increasing function of α. The case $\lambda = 1$ implies $I_1(\lambda, \alpha) = 0$ for any $\alpha > 0$.

The quantile function of the LMOEBS(λ, α, μ) distribution takes the form

$$y(u) = \mu + 2\,\text{arcsinh}(m(u)), \quad u \in (0, 1),$$

where $m(u) = m(u; \lambda, \alpha) = \alpha\,\Phi^{-1}(\lambda\,u/[1-\bar{\lambda}\,u])/2$. The quantile function can also be expressed as $y(u) = \mu + 2\,\log\left(m(u) + \sqrt{m(u)^2+1}\right)$. The new model is easily simulated as follows: if $U \sim \mathcal{U}(0, 1)$, then $Y = \mu + 2\,\text{arcsinh}(m(U))$ has the LMOEBS(λ, α, μ) distribution. This scheme is useful because of the existence of fast generators for uniform random variables and the standard normal quantile function is available in most statistical packages.

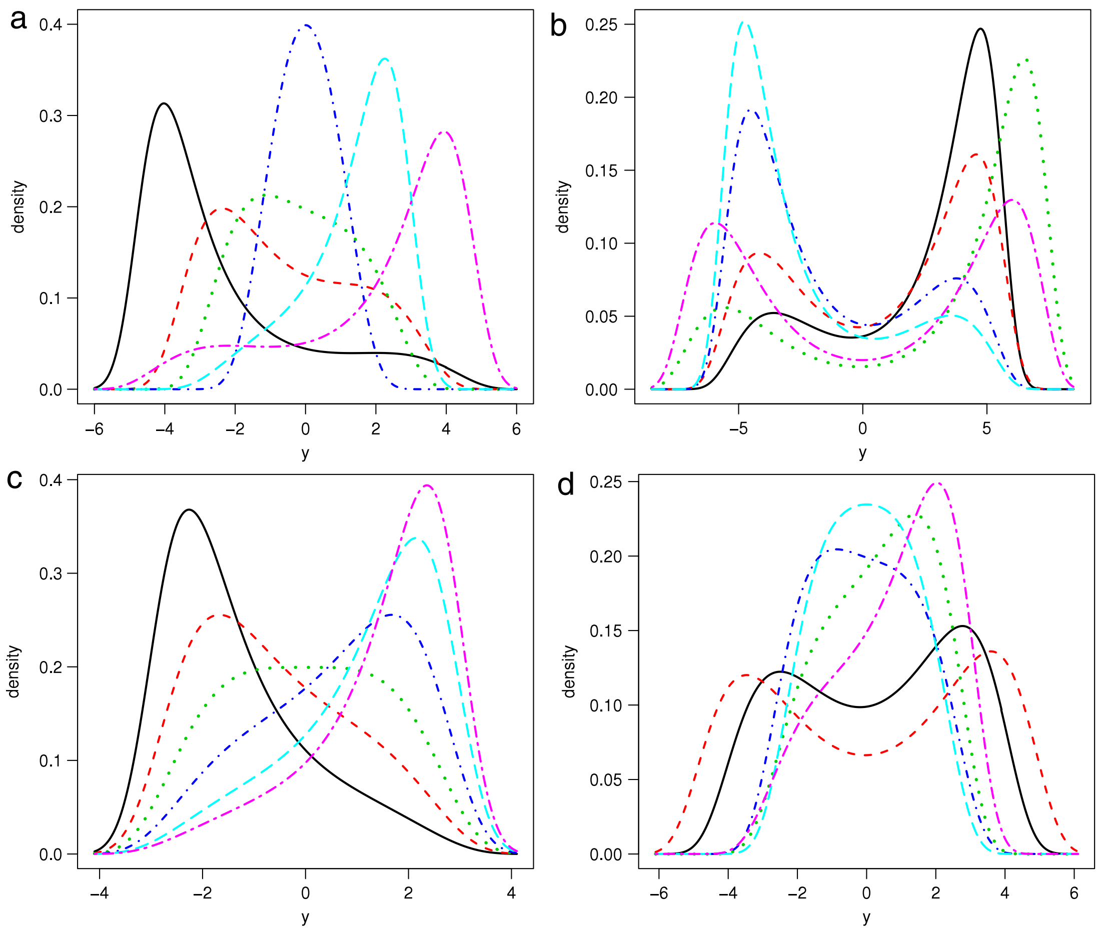

Fig. 1. Plots of the LMOEBS density function for some parameter values ($\mu = 0$): (a) $(\lambda, \alpha) = \{(0.2, 5), (0.6, 3), (0.8, 2), (1.0, 1), (4.8, 2), (4, 5)\}$; (b) $(\lambda, \alpha) = \{(3.2, 8), (1.5, 9), (2.8, 20), (0.5, 8), (0.3, 8), (1.1, 20)\}$; (c) $(\lambda, \alpha) = \{(0.2, 2), (0.5, 2), (1.0, 2), (2.0, 2), (4.0, 2), (6.0, 2)\}$; (d) $(\lambda, \alpha) = \{(1.2, 4), (1.1, 6), (1.5, 2), (0.9, 2), (1, 1.7), (2.2, 2.3)\}$.

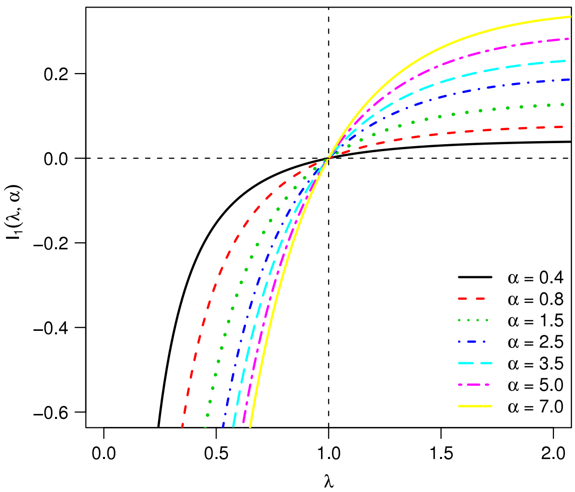

Fig. 2. Plot of the function $I_1(\lambda, \alpha)$.

3. The model and estimation

The extended BS regression model (that is, the LMOEBS regression model) is defined in the form

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{4}$$

where y_i is the observed log-lifetime or log-censoring time for the ith individual, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ is a vector of known explanatory variables associated with y_i, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a p-vector (where $p < n$ and it is fixed) of unknown regression parameters to be estimated and $\varepsilon_i \sim \text{LMOEBS}(\lambda, \alpha, 0)$. It is also assumed that the random variables ε_i's are independent and identically distributed. The extended BS regression model (4) opens new possibilities for fitting many different types of data. It is an extension of the accelerated failure time model based on the BS distribution for censored data, which arises when $\lambda = 1$ (see, for example, Leiva et al., 2007).

In the following, we will assume right random (or non-informative) censoring and that the observed lifetime and censoring time are independent. Consider the situation where the time to the event (T) is not completely observed and is subjected to right censoring. Let C denote the censoring time. We then observe $z_i = \min\{T_i, C_i\}$ and $\delta_i = \mathbb{I}(T_i \le C_i)$, for $i = 1, \ldots, n$, where $\mathbb{I}(\cdot)$ is the indicator function, that is, $\delta_i = 1$ if T_i is the observed time to the event and $\delta_i = 0$ if it is right censored, for $i = 1, \ldots, n$. From n pairs $(y_1, \delta_1), \ldots, (y_n, \delta_n)$, where $y_i = \log(z_i)$ (that is, the ith log-lifetime or log-censoring time), the total log-likelihood function for $\boldsymbol{\theta} = (\lambda, \alpha, \boldsymbol{\beta}^\top)^\top$ under non-informative censoring can be expressed, apart from an unimportant constant, as

$$\ell(\boldsymbol{\theta}) = n\log(\lambda) + \sum_{i=1}^{n} \delta_i \left\{\log(\xi_{i1}) - (1/2)\xi_{i2}^2 - \log[\Phi(-\xi_{i2})] - \log[1 - \bar{\lambda}\,\Phi(-\xi_{i2})]\right\}$$
$$+ \sum_{i=1}^{n} \log\left[\frac{\Phi(-\xi_{i2})}{1 - \bar{\lambda}\,\Phi(-\xi_{i2})}\right], \tag{5}$$

where

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

with $\mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, for $i = 1, \ldots, n$. The maximum likelihood estimates of the unknown parameters are obtained by maximizing the log-likelihood function in (5) with respect to $\boldsymbol{\theta} = (\lambda, \alpha, \boldsymbol{\beta}^\top)^\top$.

The log-likelihood function $\ell(\boldsymbol{\theta})$ in (5) is assumed regular (Cox and Hinkley, 1974, Chapter 9) with respect to all derivatives up to second order. The score functions for the parameters λ, α and $\boldsymbol{\beta}$ are obtained by taking the partial derivatives of (5) with respect to these parameters. They are given by

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} \frac{(1+\delta_i)\,\Phi(-\xi_{i2})}{1 - \bar{\lambda}\,\Phi(-\xi_{i2})}, \qquad \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{s},$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = -\frac{q}{\alpha} + \frac{1}{\alpha}\sum_{i=1}^{n} \delta_i \left\{\xi_{i2}^2 - \frac{\xi_{i2}\,\phi(\xi_{i2})}{\Phi(-\xi_{i2})} + \frac{\bar{\lambda}\,\xi_{i2}\,\phi(\xi_{i2})}{1 - \bar{\lambda}\,\Phi(-\xi_{i2})}\right\} + \frac{1}{\alpha}\sum_{i=1}^{n}\left\{\frac{\xi_{i2}\,\phi(\xi_{i2})}{\Phi(-\xi_{i2})} + \frac{\bar{\lambda}\,\xi_{i2}\,\phi(\xi_{i2})}{1 - \bar{\lambda}\,\Phi(-\xi_{i2})}\right\},$$

where q is the number of uncensored observations (failures), $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$ is a known model matrix of full rank (i.e. $\text{rank}(\boldsymbol{X}) = p$) and $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{\theta}) = (s_1, \ldots, s_n)^\top$ with

$$s_i = s_i(\boldsymbol{\theta}) = \frac{\delta_i}{2}\left\{\xi_{i1}\,\xi_{i2} - \frac{\xi_{i2}}{\xi_{i1}} - \frac{\xi_{i1}\,\phi(\xi_{i2})}{\Phi(-\xi_{i2})} + \frac{\bar{\lambda}\,\xi_{i1}\,\phi(\xi_{i2})}{1 - \bar{\lambda}\,\Phi(-\xi_{i2})}\right\} + \frac{1}{2}\left\{\frac{\xi_{i1}\,\phi(\xi_{i2})}{\Phi(-\xi_{i2})} + \frac{\bar{\lambda}\,\xi_{i1}\,\phi(\xi_{i2})}{1 - \bar{\lambda}\Phi(-\xi_{i2})}\right\}.$$

The maximum likelihood estimate $\widehat{\boldsymbol{\theta}} = (\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\beta}}^\top)^\top$ of $\boldsymbol{\theta} = (\lambda, \alpha, \boldsymbol{\beta}^\top)^\top$ can be obtained by solving the likelihood equations

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda} = 0, \qquad \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = 0, \qquad \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \boldsymbol{0},$$

simultaneously. There is no closed-form expression for the maximum likelihood estimator and its computation has to be performed numerically using a non-linear optimization algorithm. The Newton–Raphson iterative technique could be applied to solve the likelihood equations and obtain the estimate $\widehat{\boldsymbol{\theta}} = (\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\beta}}^\top)^\top$ numerically.

