Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics and Data Analysis 52 (2008) 3820–3842
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Log-Burr XII regression models with censored data

Giovana Oliveira Silva[a], Edwin M.M. Ortega[a,*], Vicente G. Cancho[b], Mauricio Lima Barreto[c]

[a] ESALQ, Universidade de São Paulo, Piracicaba, Brazil
[b] ICMC, Universidade de São Paulo, São Carlos, Brazil
[c] ISC, Universidade Federal da Bahia, Salvador, Brazil

Received 13 April 2007; received in revised form 7 January 2008; accepted 9 January 2008
Available online 17 January 2008

Abstract

In survival analysis applications, the failure rate function may frequently present a unimodal shape. In such case, the log-normal or log-logistic distributions are used. In this paper, we shall be concerned only with parametric forms, so a location-scale regression model based on the Burr XII distribution is proposed for modeling data with a unimodal failure rate function as an alternative to the log-logistic regression model. Assuming censored data, we consider a classic analysis, a Bayesian analysis and a jackknife estimator for the parameters of the proposed model. For different parameter settings, sample sizes and censoring percentages, various simulation studies are performed and compared to the performance of the log-logistic and log-Burr XII regression models. Besides, we use sensitivity analysis to detect influential or outlying observations, and residual analysis is used to check the assumptions in the model. Finally, we analyze a real data set under log-Burr XII regression models.

1. Introduction

In this paper, we consider a data set provided by Instituto de Saúde Coletiva — Universidade Federal da Bahia. This data set was designed to evaluate the effect of vitamin A supplementation on recurrent diarrheal episodes in small children (see Barreto et al. (1994)). Censoring times are random, and we aimed at modeling the treatment effect in time until the first occurrence of diarrheal episodes, this can be done by means of an appropriate regression model with censored data. By analyzing the total-time-on-test (TTT) curve (Aarset, 1987) of the survival times for the previously described data, it was observed that the failure rate function had a unimodal shape. It is known that the log-normal distribution is a popular model for survival time when the failure rate function is unimodal (see, for example, Nelson (1982)), and the log-logistic distribution is often used as an alternative to the former. Another possibility would be the use of the Burr XII distribution (see, for example, Zimmer et al. (1998) instead of usual distributions, since one of its advantages is that its survival function can be written in closed form. Additionally, the log-logistic distribution is a special case of the Burr XII distribution.

* Corresponding address: Departamento de Ciências Exatas, USP Av. Pádua Dias 11 - Caixa Postal 9, 13418-900 Piracicaba, São Paulo, Brazil.
E-mail addresses: gosilva@esalq.usp.br (G.O. Silva), edwin@esalq.usp.br (E.M.M. Ortega), garibay@icmc.usp.br (V.G. Cancho), mauricio@ufba.br (M.L. Barreto).

0167-9473/$ - see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.csda.2008.01.003

Regression models can be proposed in different forms in survival analysis. Among them, the location-scale regression model (Lawless, 2003) is distinguished and it is frequently used in clinical trials. In this paper, we propose a location-scale regression model using the Burr XII distribution, denoted by a log-Burr XII regression model, for survival times analysis as a feasible alternative to the log-logistic regression model. Considering that the log-logistic and log-Burr regression models are embedded models, the likelihood-ratio test, for instance, can be used to discriminate such models.

We considered a classic analysis for the log-Burr regression model. The inferential part was carried out using the asymptotic distribution of the maximum likelihood estimators, which, in situations when the sample is small, may present difficult results to be justified. As an alternative to classic analysis we explored the use of Markov Chain Monte Carlo (MCMC) techniques to develop a Bayesian inference and the jackknife estimator for the log-Burr XII regression model. In both cases, Bayesian and jackknife, it is not necessary to use the asymptotic distribution of the maximum likelihood estimators.

Studies were conducted via Monte Carlo simulation in order to evaluate the performance of the log-Burr XII and log-logistic regression models by means of variance, mean squared error and the size and power of the likelihood-ratio test for both models.

After modeling, it is important to check assumptions in the model as well as to conduct a robustness study in order to detect influential or extreme observations that can cause distortions in the results of the analysis. Numerous approaches have been proposed in the literature to detect influential or outlying observations. An efficient way to detect influential observations was proposed by Cook (1986). He suggested that more confidence can be put in a model which is relatively stable under small modifications. The best known perturbation schemes are based on case-deletion introduced by Cook (1977), in which the effect of completely removing cases from the analysis is studied. This reasoning will form the basis for our global influence introduced in Section 4.1, and in doing so, it will be possible to determine which subjects might be influential for the analysis (see, for example, Cook and Weisberg (1982) and Xie and Wei (2007)).

On the other hand, when using case deletion, all information from a single subject is deleted at once and, therefore, it is hard to tell whether that subject has any influence on a specific aspect of the model. A solution for the earlier problem can be found in a quite different paradigm, being a local influence approach where one again investigates how the results of an analysis are changed under small perturbations in the model, and where these perturbations can be specific interpretations. Also, some authors have investigated the assessment of local influence in survival analysis models: for instance, Pettit and Bin Daud (1989) investigated local influence in proportional hazard regression models; Escobar and Meeker (1992) adapted local influence methods to regression analysis with censoring; Ortega et al. (2003) considered the problem of assessing local influence in generalized log-gamma regression models with censored observations; Ortega et al. (2006) derived curvature calculations under various perturbation schemes in exponentiated-Weibull regression models with censored data, and more recently, Leiva-Sanchez et al. (2006) investigated local influence in log-Birnbaum–Saunders regression models with censored data. We developed a similar methodology to detect influential subjects in log-Burr XII regression models with censored data. Finally, the generalized leverage methodology developed by Wei et al. (1998) was applied. Additionally, the examination of residuals was used to check assumptions in the model.

In Section 2, this article considers a brief study on the Burr XII distribution besides the inferential part of this model. In Section 3, we suggest a log-Burr XII regression model of location-scale form, in addition to the maximum likelihood estimators, Bayesian inference, the jackknife estimator and the results from various simulation studies are displayed and commented. In the Section 4, we use several diagnostics measures considering three perturbation schemes, case-deletion and the generalized leverage in log-Burr XII regression models with censored observations. We present residuals from a fitted model using the Martingale residual proposed by Barlow and Prentice (1988) in the Section 5. Finally, in Section 6, the real data set is analyzed and the conclusion appears in Section 7.

2. The Burr XII distribution

The Burr XII distribution used in Zimmer et al. (1998) with parameters s, c and k considers that life time T has a density function given by

$$f(t; s, k, c) = ck\left(1 + \left(\frac{t}{s}\right)^{c}\right)^{(-k-1)} \frac{t^{c-1}}{s^{c}}, \quad t > 0, \tag{1}$$

where $k > 0$ and $c > 0$ are shape parameters and $s > 0$ is a scale parameter. The survival function corresponding to random variable T with Burr XII density is given by

$$S(t; s, k, c) = P(T \geq t) = \left(1 + \left(\frac{t}{s}\right)^c\right)^{-k}.$$

The corresponding failure rate function has the following form

$$h(t; s, k, c) = \frac{ck\left(\frac{t}{s}\right)^{c-1}}{s\left(1 + \left(\frac{t}{s}\right)^c\right)}.$$

2.1. Characterizing the failure rate function

According to Zimmer et al. (1998), the failure rate function of the Burr XII distribution can be decreased when $c \leq 1$ and when $c > 2$ the failure rate function reaches a maximum and the decreases, where the range of values in which the failure rate function is increasing can be manipulated using s. When c values are between 1 and 2, the failure rate function can be made to be essentially constant over much of the range of the distribution, which depends on s values. To study the shape of the failure rate function, we have found its derivative, which can be written as

$$h'(t; c, k, s) = \frac{ckt^{c-2}}{s^c\left(1 + \left(\frac{t}{s}\right)^c\right)^2}\left[c - 1 - \left(\frac{t}{s}\right)^c\right].$$

In order to better study this function, one can note that two situations can be considered:

- $c \leq 1$

  For any $t > 0$, $h'(t) < 0$ and, therefore, $h(t)$ is a decreasing function.
- $c > 1$

  When $h'(t^*) = 0$, we have $c - 1 - \left(\frac{t^*}{s}\right)^c = 0$, hence the critic point is given by $t^* = s(c-1)^{\frac{1}{c}}$. When $t < t^*$, $h'(t) > 0$, the failure rate function is increasing and when $t > t^*$, $h'(t) < 0$, the failure rate function is decreasing. Hence, t^* is an inflexion point and the failure rate function has a unimodal shape property. Besides, $h(t) \to 0$ for $t \to 0$ or $t \to \infty$.

Fig. 1 shows the plots of the failure rate function for some different parameter combinations.

From Fig. 1, it can be seen that the failure rate function is a decreasing function when $c \leq 1$ and $h(t)$ is a unimodal-shaped function and when $c > 1$.

2.2. Moments for the failure time

The qth moment for the failure time is given by:

$$E(T^q) = s^q kB\left[\frac{q}{c} + 1, k - \frac{q}{c}\right], \quad \text{if } ck > q,$$

where $B(a, b)$ is the complete beta function (see Lawless (2003)).

2.3. Relation to other distributions

The log-logistic distribution is a special case of the Burr XII distribution. When $\frac{1}{s} = m$ and $k = 1$, the Burr XII distribution is reduced to the log-logistic distribution, where the survival function can be written as $S(t; m, c) = \frac{1}{1+(tm)^c}$.

Besides, Rodriguez (1977) shows that the Burr coverage area on a specific plane is occupied by various well-known and useful distributions, including the normal, log-normal, gamma, logistic and extreme-value type-I distributions.

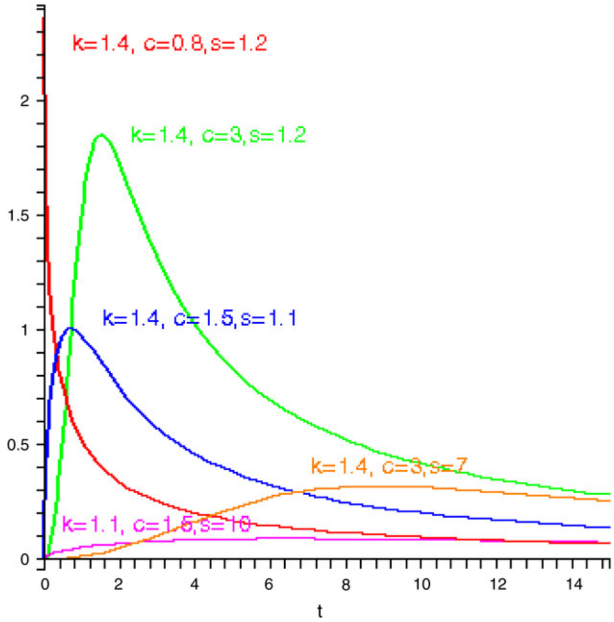

Fig. 1. Plots of the failure rate function for Burr XII distribution.

