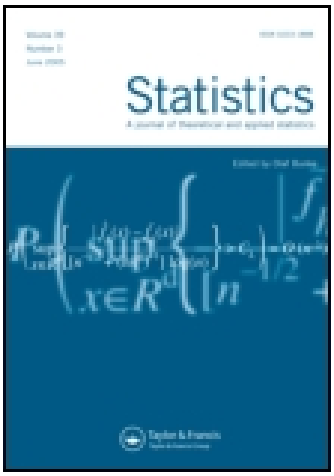

Statistics: A Journal of Theoretical and Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gsta20

The Poisson Birnbaum-Saunders model with long-term survivors

Elizabeth M. Hashimoto[a], Edwin M.M. Ortega[a], Gauss M. Cordeiro[b] & Vicente G. Cancho[c]

[a] Departamento de Ciências Exatas, Universidade de São Paulo, 13418-900, Piracicaba, SP, Brazil

[b] Departamento de Estatística Universidade Federal de Pernambuco, 50740-540, Recife, PE, Brazil

[c] Departamento de Matemática Aplicada e Estatística Universidade de São Paulo, 13560-970, São Carlos, SP, Brazil

Published online: 07 Oct 2013.

To cite this article: Elizabeth M. Hashimoto, Edwin M.M. Ortega, Gauss M. Cordeiro & Vicente G. Cancho (2014) The Poisson Birnbaum-Saunders model with long-term survivors, Statistics: A Journal of Theoretical and Applied Statistics, 48:6, 1394-1413, DOI: 10.1080/02331888.2013.822503

To link to this article: http://dx.doi.org/10.1080/02331888.2013.822503

Statistics, 2014
Vol. 48, No. 6, 1394–1413, http://dx.doi.org/10.1080/02331888.2013.822503

The Poisson Birnbaum–Saunders model with long-term survivors

Elizabeth M. Hashimoto[a], Edwin M.M. Ortega[a]*, Gauss M. Cordeiro[b] and Vicente G. Cancho[c]

[a]*Departamento de Ciências Exatas, Universidade de São Paulo, 13418–900, Piracicaba, SP, Brazil;*
[b]*Departamento de Estatística Universidade Federal de Pernambuco, 50740–540, Recife, PE, Brazil;*
[c]*Departamento de Matemática Aplicada e Estatística Universidade de São Paulo, 13560–970, São Carlos, SP, Brazil*

(*Received 17 January 2012; accepted 20 June 2013*)

In this paper, we propose a cure rate survival model by assuming that the number of competing causes of the event of interest follows the Poisson distribution and the time to event has the Birnbaum–Saunders (BS) distribution. We define the Poisson BS distribution and provide two useful representations for its density function which facilitate to obtain some mathematical properties. Two closed-form expressions for the moments of the new distribution are given. We estimate the parameters of the model with cure rate using maximum likelihood. For different parameter settings, sample sizes and censoring percentages, several simulations are performed. We derive the appropriate matrices for assessing local influence on the parameter estimates under different perturbation schemes and present some ways to perform a global influence study. We analyse a real data set from the medical area.

Keywords: Birnbaum–Saunders distribution; cure fraction model; lifetime data; Poisson distribution; sensitivity analysis

1. Introduction

The Birnbaum–Saunders (BS) distribution [1] is a positively skewed model with non-negative support that has received considerable attention in the last two decades. This is primarily due to its derivation that is based on physical consideration, its attractive properties and close relationship to the normal distribution and its applicability in a wide variety of fields. For details about various applications of the BS distribution, including the medical area, see [2–6].