For computing the maximum likelihood estimates, starting values $\lambda^{(0)}$, $\alpha^{(0)}$ and $\boldsymbol{\beta}^{(0)}$ for the algorithm are required. Our suggestion is to use $\boldsymbol{\beta}^{(0)} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$, the ordinary least squares estimate of this parameter vector, as an initial point estimate for $\boldsymbol{\beta}$. We suggest for α the initial value

$$\alpha^{(0)} = \left[\frac{4}{n}\sum_{i=1}^{n}\sinh^2\left(\frac{y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}^{(0)}}{2}\right)\right]^{1/2}.$$

We suggest $\lambda^{(0)} = 1$, which corresponds to the usual BS regression model for censored data. These initial guesses worked well in the applications described in Section 6. The Ox matrix programming language (Doornik, 2009) and the R program (R Development Core Team, 2012) can be used to compute $\widehat{\boldsymbol{\theta}}$ numerically. It should be pointed out that like the SN density function, the bi-modality of the LMOEBS density function may cause multiple maxima of the likelihood function for the LMOEBS regression model. However, we believe that in most cases when the regression model (4) is appropriate the likelihood will have a unique maximum.

The asymptotic inference for the parameter vector $\boldsymbol{\theta} = (\lambda, \alpha, \boldsymbol{\beta}^\top)^\top$ can be based on the normal approximation of $\widehat{\boldsymbol{\theta}} = (\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\beta}}^\top)^\top$. Let $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ be the asymptotic variance–covariance matrix for $\widehat{\boldsymbol{\theta}}$. Then, under some regular conditions stated in Cox and Hinkley (1974, Chapter 9), for n large, $\widehat{\boldsymbol{\theta}} \overset{a}{\sim} \mathcal{N}_{p+2}(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$, where $\overset{a}{\sim}$ denotes approximately distributed. Additionally, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ may be approximated by $-\ddot{\boldsymbol{L}}_{\widehat{\boldsymbol{\theta}}\widehat{\boldsymbol{\theta}}}^{-1}$, where $-\ddot{\boldsymbol{L}}_{\widehat{\boldsymbol{\theta}}\widehat{\boldsymbol{\theta}}}$ is the $(p+2) \times (p+2)$ observed information matrix evaluated at $\widehat{\boldsymbol{\theta}}$ which is obtained from

$$\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{bmatrix} \mathrm{tr}(\boldsymbol{K}_1) & \mathrm{tr}(\boldsymbol{K}_2) & \boldsymbol{s}_1^\top \boldsymbol{X} \\ \mathrm{tr}(\boldsymbol{K}_2) & \mathrm{tr}(\boldsymbol{K}_3) & \boldsymbol{s}_2^\top \boldsymbol{X} \\ \boldsymbol{X}^\top \boldsymbol{s}_1 & \boldsymbol{X}^\top \boldsymbol{s}_2 & \boldsymbol{X}^\top \boldsymbol{M}\boldsymbol{X} \end{bmatrix},$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator, $\boldsymbol{K}_1 = \boldsymbol{K}_1(\boldsymbol{\theta}) = \mathrm{diag}\{k_{11}, \ldots, k_{n1}\}$, $\boldsymbol{K}_2 = \boldsymbol{K}_2(\boldsymbol{\theta}) = \mathrm{diag}\{k_{12}, \ldots, k_{n2}\}$, $\boldsymbol{K}_3 = \boldsymbol{K}_3(\boldsymbol{\theta}) = \mathrm{diag}\{k_{13}, \ldots, k_{n3}\}$, $\boldsymbol{s}_1 = \boldsymbol{s}_1(\boldsymbol{\theta}) = (s_{11}, \ldots, s_{n1})^\top$, $\boldsymbol{s}_2 = \boldsymbol{s}_2(\boldsymbol{\theta}) = (s_{12}, \ldots, s_{n2})^\top$ and $\boldsymbol{M} = \boldsymbol{M}(\boldsymbol{\theta}) = \mathrm{diag}\{m_1, \ldots, m_n\}$. After some algebraic manipulations, we obtain

$$k_{i1} = k_{i1}(\boldsymbol{\theta}) = -\frac{1}{\lambda^2} + \frac{(1+\delta_i)\,\Phi(-\xi_{i2})^2}{[1-\bar{\lambda}\,\Phi(-\xi_{i2})]^2},$$

$$k_{i2} = k_{i2}(\boldsymbol{\theta}) = -\frac{(1+\delta_i)\,\xi_{i2}\,\phi(\xi_{i2})}{\alpha\,[1-\bar{\lambda}\,\Phi(-\xi_{i2})]^2}, \qquad s_{i1} = s_{i1}(\boldsymbol{\theta}) = -\frac{(1+\delta_i)\,\xi_{i1}\,\phi(\xi_{i2})}{2\,[1-\bar{\lambda}\,\Phi(-\xi_{i2})]^2},$$

$$\begin{aligned} k_{i3} &= k_{i3}(\boldsymbol{\theta}) = \frac{\delta_i}{\alpha^2}\left(1 - 3\,\xi_{i2}^2\right) + \frac{(\delta_i - 1)\,\xi_{i2}\,\phi(\xi_{i2})}{\alpha^2\,\Phi(-\xi_{i2})}\left\{2 - \xi_{i2}^2 + \frac{\xi_{i2}\,\phi(\xi_{i2})}{\Phi(-\xi_{i2})}\right\} \\ &\quad + \frac{\bar{\lambda}\,(\delta_i+1)\,\xi_{i2}\,\phi(\xi_{i2})}{\alpha^2\,[1-\bar{\lambda}\,\Phi(-\xi_{i2})]}\left\{-2 + \xi_{i2}^2 + \frac{\bar{\lambda}\,\xi_{i2}\,\phi(\xi_{i2})}{1-\bar{\lambda}\,\Phi(-\xi_{i2})}\right\}, \\ s_{i2} &= s_{i2}(\boldsymbol{\theta}) = -\frac{\delta_i\,\xi_{i1}\,\xi_{i2}}{\alpha} + \frac{(\delta_i - 1)\,\xi_{i1}\,\phi(\xi_{i2})}{2\,\alpha\,\Phi(-\xi_{i2})}\left\{1 - \xi_{i2}^2 + \frac{\xi_{i2}\,\phi(\xi_{i2})}{\Phi(-\xi_{i2})}\right\} \\ &\quad + \frac{\bar{\lambda}\,(\delta_i+1)\,\xi_{i1}\,\phi(\xi_{i2})}{2\,\alpha\,[1-\bar{\lambda}\,\Phi(-\xi_{i2})]}\left\{-1 + \xi_{i2}^2 + \frac{\bar{\lambda}\,\xi_{i2}\,\phi(\xi_{i2})}{1-\bar{\lambda}\,\Phi(-\xi_{i2})}\right\}, \\ m_i &= m_i(\boldsymbol{\theta}) = \frac{\delta_i}{4}\left\{1 - \xi_{i1}^2 - \xi_{i2}^2 - \frac{\xi_{i2}^2}{\xi_{i1}^2}\right\} + \frac{(\delta_i - 1)\,\phi(\xi_{i2})}{4\,\Phi(-\xi_{i2})}\left\{\xi_{i2} - \xi_{i1}^2\,\xi_{i2} + \frac{\xi_{i1}^2\,\phi(\xi_{i2})}{\Phi(-\xi_{i2})}\right\} \\ &\quad + \frac{\bar{\lambda}\,(\delta_i+1)\,\phi(\xi_{i2})}{4\,[1-\bar{\lambda}\,\Phi(-\xi_{i2})]}\left\{-\xi_{i2} + \xi_{i1}^2\,\xi_{i2} + \frac{\bar{\lambda}\,\xi_{i1}^2\,\phi(\xi_{i2})}{1-\bar{\lambda}\,\Phi(-\xi_{i2})}\right\}. \end{aligned}$$

Besides estimation of the model parameters, hypotheses tests can be taken into account. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are disjoint subsets of $\boldsymbol{\theta}$. Consider the test of the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{01}$ against $\mathcal{H}_a : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{01}$, where $\boldsymbol{\theta}_{01}$ is a specified vector. Let $\widetilde{\boldsymbol{\theta}}$ be the restricted maximum likelihood estimator of $\boldsymbol{\theta}$ obtained under $\mathcal{H}_0$. The likelihood ratio (LR) statistic to test $\mathcal{H}_0$ is given by $\Lambda = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widetilde{\boldsymbol{\theta}})\}$. Under $\mathcal{H}_0$ and some regularity conditions, the LR statistic converges in distribution to a chi-square distribution with $\dim(\boldsymbol{\theta}_1)$ degrees of freedom. In particular, the LR statistic to test the null hypothesis $\mathcal{H}_0 : \lambda = 1$ against $\mathcal{H}_a : \lambda \neq 1$ takes the form

$$\Lambda = 2\{\ell(\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\beta}}) - \ell(1, \widetilde{\alpha}, \widetilde{\boldsymbol{\beta}})\},$$

where $\widetilde{\alpha}$ and $\widetilde{\boldsymbol{\beta}}$ are the restricted maximum likelihood estimators of α and $\boldsymbol{\beta}$, respectively, obtained from the maximization of (5) under $\mathcal{H}_0 : \lambda = 1$. The limiting distribution of this statistic is χ_1^2 under the null hypothesis. The null hypothesis is rejected if the test statistic exceeds the upper $100(1-\gamma)\%$ quantile of the χ_1^2 distribution.