2.4. Maximum likelihood estimation

We assume that the lifetimes are independently distributed and also independent from the censoring mechanism. Considering right-censored lifetime data, we observe $t_i = \min(T_i, C_i)$, where T_i is the lifetime and C_i is the censoring time, both for the ith individual $i = 1, \ldots, n$. Assuming that $T_1, T_2, \ldots, T_n$ is a random sample of the random variable T with Burr XII distribution (1). The likelihood function of c, k and s corresponding to the observed sample is given by

$$L(c,k,s) = (kc)^r \prod_{i\in F}\left[\left(1+\left(\frac{t_i}{s}\right)^c\right)^{-(k+1)} \frac{t_i^{c-1}}{s^c}\right] \prod_{i\in C}\left[\left(1+\left(\frac{t_i}{s}\right)^c\right)^{-k}\right], \tag{2}$$

where r is the observed number of failures, F denotes the set of uncensored observations and C denotes the set of censored observations. The log-likelihood function is given by:

$$l(c,k,s) = r\log(k) + r\log(c) - (k+1)\sum_{i\in F}\log\left(1+\left(\frac{t_i}{s}\right)^c\right) + \sum_{i\in F}\log\left(\frac{t_i^{c-1}}{s^c}\right) - k\sum_{i\in C}\log\left(1+\left(\frac{t_i}{s}\right)^c\right).$$

The maximum likelihood estimators $\hat{c}$, $\hat{k}$ and $\hat{s}$ of c, k and s are obtained by maximizing the log-likelihood, which results in solving equations

$$\frac{\partial l(c,k,s)}{\partial c} = \frac{r}{c} - (k+1)\sum_{i\in F}\frac{\left(\frac{t_i}{s}\right)^c \log\left(\frac{t_i}{s}\right)}{\left(1+\left(\frac{t_i}{s}\right)^c\right)} + \sum_{i\in F}\log\left(\frac{t_i}{s}\right) - k\sum_{i\in C}\frac{\left(\frac{t_i}{s}\right)^c \left(\log\frac{t_i}{s}\right)}{\left(1+\left(\frac{t_i}{s}\right)^c\right)}$$

$$\frac{\partial l(c,k,s)}{\partial k} = \frac{r}{k} - \sum_{i\in F}\log\left(1+\left(\frac{t_i}{s}\right)^c\right) - \sum_{i\in C}\log\left(1+\left(\frac{t_i}{s}\right)^c\right)$$

$$\frac{\partial l(c,k,s)}{\partial s} = c(k+1)\sum_{i\in F}\frac{t_i^c s^{-(c+1)}}{\left(1+\left(\frac{t_i}{s}\right)^c\right)} - \frac{rc}{s} + c\sum_{i\in C}\frac{t_i^c s^{-(c+1)}}{\left(1+\left(\frac{t_i}{s}\right)^c\right)}.$$

These equations cannot be solved analytically, so statistical software such as Ox or R can be used to solve them. In this paper, software Ox (Doornik, 2001) through MaxBFGS subroutine is used to compute the maximum likelihood estimator (ML estimator), but reparametrization is necessary, for example, $c = \frac{1}{\sigma}$ and $s = \exp(\mu)$ can be used.

3. Log-Burr XII regression models

3.1. Log-location-scale regression model

In many practical applications, lifetimes are affected by variables, which are referred to as explanatory variables or covariates, such as the cholesterol level, blood pressure and many others. So, it is important to explore the relationship between the lifetime and explanatory variables. An approach based on a regression model can be used. This paper considers the class of location-scale models.

The covariates vector is denoted by $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)^{\mathrm{T}}$, which is related to responses $Y = \log(T)$ through a regression model.

Considering reparametrization, $c = \frac{1}{\sigma}$ and $s = \exp(\mu)$. Hence, it follows that the density function of Y can be written as

$$f(y; k, \sigma, \mu) = \frac{k}{\sigma}\left(1 + \exp\left(\frac{y-\mu}{\sigma}\right)\right)^{-(k+1)} \exp\left(\frac{y-\mu}{\sigma}\right), \quad -\infty < y < \infty, \tag{3}$$

where $k > 0$, $\sigma > 0$, and $-\infty < \mu < \infty$. This new distribution will be referred to as the log-Burr XII. The survival function is given by

$$S(y) = \left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^{-k}.$$

Besides, we have the following important theorem.

Theorem 1. *For variable Y, the moment-generating function (mgf) is given by*

$$M_Y(t) = ks^t B\left[\frac{t}{c} + 1, k - \frac{t}{c}\right], \quad \text{if } kc > t,$$

where $B[a, b]$ is the complete beta function (proof given in Appendix B).

Hence, the mean of Y is given by

$$E(Y) = s + \sigma[\psi(1) - \psi(k)], \quad \text{if } kc > t,$$

where $\psi(a)$ is the digamma function (see Lawless (2003)).

We can write the above model as a log-linear model

$$Y = \mu + \sigma Z, \tag{4}$$

where variable Z follows the density

$$f(z) = k(1 + \exp(z))^{-(k+1)} \exp(z), \quad -\infty < z < \infty \quad \text{and} \quad k > 0. \tag{5}$$

Now, it is also considered that the scale parameter μ of the log-Burr XII model depends on the matrix of explanatory variables X, this is, $\mu_i = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}$. We also consider the regression model based on the log-Burr XII given in (3) relating response Y and covariates vector $\boldsymbol{x}$, so that distribution $Y|\boldsymbol{x}$ can be represented as

$$y_i = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \sigma z_i, \quad i = 1, \ldots, n, \tag{6}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$, $\sigma > 0$ and $k > 0$ are unknown parameters, $\boldsymbol{x}_i^{\mathrm{T}} = (x_{i1}, x_{i2}, \ldots, x_{ip})$ is the explanatory vector and Z follows the distribution in (5).

In this case, the survival function of $Y|\boldsymbol{x}$ is given by

$$S(y|\boldsymbol{x}) = \left[1 + \exp\left(\frac{y - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}}{\sigma}\right)\right]^{-k}.$$

It is observed that when $k = 1$ the log-logistic regression models is obtained.

3.2. Estimation by maximum likelihood

For the corresponding values to the sample $(y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \ldots, (y_n, \boldsymbol{x}_n)$ of n observations, where y_i represents the logarithm of the survival time that has distribution (3) and $\boldsymbol{x}_i$ the covariate vector associated with the ith individual, the log-likelihood function can be written as

$$l(\boldsymbol{\theta}) = r\log(k) - r\log(\sigma) + \sum_{i \in F} z_i - (k+1)\sum_{i \in F}\log(1+\exp(z_i)) - k\sum_{i \in C}\log(1+\exp(z_i)), \tag{7}$$

where r is the number of uncensored observations (failures) and $z_i = \frac{y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}}{\sigma}$. Maximum likelihood estimates for parameter vector $\boldsymbol{\theta} = (k, \sigma, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$ can be obtained by maximizing the likelihood function. In this paper, software Ox (see, Doornik (2001)) through MAXBFGS subroutine was used to compute maximum likelihood estimates. Covariance matrix estimates for maximum likelihood estimators $\widehat{\boldsymbol{\theta}}$ can be obtained using the Hessian matrix. Confidence intervals and hypothesis testing can be conducted using the large sample distribution of the ML estimators, which is a normal distribution with the covariance matrix as the inverse of the Fisher information as long as regularity conditions are satisfied. More specifically, the asymptotic covariance matrix is given by $\mathbf{I}^{-1}(\boldsymbol{\theta})$ with $\mathbf{I}(\boldsymbol{\theta}) = E[\ddot{\mathbf{L}}(\boldsymbol{\theta})]$ such that $\ddot{\mathbf{L}}(\boldsymbol{\theta}) = -\left\{\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\right\}$.

It is difficult to compute the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ due to the censored observations (censoring is random and noninformative), but it is possible to use the matrix of second derivatives of the log-likelihood, $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$, evaluated at the ML estimator $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, which is consistent. The asymptotic normal approximation for $\widehat{\boldsymbol{\theta}}$ may be expressed as $\widehat{\boldsymbol{\theta}}^{\mathrm{T}} \sim N_{(p+2)}\{\boldsymbol{\theta}^{\mathrm{T}}; \ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1}\}$, where $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ is the $(p+2)(p+2)$ observed information matrix, obtained from:

$$-\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{L}_{kk} & \mathbf{L}_{k\sigma} & \mathbf{L}_{k\beta_j} \\ \cdot & \mathbf{L}_{\sigma\sigma} & \mathbf{L}_{\sigma\beta_j} \\ \cdot & \cdot & \mathbf{L}_{\beta_j\beta_s} \end{pmatrix}$$

with the submatrices given in Appendix A.

As to the interpretation of the estimated coefficients, a possible proposal is based on the ratio of median times (see Hosmer and Lemeshow (1999)). Hence, when the covariable is binary (1 or 0), and considering the ratio of median times with $x = 1$ in the numerator, if $\hat{\beta}$ is negative, it implies that individuals with $x = 1$ present reduced median survival time (increased) in $[\exp\{\hat{\beta}\} \times 100\%]$ as compared to that of individuals in the group with $x = 0$ by fixing the other covariables. This interpretation can be extended to continuous or categorical covariables.

Another interest is to investigate the use of the log-logistic regression model, which is a simpler model than the proposed one. Since the log-Burr and log-logistic regression models are embedded, the likelihood-ratio test can be used to discriminate such models. In this case, the hypotheses are given by $H_0 : k = 1$ versus $H_1 : k \neq 1$. The test statistic is given by $\lambda = -2 \times \log\left(\frac{L(\widehat{\boldsymbol{\theta}}_0)}{L(\widehat{\boldsymbol{\theta}})}\right)$, where $\widehat{\boldsymbol{\theta}}_0$ is the maximum likelihood estimator for $\boldsymbol{\theta}$ under H_0, and the null hypothesis is rejected when $\lambda > \chi^2_{1-\alpha}(1)$, which is the quantile of the chi-square distribution with one degree of freedom.

3.3. A Bayesian analysis

Besides being an alternative analysis, the use of the Bayesian method allows for the incorporation of previous knowledge of the parameters through informative prior densities. When this information is not available, one considers noninformative prior. In the Bayesian approach, the information referring to the model parameters is obtained through posterior marginal distribution. In this way, two difficulties arise. The first refers to attaining marginal posterior distribution, and the second to the calculation of the interest moments. Both cases require integral resolutions that, many times, do not present an analytical solution. In this paper, we have used the simulation method of Markov Chain Monte Carlo, such as the Gibbs sampler and Metropolis–Hasting algorithm.

Consider the Burr XII distribution (1), censored data and the likelihood function (2) for k, c and s. For a Bayesian analysis, we assume the following prior densities for k, s and c

- $k \sim \Gamma(a_1, b_1)$, a_1 and b_1 known;
- $s \sim \Gamma(a_2, b_2)$, a_2 and b_2 known;
- $c \sim \Gamma(a_3, b_3)$, a_3 and b_3 known;

where $\Gamma(a_i, b_i)$ denotes a gamma distribution with mean $\frac{a_i}{b_i}$, variance $\frac{a_i}{b_i^2}$ and density function given by

$$f(v; a_i, b_i) = \frac{b_i^{a_i} v^{a_i-1} \exp\{-v b_i\}}{\Gamma(a_i)},$$

where $v > 0$, $a_i > 0$ and $b_i > 0$.

In the special case where $a_1 = b_1 = a_2 = b_2 = a_3 = b_3 = 0$, the noninformative case follows, and it assumes independence among the parameters, and the prior densities for k, s and c are written as

$$\pi(k, s, c) \propto \frac{1}{ksc}.$$

We further assume independence among parameters k, s and c. The joint posteriori distributions for k, s and c is given by

$$\pi(k, s, c|D) \propto k^{a_1-1} \exp\{-kb_1\} s^{a_2-1} \exp\{-sb_2\} c^{a_3-1} \exp\{-cb_3\} \\ \times \left(\frac{kc}{s^c}\right)^r \prod_{i\in F} \left[\left(1 + \left(\frac{t_i}{s}\right)^c\right)^{-(k+1)}\right] \prod_{i\in F} t_i^{c-1} \prod_{i\in C} \left[\left(1 + \left(\frac{t_i}{s}^c\right)\right)^{-k}\right],$$

where D denotes the data sets.