The BS survival function (for $t > 0$) is given by

$$S_{\mathrm{BS}}(t) = \Phi\left[-\frac{1}{\alpha}\left(\sqrt{\frac{t}{\lambda}} - \sqrt{\frac{\lambda}{t}}\right)\right], \tag{1}$$

where $\Phi(\cdot)$ is the standard normal cumulative function, $\alpha > 0$ and $\lambda > 0$ are shape and scale parameters, respectively. The cumulative distribution function and probability density function (pdf) of the BS distribution are obtained from Equation (1) as

$$F_{\mathrm{BS}}(t) = 1 - S_{\mathrm{BS}}(t) = \Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{t}{\lambda}} - \sqrt{\frac{\lambda}{t}}\right)\right] \tag{2}$$

*Corresponding author. Email: edwin@esalq.usp.br, edwin@usp.br

and

$$f_{\mathrm{BS}}(t) = \frac{t^{-3/2}(t+\lambda)}{2\alpha\sqrt{2\pi\lambda}} \exp\left[-\frac{1}{2\alpha^2}\left(\frac{t}{\lambda} + \frac{\lambda}{t} - 2\right)\right], \tag{3}$$

respectively. The mean and the variance of the BS distribution are

$$\mathrm{E}(T) = \lambda\left(1 + \frac{\alpha^2}{2}\right) \quad \text{and} \quad \mathrm{Var}(T) = \alpha^2\lambda^2\left(\frac{5}{4}\alpha^2 + 1\right).$$

Some proposals have been made recently in the literature by replacing the relationship between the BS and normal distributions by more general classes of distributions. For instance, Díaz-García and Leiva-Sánchez [7] introduced the generalized Birnbaum-Saunders (GBS) distribution by considering the elliptical family of distributions. The main motivation for the use of the GBS distribution is to make the kurtosis more flexible (compared with the BS model). Sanhueza et al. [8] presented a complete compilation of the results of the GBS distribution and Gómez et al. [9] proposed an extension of the GBS model based on the slash-elliptical distributions.

Cure rate models for survival data (also called 'survival models with a surviving fraction' or 'long-term survival models') are often used to model cured proportions of subjects who may not remain susceptible to the event of interest. These models have become very popular due to significant progress and advancements in treatment therapies leading to enhanced cure rates. The proportion of these 'cured' units is termed the cured fraction. In clinical studies, the event of interest can be the death of a patient (which can happen due to different competing causes) or a tumour recurrence (which can be attributed to metastasis-component tumour cells left active after an initial treatment).

Models to accommodate a cured fraction have been widely developed. Perhaps the most popular type of cure rate model is the mixture distribution introduced by Boag [10] and Berkson and Gage.[11] Further, a mixture model is based on the assumption that only a cause is responsible for the occurrence of the event of interest. However, in clinical studies, the patient's death, which is the event of interest, may happen due to different latent competing causes, in the sense that there is no information about which cause was responsible for the individual death. A tumour recurrence can be attributed to metastasis-component tumour cells left active after initial treatment. A metastasis-component tumour cell is a tumour cell with potential to metastasize.[12] The literature on distributions which accommodates different latent competing causes is rich and growing rapidly. The book by Ibrahim et al.,[13] as well as the review paper by Tsodikov et al. [14] and the works by Cooner et al.,[15] Ortega et al. [16] and Cancho et al.,[17] can be mentioned as key references.

In this context, we propose a new model called the *Poisson Birnbaum–Saunders cure rate* (PBScr) model, conceived inside a latent competing causes scenario with cure fraction, where there is no information about which cause was responsible for the individual death or tumour recurrence, but only the minimum lifetime value among all risks is observed and a part of the population is not susceptible to the event of interest. For the assessment of model adequacy, we develop diagnostic studies to detect possible influential or extreme observations that can cause distortions on the results of the analysis using the local influence approach, where we investigate how the results of the analysis are changed under small perturbations in the model or data.