4. Diagnostic analysis

Since regression models are sensitive to the underlying model assumptions, generally performing a sensitivity analysis is strongly advisable. In order to assess the sensitivity of the maximum likelihood estimates of the parameters of the regression model (4), the local influence method under three perturbation schemes is carried out. In order to assess departures from the underlying LMOEBS distribution as well as to detect outlying observations, a kind of deviance residual will be considered.

4.1. Local influence

The local influence approach based on normal curvature is an important diagnostic tool for assessing local influence of minor perturbations to a statistical model. Assessing local influence of perturbing a statistical model has been an active area of statistical research in the past twenty years since the seminal work of Cook (1986). Let $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ be a k-dimensional vector of perturbations, where $\boldsymbol{\Omega} \subset \mathbb{R}^k$ is an open set. The perturbed log-likelihood function is denoted by $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$. The vector of no perturbation is $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$, such that $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \ell(\boldsymbol{\theta})$. The Cook's idea for assessing local influence is essentially analyzing the local behavior of the log-likelihood displacement $LD_{\omega} = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{\omega})\}$, where $\widehat{\boldsymbol{\theta}}_{\omega}$ denotes the maximum likelihood estimate under $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$, around $\boldsymbol{\omega}_0$ by evaluating the curvature of the plot of $LD_{\omega_0+a\boldsymbol{d}}$ against a, where $a \in \mathbb{R}$ and $\boldsymbol{d}$ is a unit norm direction. One of the measures of particular interest is the direction $\boldsymbol{d}_{\max}$ corresponding to the largest curvature $C_{\boldsymbol{d}_{\max}}$. Cook (1986) showed that the normal curvature at the direction $\boldsymbol{d}$ is given by

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

where $\boldsymbol{\Delta} = \partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top$ and $-\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}$ is the observed information matrix, both $\boldsymbol{\Delta}$ and $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}$ are evaluated at $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega}_0$. So, the quantity $(1/2)C_{\boldsymbol{d}_{\max}}$ is the largest eigenvalue of $\boldsymbol{B} = -\boldsymbol{\Delta}^\top \ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} \boldsymbol{\Delta}$ and $\boldsymbol{d}_{\max}$ is the corresponding unit norm eigenvector ($\|\boldsymbol{d}_{\max}\| = 1$). The index plot of $\boldsymbol{d}_{\max}$ for the matrix $\boldsymbol{B}$ may show how to perturb the model (or data) to obtain large changes in the estimate of the parameter vector $\boldsymbol{\theta} = (\lambda, \alpha, \boldsymbol{\beta}^\top)^\top$. If the interest lies in computing the local influence for $\boldsymbol{\beta}$, the normal curvature in the direction of the vector $\boldsymbol{d}$ is $C_{\boldsymbol{d};\boldsymbol{\beta}}(\boldsymbol{\theta}) = 2|\boldsymbol{d}^\top \boldsymbol{\Delta}^\top (\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} - \ddot{\boldsymbol{L}}_{22})\boldsymbol{\Delta}\boldsymbol{d}|$, where

$$\ddot{\boldsymbol{L}}_{22} = \begin{bmatrix} \ddot{\boldsymbol{L}}_{11}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \qquad \ddot{\boldsymbol{L}}_{11} = \begin{bmatrix} \mathrm{tr}(\boldsymbol{K}_1) & \mathrm{tr}(\boldsymbol{K}_2) \\ \mathrm{tr}(\boldsymbol{K}_2) & \mathrm{tr}(\boldsymbol{K}_3) \end{bmatrix}.$$

Here, $\boldsymbol{d}_{\max;\boldsymbol{\beta}}$ is the unit norm eigenvector corresponding to the largest eigenvalue of the matrix $\boldsymbol{B}_1 = -\boldsymbol{\Delta}^\top(\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} - \ddot{\boldsymbol{L}}_{22})\boldsymbol{\Delta}$. The index plot of the largest eigenvector of $\boldsymbol{B}_1$ may reveal those influential observations on $\widehat{\boldsymbol{\beta}}$. On the other hand, if the interest lies in computing the local influence for (λ, α), the normal curvature in the direction of the vector $\boldsymbol{d}$ is $C_{\boldsymbol{d};(\lambda,\alpha)}(\boldsymbol{\theta}) = 2|\boldsymbol{d}^\top \boldsymbol{\Delta}^\top (\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} - \ddot{\boldsymbol{L}}_{33})\boldsymbol{\Delta}\boldsymbol{d}|$, where

$$\ddot{\boldsymbol{L}}_{33} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & (\boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{X})^{-1} \end{bmatrix}.$$

Here, $\boldsymbol{d}_{\max;(\lambda,\alpha)}$ is the unit norm eigenvector corresponding to the largest eigenvalue of the matrix $\boldsymbol{B}_2 = -\boldsymbol{\Delta}^\top(\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} - \ddot{\boldsymbol{L}}_{33})\boldsymbol{\Delta}$. The index plot of the largest eigenvector of $\boldsymbol{B}_2$ may reveal those influential observations on $(\widehat{\lambda}, \widehat{\alpha})$.

In the following, we shall obtain the matrix $\boldsymbol{\Delta}$ for the extended BS regression model under three different perturbation schemes, namely: case weighting, response perturbation and covariate perturbation. So, we derive for these three perturbation schemes, the matrix

$$\boldsymbol{\Delta} = \left.\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \left[\boldsymbol{\Delta}_\lambda^\top \quad \boldsymbol{\Delta}_\alpha^\top \quad \boldsymbol{\Delta}_{\boldsymbol{\beta}}^\top\right]^\top.$$

The quantities evaluated at $\widehat{\boldsymbol{\theta}} = (\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\beta}}^\top)^\top$ are written with a circumflex. First, we consider a case weight perturbation which modifies the weight given to each subject in the log-likelihood. After some algebra, we have

$$\boldsymbol{\Delta}_\lambda = (\widehat{w}_{11}, \ldots, \widehat{w}_{n1}), \qquad \boldsymbol{\Delta}_\alpha = (\widehat{w}_{12}, \ldots, \widehat{w}_{n2}), \qquad \boldsymbol{\Delta}_{\boldsymbol{\beta}} = \boldsymbol{X}^\top \widehat{\boldsymbol{S}},$$

where $\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{\theta}) = \mathrm{diag}\{s_1, \ldots, s_n\}$,

$$w_{i1} = w_{i1}(\boldsymbol{\theta}) = \frac{1}{\lambda} - \frac{(1+\delta_i)\,\Phi(-\xi_{i2})}{1 - \bar{\lambda}\,\Phi(-\xi_{i2})},$$

$$w_{i2} = w_{i2}(\boldsymbol{\theta}) = \frac{\delta_i}{\alpha}\left\{-1 + \xi_{i2}^2 - \frac{\xi_{i2}\,\phi(\xi_{i2})}{\Phi(-\xi_{i2})} + \frac{\bar{\lambda}\,\xi_{i2}\,\phi(\xi_{i2})}{1 - \bar{\lambda}\,\Phi(-\xi_{i2})}\right\} + \frac{1}{\alpha}\left\{\frac{\xi_{i2}\,\phi(\xi_{i2})}{\Phi(-\xi_{i2})} + \frac{\bar{\lambda}\,\xi_{i2}\,\phi(\xi_{i2})}{1 - \bar{\lambda}\,\Phi(-\xi_{i2})}\right\}.$$

In the response perturbation scheme, each y_i is perturbed as $y_{i\omega} = y_i + s_y\,\omega_i$, where s_y is a scale factor that may be estimated by the standard deviation of $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$. We obtain

$$\boldsymbol{\Delta}_\lambda = -s_y \widehat{\boldsymbol{s}}_1^\top, \qquad \boldsymbol{\Delta}_\alpha = -s_y \widehat{\boldsymbol{s}}_2^\top, \qquad \boldsymbol{\Delta}_{\boldsymbol{\beta}} = -s_y \boldsymbol{X}^\top \widehat{\boldsymbol{M}}.$$

Finally, under perturbation on a particular continuous explanatory variable, say $\boldsymbol{x}_j$ ($j = 1, \ldots, p$), we follow Thomas and Cook (1990) and replace x_{ij} with $x_{ij\omega} = x_{ij} + s_x\,\omega_i$, where s_x is a scale factor that may be estimated by the standard deviation of $\boldsymbol{x}_j$. It follows that

$$\boldsymbol{\Delta}_\lambda = s_x \widehat{\beta}_j \widehat{\boldsymbol{s}}_1^\top, \qquad \boldsymbol{\Delta}_\alpha = s_x \widehat{\beta}_j \widehat{\boldsymbol{s}}_2^\top, \qquad \boldsymbol{\Delta}_{\boldsymbol{\beta}} = s_x \widehat{\beta}_j \boldsymbol{X}^\top \widehat{\boldsymbol{M}} + s_x \boldsymbol{c}_j \widehat{\boldsymbol{s}}^\top,$$

where $\boldsymbol{c}_j$ denotes a $p \times 1$ vector with 1 at the jth position and zero elsewhere, and $\widehat{\beta}_j$ denotes the jth element of $\widehat{\boldsymbol{\beta}}$, for $j = 1, \ldots, p$.

Table 1
Descriptive measures of R_{MD_i} for non-censored observations and $n = 90$.

α	λ	Skewness	Kurtosis	Q_{99}	$Q_{99.5}$
0.5	0.5	−0.0323	2.8054	2.5066	2.6638
	1.5	−0.0131	2.8083	2.5186	2.6798
	3.0	−0.0048	2.8177	2.5297	2.6861
	4.0	0.0030	2.8073	2.5360	2.6894
1.0	0.5	−0.0314	2.8032	2.5046	2.6613
	1.5	−0.0147	2.7965	2.5182	2.6734
	3.0	−0.0042	2.8115	2.5342	2.6901
	4.0	0.0006	2.8135	2.5386	2.6962
1.5	0.5	−0.0315	2.7979	2.5047	2.6583
	1.5	−0.0131	2.7964	2.5189	2.6783
	3.0	−0.0026	2.8094	2.5341	2.6939
	4.0	0.0001	2.8166	2.5407	2.6965

4.2. Residual analysis

We initially consider the martingale residual proposed by Barlow and Prentice (1988). In parametric lifetime models, the martingale residual can be expressed as $R_{M_i} = \delta_i + \log[S(y_i; \widehat{\boldsymbol{\theta}})]$, for $i = 1, \ldots, n$, where $S(y_i; \widehat{\boldsymbol{\theta}})$ denotes the survival function available at $\widehat{\boldsymbol{\theta}}$. For the extended BS regression model defined in (4), we have

$$R_{M_i} = \delta_i + \log\left[\frac{\widehat{\lambda}\,\Phi(-\widehat{\xi}_{i2})}{1 - \widehat{\bar{\lambda}}\,\Phi(-\widehat{\xi}_{i2})}\right], \quad i = 1, \ldots, n,$$

where $\widehat{\bar{\lambda}} = 1 - \widehat{\lambda}$ and $\widehat{\xi}_{i2} = \xi_{i2}(\widehat{\boldsymbol{\theta}})$. The martingale residuals are skewed, have maximum value $+1$ and minimum value $-\infty$.