It can be shown that the conditional posteriori densities are given by

$$\pi(k|s, c, D) \propto k^{a_1+r-1} \exp\{-kb_1\} \prod_{i\in F} \left[\left(1 + \left(\frac{t_i}{s}\right)^c\right)^{-(k+1)}\right] \prod_{i\in C} \left[\left(1 + \left(\frac{t_i}{s}^c\right)\right)^{-k}\right]$$

$$\pi(s|k, c, D) \propto s^{a_2-cr-1} \exp\{-sb_2\} \prod_{i\in F} \left[\left(1 + \left(\frac{t_i}{s}\right)^c\right)^{-(k+1)}\right] \prod_{i\in C} \left[\left(1 + \left(\frac{t_i}{s}^c\right)\right)^{-k}\right]$$

$$\pi(c|k, s, D) \propto c^{a_3-1} \exp\{-cb_3\} c^r s^{-cr} \left(\frac{kc}{s^c}\right)^r \prod_{i\in F} \left[\left(1 + \left(\frac{t_i}{s}\right)^c\right)^{-(k+1)}\right] \prod_{i\in F} t_i^{c-1} \prod_{i\in C} \left[\left(1 + \left(\frac{t_i}{s}^c\right)\right)^{-k}\right].$$

Observe that we need to use the Metropolis–Hastings algorithm to generate variables k, s and c from the respective conditional posteriori densities since their forms are somewhat complex.

For Bayesian inference, considering model (6), assume the following prior densities for σ, k and $\boldsymbol{\beta}^{\mathrm{T}}$:

- $k \sim \Gamma(c_1, d_1)$, c_1 and d_1 known;
- $\sigma \sim$ Inverse $\Gamma(c_2, d_2)$, c_2 and d_2 known;
- $\beta_j \sim N(\mu_{0j}, \sigma_{0j}^2)$, μ_{0j} and σ_{0j}^2 known, $j = 0, \ldots, p$.

The noninformative case follows, and it assumes independence among the parameters, by considering $c_1 = c_2 = d_1 = d_2 = 0$ and σ_{0j}^2 large.

We again assume independence among the parameters. The joint posteriori distribution for σ, k and $\boldsymbol{\beta}$ is given by:

$$\pi(\sigma, k, \boldsymbol{\beta}^{\mathrm{T}}|D) \propto k^{c_1-1} \exp\{-kd_1\} \sigma^{-(c_2+1)} \exp\left\{-\frac{d_2}{\sigma}\right\} \exp\left\{-\frac{1}{2}\sum_{j=0}^{p}\left(\frac{\beta_j - \mu_{0j}}{\sigma_{0j}}\right)^2\right\} \\ \times \left(\frac{k}{\sigma}\right)^r \exp\left\{\sum_{i\in F} z_i\right\} \prod_{i\in F}[(1 + \exp\{z_i\})^{-(k+1)}] \prod_{i\in c}[(1 + \exp\{z_i\})^{-k}],$$

where $z_i = \frac{y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}}{\sigma}$.

It can be shown that the conditional marginal distributions are given by:

$$\pi(k|\sigma, \boldsymbol{\beta}^{\mathrm{T}}, D) \propto k^{c_1+r-1} \exp\{-kd_1\} \exp \prod_{i\in F}[(1+\exp\{z_i\})^{-(k+1)}] \prod_{i\in c}[(1+\exp\{z_i\})^{-k}]$$

$$\pi(\sigma|k, \boldsymbol{\beta}^{\mathrm{T}}, D) \propto \sigma^{-c_2-r-1} \exp\left\{-\frac{d_2}{\sigma}\right\} \exp\left\{\sum_{i\in F} z_i\right\} \prod_{i\in F}[(1+\exp\{z_i\})^{-(k+1)}] \prod_{i\in c}[(1+\exp\{z_i\})^{-k}]$$

$$\pi(\beta_j|k, \sigma, \boldsymbol{\beta}_{-j}, D) \propto \exp\left\{-\frac{1}{2}\sum_{j=0}^{p}\left(\frac{\beta_j-\mu_{0j}}{\sigma_{0j}}\right)^2\right\} \exp\left\{\sum_{i\in F} z_i\right\} \times \prod_{i\in F}[(1+\exp\{z_i\})^{-(k+1)}] \prod_{i\in c}[(1+\exp\{z_i\})^{-k}].$$

Observe that we need to use the Metropolis–Hastings algorithm to generate from the posteriori conditional distributions of k, σ and β_j $(j = 0, \ldots, p)$.

3.4. The jackknife estimator for the model

The idea of jackknifing is to transform the problem of estimating any population parameter into the problem of estimating a population mean. So, the procedure used to estimate a mean value is performed in this method, but from an unusual point of view. In this paper, we used this method as an alternative method to estimate the population parameter.

Suppose that $T_1, T_2, \ldots, T_n$ is a random sample of n values and the sample mean is given by

$$\bar{T} = \sum_{i=1}^{n} \frac{T_i}{n}$$

and is used to estimate the population mean.

Now, the sample mean is calculated with the lth observation missed out,

$$\bar{T}_{-l} = \frac{\sum_{i=1}^{n} T_i - T_l}{n-1}.$$

Then from the two expressions above the following is obtained

$$T_l = n\bar{T} - (n-1)\bar{T}_{-l}. \tag{8}$$

In a general situation, consider that θ is a parameter estimated by $\hat{E}(T_1, T_2, \ldots, T_n)$, and for ease of notation, drop $(T_1, T_2, \ldots, T_n)$. Finally, $\hat{E}_{-l}$ is calculated, which is obtained with the T_l observation missed out. It follows, by Eq. (8), that pseudo-values can be calculated, which is obtained by

$$\hat{E}_l^* = n\hat{E} - (n-1)\hat{E}_{-l}, \quad l = 1, \ldots, n.$$

The mean of the pseudo-values is given by

$$\hat{E}^* = \frac{\sum_{l=1}^{n} \hat{E}_l^*}{n}$$

which is the jackknife estimate of θ.

Manly (1997) suggests that an approximate $100(1-\alpha)\%$ confidence interval for θ is given by $\hat{E}^* \pm t_{\alpha/2,n-1}s/\sqrt{n}$, where $t_{\alpha/2,n-1}$ is the value that is exceeded with probability $\alpha/2$ for the t distribution with $(n-1)$ degrees of freedom and the jackknife estimator had the effect of removing bias of order $1/n$.

By following the theory described, it was possible to obtain the estimates and their respective confidence intervals for the parameter vector $\boldsymbol{\theta}$ of the log-Burr XII regression model using the jackknife method. The Ox matrix programming language was used.

3.5. Simulations study

In order to investigate the performance of the log-Burr XII regression model, we performed various simulation studies for different settings of n and censoring percentages. The lifetimes denoted by $T_1, \ldots, T_n$ were generated from the Burr XII distribution given in (1), again considering the following reparametrization $c = \frac{1}{\sigma}$ and $s = \exp\{\mu\}$, for values $\sigma = 0.36$ and 0.8 (failure rate function is unimodal), $k = 0.15, 0.27$ and 1.00 and by assuming $\mu_i = \beta_0 + \beta_1 x_i$, with x_i being generated from a uniform distribution on the range [0, 1], β_0 and β_1 fixed. The censoring times denoted by $C_1, \ldots, C_n$ were generated from a uniform distribution $[0, \theta]$, where θ was adjusted until the censoring percentages, 0 or 0.10 or 0.30, were reached. The lifetimes considered in each fit were calculated as $\min\{C_i, T_i\}$. For each setting of n, k, σ and censoring percentages, 1000 samples were generated each one being fitted under the log-Burr XII regression model (4) with $\mu_i = \beta_0 + \beta_1 x_i$. For each fit the likelihood-ratio test for hypotheses $H_0 : k = 1$ versus $H_1 : k \neq 1$ was performed. Then the proportion of times which rejected the null hypothesis was just the simulated value power. Here, all the statistics were compared with the χ_1^2 critical value at an $\alpha = 0.05$ level. The simulations were performed for different $n = 50$ and 100 and different values of k to obtain the simulated sizes and powers for testing. From the results of simulations, given in Table 1, the results for testing $k = 1$ indicated that the actual sizes of the test were close to 0.05, and the powers of tests were increased as k and % censuring decreased and/or n was increased. Besides, the σ values did not influence the results of simulations.

Additionally, the variance and the mean squared error (MSE) (Cox and Hinkley, 1974) of the maximum likelihood estimators were also calculated for simulated samples in the same conditions as those in previous simulations. Again, for each situation, 1000 samples were generated and the log-logistic and log-Burr XII regression models were fitted to the generated samples. From the simulation results, the data of which are shown in Table 2, it was observed that when $k = 1$, variance and MSE showed similar values for both models. However, when $k \neq 1$, variance and MSE were smaller for the log-Burr regression model and they decreased when n increased and were larger when σ approached 1. In general, for the log-Burr regression model, variance and MSE increased when the censoring percentage increased.

4. Sensitivity analysis

4.1. Global influence

A first tool to perform sensitivity analysis, as stated before, is by means of global influence starting from case-deletion. Case-deletion is a common approach to study the effect of dropping the ith case from the data set. The case-deletion model for model (6) is given by

$$Y_l = \boldsymbol{x}_l^{\mathrm{T}} \boldsymbol{\beta} + \sigma Z_l, \quad l = 1, 2, \ldots, n, \ l \neq i. \tag{9}$$

In the following, a quantity with subscript "(i)" means the original quantity with the ith case deleted. For model (9), the log-likelihood function of $\boldsymbol{\theta}$ is denoted by $l_{(i)}(\boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}_{(i)} = (\hat{k}_{(i)}, \hat{\sigma}_{(i)}, \hat{\boldsymbol{\beta}}_{(i)}^{\mathrm{T}})^{\mathrm{T}}$ be the ML estimator of $\boldsymbol{\theta}$ from $l_{(i)}(\boldsymbol{\theta})$. To assess the influence of the ith case on the ML estimator $\hat{\boldsymbol{\theta}} = (\hat{k}, \hat{\sigma}, \hat{\boldsymbol{\beta}})^{\mathrm{T}}$, the basic idea is to compare the difference between $\hat{\boldsymbol{\theta}}_{(i)}$ and $\hat{\boldsymbol{\theta}}$. If deletion of a case seriously influences the estimates, more attention should be paid to that case. Hence, if $\hat{\boldsymbol{\theta}}_{(i)}$ is far from $\hat{\boldsymbol{\theta}}$, then the ith case is regarded as an influential observation. A first measure of the global influence is defined as the standardized norm of $\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}$ (generalized Cook distance)

$$GD_i(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^{\mathrm{T}} \left[\ddot{\mathbf{L}}(\boldsymbol{\theta})\right]^{-1} (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}).$$

Another alternative is to assess values $GD_i(\boldsymbol{\beta})$ and $GD_i(k, \sigma)$. Such values reveal the impact of the ith case on the estimates of $\boldsymbol{\beta}$ and (k, σ), respectively. Another popular measure of the difference between $\hat{\boldsymbol{\theta}}_{(i)}$ and $\hat{\boldsymbol{\theta}}$ is the likelihood distance

$$LD_i(\boldsymbol{\theta}) = 2\left\{l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{(i)})\right\}.$$

Besides, we can also compute $\beta_j - \beta_{j(i)}$ $(j = 1, 2, \ldots, p)$ to see the difference between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{(i)}$. Alternative global influence measures are possible. One could think of the behavior of a test statistics, such as the Wald test for covariate or censoring effect, under a case-deletion scheme.