Cook [18] proposed a general framework to detect the influence of the observations to indicate how sensitive the analysis is when small perturbations in the data or model occur. Several authors have applied the local influence methodology in regression analysis with censoring. Ortega et al. [19] considered the problem of assessing local influence in generalized log-gamma regression models with censored observations; Silva et al. [20] investigated local influence in log-Burr XII regression models with censored data; Fachini et al. [21] adapted local influence methods to poly-hazard models under the presence of covariates. Cancho et al. [17] derived curvature calculations

under various perturbation schemes in log-exponentiated Weibull regression models with cure rate and Hashimoto et al. [22] determined the appropriate matrices for assessing local influences on the parameter estimates under different perturbation schemes in the log-exponentiated Weibull regression model for interval-censored data. Here, we propose a similar methodology to detect influential subjects on the PBScr model.

The paper is organized as follows. In Section 2, we formulate the PBScr model by defining the density, cumulative distribution and hazard rate functions of the Poisson Birnbaum–Saunders (PBS) distribution. In Section 3, we provide two useful expansions for the density function in order to obtain its mathematical properties. In Section 4, we calculate its moments and generating function. Section 5 deals with the mean residual lifetime (MRL) function. In Section 6, we describe the maximum-likelihood estimation procedure. The bootstrap re-sampling method [23] is reviewed in Section 7. In Section 8, we evaluate the performance of the parameter estimates using Monte Carlo simulation. In Section 9, we obtain the normal curvatures of local influence and derive the global influence under some usual perturbations. An application to a real data set on breast cancer is given in Section 10. Finally, Section 11 provides some general remarks.

2. Model formulation

The PBScr model is derived as follows. For an individual in the population, let N denote the unobservable number of causes of the event of interest for this individual. We assume that N has a Poisson distribution with mean φ. The time for the jth cause to produce the event of interest is denoted by Z_j, $j = 1, \ldots, N$. Further, we consider that, conditional on N, the Z_j's are i.i.d. random variables having the BS cumulative function given by Equation (2) and that $Z_1, Z_2, \ldots$ are independent of N. The observable time to the event of interest is defined by $T = \min\{Z_1, \ldots, Z_N\}$, and $T = \infty$ if $N = 0$ with $P(T = \infty \mid N = 0) = 1$. Under this set-up, the survival function (not a proper function) for the population is

$$S_{\text{pop}}(t) = \exp\left\{-\varphi\Phi\left[\alpha^{-1}\left(\sqrt{\frac{t}{\lambda}} - \sqrt{\frac{\lambda}{t}}\right)\right]\right\} \tag{4}$$

and the cured fraction is defined by $p_0 = S_{\text{pop}}(\infty) = \exp(-\varphi)$. The corresponding pdf reduces to

$$f_{\text{pop}}(t) = \varphi f_{\text{BS}}(t) \exp\left\{-\varphi\Phi\left[\alpha^{-1}\left(\sqrt{\frac{t}{\lambda}} - \sqrt{\frac{\lambda}{t}}\right)\right]\right\}, \tag{5}$$

where $f_{\text{BS}}(t)$ is given by Equation (3). The hazard rate function for the population is

$$h_{\text{pop}}(t) = \varphi f_{\text{BS}}(t). \tag{6}$$

The function $h_{\text{pop}}(t)$ is multiplicative in φ and $f_{\text{BS}}(t)$, and thus it has the proportional hazard structure when the covariates are modelled through φ. The functions $f_{\text{pop}}(t)$ and $h_{\text{pop}}(t)$ are improper functions, since $S_{\text{pop}}(t)$ is not a proper survival function.

The model (4) can be expressed as a mixture model [10,11] given by

$$S_{\text{pop}}(t) = \exp(-\varphi) + [1 - \exp(-\varphi)]\left\{\frac{\exp\{-\varphi\Phi[(1/\alpha)(\sqrt{t/\lambda} - \sqrt{\lambda/t})]\} - \exp(-\varphi)}{1 - \exp(-\varphi)}\right\}. \tag{7}$$