Due to the skewness distributional form of R_{M_i}, transformations to achieve a more normal-shaped form would be appropriate for residual analysis. For example, the deviance component residual proposed by Therneau et al. (1990) is a transformation of the martingale residual to attenuate the skewness. It can be expressed in the form

$$R_{MD_i} = \operatorname{sign}(R_{M_i})\left[-2\{R_{M_i} + \delta_i \log(\delta_i - R_{M_i})\}\right]^{1/2}, \quad i = 1, \ldots, n.$$

This transformation was motivated by the deviance component residuals found in generalized linear models. As pointed out by Therneau et al. (1990), the log function inflates the martingale residual close to one, while the square root contracts the large negative values. Also, this transformation leads to the deviance component residual for the Cox's proportional hazard model with no time-dependent variable (Therneau et al., 1990).

We will name R_{MD_i} as the martingale-type residual. It seems reasonable to expect that R_{MD_i} may work well in residual analysis similar to those applied in normal linear regression models. However, since the residuals R_{MD_i} are in fact neither independent nor normal, it is usual to adding envelopes as suggested by Atkinson (1981) into the normal probability plots for R_{MD_i}. If $\lambda = 1$, then the martingale-type residual proposed in this section reduces to the one proposed by Leiva et al. (2007).

The empirical distribution of the martingale-type residual R_{MD_i} for the LMOEBS regression model was investigated by using Monte Carlo simulation experiments for different sample sizes, parameter values and censoring proportions; that is, we conduct a small simulation study to investigate the form of the empirical distribution of the residual R_{MD_i}. We consider the regression model $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, where $\beta_1 = \beta_2 = 1$ and $\varepsilon_i \sim \text{LMOEBS}(\lambda, \alpha, 0)$, for $i = 1, \ldots, n$. The covariate values were selected as random draws from the $\mathcal{U}(0, 1)$ distribution and were kept constant throughout the experiment. The number of Monte Carlo replications was 5000. For each combination of sample sizes, parameter values and censoring proportions, we compute the mean, standard deviation, skewness, kurtosis, quantile 97.5%, quantile 99% and quantile 99.5% of the empirical distribution of R_{MD_i}. To save space, we only show some results of the Monte Carlo experiment for the no censoring case when $n = 90$, $\alpha = 0.5, 1.5$, and $\lambda = 0.5, 1.5, 3.0, 4.0$. Table 1 lists the skewness, kurtosis, quantile 99% and quantile 99.5% of the empirical distribution of R_{MD_i}.

In general, from the Monte Carlo simulation experiments, we observe that the martingale-type residual R_{MD_i} has approximately zero mean and unit standard deviation. It has skewness close to zero, which indicates that it is approximately symmetrical. Also, the kurtosis is near three and the quantiles are close to the quantiles of the standard normal distribution. It suggests a good agreement of the empirical distribution of R_{MD_i} with the standard normal distribution. It should be mentioned that, as the censoring proportion decreases, the empirical distribution of R_{MD_i} approaches faster to the standard normal distribution. Additionally, as λ increases, the empirical distribution of R_{MD_i} becomes less skewed. Finally, generalizations of these results for more general scenarios is not straightforward, and therefore, as suggested by Atkinson (1981), the use of normal probability plots for R_{MD_i} with envelope is recommended.

5. Testing the homogeneity of the shape parameters

In the extended BS regression model introduced in Section 3, the homogeneity of the shape parameters λ and α is a standard assumption. This assumption, however, is not necessarily appropriate, because the actual shape parameters of the response variable y_i may be related to the ith observation. In this case, the inference would be much difficult to deal with. Hence, this assumption usually need to be checked. In this section, we consider a LR test statistic to verify the homogeneity of the shape parameters in the extended BS regression model. This problem has been mentioned by Rieck and Nedelman (1991) for the BS regression model and investigated by Xie and Wei (2007) and Qu and Xie (2011).

We assume that the homogeneous extended BS regression model (4) takes the form

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{6}$$

where $\varepsilon_i \sim \text{LMOEBS}(\lambda_i, \alpha_i, 0)$, $\lambda_i = \lambda\, k_1(\boldsymbol{w}_i, \boldsymbol{\rho})$ and $\alpha_i = \alpha\, k_2(\boldsymbol{z}_i, \boldsymbol{\psi})$. The parameters λ and α are the factors of the shape parameters with weight functions $k_1(\boldsymbol{w}_i, \boldsymbol{\rho})$ and $k_2(\boldsymbol{z}_i, \boldsymbol{\psi})$, respectively. Also, $\boldsymbol{w}_i$ and $\boldsymbol{z}_i$ are $d_1 \times 1$ and $d_2 \times 1$ vectors, respectively, of nonstochastic variables. Notice that $\boldsymbol{x}_i$, $\boldsymbol{w}_i$ and $\boldsymbol{z}_i$ may have common components. The parameter vectors $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{d_1})^\top$ and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{d_2})^\top$ of dimensions d_1 and d_2, respectively, indicate the heterogeneity of the shape parameters. Under the regression model (6), the log-likelihood function for the parameter vector $\boldsymbol{\theta} = (\lambda, \alpha, \boldsymbol{\rho}^\top, \boldsymbol{\psi}^\top, \boldsymbol{\beta}^\top)^\top$ can be expressed, apart from an unimportant constant, in the form

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= n\log(\lambda) + \sum_{i=1}^n \log[k_1(\boldsymbol{w}_i, \boldsymbol{\rho})] + \sum_{i=1}^n \log\left[\frac{\Phi(-\xi_{i2}^*)}{1 - \bar{\lambda}_i\, \Phi(-\xi_{i2}^*)}\right] \\ &\quad + \sum_{i=1}^n \delta_i \left\{\log(\xi_{i1}^*) - (1/2)\xi_{i2}^{*2} - \log[\Phi(-\xi_{i2}^*)] - \log[1 - \bar{\lambda}_i\, \Phi(-\xi_{i2}^*)]\right\},\end{aligned} \tag{7}$$

where $\bar{\lambda}_i = 1 - \lambda\, k_1(\boldsymbol{w}_i, \boldsymbol{\rho})$,

$$\xi_{i1}^* = \xi_{i1}^*(\boldsymbol{\theta}) = \frac{2}{\alpha_i}\cosh\left(\frac{y_i - \mu_i}{2}\right), \qquad \xi_{i2}^* = \xi_{i2}^*(\boldsymbol{\theta}) = \frac{2}{\alpha_i}\sinh\left(\frac{y_i - \mu_i}{2}\right),$$

with $\mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, for $i = 1, \ldots, n$.

The test of homogeneity of the shape parameters lies in testing the null hypothesis $\mathcal{H}_0 : (\boldsymbol{\rho}, \boldsymbol{\psi}) = (\boldsymbol{\rho}_0, \boldsymbol{\psi}_0)$ against $\mathcal{H}_a : (\boldsymbol{\rho}, \boldsymbol{\psi}) \neq (\boldsymbol{\rho}_0, \boldsymbol{\psi}_0)$, where $\boldsymbol{\rho}_0$ and $\boldsymbol{\psi}_0$ are vectors of known scalars. It is assumed that there is a unique value of $\boldsymbol{\rho}_0$ and $\boldsymbol{\psi}_0$ such that $k_1(\boldsymbol{w}_i, \boldsymbol{\rho}_0) = 1$ and $k_2(\boldsymbol{z}_i, \boldsymbol{\psi}_0) = 1$. Let $\widehat{\boldsymbol{\theta}} = (\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\rho}}^\top, \widehat{\boldsymbol{\psi}}^\top, \widehat{\boldsymbol{\beta}}^\top)^\top$ and $\widetilde{\boldsymbol{\theta}} = (\widetilde{\lambda}, \widetilde{\alpha}, \boldsymbol{\rho}_0^\top, \boldsymbol{\psi}_0^\top, \widetilde{\boldsymbol{\beta}}^\top)^\top$ be the unrestricted and restricted (obtained from the maximization of (7) under $\mathcal{H}_0$) maximum likelihood estimators of $\boldsymbol{\theta} = (\lambda, \alpha, \boldsymbol{\rho}^\top, \boldsymbol{\psi}^\top, \boldsymbol{\beta}^\top)^\top$, respectively. Here, λ, α and $\boldsymbol{\beta}$ act as nuisance parameters. The LR statistic for testing $\mathcal{H}_0$ is given by $\Upsilon = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widetilde{\boldsymbol{\theta}})\}$. Under the usual regularity conditions and under $\mathcal{H}_0$, $\Upsilon \xrightarrow{D} \chi^2_{d_1+d_2}$, where $\xrightarrow{D}$ denotes convergence in distribution, so that a test can be performed using approximate critical values from the asymptotic $\chi^2_{d_1+d_2}$ distribution. After some algebra, we can write

$$\begin{aligned}\Upsilon &= 2n\log\left(\frac{\widehat{\lambda}}{\widetilde{\lambda}}\right) + 2\sum_{i=1}^n \log\left[\frac{k_1(\boldsymbol{w}_i, \widehat{\boldsymbol{\rho}})}{k_1(\boldsymbol{w}_i, \boldsymbol{\rho}_0)}\right] + 2\sum_{i=1}^n \log\left[\frac{\Phi(-\widehat{\xi}_{i2}^*)[1 - \widetilde{\bar{\lambda}}_i\, \Phi(-\widetilde{\xi}_{i2}^*)]}{\Phi(-\widetilde{\xi}_{i2}^*)[1 - \widehat{\bar{\lambda}}_i\, \Phi(-\widehat{\xi}_{i2}^*)]}\right] \\ &\quad + 2\sum_{i=1}^n \delta_i \left\{\log\left(\frac{\widehat{\xi}_{i1}^*}{\widetilde{\xi}_{i1}^*}\right) - \frac{1}{2}\left(\widehat{\xi}_{i2}^{*2} - \widetilde{\xi}_{i2}^{*2}\right) - \log\left[\frac{\Phi(-\widehat{\xi}_{i2}^*)}{\Phi(-\widetilde{\xi}_{i2}^*)}\right] - \log\left[\frac{1 - \widehat{\bar{\lambda}}_i\, \Phi(-\widehat{\xi}_{i2}^*)}{1 - \widetilde{\bar{\lambda}}_i\, \Phi(-\widetilde{\xi}_{i2}^*)}\right]\right\},\end{aligned}$$

where $\widehat{\bar{\lambda}}_i = 1 - \widehat{\lambda}\, k_1(\boldsymbol{w}_i, \widehat{\boldsymbol{\rho}})$, $\widetilde{\bar{\lambda}}_i = 1 - \widetilde{\lambda}\, k_1(\boldsymbol{w}_i, \boldsymbol{\rho}_0)$, $\widehat{\xi}_{i1}^* = \xi_{i1}^*(\widehat{\boldsymbol{\theta}})$, $\widehat{\xi}_{i2}^* = \xi_{i2}^*(\widehat{\boldsymbol{\theta}})$, $\widetilde{\xi}_{i1}^* = \xi_{i1}^*(\widetilde{\boldsymbol{\theta}})$ and $\widetilde{\xi}_{i2}^* = \xi_{i2}^*(\widetilde{\boldsymbol{\theta}})$. The null hypothesis is rejected if the observed value of Υ exceeds the upper $100(1-\gamma)\%$ quantile of the $\chi^2_{d_1+d_2}$ distribution.