Table 1
Simulated sizes and powers of the likelihood-ratio test for hypotheses $H_0 : k = 1$ versus $H_1 : k \neq 1$

k value	σ value	n	% censoring	% of significant test results at the 5% level
0.15	0.36	50	0	96.6
			10	95.4
			30	89.2
		100	0	99.9
			10	99.9
			30	99.3
	0.80	50	0	95.0
			10	95.2
			30	89.7
		100	0	99.9
			10	99.9
			30	99.3
0.27	0.36	50	0	79.8
			10	75.5
			30	65.5
		100	0	97.5
			10	96.1
			30	92.6
	0.80	50	0	79.9
			10	76.0
			30	67.8
		100	0	97.5
			10	96.3
			30	93.5
1.00	0.36	50	0	5.30
			10	4.30
			30	3.30
		100	0	6.50
			10	5.90
			30	5.40
	0.80	50	0	5.20
			10	4.60
			30	3.90
		100	0	6.50
			10	5.70
			30	5.30

4.2. Local influence

As a second tool for sensitivity analysis the local influence method will now be described for the log-Burr XII regression model with censored data. Local influence calculation can be carried out in model (6). If likelihood displacement $LD(\boldsymbol{\omega}) = 2\{l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}$ is used, where $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denotes the ML estimator under the perturbed model, the normal curvature for $\boldsymbol{\theta}$ at direction $\mathbf{d}$, $\|\mathbf{d}\| = 1$, is given by $C_{\mathbf{d}}(\boldsymbol{\theta}) = 2|\mathbf{d}^{\mathrm{T}}\boldsymbol{\Delta}^{\mathrm{T}}\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1}\boldsymbol{\Delta}\mathbf{d}|$, where $\boldsymbol{\Delta}$ is a $(p+2)\times n$ matrix that depends on the perturbation scheme and whose elements are given by $\Delta_{ji} = \partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\theta_j\partial\omega_i$, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p+2$ evaluated at $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega}_0$, where $\boldsymbol{\omega}_0$ is the no perturbation vector (see Cook (1986)). For the log-Burr XII regression model the elements of $-\ddot{\mathbf{L}}(\hat{\theta})$ are given in Appendix A. We can calculate normal curvatures $C_{\mathbf{d}}(\boldsymbol{\theta})$, $C_{\mathbf{d}}(k)$, $C_{\mathbf{d}}(\sigma)$ and $C_{\mathbf{d}}(\boldsymbol{\beta})$ to perform various index plots, for instance, the index plot of $\mathbf{d}_{\max}$, the eigenvector corresponding to $C_{\mathbf{d}_{\max}}$, the largest eigenvalue of the matrix $\mathbf{B} = \boldsymbol{\Delta}^{\mathrm{T}}\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1}\boldsymbol{\Delta}$ and the index plots of $C_{\mathbf{d}_i}(\boldsymbol{\theta})$, $C_{\mathbf{d}_i}(k)$, $C_{\mathbf{d}_i}(\sigma)$ and $C_{\mathbf{d}_i}(\boldsymbol{\beta})$ named total local influence (see, for example, Lesaffre and Verbeke (1998)), where $\mathbf{d}_i$ denotes an $n \times 1$ vector of zeros with one at the ith position. Thus, the curvature at direction $\mathbf{d}_i$ assumes form $C_i = 2|\boldsymbol{\Delta}_i^{\mathrm{T}}\ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1}\boldsymbol{\Delta}_i|$ where $\boldsymbol{\Delta}_i^{\mathrm{T}}$ denotes the ith row of $\boldsymbol{\Delta}$. It is usual to point out those cases such that

Table 2
Variance and mean squared error of β_1 for the log-logistic and log-Burr regression models

k value	σ value	n	% censoring	Variance		MSE	
				Log-logistic	Log-Burr	Log-logistic	Log-Burr
0.15	0.36	50	0	1.14	0.50	2.25	1.04
			10	1.06	0.52	2.16	1.14
			30	1.15	0.56	2.34	1.21
		100	0	0.56	0.23	1.09	0.48
			10	0.53	0.23	1.01	0.50
			30	0.50	0.24	0.99	0.49
	0.80	50	0	5.69	2.50	11.24	5.39
			10	5.13	2.52	10.23	5.54
			30	5.08	2.61	10.39	5.70
		100	0	2.79	1.16	5.37	2.38
			10	2.53	1.15	4.86	2.40
			30	2.51	1.16	4.98	2.35
0.27	0.36	50	0	0.43	0.27	0.84	0.57
		50	10	0.41	0.28	0.83	0.60
			30	0.47	0.33	0.98	0.72
		100	0	0.21	0.13	0.40	0.28
			10	0.20	0.14	0.39	0.28
			30	0.22	0.15	0.44	0.31
	0.80	50	0	2.10	1.33	4.14	2.82
			10	1.97	1.36	3.98	2.92
			30	2.18	1.48	4.48	3.15
		100	0	1.02	0.65	1.99	1.38
			10	0.96	0.66	1.85	1.35
			30	0.95	0.68	1.89	1.39
1.00	0.36	50	0	0.09	0.09	0.19	0.20
			10	0.10	0.10	0.21	0.22
			30	0.13	0.13	0.27	0.29
		100	0	0.05	0.05	0.09	0.10
			10	0.05	0.05	0.00	0.10
			30	0.07	0.07	0.13	0.13
	0.80	50	0	0.47	0.47	0.95	0.97
			10	0.49	0.50	1.01	1.05
			30	0.59	0.60	1.25	1.32
		100	0	0.23	0.23	0.47	0.48
			10	0.24	0.24	0.47	0.48
			30	0.29	0.29	0.56	0.58

$$C_i \geq 2\bar{C}, \qquad \bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i.$$

4.3. Curvature calculations

Next, we calculate, for three perturbation schemes, the matrix

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{ji})_{(p+2)\times n} = \left(\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial\theta_i \boldsymbol{\omega}_j}\right)_{(p+2)\times n}, \qquad j = 1, 2, \ldots, p+2 \text{ and } i = 1, 2, \ldots, n,$$

considering the model defined in (6) and its log-likelihood function given by (7). Consider the vector of weights $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)^{\mathrm{T}}$.

4.3.1. Case-weight perturbation

In this case, the log-likelihood function takes the form

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = \left[\log(k) - \log(\sigma)\right]\sum_{i\in F}\omega_i + \sum_{i\in F}\omega_i z_i + (k+1)\sum_{i\in F}\omega_i \log\left[1+\exp\{z_i\}\right] - \sum_{i\in C}, \omega_i \log\left[1+\exp\{z_i\}\right],$$

where $0 \leq \omega_i \leq 1$ and $\boldsymbol{\omega} = (1, \ldots, 1)^{\mathrm{T}}$. Let us denote $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_{p+2})^{\mathrm{T}}$.

Then the elements of vector $\boldsymbol{\Delta}_1$ take the form

$$\Delta_{1i} = \begin{cases} \hat{k}^{-1} + \log\left[1+\exp\{\hat{z}_i\}\right] & \text{if } i \in F \\ \log\left[1+\exp\{\hat{z}_i\}\right] & \text{if } i \in C. \end{cases}$$

On the other hand, the elements of vector $\boldsymbol{\Delta}_2$ can be shown to be given by

$$\Delta_{2i} = \begin{cases} -\hat{\sigma}^{-1}\left\{1 + \hat{z}_i + (\hat{k}+1)\hat{z}_i \exp\{\hat{z}_i\}\left[1+\exp\{z_i\}\right]^{-1}\right\} & \text{if } i \in F \\ \hat{k}\hat{\sigma}^{-1}\hat{z}_i \exp\{\hat{z}_i\}\left[1+\exp\{z_i\}\right]^{-1} & \text{if } i \in C. \end{cases}$$

The elements of vector $\boldsymbol{\Delta}_j$, for $j = 3, \ldots, p+2$, can be expressed as

$$\Delta_{ji} = \begin{cases} -x_{ij}\hat{\sigma}^{-1}\left\{1 + (\hat{k}+1)\exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-1}\right\} & \text{if } i \in F \\ x_{ij}\hat{k}\hat{\sigma}^{-1}\exp\{z_i\}\left[1+\exp\{z_i\}\right]^{-1} & \text{if } i \in C. \end{cases}$$

4.3.2. Response perturbation

We will consider here that each y_i is perturbed as $y_{iw} = y_i + \omega_i S_y$, where S_y is a scale factor that may be the estimated standard deviation of Y and $\omega_i \in \mathbf{R}$.

Here the perturbed log-likelihood function becomes expressed as

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = r\left[\log(k) - \log(\sigma)\right] + \sum_{i\in F} z_i^* - (k+1)\sum_{i\in F}\log\left[1+\exp\{z_i^*\}\right] - k\sum_{i\in C}\log\left[1+\exp\{z_i^*\}\right],$$

where $z_i^* = \frac{(y_i+\omega_i S_y) - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}}{\sigma}$. In addition, the elements of vector $\boldsymbol{\Delta}_1$ take the form

$$\Delta_{1i} = \begin{cases} -S_y\hat{\sigma}^{-1}\hat{z}_i\left[1+\exp\{z_i\}\right]^{-1} & \text{if } i \in F \\ -S_y\hat{\sigma}^{-1}\hat{z}_i\left[1+\exp\{z_i\}\right]^{-1} & \text{if } i \in C. \end{cases}$$

On the other hand, the elements of vector $\boldsymbol{\Delta}_2$ can be shown to be given by

$$\Delta_{2i} = \begin{cases} -S_y\sigma^{-2}\left\{1 - (\hat{k}+1)\exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-1}\left(\hat{z}_i\left[1+\exp\{\hat{z}_i\}\right]^{-1} + 1\right)\right\} & \text{if } i \in F \\ S_y\hat{k}\hat{\sigma}^{-2}\exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-1}\left\{\hat{z}_i\left[1+\exp\{\hat{z}_i\}\right]^{-1} + 1\right\} & \text{if } i \in C. \end{cases}$$

The elements of vector $\boldsymbol{\Delta}_j$, for $j = 3, \ldots, p+2$, can be expressed as

$$\Delta_{ji} = \begin{cases} x_{ij}S_y(\hat{k}+1)\hat{\sigma}^{-2}\exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-2} & \text{if } i \in F \\ x_{ij}S_y\hat{k}\hat{\sigma}^{-2}\exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-2} & \text{if } i \in C. \end{cases}$$

4.3.3. Explanatory variable perturbation

Consider now an additive perturbation on a particular continuous explanatory variable, namely X_t, by making $x_{it\omega} = x_{it} + \omega_i S_t$, where S_t is a scaled factor, $\omega_i \in \mathbf{R}$. This perturbation scheme leads to the following expressions for the log-likelihood function and for the elements of matrix $\boldsymbol{\Delta}$.

In this case the log-likelihood function takes the form

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = r\left[\log(k) - \log(\sigma)\right] + \sum_{i\in F} z_i^* - (k+1)\sum_{i\in F}\log\left[1+\exp\{z_i^*\}\right] - k\sum_{i\in C}\log\left[1+\exp\{z_i^*\}\right],$$

where $z_i^* = \frac{y_i - x_i^{*T}\boldsymbol{\beta}}{\sigma}$ and $\boldsymbol{x}_i^{*T} = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_t(x_{it} + \omega_i S_t) + \cdots + \beta_p x_{ip}$.