Then, from Equation (7), the survival function for the non-cured population, so-called the PBS survival function, is given by

$$S_{\text{PBS}}(t) = P(T > t \mid N \geq 1) = \frac{\exp\{-\varphi\Phi[(1/\alpha)(\sqrt{t/\lambda} - \sqrt{\lambda/t})]\} - \exp(-\varphi)}{1 - \exp(-\varphi)}. \tag{8}$$

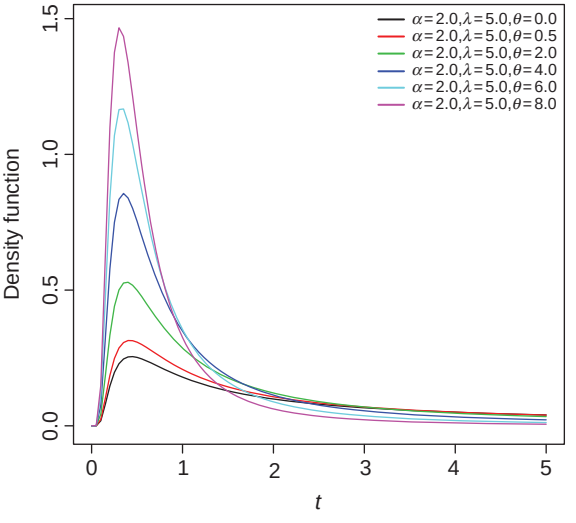

Figure 1. The PBS density function. The parameters are fixed: $\varphi = 0, 0.5, 2, 4, 6, 8$ and $\lambda = 5$, $\alpha = 0.2$ (left panel), $\alpha = 2$ (right panel).

We note that $S_{\mathrm{PBS}}(0) = 1$ and $S_{\mathrm{PBS}}(\infty) = 0$, so that it is a proper survival function. Henceforth, the model (8) is referred to as the PBS survival function, whereas Equation (5) is called the PBS model with long-term survivors in competitive-risk structure. The new density function for the non-cured population reduces to

$$f_{\mathrm{PBS}}(t) = \frac{\varphi t^{-3/2}(t+\lambda)}{2\alpha\sqrt{2\pi\lambda}(1-\mathrm{e}^{-\varphi})} \times \exp\left\{-\frac{1}{2\alpha^2}\left(\frac{t}{\lambda}+\frac{\lambda}{t}-2\right) - \varphi\Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{t}{\lambda}}-\sqrt{\frac{\lambda}{t}}\right)\right]\right\}. \quad (9)$$

In Equation (9), the parameter λ controls the scale of the distribution and the parameters α and φ control its shape. As φ approaches zero, the PBS distribution converges to the BS distribution. In Figure 1, we plot the PBS density function for selected values of φ. So, the PBS distribution can be used to model data in survival analysis.

From Equations (8) and (9), it is easy to verify that the hazard rate function of the non-cured population reduces to

$$h_{\mathrm{PBS}}(t) = \frac{\varphi \exp\{-(1/2\alpha^2)(t/\lambda+\lambda/t-2) - \varphi\Phi[(1/\alpha)(\sqrt{t/\lambda}-\sqrt{\lambda/t})]\}t^{-3/2}(t+\lambda)}{2\alpha\sqrt{2\pi\lambda}\{\exp\{-\varphi\Phi[(1/\alpha)(\sqrt{t/\lambda}-\sqrt{\lambda/t})]\} - \exp(-\varphi)\}}, \quad t > 0. \quad (10)$$

Numerical examples allow to induce that the hazard rate function (10) is either increasing or unimodal. In the right panel of Figure 2, we plot the hazard function for selected parameter values. For $\alpha = 2$ and $\lambda = 5$, the hazard rate function is unimodal for large values of φ. When φ decreases, $h_{\mathrm{PBS}}(t)$ increases. On the other hand, for $\alpha = 0.2$, this function is unimodal for all values of φ. Furthermore, $h_{\mathrm{BS}}(t) \leq h_{\mathrm{PBS}}(t)$ and $\lim_{t\to\infty} h_{\mathrm{PBS}}(t) = (2\alpha^2\lambda)^{-1}$ and then the limit behaviour of the PBS hazard ratio function is the same behaviour as that of the BS hazard ratio function.