In practical applications, explicit forms for $k_1(\boldsymbol{w}_i, \boldsymbol{\rho})$ and $k_2(\boldsymbol{z}_i, \boldsymbol{\psi})$ need to be provided. As suggested by Cook and Weisberg (1982), the exponential function is usually employed in practice. For example, we can assume $k_1(\boldsymbol{w}_i, \boldsymbol{\rho}) = \exp(\boldsymbol{w}_i^\top \boldsymbol{\rho})$ and $k_2(\boldsymbol{z}_i, \boldsymbol{\psi}) = \exp(\boldsymbol{\psi}^\top \boldsymbol{z}_i)$. Note that $\boldsymbol{\rho} = \boldsymbol{0}$ and $\boldsymbol{\psi} = \boldsymbol{0}$ imply $k_1(\boldsymbol{w}_i, \boldsymbol{\rho}) = k_2(\boldsymbol{z}_i, \boldsymbol{\psi}) = 1$, and therefore, $\lambda_i = \lambda$ and $\alpha_i = \alpha$ for all $i = 1, \ldots, n$.

6. Real data illustrations

In this section, we use two real data sets to show the flexibility and applicability of the extended BS regression model in practice. We will consider real data with and without censoring. All the computations presented in this section were done using the `Ox` matrix programming language (Doornik, 2009), which is freely distributed for academic purposes and available at http://www.doornik.com. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) method with analytical derivatives through the subroutine `maxBFGS` has been used for maximizing the log-likelihood function.

Table 2
Maximum likelihood estimates; first data set.

Parameter	LBS model		LMOEBS model	
	Estimate	SE	Estimate	SE
β_1	0.0978	0.1707	−0.7899	0.4317
β_2	−14.1164	1.5714	−14.4809	1.3242
α	1.2791	0.1438	1.4249	0.2505
λ			4.9578	2.7537

Table 3
AIC, BIC and HQIC for the fitted models; first data set.

Criterion	Model	
	LBS	LMOEBS
AIC	55.72	49.16
BIC	60.79	55.92
HQIC	57.55	51.60

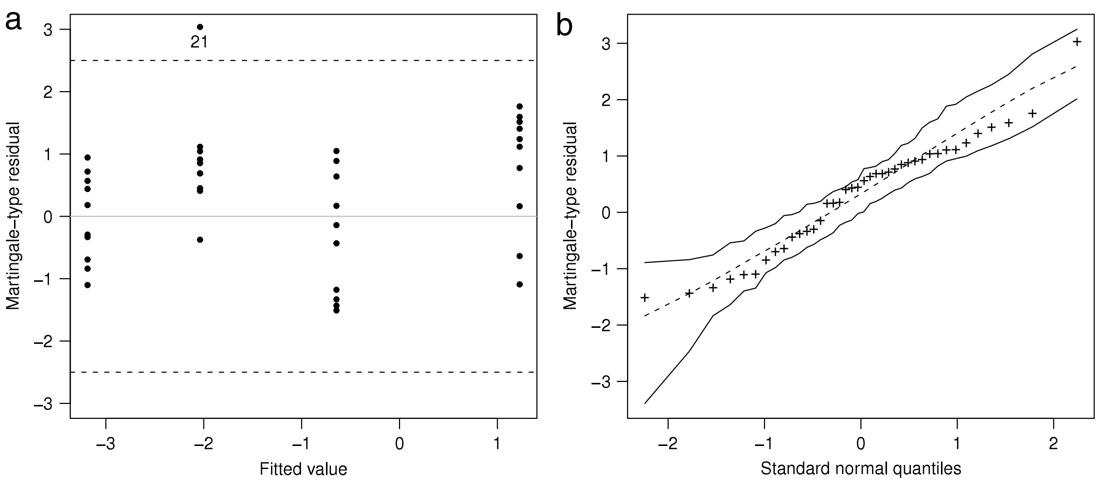

Fig. 3. First data set: (a) martingale-type residuals against the fitted values; (b) normal probability plot with envelope.

First, we consider the real data provided by McCool (1980) and recently analyzed in Chan et al. (2008) by using an extreme-value regression model. These data consist of times to failure (T) in rolling contact fatigue of ten hardened steel specimens tested at each of four values of four contact stress (x_2). The data were obtained using a 4-ball rolling contact test rig at the Princeton Laboratories of Mobil Research and Development Co. It should be noticed that the fatigue processes are by excellence ideally modeled by the BS distribution due to its genesis. So, based on this argument, we will use the extended BS regression model for analyzing these data. We follow Chan et al. (2008) and consider the regression model

$$y_i = \beta_1 x_{i1} + \beta_2 \log(x_{i2}) + \varepsilon_i, \quad i = 1, \ldots, 40, \tag{8}$$

where y_i is the logarithm of the failure time, $x_{i1} = 1$ and the ε_i's are independent and identically distributed such that $\varepsilon_i \sim \text{LMOEBS}(\lambda, \alpha, 0)$.

Table 2 lists the maximum likelihood estimates of the model parameters and the asymptotic standard errors (SE) for the extended BS and BS regression models. To compare these regression models, we consider selection criteria on the candidate models (see Table 3). According to the AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and HQIC (Hannan–Quinn Information Criterion) criteria, the new extended BS regression model outperforms the BS regression model and therefore should be preferred. Additionally, the observed value of the LR statistic for testing the null hypothesis $\mathcal{H}_0 : \lambda = 1$ against $\mathcal{H}_a : \lambda \neq 1$, is $\Lambda = 8.56$ (p-value $= 0.0034$). So, the null hypothesis $\mathcal{H}_0 : \lambda = 1$ is strongly rejected at any usual significance level. Therefore, we select the extended BS regression model as our working model.

In order to detect possible departures from the assumption of LMOEBS errors in the model (8) as well as outlying observations, we present in Fig. 3(a) the martingale-type residuals R_{MD_i} against the fitted values and the corresponding normal probability plot with generated envelopes for R_{MD_i} in Fig. 3(b). Fig. 3(a) indicates a large positive residual (case #21), whereas Fig. 3(b) reveals that the assumption of LMOEBS error seems to be suitable, since there are no observations falling outside the envelope. The observation #21 was highlighted by Fig. 3(a). This case corresponds to the smallest value of the time to failure. In Fig. 4 we present the index plots of the absolute value of $\boldsymbol{d}_{\max}$ (that is, $|\boldsymbol{d}_{\max}|$) for $(\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\beta}})$, $\widehat{\boldsymbol{\beta}}$ and $(\widehat{\lambda}, \widehat{\alpha})$

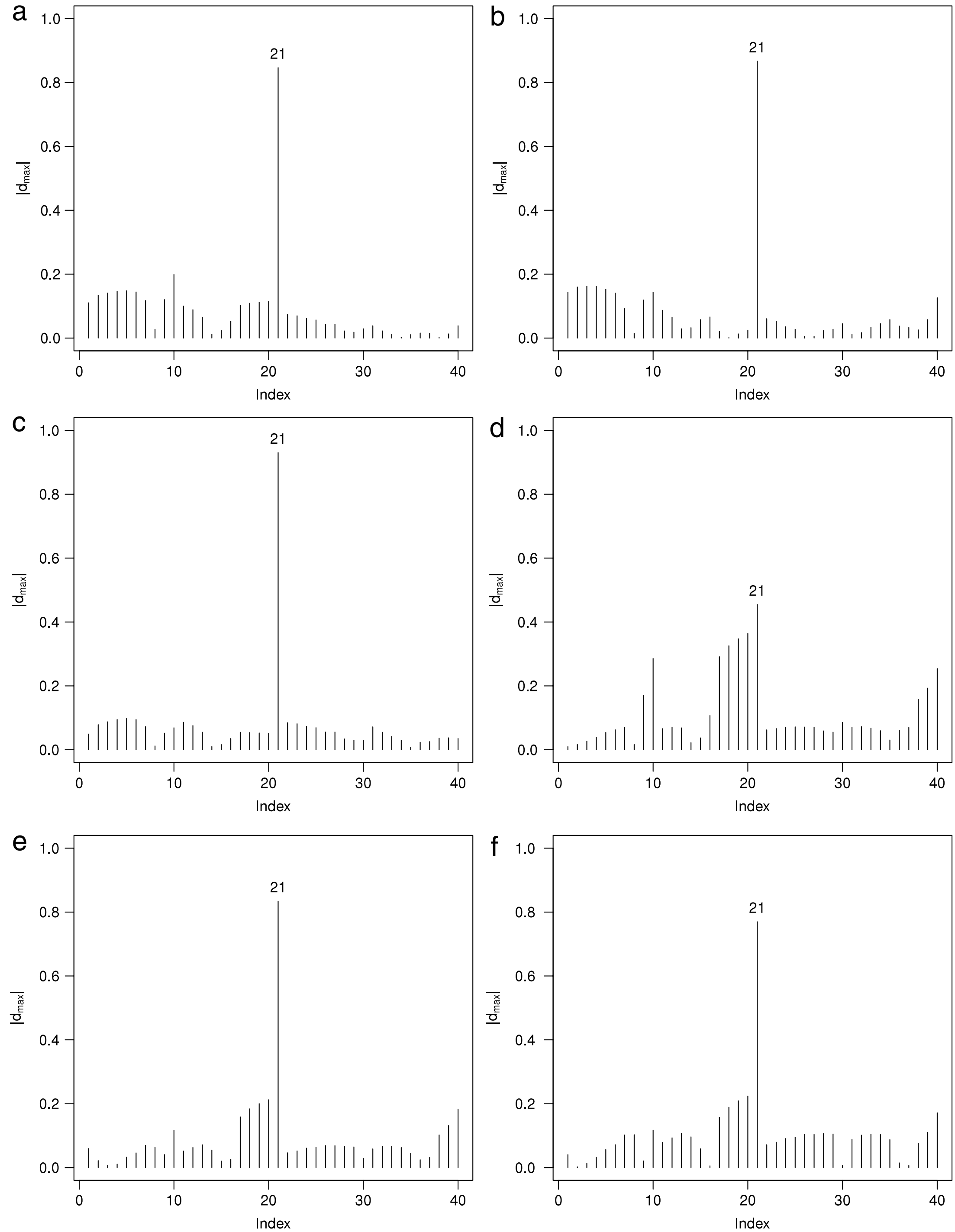

Fig. 4. Index plots of $|\boldsymbol{d}_{\max}|$: case-weight perturbation for $(\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\beta}})$ (a), $\widehat{\boldsymbol{\beta}}$ (b) and $(\widehat{\lambda}, \widehat{\alpha})$ (c); response perturbation for $(\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\beta}})$ (d), $\widehat{\boldsymbol{\beta}}$ (e) and $(\widehat{\lambda}, \widehat{\alpha})$ (f); first data set.

under the perturbation scheme indicated. Note that the observation #21 appears as the most influential in all the graphs. Therefore, the diagnostic analysis detected as potentially influential on the parameter estimates the case #21.