In addition, the elements of vector $\mathbf{\Delta}_1$ are expressed as

$$\Delta_{1i} = \begin{cases} S_x\hat{\beta}_t\hat{\sigma}^{-1}\exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-1} & \text{if } i \in F \\ S_x\hat{\beta}_t\hat{\sigma}^{-1}\exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-1} & \text{if } i \in C, \end{cases}$$

the elements of vector $\mathbf{\Delta}_2$ are expressed as

$$\Delta_{2i} = \begin{cases} \hat{\beta}_t S_x\hat{\sigma}^{-2}\left\{1-(k+1)\exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-1}\left(1+z_i\left[1+\exp\{\hat{z}_i\}\right]^{-1}\right)\right\} & \text{if } i \in F \\ -\hat{\beta}_t\hat{k}S_x\hat{\sigma}^{-2}\exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-1}\left(1+\hat{z}_i\left[1+\exp\{\hat{z}_i\}\right]^{-1}\right) & \text{if } i \in C, \end{cases}$$

the elements of vector $\mathbf{\Delta}_j$, for $j = 3, \ldots, p+2$ and $j \neq t$, take the forms

$$\Delta_{ji} = \begin{cases} -x_{ij}S_x\beta_t(\hat{k}+1)\hat{\sigma}^{-2}\exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-2} & \text{if } i \in F \\ -x_{ij}S_x\beta_t\hat{k}\hat{\sigma}^{-2}\exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-2} & \text{if } i \in C, \end{cases}$$

and the elements of vector $\mathbf{\Delta}_t$ are given by

$$\Delta_{ti} = \begin{cases} S_x\hat{\sigma}^{-1}+(\hat{k}+1)S_x\hat{\sigma}^{-1}\exp\{z_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-1}\left[x_{it}\hat{\beta}_t-1\right] & \text{if } i \in F \\ \hat{k}S_x\hat{\sigma}^{-1}\exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-1}\left[x_{it}\hat{\beta}_t-1\right] & \text{if } i \in C. \end{cases}$$

4.4. Generalized leverage

Let $l(\boldsymbol{\theta})$ denote the log-likelihood function from the postulated model in Eq. (6), $\widehat{\boldsymbol{\theta}}$ the ML estimator of $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$ the expectation of $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^{\mathrm{T}}$, then, $\widehat{\mathbf{y}} = \boldsymbol{\mu}(\widehat{\boldsymbol{\theta}})$ will be the predicted response vector.

The main idea behind the concept of leverage (see, for instance, Cook and Weisberg (1982) and Wei et al. (1998)) is that of evaluating the influence of y_i on its own predicted value. This influence may well be represented by the derivative $\frac{\partial\widehat{y_i}}{\partial yi}$ that, when equal to h_{ii}, is the ith principal diagonal element of projection matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$ and $\mathbf{X}$ is the model matrix. Extensions to more general regression models have been given, for instance, by St. Laurent and Cook (1992) and Paula (1999), when $\boldsymbol{\theta}$ is restricted with inequalities. Hence, it follows from Wei et al. (1998) that the $n \times n$ matrix $(\frac{\partial\widehat{\mathbf{y}}}{\partial\mathbf{y}})$ of generalized leverage can be expressed as:

$$\mathbf{GL}(\widehat{\boldsymbol{\theta}}) = \left\{\mathbf{D}_{\boldsymbol{\theta}}\left[\ddot{\mathbf{L}}(\boldsymbol{\theta})\right]^{-1}\ddot{\mathbf{L}}_{\boldsymbol{\theta}\mathbf{y}}\right\}$$

evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and where $\mathbf{D}_{\boldsymbol{\theta}} = \left(\frac{\partial[E(Y_i)]}{\partial k}, \frac{\partial[E(Y_i)]}{\partial\sigma}, x_{ij}\right)$ and

$$\ddot{\mathbf{L}}_{\boldsymbol{\theta}\mathbf{y}} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\mathbf{y}^{\mathrm{T}}} = \left(\ddot{\mathbf{L}}_{ky_i}, \ddot{\mathbf{L}}_{\sigma y_i}, \ddot{\mathbf{L}}_{\beta_j y_i}\right)^{\mathrm{T}}$$

with

$$\ddot{\mathbf{L}}_{\hat{k}y_i} = \begin{cases} -\hat{\sigma}^{-1}\hat{h}_i & \text{if } i \in F \\ -\hat{\sigma}^{-1}\exp\{\hat{h}_i\} & \text{if } i \in C, \end{cases}$$

$$\ddot{\mathbf{L}}_{\hat{\sigma}y_i} = \begin{cases} \hat{\sigma}^{-2}\left\{-1+(\hat{k}+1)\hat{h}_i\left[1+\hat{z}_i+\exp\{\hat{z}_i\}\right]\left[1+\exp\{\hat{z}_i\}\right]^{-1}\right\} & \text{if } i \in F \\ \hat{\sigma}^{-2}\hat{k}\hat{h}_i\left[1+\hat{z}_i+\exp\{\hat{z}_i\}\right]\left[1+\exp\{\hat{z}_i\}\right]^{-1} & \text{if } i \in C, \end{cases}$$

$$\ddot{\mathbf{L}}_{\hat{\beta}y_i} = \begin{cases} x_{ij}\hat{\sigma}^{-2}(\hat{k}+1)\hat{h}_i\left[1+\exp\{\hat{z}_i\}\right]^{-1} & \text{if } i \in F \\ x_{ij}\hat{\sigma}^{-2}\hat{k}\hat{h}_i\left[1+\exp\{\hat{z}_i\}\right]^{-1} & \text{if } i \in C, \end{cases}$$

where $\hat{h}_i = \exp\{\hat{z}_i\}\left[1+\exp\{\hat{z}_i\}\right]^{-1}$.

5. Residual analysis

In order to study departures from the error assumption as well as presence of outliers we will consider the martingale residual proposed by Barlow and Prentice (1988) and transformations in this residual.

5.1. Martingale residual

This residual was introduced in the counting process (see Fleming and Harrington (1991)) and can be written in log-Burr XII regression models as

$$r_{M_i} = \begin{cases} 1 - \hat{k}\log(1+\exp\{\hat{z}_i\}) & \text{if } i \in F \\ -\hat{k}\log(1+\exp\{\hat{z}_i\}) & \text{if } i \in C, \end{cases}$$

where $\hat{z}_i = \frac{y_i - \hat{\mu}}{\hat{\sigma}}$. The distributional form of r_{M_i} is skewness, and it has maximum value $+1$ and minimum value $-\infty$. Transformations to achieve a more normal shaped form would be more appropriate for residual analysis.

5.2. Martingale-type residual

Another possibility is to use a transformation of the martingale residual based on the deviance residuals for the Cox model with no time-dependent covariates introduced by Therneau et al. (1990). We will use this transformation of the martingale residual in order to have a new residual symmetrically distributed around zero (see Ortega et al. (in press)). Thus, a martingale-type residual for the log-Burr XII regression model can be expressed as:

$$r_{D_i} = \begin{cases} \operatorname{sign}\left[1 - \hat{k}\log(1+\exp\{\hat{z}_i\})\right] & \\ \quad \times \left[-2\left[1 - \hat{k}\log(1+\exp\{\hat{z}_i\}) + \log(1 - \hat{k}\log(1+\exp\{\hat{z}_i\}))\right]\right]^{\frac{1}{2}} & \text{if } i \in F \\ \operatorname{sign}\left[-\hat{k}\log(1+\exp\{\hat{z}_i\})\right]\left[2\hat{k}\log(1+\exp\{\hat{z}_i\})\right]^{\frac{1}{2}} & \text{if } i \in C. \end{cases}$$

5.3. Modified martingale-type residual

We have proposed a change in the martingale-type residual, and it can be written as

$$r_{MD_i} = (1 - \delta_i) + r_{D_i},$$

where $\delta_i = 0$ denotes censored observation and $\delta_i = 1$ uncensored and r_{D_i} is the martingale-type residual that is defined in Section 5.2. In the log-Burr XII regression models, the modified martingale-type residual is given by

$$r_{MM_i} = \begin{cases} \operatorname{sign}\left[1 - \hat{k}\log(1+\exp\{\hat{z}_i\})\right] & \\ \quad \times \left[-2\left[1 - \hat{k}\log(1+\exp\{\hat{z}_i\}) + \log(1 - \hat{k}\log(1+\exp\{\hat{z}_i\}))\right]\right]^{\frac{1}{2}} & \text{if } i \in F \\ 1 + \operatorname{sign}\left[-\hat{k}\log(1+\exp\{\hat{z}_i\})\right]\left[2\hat{k}\log(1+\exp\{\hat{z}_i\})\right]^{\frac{1}{2}} & \text{if } i \in C. \end{cases}$$

5.4. Impact of the detected influential observations

To reveal the impact of the detected influential observations, we estimated the parameters again without the influential observations. Let $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^0$ be the maximum likelihood estimator of the parameters models that are obtained from the data sets with and without the influential observations, respectively. Lee et al. (2006) define the following two quantities to measure the difference between $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^0$:

$$\text{TRC} = \sum_{i=1}^{n_p}\left|\frac{\hat{\theta}_i - \hat{\theta}_i^0}{\hat{\theta}_i}\right| \quad \text{and} \quad \text{MRC} = \max_i \left|\frac{\hat{\theta}_i - \hat{\theta}_i^0}{\hat{\theta}_i}\right|,$$

where TRC is total relative changes, MRC maximum relative changes and n_p is the number of parameters, and likelihood displacement: $LD_I(\boldsymbol{\theta}) = 2\{l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{(I)})\}$, where $\hat{\boldsymbol{\theta}}_{(I)}$ denotes the ML estimator of $\boldsymbol{\theta}$ after the set (I) of influential observations has been removed (see Cook et al. (1988)).

Now, the same number of the influential observations are randomly selected from the non influential observations and TRC, MRC and LD_I are again calculated. After this, the results can be compared, and if there is difference between them, the observations are influential.

6. Application

We provide an application of the results derived in the previous sections using real data. The required numerical evaluations were implemented using program Ox (see, Doornik (2001)).

6.1. Application of vitamin A data

We illustrate the proposed model using data from a randomized community trial that was designed to evaluated the effect of vitamin A supplementation on diarrheal episodes in 1207 pre-school children, aged 6–48 months at the baseline, who were assigned to receive either placebo or vitamin A in a small city in the Northeast of Brazil from December 1990 to December 1991.

The vitamin A dosage was 100,000 IU for children younger than 12 months and 200,000 IU for older children, which is the highest dosage guideline established by the World Health Organization (WHO) for the prevention of vitamin A deficiency.

The total time was defined as the time from the first dose of vitamin A until the occurrence of an episode of diarrhea. An episode of diarrhea was defined as a sequence of days with diarrhea and a day with diarrhea was defined when 3 or more liquid or semi-liquid motions were reported in a 24 h period. The information on the occurrence of diarrhea collected at each visit corresponds to a recall period of 48–72 h. The number of liquid and semi-liquid motions per 24 hours was recorded.

The covariates considered in the models are:

- x_{i1}: age at baseline (in months);
- x_{i2}: treatment (0 = placebo, 1 = vitamin A);
- x_{i3}: gender (0 = girl, 1 = boy).