In order to reveal the impact of the observation #21 on the parameter estimates, Table 4 shows the absolute changes (AC) in the parameter estimates after dropping the case #21 from the real data set. We also present the corresponding p-values (in parentheses) for the new estimates in this table. The AC of each estimate is defined as $\mathrm{AC}_{\theta_j} = |\widehat{\theta}_j - \widehat{\theta}_{j(I)}|$, where $\widehat{\theta}_{j(I)}$ denotes the maximum likelihood estimate of θ_j, with $j = 1, \ldots, k$ (where k is the total number of parameters), after the set

Table 4
AC and the corresponding p-values in parentheses; first data set.

Estimated parameter	Dropped observation	
	None	#21
$\widehat{\beta}_1$	–	1.711
	(0.067)	(0.028)
$\widehat{\beta}_2$	–	3.414
	(0.000)	(0.000)
$\widehat{\lambda}$	–	4.677
$\widehat{\alpha}$	–	0.465

Table 5
Maximum likelihood estimates; second data set.

Parameter	LBS model		LMOEBS model	
	Estimate	SE	Estimate	SE
β_1	1.1417	0.6375	0.9004	0.6084
β_2	0.0405	0.0050	0.0317	0.0053
β_3	−0.0027	0.0086	0.0004	0.0080
β_4	0.0218	0.0080	0.0125	0.0084
β_5	−0.0021	0.0227	−0.0001	0.0212
β_6	−0.2800	0.3032	0.3629	0.2717
β_7	−0.7053	0.3009	−0.4101	0.2451
β_8	−0.6909	0.3667	−0.6902	0.2707
β_9	−0.3831	0.1926	−0.3407	0.1829
α	1.2619	0.0792	1.4334	0.1650
λ			4.1969	1.7436

I of observations has been removed. The figures in Table 4 reveal that the significance of the parameters β_1 and β_2 are not modified when the observation #21 is removed from the data set, that is, the case #21 does not change the significance of these parameters in the regression model (8). Therefore, the LMOEBS regression model is suitable to model these real data.

Next, as a second application, we shall consider the lung cancer data presented by Kalbfleisch and Prentice (2002, p. 378). In this trial, males with advanced inoperable lung cancer were randomized to either standard or test chemotherapy. Only 9 of the 137 survival times were censored. The variables considered in this study are: the survival time (T, in days) of the patients with lung cancer; a measure, at randomization, of the patient's performance status (Karnofsky rating) (x_2), where 10–30 is completely hospitalized, 40–60 is partial confinement and 70–90 is able to take care of self; time in months from diagnosis to randomization (x_3); age in years (x_4); prior therapy (x_5), a dichotomous variable taking the value 10 for yes and 0 for no; histological type of tumor, which has the categories squamous, small cell, adeno and large cell, making necessary the use of dummy variables given by $x_6 = 1$, $x_7 = 1$ and $x_8 = 1$ if the type of cancer cell is squamous, small and adeno, respectively, and 0 otherwise; type of treatment (x_9),which takes the value 0 for standard chemotherapy and 1 for test chemotherapy. One of the objectives of this study was to explain the survival time T by using a regression model with the explanatory variables described above.

Lee and Wang (2003) applied various models to fit this data set with different error distributions, such as the exponential, generalized gamma, log–logistic, log-normal and Weibull. More recently, Barros et al. (2008) considered these data in the BS context. According to these authors, an argument for the use of the BS distribution is the possibility of relating the propagation lifetimes that lead to a fatigue process with some cumulative damage. On the basis of the same argument, we will use the extended BS regression model for analyzing these data.

Firstly, we consider the following regression model:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_9 x_{i9} + \varepsilon_i, \quad i = 1, \ldots, 137, \tag{9}$$

where y_i is the logarithm of the survival or censoring time, $x_{i1} = 1$ and the ε_i's are independent and identically distributed such that $\varepsilon_i \sim \text{LMOEBS}(\lambda, \alpha, 0)$. Table 5 lists the maximum likelihood estimates of the model parameters and the asymptotic SE for the extended BS and BS regression models. The AIC, BIC and HQIC criteria for both the models are presented in Table 6. From this table, it is evident that the new extended BS regression model outperforms the BS regression model irrespective of the criteria and therefore should be preferred. Additionally, the observed value of the LR statistic for testing the null hypothesis $\mathcal{H}_0 : \lambda = 1$ against $\mathcal{H}_a : \lambda \neq 1$ is in accordance with the information criteria, that is, we obtain $\Lambda = 7.96$ and the associated critical point of the χ^2_1 distribution at the 5% significance level, for instance, is 3.84, which yields a p-value of less than 0.005. It implies that the null hypothesis $\mathcal{H}_0 : \lambda = 1$ is strongly rejected at any usual significance level. Therefore, we select the extended BS regression model as our working model. We note that the predictors x_3, x_4, x_5 and x_6 are not marginally significant at the level of 10% in model (9).

In order to detect possible departures from the assumption of LMOEBS errors in the model (9) as well as outlying observations, we present in Fig. 5(a) the martingale-type residuals R_{MD_i} against the fitted values and the corresponding normal probability plot with generated envelopes for R_{MD_i} in Fig. 5(b). Fig. 5(a) reveals two large positive residuals (patients

Table 6
AIC, BIC and HQIC for the fitted models; second data set.

Criterion	Model	
	LBS	LMOEBS
AIC	189.39	183.45
BIC	218.59	215.57
HQIC	201.26	196.50

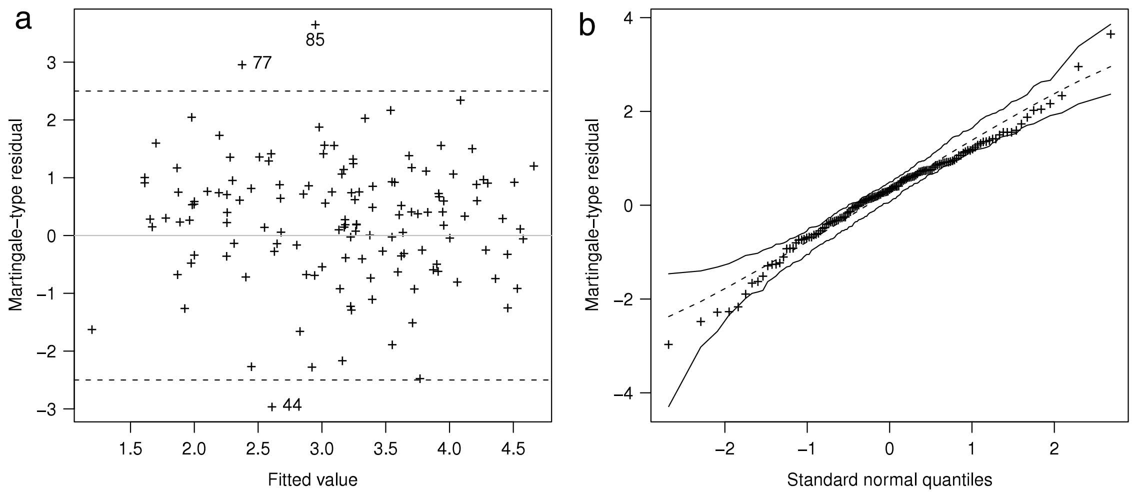

Fig. 5. Second data set: (a) martingale-type residuals against the fitted values; (b) normal probability plot with envelope.

#77 and #85) and a large negative residual (patient #44). Fig. 5(b) indicates that the assumption of LMOEBS error seems to be suitable, since there are no observations falling outside the envelope. As can be observed, Fig. 5(a) highlights strongly the observation #85. It corresponds to a patient who had waited for 7 months until randomization. Also, he did not have any prior therapy and his performance was partial confinement. He received the test chemotherapy treatment and his histological tumor type was squamous. He corresponds to a 35-year-old patient whose survival time was one day. The case #85 is the youngest patient with survival time less than 4 days. As we will see in the following, the patients #44, #77 and #85 appear as potentially influential on the parameter estimates.

Fig. 6 shows the index plots of $|\boldsymbol{d}_{\max}|$ for $(\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\beta}})$, $\widehat{\boldsymbol{\beta}}$ and $(\widehat{\lambda}, \widehat{\alpha})$ under the perturbation scheme indicated. From this figure, notice that the observations #44 and #85 appear as the most influential in all the graphs. Other observations also appear with some outstanding influence on the parameter estimates. For example, observations #17 and #77 appear as possible influential on $(\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\beta}})$ and on $\widehat{\boldsymbol{\beta}}$ under the perturbation scheme indicated. Therefore, the diagnostic analysis detected as potentially influential the following four cases: #17, #44, #77 and #85.

In order to reveal the impact of these four observations (cases #17, #44, #77 and #85) on the parameter estimates, Table 7 shows the AC in the estimates after dropping one of the four cases with outstanding influence and also when all of them are dropped at once (represented by the set $\mathcal{A} = \{17, 44, 77, 85\}$). We also present the corresponding p-values (in parentheses) for the new estimates in this table. From the figures in Table 7, note that observations #17 and #77 are highly influential for the estimate of β_4; indeed, the age in years becomes significant at the 5% and 10% nominal levels when these observations are not in the data, respectively. Observation #44 is highly influential for the estimate of β_6, which makes this parameter significant at the 10% nominal level when this observation is not in the data. Additionally, looking at Table 7, we can notice that the elimination of observations #44, #85 and the set $\mathcal{A}$ make the explanatory variable x_9 (type of treatment) non-significant; that is, the significance of this variable in the extended BS regression model was masked by these observations, so it should be removed from the model. Finally, Table 7 reveals that the highest values of the AC correspond to the estimate of the shape parameter λ.

Therefore, on the basis of the above analysis, the survival time (in days) of the patients with lung cancer depends on the performance status and on the histological type of tumor, that is, the final selected model takes the form

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8} + \varepsilon_i, \quad i = 1, \ldots, 137, \tag{10}$$

where $x_{i1} = 1$ and $\varepsilon_i \sim$ LMOEBS$(\lambda, \alpha, 0)$. The maximum likelihood estimates of the parameters (estimated SE in parenthesis) are: $\widehat{\lambda} = 5.0287\,(1.7232)$, $\widehat{\alpha} = 1.5573\,(0.1771)$, $\widehat{\beta}_1 = 1.4423\,(0.3977)$, $\widehat{\beta}_2 = 0.0288\,(0.0048)$, $\widehat{\beta}_6 = 0.2858\,(0.2517)$, $\widehat{\beta}_7 = -0.3541\,(0.2409)$ and $\widehat{\beta}_8 = -0.7820\,(0.2527)$. From these final estimates, note that there is no significant difference (at the 10% nominal level) among the squamous, small and large types of tumors, whereas there is significant difference (at the 1% nominal level) between the adeno and large types of tumors. We may interpret the estimated coefficients of the

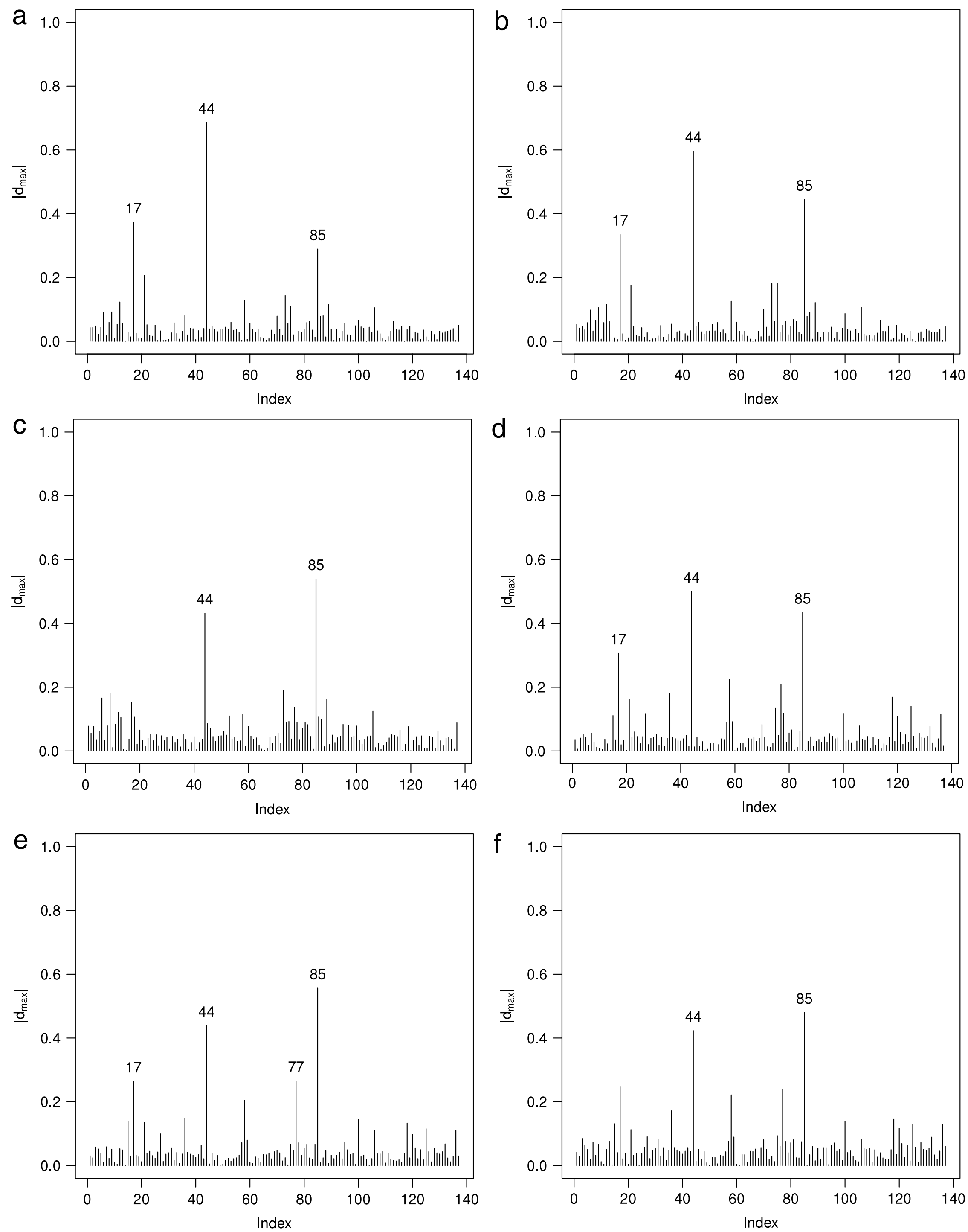

Fig. 6. Index plots of $|\boldsymbol{d}_{\max}|$: case-weight perturbation for $(\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\beta}})$ (a), $\widehat{\boldsymbol{\beta}}$ (b) and $(\widehat{\lambda}, \widehat{\alpha})$ (c); response perturbation for $(\widehat{\lambda}, \widehat{\alpha}, \widehat{\boldsymbol{\beta}})$ (d), $\widehat{\boldsymbol{\beta}}$ (e) and $(\widehat{\lambda}, \widehat{\alpha})$ (f); second data set.

final model (10) as the following. The expected survival time should increase with the performance status and no significance appears among the squamous, small and large types of tumors. On the other hand, assuming that the performance status is fixed, the survival time is expected to decrease 119% for the adeno tumor type with respect to the large one.

Now, we shall verify if the assumption of homogeneity of the shape parameters in the regression models (8) and (10) are appropriate to model the first and second real data sets, respectively. For the regression model (8), we assume $k_1(\boldsymbol{w}_i, \boldsymbol{\rho}) = \exp(\rho \log(x_{i2}))$ and $k_2(\boldsymbol{z}_i, \boldsymbol{\psi}) = \exp(\psi \log(x_{i2}))$. Additionally, we assume $k_1(\boldsymbol{w}_i, \boldsymbol{\rho}) = \exp(\rho\, x_{i2})$ and $k_2(\boldsymbol{z}_i, \boldsymbol{\psi}) = \exp(\psi\, x_{i2})$

Table 7
AC and the corresponding p-values in parentheses; second data set.

Estimated parameter	Dropped observation					
	None	#17	#44	#77	#85	Set $\mathcal{A}$
$\widehat{\beta}_1$	–	0.311	0.002	0.091	1.031	0.829
	(0.140)	(0.324)	(0.125)	(0.094)	(0.006)	(0.006)
$\widehat{\beta}_2$	–	0.001	0.001	0.001	0.002	0.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\widehat{\beta}_3$	–	0.001	0.004	0.000	0.002	0.001
	(0.963)	(0.866)	(0.560)	(0.989)	(0.833)	(0.867)
$\widehat{\beta}_4$	–	0.004	0.006	0.002	0.008	0.005
	(0.137)	(0.037)	(0.451)	(0.077)	(0.562)	(0.299)
$\widehat{\beta}_5$	–	0.004	0.001	0.007	0.020	0.002
	(0.995)	(0.854)	(0.965)	(0.766)	(0.358)	(0.920)
$\widehat{\beta}_6$	–	0.033	0.057	0.007	0.158	0.049
	(0.182)	(0.128)	(0.083)	(0.198)	(0.514)	(0.216)
$\widehat{\beta}_7$	–	0.142	0.146	0.038	0.104	0.351
	(0.094)	(0.019)	(0.011)	(0.068)	(0.067)	(0.001)
$\widehat{\beta}_8$	–	0.027	0.005	0.013	0.067	0.108
	(0.011)	(0.005)	(0.005)	(0.009)	(0.013)	(0.002)
$\widehat{\beta}_9$	–	0.031	0.112	0.004	0.122	0.276
	(0.062)	(0.072)	(0.175)	(0.060)	(0.230)	(0.685)
$\widehat{\lambda}$	–	0.861	2.168	0.475	2.176	1.297
$\widehat{\alpha}$	–	0.021	0.095	0.088	0.253	0.362

for the regression model (10). Note that $\rho = 0$ and $\psi = 0$ imply $k_1(\boldsymbol{w}_i, \boldsymbol{\rho}) = k_2(\boldsymbol{z}_i, \boldsymbol{\psi}) = 1$, and therefore, $\lambda_i = \lambda$ and $\alpha_i = \alpha$ $(i = 1, \ldots, n)$ for both the regression models. Hence, the test for the homogeneity of the shape parameters becomes the test of the null hypothesis $\mathcal{H}_0 : (\rho, \psi) = (0, 0)$ against the alternative hypothesis $\mathcal{H}_a : (\rho, \psi) \neq (0, 0)$. By a little computation, we have that the LR test statistics are $\Upsilon = 7.059$ (p-value $= 0.029$) and $\Upsilon = 12.297$ (p-value $= 0.002$) for the regression models (8) and (10), respectively. Therefore, we reject the null hypothesis $\mathcal{H}_0 : (\rho, \psi) = (0, 0)$ at the 5% nominal level for the both regression models and the assumption of homogeneity of the shape parameters seems not suitable for the first real data as well as for the second real data.

Finally, it should be mentioned that the lung cancer data (i.e. second data set) have also been analyzed in Li and Xie (2012). They showed that the assumption of homogeneity for the shape parameter of the BS Student-t regression model is not suitable for these data by using a score test statistic. Therefore, on the basis of the above discussions, it is evident that a heterogeneous extended BS regression model needs to be introduced and of course deserve a separate paper. This model will be a generalization of the proposed extended BS regression model and the inference in this case is much more difficult to deal with. Future research regarding a heterogeneous extended BS regression model will be discussed in a separate paper elsewhere.