In many applications, there is qualitative information about the failure rate function shape, which can help with selecting a particular model. In this context, a device called the total time on test (TTT) plot (Aarset, 1987) is useful. The TTT plot is obtained by plotting $\mathsf{G}(r/n) = [(\sum_{i=1}^{r} T_{i:n}) + (n-r)T_{r:n}]/(\sum_{i=1}^{n} T_{i:n})$, where $r = 1, \ldots, n$ and $T_{i:n}, i = 1, \ldots, n$, are the order statistics of the sample, against r/n (Mudholkar et al., 1996). The TTT plot for these data is in Fig. 2 and indicates a unimodal-shaped failure rate function.

We now present results on fitting the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \sigma z_i$$

where variable Y_i follows the log-Burr XII distribution given in (3), $i = 1, 2, \ldots, 1207$.

6.1.1. Maximum likelihood estimation

To obtain the maximum likelihood estimates for the parameters in the model, we used subroutine MAXBFGS in Ox, whose results are given in the Table 3. We can observe that variable $x1$ is significant for the model.

6.1.2. Bayesian analysis

The following independent priors were considered to perform the Gibbs sampler. $\beta_j \sim (0, 1000)$ $j = 0, 1, 2, 3, \sigma \sim IG(0.01, 0.01)$ and $k \sim G(0.01, 0.01)$, so that we have a vague prior distribution. Considering these prior densities, we generated two parallel independent runs of the Metropolis–Hasting algorithm chain with size 35,000 for each parameter, disregarding the first 5000 iterations to eliminate the effect of the initial values, and in order to avoid correlation problems, we considered spacing of size 10, obtaining a sample of size 3000 from each chain. To monitor the convergence of the Gibbs samples, we used the between and within sequence information, following the approach

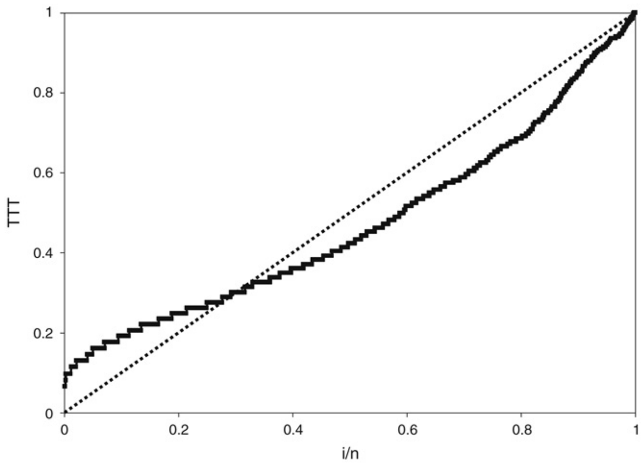

Fig. 2. TTT-plot on Vitamin A data.

Table 3
Maximum likelihood estimates for the parameters from the log-Burr XII regression model on the complete Vitamin A data set

Parameter	Estimate	SE	p-value
k	0.2764	0.0387	–
σ	0.3567	0.0305	–
β_0	2.2522	0.0926	0
β_1	0.0221	0.0029	<0.01
β_2	0.0898	0.0600	0.1346
β_3	0.0441	0.0598	0.4601

Table 4
Posterior summaries for the parameters from the log-Burr XII regression model in the complete Vitamin A data set

Parameter	Mean	Median	S.D.	2.5%	97.5%	$\hat{R}$
k	0.2853	0.2817	0.04085	0.2171	0.3763	1.001
σ	0.3628	0.3617	0.0308	0.3064	0.4271	1.009
β_0	2.2551	2.2534	0.0948	2.0693	2.4451	1.000
β_1	0.0224	0.0223	0.0028	0.0169	0.0281	1.002
β_2	0.0905	0.0904	0.0602	−0.0267	0.2098	1.006
β_3	0.0461	0.0458	0.0605	−0.0743	0.16428	1.004

developed in Gelman and Rubin (1992) to obtain the potential scale reduction, $\hat{R}$. In all cases, these values were close to one, indicating the convergence of the chain. The histograms with the approximate posterior marginal density of the parameters are presented in Fig. 3.

In Table 4, we report posterior summaries for the parameters of the log-Burr regression model. We can observe that variable $x1$ is significant for the model.

6.1.3. Jackknife estimator

In Table 5, we report the jackknife estimates for the parameters of the log-Burr XII regression model. From Table 5, we can observe that variable $x1$ is significant for the model when the jackknife estimator is used.

6.2. Global influence analysis

In this sub-section, we use Ox to compute case-deletion measures $GD_i(\boldsymbol{\theta})$ and $LD_i(\boldsymbol{\theta})$, presented in Sub-Section 4.1. The results of such influence measures index plots are displayed in Fig. 4.

From the figure, we can see that cases 825 and 1192 are possible influential observations.

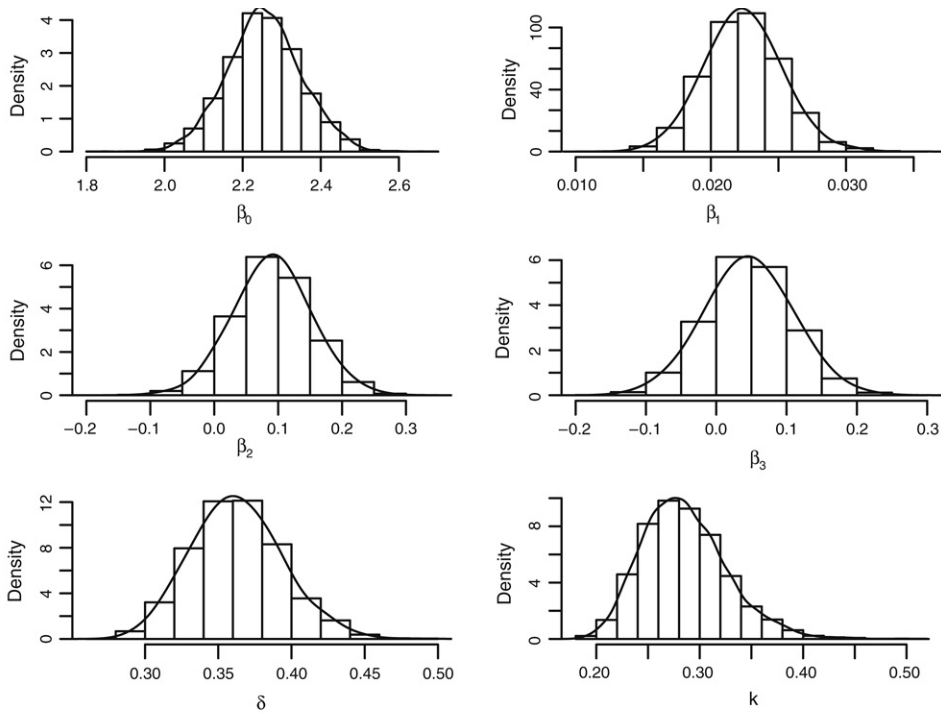

Fig. 3. Approximate posterior marginal densities for the parameters from the log-Burr XII regression model on the complete Vitamin A data set.

Table 5
Jackknife estimates for the parameters from the log-Burr XII regression model in the complete Vitamin A data set

Parameter	Estimate	SE	95% confidence interval
k	0.26641	0.0477	(0.1728; 0.3600)
σ	0.3599	0.0364	(0.2885; 0.4313)
β_0	2.2464	0.0879	(2.0739; 2.4189)
β_1	0.0255	0.0035	(0,0186; 0.0324)
β_2	0.0921	0.0622	(−0.0299; 0.2141)
β_3	0.0482	0.0616	(−0.0727; 0.1691)

6.3. Local and total influence analysis

In this section, we will make an analysis of local influence for the data set using log-Burr XII regression models.

6.3.1. Case-weight perturbation

By applying the local influence theory developed in Section 4, where case-weight perturbation is used, value $C_{\mathbf{d}_{\max}} = 2.0230$ was obtained as maximum curvature. In Fig. 5(a), the graph of eigenvector corresponding to $C_{\mathbf{d}_{\max}}$ is presented, and total influence C_i is shown in Fig. 5(b). Observation 1, 192 is the most distinguished in relation to the others.

6.3.2. Response variable perturbation

Next, the influence of perturbations in the observed survival times will be analyzed. The value for the maximum curvature calculated was $C_{\mathbf{d}_{\max}} = 5.875$. Fig. 6(a) containing the graph for $|\mathbf{d}_{\max}|$ versus the observation index shows that no point is salient in relation to the others. The same applies to Fig. 6(b), which corresponds to total local influence (C_i).

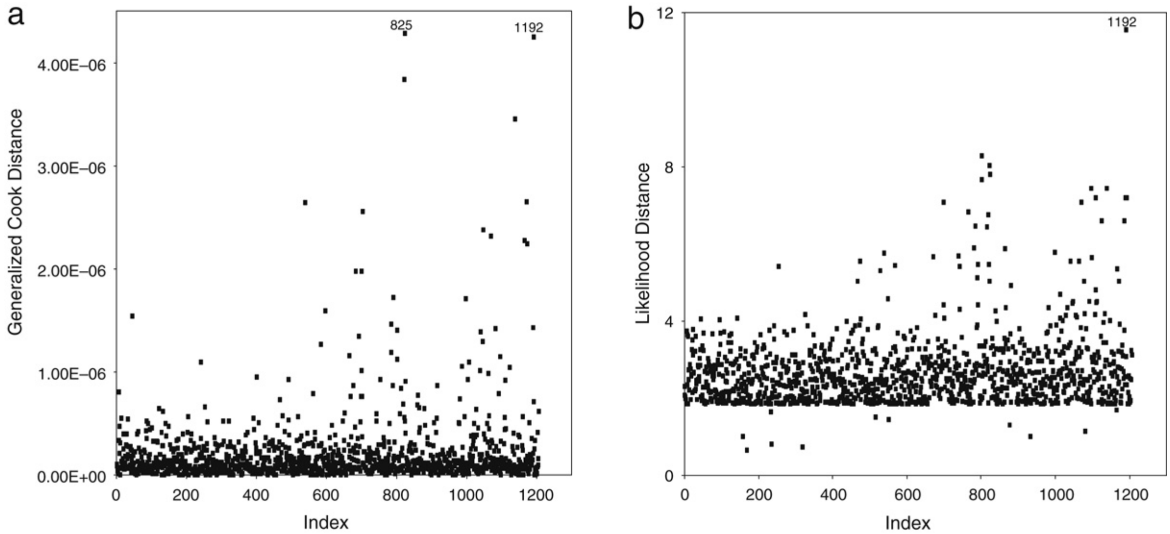

Fig. 4. Index plot for $\boldsymbol{\theta}$: (a) $GD_i(\boldsymbol{\theta})$ (Generalized Cook's distance) and (b) $LD_i(\boldsymbol{\theta})$ (Likelihood distance).