7. Concluding remarks

The BS distribution has many attractive properties and has found several applications in the literature including lifetime, survival and environmental data analysis. It has received significant attention over the last few years and some generalizations and extensions of this distribution have been proposed by many researchers. Based on the BS distribution, Rieck and Nedelman (1991) introduced the BS regression model, which has been studied by several authors. Their regression model is becoming increasingly popular to model times to failure for materials subject to fatigue and for modeling lifetime data. In this paper, we proposed a new class of extended BS regression models on the basis of the extended BS distribution introduced by Lemonte (2013), which generalizes the BS regression model in Rieck and Nedelman (1991) and in Leiva et al. (2007). The new class of regression models can serve as a good alternative for lifetime data analysis and it is much more flexible than the usual BS regression model in analyzing lifetime data in many practical situations. The parameter estimation of the new regression model is approached by maximum likelihood and the observed information matrix is derived. We discuss diagnostic techniques in the extended BS regression model. Diagnostic methods have been an important tool in regression analysis to detect anomalies with the fitted model, such as departures from the error assumptions, presence of outliers and presence of influential observations. In particular, appropriate matrices for assessing local influence on the parameter estimates under different perturbation schemes are obtained, which are quite simple, compact and can be easily implemented into any mathematical or statistical/econometric programming environment with numerical linear algebra facilities. We investigate a test of homogeneity of the shape parameters in the extended BS regression model based on the likelihood ratio statistic. Two applications to real data sets are presented to show that the new extended BS regression model provides a better fit than the usual BS regression model. It illustrates the fact that there is still room for improving the usual BS regression model. We hope that the proposed regression model may attract wider applications in survival analysis and fatigue life modeling.

Acknowledgments

The author thanks the associate editor and three anonymous referees for useful suggestions and comments that aided in improving the first version of the manuscript. The author gratefully acknowledges grants from FAPESP (Brazil).

References

Atkinson, A.C., 1981. Two graphical display for outlying and influential observations in regression. Biometrika 68, 13–20.
Balakrishnan, N., Leiva, V., López, J., 2007. Acceptance sampling plans from truncated life tests from generalized Birnbaum–Saunders distribution. Communications in Statistics—Simulation and Computation 36, 643–656.
Barlow, W.E., Prentice, R.L., 1988. Residuals for relative risk regression. Biometrika 75, 65–74.
Barros, M., Paula, G.A., Leiva, V., 2008. A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. Lifetime Data Analysis 14, 316–332.
Birnbaum, Z.W., Saunders, S.C., 1969. A new family of life distributions. Journal of Applied Probability 6, 319–327.
Chan, P.S., Ng, H.K.T., Balakrishnan, N., Zhou, Q., 2008. Point and interval estimation for extreme-value regression model under Type-II censoring. Computational Statistics ans Data Analysis 52, 4040–4058.
Cook, R.D., 1977. Detection of influential observation in linear regression. Technometrics 19, 15–18.
Cook, R.D., 1986. Assessment of local influence. Journal of the Royal Statistical Society B 48, 133–169 (with discussion).
Cook, R.D., Weisberg, S., 1982. Residuals and Influence in Regression. Chapman and Hall, London.
Cox, D.R., Hinkley, D.V., 1974. Theoretical Statistics. Chapman and Hall, London.
Desmond, A.F., 1985. Stochastic models of failure in random environments. Canadian Journal of Statistics 13, 171–183.
Desmond, A.F., 1986. On the relationship between two fatigue-life models. IEEE Transactions on Reliability 35, 167–169.
Desmond, A.F., González, C.L.C., Singh, R.S., Lu, X., 2012. A mixed effects log-linear model based on the Birnbaum–Saunders distribution. Computational Statistics and Data Analysis 56, 399–407.
Desmond, A.F., Rodríguez-Yam, G.A., Lu, X., 2008. Estimation of parameters for a Birnbaum–Saunders regression model with censored data. Journal of Statistical Computation and Simulation 78, 983–997.
Doornik, J.A., 2009. An Object-Oriented Matrix Language—Ox 6. Timberlake Consultants Press, London.
Espinheira, P.L., Ferrari, S.L.P., Cribari-Neto, F., 2008. Influence diagnostics in beta regression. Computational Statistics and Data Analysis 52, 4417–4431.
Gijbels, I., 2010. Censored data. Wiley Interdisciplinary Reviews: Computational Statistics 2, 178–188.
Jones, M.C., 2012. Relationships between distributions with certain symmetries. Statistics and Probability Letters 82, 1737–1744.
Kalbfleisch, J.D., Prentice, R.L., 2002. The Statistical Analysis of Failure Time Data. Wiley, New York.
Kundu, D., Kannan, N., Balakrishnan, N., 2008. On the hazard function of Birnbaum–Saunders distribution and associated inference. Computational Statistics and Data Analysis 52, 2692–2702.
Lee, E.T., Wang, J.W., 2003. Statistical Methods for Survival Data Analysis. Wiley, New York.
Leiva, V., Barros, M., Paula, G.A., Galea, M., 2007. Influence diagnostics in log-Birnbaum–Saunders regression models with censored data. Computational Statistics and Data Analysis 51, 5694–5707.
Leiva, V., Barros, M., Paula, G.A., Sanhueza, A., 2008. Generalized Birnbaum–Saunders distributions applied to air pollutant concentration. Environmetrics 19, 235–249.
Leiva, V., Sanhueza, A., Angulo, J.M., 2009. A length-biased version of the Birnbaum–Saunders distribution with application in water quality. Stochastic Environmental Research and Risk Assessment 23, 299–307.
Lemonte, A.J., 2011. Covariance matrix formula for Birnbaum–Saunders regression models. Journal of Statistical Computation and Simulation 81, 899–908.
Lemonte, A.J., 2012. A log-Birnbaum–Saunders regression model with asymmetric errors. Journal of Statistical Computation and Simulation 82, 1775–1787.
Lemonte, A.J., 2013. A new extension of the Birnbaum–Saunders distribution. Brazilian Journal of Probability and Statistics 27, 133–149.
Lemonte, A.J., Cordeiro, G.M., 2009. Birnbaum–Saunders nonlinear regression models. Computational Statistics and Data Analysis 53, 4441–4452.
Lemonte, A.J., Cribari-Neto, F., Vasconcellos, K.L.P., 2007. Improved statistical inference for the two-parameter Birnbaum–Saunders distribution. Computational Statistics and Data Analysis 51, 4656–4681.
Lemonte, A.J., Ferrari, S.L.P., 2011a. Size and power properties of some tests in the Birnbaum–Saunders regression model. Computational Statistics and Data Analysis 55, 1109–1117.
Lemonte, A.J., Ferrari, S.L.P., 2011b. Small-sample corrections for score tests in Birnbaum–Saunders regressions. Communications in Statistics—Theory and Methods 40, 232–243.
Lemonte, A.J., Ferrari, S.L.P., 2011c. Signed likelihood ratio tests in the Birnbaum–Saunders regression model. Journal of Statistical Planning and Inference 141, 1031–1040.
Lemonte, A.J., Ferrari, S.L.P., Cribari-Neto, F., 2010. Improved likelihood inference in Birnbaum–Saunders regressions. Computational Statistics and Data Analysis 54, 1307–1316.
Lemonte, A.J., Patriota, A.G., 2011. Influence diagnostics in Birnbaum–Saunders nonlinear regression models. Journal of Applied Statistics 38, 871–884.
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.
Li, A.P., Chen, Z.X., Xie, F.C., 2012. Diagnostic analysis for heterogeneous log-Birnbaum–Saunders regression models. Statistics and Probability Letters 82, 1690–1698.
Li, A.P., Xie, F.C., 2012. Diagnostics for a class of survival regression models with heavy-tailed errors. Computational Statistics and Data Analysis 56, 4204–4214.
Marshall, A.W., Olkin, I., 1997. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. Biometrika 84, 641–652.
Marshall, A.W., Olkin, I., 2007. Life Distributions, Structure of Nonparametric, Semiparametric and Parametric Families. Springer, New York.
Matos, L.A., Lachos, V.H., Balakrishnan, N., Labra, F.V., 2013. Influence diagnostics in linear and nonlinear mixed-effects models with censored data. Computational Statistics and Data Analysis 57, 450–464.
McCool, J.I., 1980. Confidence limits for Weibull regression with censored data. IEEE Transactions on Reliability 29, 145–150.
Patriota, A.G., Lemonte, A.J., de Castro, M., 2010. Influence diagnostics in a multivariate normal regression model with general parameterization. Statistical Methodology 7, 644–654.
Qu, H., Xie, F.C., 2011. Diagnostics analysis for log-Birnbaum–Saunders regression models with censored data. Statistica Neerlandica 65, 1–21.
R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.
Rieck, J.R., Nedelman, J.R., 1991. A log-linear model for the Birnbaum–Saunders distribution. Technometrics 33, 51–60.
Santana, L., Vilca, F., Leiva, V., 2011. Influence analysis in skew-Birnbaum–Saunders regression models and applications. Journal of Applied Statistics 38, 1633–1649.
Saunders, S.C., 2007. Reliability, Life Testing, and Prediction of Service Lives. Springer, New York.
Therneau, T.M., Grambsch, P.M., Fleming, T.R., 1990. Martingale-based residuals for survival models. Biometrika 77, 147–160.
Thomas, W., Cook, R.D., 1990. Assessing influence on predictions from generalized linear models. Biometrika 76, 741–749.

Vasconcellos, K.L.P., Fernandez, L.M.Z., 2009. Influence analysis with homogeneous linear restrictions. Computational Statistics and Data Analysis 53, 3787–3794.
Villegas, C., Paula, G.A., Leiva, V., 2011. Birnbaum–Saunders mixed models for censored reliability data analysis. IEEE Transactions on Reliability 60, 748–758.
Wang, B.X., 2012. Generalized interval estimation for the Birnbaum–Saunders distribution. Computational Statistics and Data Analysis 56, 4320–4326.
Wu, J., Wong, A.C.M., 2004. Improved interval estimation for the two-parameter Birnbaum–Saunders distribution. Computational Statistics and Data Analysis 47, 809–821.
Xiao, Q., Liu, Z., Balakrishnan, N., Lu, X., 2010. Estimation of the Birnbaum–Saunders regression model with current status data. Computational Statistics and Data Analysis 54, 326–332.
Xie, F.C., Wei, B.C., 2007. Diagnostics analysis for log-Birnbaum–Saunders regression models. Computational Statistics and Data Analysis 51, 4692–4706.
Zevallos, M., Santos, B., Hotta, L.K., 2012. A note on influence diagnostics in AR(1) time series models. Journal of Statistical Planning and Inference 142, 2999–3007.