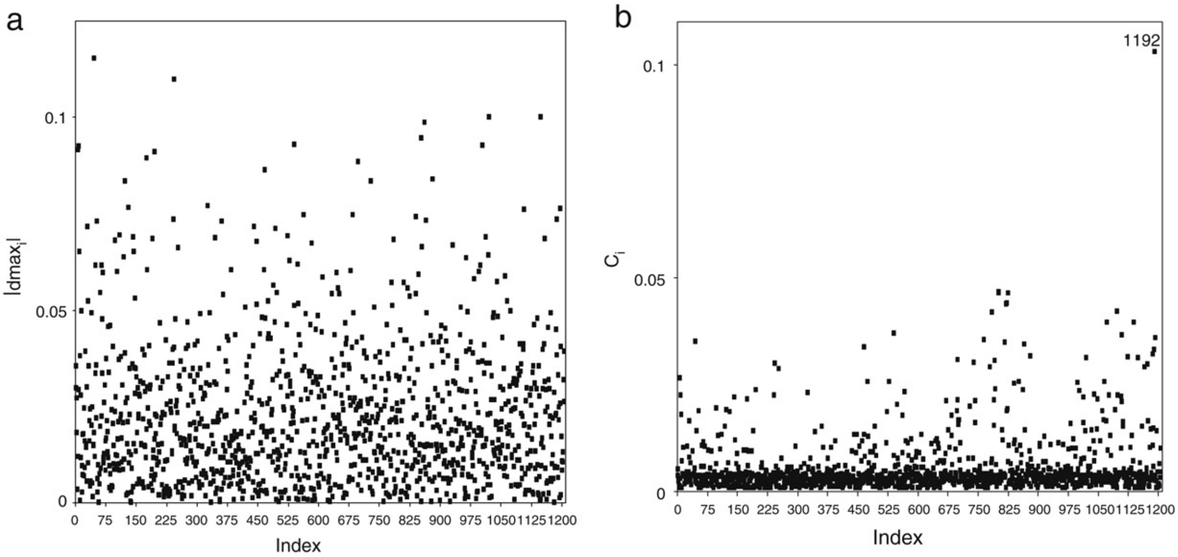

Fig. 5. Index plot for $\boldsymbol{\theta}$ (case-weight perturbation): (a) $\mathbf{d}_{\max}$ and (b) Total local influence.

6.3.3. Explanatory variable perturbation

The perturbation of the vector for covariable age ($x1$) is investigated here. For perturbation of covariable age, value $C_{\mathbf{d}_{\max}} = 0.0086$ was obtained as maximum curvature. The respective graphs of $|\mathbf{d}_{\max}|$ as well as total local influence C_i against the observation index are shown in Fig. 7(a) and 7(b), respectively. In these two graphs, we can see no influential observation.

6.3.4. Generalized leverage analysis

Fig. 8 exhibits the index plot of $GL(\boldsymbol{\theta})$, using the model given in Eq. (6). The generalized leverage graph presented in Fig. 8 shows no points as possible leverage points. We can notice that all the observations have been well shaped.

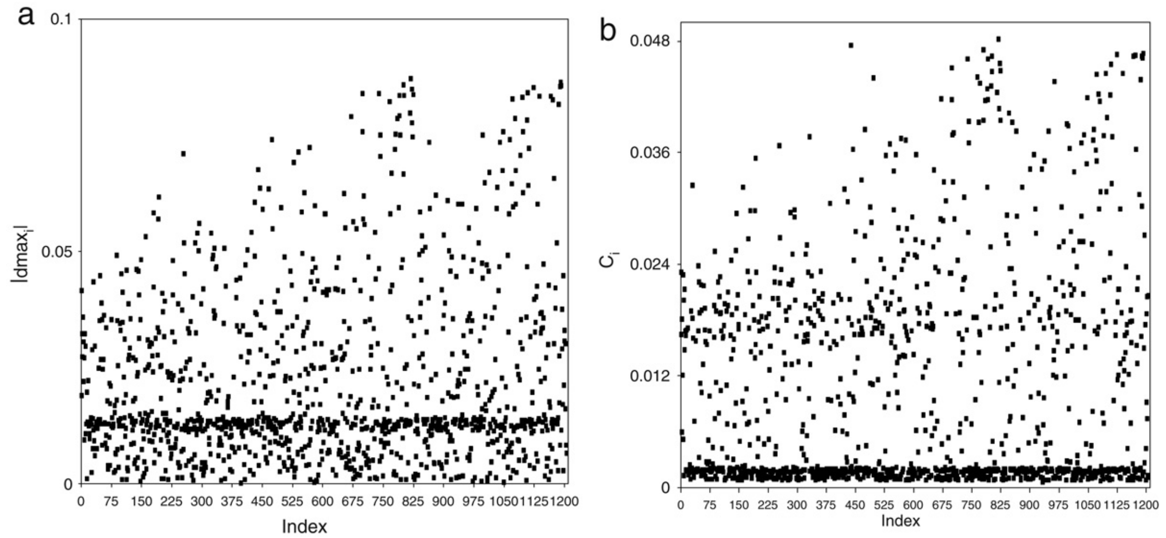

Fig. 6. Index plot for $\boldsymbol{\theta}$ (response perturbation): (a) $\mathbf{d}_{\max}$ and (b) Total local influence.

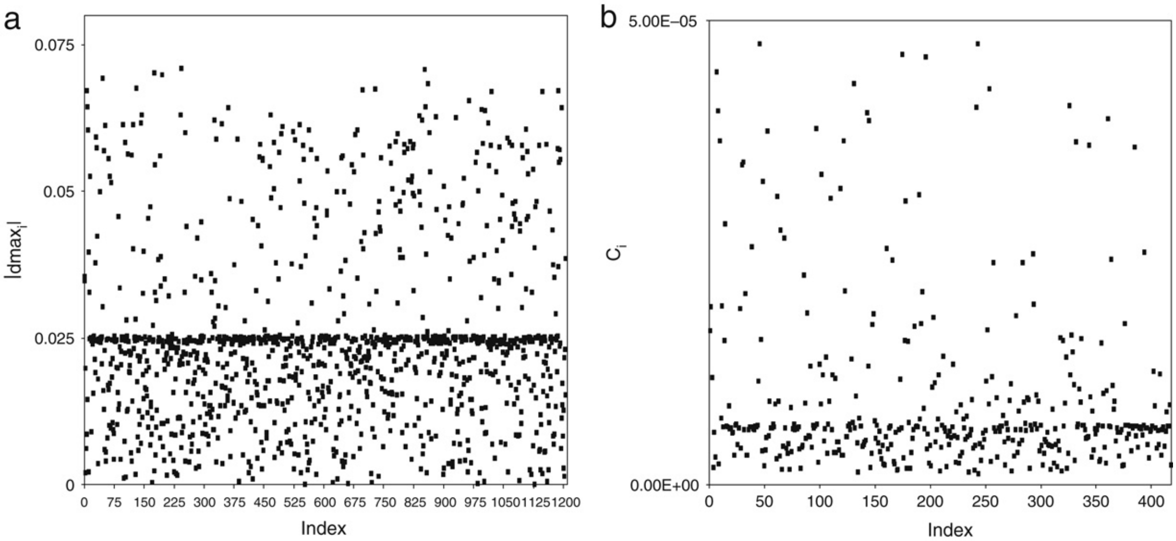

Fig. 7. Index plot for $\boldsymbol{\theta}$ (age explanatory variable perturbation): (a) $\mathbf{d}_{\max}$ and (b) Total local influence.

6.4. Residual analysis

In order to detect possible outlying observations as well as departures from the assumptions of log-Burr XII regression model, we present, in Fig. 9, the graphs of r_{Mi} and r_{MMi} against the order observations. By analyzing these graphs, asymmetry is observed; however, the modified martingale-type residual graph presents a reasonably random pattern and case 1192 exhibits an atypical residual value.

6.5. Impact of the detected influential observations

In concluding previous sections, we can consider case 1192 as a possible influence point or outlier observation. Case 1192 has the lowest time.

We found that TRC $= 0.2246$, MRC $= 0.1226$ and $LD_{(I)} = -5.58$. In order to compare the impact of the influential observations, we repeated the analysis by removing the same number (1 observation) randomly selected

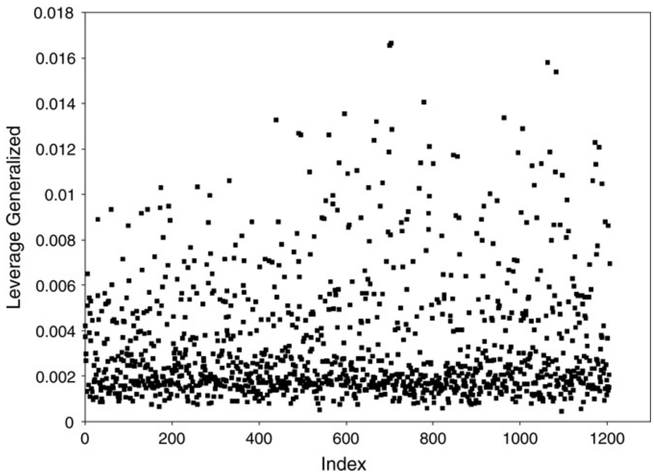

Fig. 8. Index plot of generalized leverage on fitting the log-Burr XII regression model for Vitamin A data.

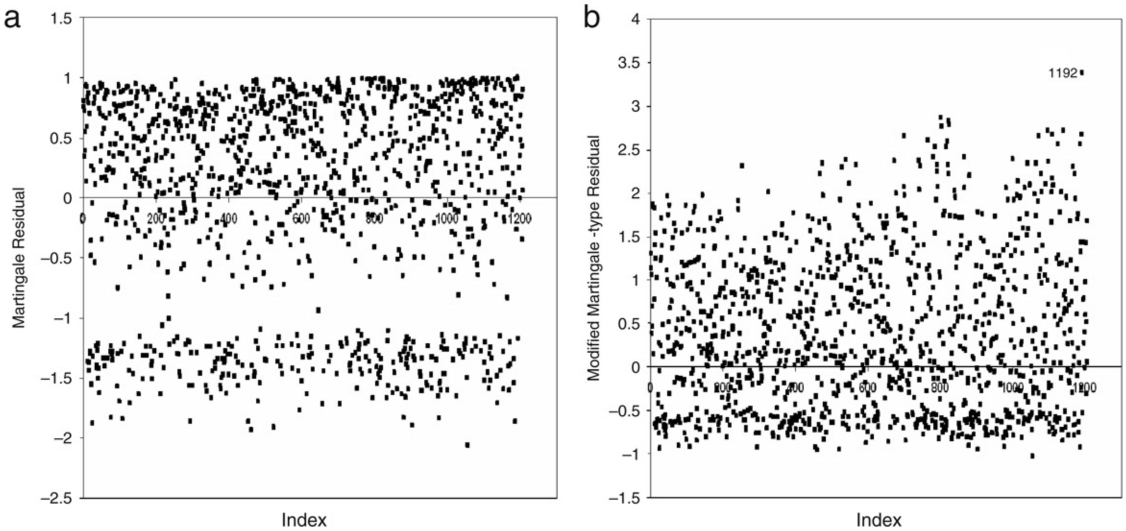

Fig. 9. Index plot of residuals on fitting the log-Burr XII regression model for Vitamin A data: (a) martingale residual (r_{M_i}) and (b) modified martingale-type residual (r_{MD_i}).

from the non influential observations. In this case, we found that TRC $= 0.1681$, MRC $= 0.1037$ and $LD_{(I)} = -2.53$. Hence, the results showed that case 1192 did not cause a strong impact on the parameters estimation.

Therefore, the final model becomes given by

$$y_i = \beta_0 + \beta_1 x_{i1} + \sigma z_i, \quad i = 1, 2, \ldots, 1207. \tag{10}$$

The maximum likelihood estimates for the parameters in the final model are given in Table 6. We can interpret the estimated coefficients of the final model as the following: The median survival time should increase approximately 2.18% ($e^{0.02152} \times 100\%$) as age increases one unit.

6.6. Goodness of fitting

In order to assess if the model is appropriate, the plot comparing the empirical distribution for the survival function and survival function estimated by the log-Burr XII and log-logistic models were introduced in Fig. 10. From this figure, it is noted that both models show satisfactory fitting; however, the log-Burr XII model presents better fitting

Table 6
Maximum likelihood estimates for the parameters from the log-Burr XII regression model on the complete Vitamin A data set — final model

Parameter	Estimate	SE	p-value
k	0.2698	0.0014	–
σ	0.3521	0.0009	–
β_0	2.3191	0.0065	0
β_1	0.02152	0.0000	<0.01

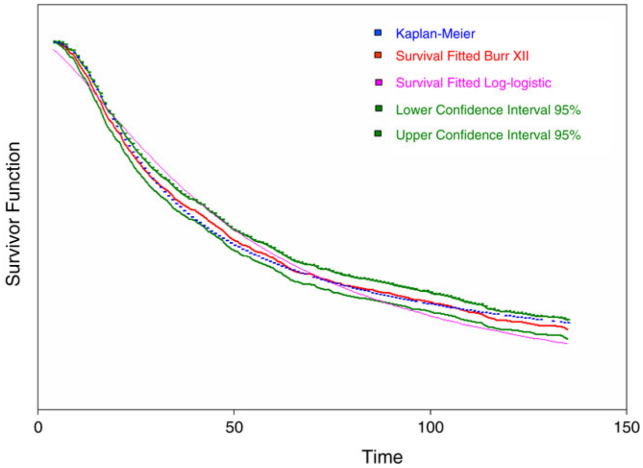

Fig. 10. Estimated survival function on fitting the log-Burr XII and log-logistic regression models with empirical survival for Vitamin A data.

to the data under analysis. Additionally, the likelihood-ratio test was performed for hypotheses $H_0 : k = 1$ or the log-logistic regression model is adequate versus $H_1 : k \neq 1$ or the log-Burr regression model is adequate. The test statistic resulted in $\lambda = 2 \times (1,639.17 - 1,611.67) = 54.99$ (p-value < 0.01), and this result provides favorable indications to the log-Burr XII regression model.

7. Concluding remarks

In this paper, a log-Burr XII regression model with the presence of censored data is proposed as an alternative to model lifetime when the failure rate function presents unimodal shape. We used three estimation methods for the parameters of the proposed model: maximum likelihood, Bayesian inference and jackknife estimator. Asymptotic tests were performed for the parameters based on the asymptotic distribution of the maximum likelihood estimators. In the applications within a real data, we observed that all estimation methods presented similar results. Furthermore, this article compared the performance of the proposed model and the log-logistic regression model based on variance, mean squared error and the likelihood-ratio test through a simulation study. These simulations suggest that log-Burr XII regression model can be used for modeling data with a unimodal failure rate function. We have also discussed applications of influence diagnostics in log-Burr XII regression models with censored data. So, we perform a general model checking analysis which makes this model a very attractive option for modeling censored and uncensored lifetime data that has a unimodal failure rate function. The approach was applied to real data sets, which indicates the usefulness of the approach.

Appendix A. Matrix of second derivatives $\ddot{\mathbf{L}}(\boldsymbol{\theta})$

Here we derive the necessary formulas to obtain the second-order partial derivatives of the log-likelihood function. After some algebraic manipulations, we obtain

$$\mathbf{L}_{kk} = -\frac{r}{k^2}$$

$$\mathbf{L}_{k\sigma} = \sum_{i=1}^{n} \frac{z_i}{\sigma} h_i$$

$$\mathbf{L}_{k\beta_j} = \sum_{i=1}^{n} \frac{x_{ij}}{\sigma} h_i$$

$$\begin{aligned}\mathbf{L}_{\sigma\sigma} = {} & \frac{r}{\sigma^2} + \frac{2}{\sigma^2}\sum_{i\in F} z_i - 2(k+1)\sum_{i\in F}\frac{z_i}{\sigma^2}h_i - (k+1)\sum_{i\in F}\left(\frac{z_i}{\sigma}\right)^2 h_i \\ & + \sum_{i\in F}\left(\frac{z_i}{\sigma}h_i\right)^2 - 2k\sum_{i\in C}\frac{z_i}{\sigma^2}h_i - k\sum_{i\in C}\left(\frac{z_i}{\sigma}\right)^2 h_i + k\sum_{i=1}^{n}\left(\frac{z_i}{\sigma}h_i\right)^2\end{aligned}$$

$$\mathbf{L}_{\sigma\beta_j} = \sum_{i\in F}\frac{x_{ij}}{\sigma^2} - (k+1)\sum_{i\in F}\left(\frac{x_{ij}}{\sigma^2} + \frac{x_{ij}z_i}{\sigma^2}\right)h_i + \sum_{i\in F}\frac{x_{ij}z_i}{\sigma^2}{h_i}^2 - k\sum_{i\in C}\left(\frac{x_{ij}}{\sigma^2} + \frac{x_{ij}z_i}{\sigma^2}\right)h_i + k\sum_{i=1}^{n}\frac{x_{ij}z_i}{\sigma^2}{h_i}^2$$

$$\mathbf{L}_{\beta_j\beta_s} = -(k+1)\sum_{i\in F}\frac{x_{ij}x_{is}}{\sigma^2}h_i + (k+1)\sum_{i\in F}\frac{x_{ij}x_{is}}{\sigma^2}{h_i}^2 - k\sum_{i\in C}\frac{x_{ij}x_{is}}{\sigma^2}h_i + k\sum_{i\in C}\frac{x_{ij}x_{is}}{\sigma^2}{h_i}^2$$

where que $j, s = 1, 2, \ldots, p$, $h_i = \frac{\exp(z_i)}{1+\exp(z_i)}$ and $z_i = \frac{y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}}{\sigma}$

Appendix B. Proof of Theorem 1

For log-Burr XII distribution (3), the moment-generating function (mgf) is given by the result from solving equation

$$M_Y(T) = E(\exp\{ty\}) = \int_{\infty}^{-\infty} \exp\{ty\}kc\left(1 + \left(\frac{\exp\{y\}}{s}\right)^c\right)^{-k-1}\left(\frac{\exp\{y\}}{s}\right)^c \mathrm{d}t.$$

Let $u = \left(\frac{\exp\{y\}}{s}\right)^c$ then $\mathrm{d}u = c\left(\frac{\exp\{y\}}{s}\right)^c \mathrm{d}y$. Hence

$$M_Y(T) = \int_{\infty}^{-\infty} \exp\{ty\}k(1+u)^{-k-1}\mathrm{d}u.$$

Now make the univariate change of variable $v = \frac{1}{1+u}$ so that $dv = -(1+u)^{-2}\mathrm{d}u$ to obtain

$$M_Y(T) = \int_{1}^{0} -s^t k(1-v)^{\frac{t}{c}} v^{-\frac{t}{c}+k-1}\mathrm{d}v ks^t B\left[\frac{t}{c}+1, k-\frac{t}{c}\right], \quad \text{if } kc > t$$

where $B(a, b)$ is the complete beta function (see Lawless (2003)). To obtain the second identity, we recognized the integrand as the kernel of the beta pdf.

References

Aarset, M.V., 1987. How to identify bathtub hazard rate. IEEE Transactions on Reliability 36, 106–108.

Barlow, W.E., Prentice, R.L., 1988. Residual for relative risk regression. Biometrika 75, 65–74.

Barreto, M.L., Santos, L.M.P., Assis, A.M.O., araújo, M.P.N., Farenzena, G.G., Santos, P.A.B., Fiaccone, R.L., 1994. Effect of vitamin A supplementation on diarrohea and acute lower-respiratory-tract infections in young children in Brazil. Lancet 344, 228–231.

Cook, R.D., Weisberg, S., 1982. Residuals and Influence in Regression. Chapman and Hill, New York.

Cook, R.D., 1977. Detection of influential observations in linear regression. Technometrics 19, 15–18.

Cook, R.D., 1986. Assesment of local influence (with discussion). Journal of the Royal Statistical Society 48 (2), 133–169.

Cook, R.D., Peña, D., Weisberg, S., 1988. The likelihood displacement: A unifying principle for influence. Communications in Statistics, Part Theory and Methods 17, 623–640.

Cox, D.R., Hinkley, D.V., 1974. Theoretical Statistics. Chapman and Hill, London, p. 511.

Doornik, J., 2001. Ox: Object-oriented Matrix Programming Using Ox, fourth ed. Timberlake Consultants Ltd, London.

Escobar, L.A., Meeker, W.Q., 1992. Assessing influence in regression analysis with censored data. Biometrics 48, 507–528.

Fleming, T.R., Harrington, D.P., 1991. Counting Process and Survival Analysis. Wiley, New York.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences (with discussion). Statistical Science 7, 457–472.

Hosmer, D.W., Lemeshow, S., 1999. Applied Survival Analysis. John Wiley and Sons, New York.

Lawless, J.F., 2003. Statistical Models and Methods for Lifetime Data. Wiley, New York.

Lee, S.Y., Lu, B., Song, X.Y., 2006. Assessing local influence for nonlinear structural equation models with ignorable missing data. Computational Statistics and Data Analysis 50, 1356–1377.

Lesaffre, E., Verbeke, G., 1998. Local influence in linear mixed models. Biometrics 54, 570–582.

Leiva-Sanchez, V., Barros, M., Paula, G.A., Galea, M., 2006. Influence diagnostics in log-birnbaum-saunders regression models with censored data. Computational Statistics and Data Analysis 51, 5694–5707.

Manly, B.F.J., 1997. Randomization, Bootstrap and Monte Carlo Methods in Biology, second ed. Chapman and Hall, London.

Mudholkar, G.S., Srivastava, D.k., Kollia, G.D., 1996. A generalization of the Weibull distribution with application to the analysis of survival data. Journal of the American Statistical Association 91, 1575–1583.

Nelson, W., 1982. Applied Life Data Analysis. John Wiley & Sons, New York.

Ortega, E.M.M., Bolfarine, H., Paula, G.A., 2003. Influence diagnostics in generalized log-gamma regression models. Computational Statistics and Data Analysis 42, 165–186.

Ortega, E.M.M., Cancho, V.G., Bolfarine, H., 2006. Influence diagnostics in exponentiated-Weibull regression models with censored data. Statistics and Operation Reserch Transactions 30 (2), 171–192.

Ortega, E.M.M., Paula, G.A., Bolfarine, H., 2008. Deviance residuals in generalized log-gamma regression models with censored observations, Journal of Statistical Computation and Simulation (in press).

Paula, G.A., 1999. Leverage in inequality constrained regression models. Journal of the Royal Statistical Society 48 (4), 529–538.

Pettit, A.N., Bin Daud, I., 1989. Case-weight measures of influence for proportional hazards regression. Applied Statistics 38, 51–67.

Rodriguez, R.N., 1977. A Guide to the Burr Type XII distributions. Biometrika 64 (1), 129–134.

St. Laurent, R.T., Cook, R.D., 1992. Leverage and superleverage in nonlinear regression. Journal of the American Statistical Association 87, 985–990.

Therneau, T.M., Grambsch, P.M, Fleming, T.R., 1990. Martingale-based residuals for survival models. Biometrika 77, 147–160.

Wei, B., Hu, Y.Q., Fung, W.K., 1998. Generalized leverage and its applications. Scandinavian Journal of statistics 25, 25–37.

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

Zimmer, W.J., Keats, J.B., Wang, F.K., 1998. The Burr XII distribution in reliability analysis. Journal of Quality Technology 30, 389–394.