3. Two useful expansions

Hereafter, let Y be a random variable having the PBS density function (9). We derive two expansions for the density function of Y, which can be useful to obtain some mathematical

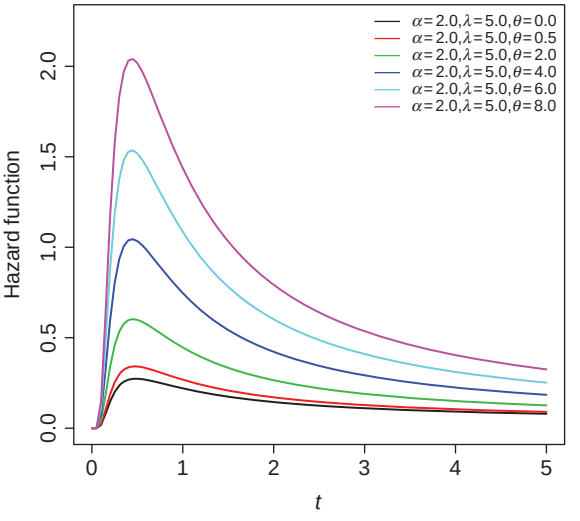

Figure 2. The PBS hazard function. The parameters are fixed: $\varphi = 0, 0.5, 2, 4, 6, 8$ and $\lambda = 5$, $\alpha = 0.2$ (left panel), $\alpha = 2$ (right panel).

measures for the PBS distribution including moments, generating function, mean deviations, Bonferroni and Lorenz curves, reliability, among others. First, the density function of Y can be expressed as

$$f_{\mathrm{PBS}}(t) = \frac{\varphi}{(1 - \mathrm{e}^{-\varphi})} f_{\mathrm{BS}}(t) \exp[-\varphi F_{\mathrm{BS}}(t)]. \tag{11}$$

Setting $F_{\mathrm{BS}}(t) = u$ in Equation (11), we obtain a simple representation given by

$$f_{\mathrm{PBS}}(t)\,\mathrm{d}t = \frac{\varphi}{(1 - \mathrm{e}^{-\varphi})} \mathrm{e}^{-\varphi u}\,\mathrm{d}u. \tag{12}$$

Equation (12) will be useful if a power series expansion for the quantile function of the BS distribution, say $Q_{\mathrm{BS}}(u)$, is available. Cordeiro and Lemonte [24] demonstrated that $t = Q_{\mathrm{BS}}(u) = F_{\mathrm{BS}}^{-1}(u)$ can be expressed as

$$t = Q_{\mathrm{BS}}(u) = \sum_{k=0}^{\infty} m_k \left(u - \frac{1}{2}\right)^k, \tag{13}$$

where $m_k = (2\pi)^{k/2} \sum_{j=0}^{\infty} p_j e_{j,k}$, $p_0 = \lambda$, $p_{2j+1} = \lambda\alpha^{2j+1} 2^{-2j} \binom{\frac{1}{2}}{j}$ for $j \geq 0$, $p_2 = \lambda\alpha^2/2$ and $p_{2j} = 0$ for $j \geq 2$ and the quantities $e_{j,k}$ are determined recursively from

$$e_{j,k} = (k d_0)^{-1} \sum_{m=1}^{k} [(j+1)m - k] d_m e_{j,k-m}$$

and $e_{j,0} = d_0^j$. Here, the d_k's are defined by $d_k = 0$ (for $k = 0, 2, 4, \ldots$) and $d_k = b_{(k-1)/2}$ (for $k = 1, 3, 5, \ldots$), where the b_k's are also calculated (for $k = 0, 1, \ldots$) recursively from

$$b_{k+1} = \frac{1}{2(2k+3)} \sum_{r=0}^{k} \frac{(2r+1)(2k-2r+1) b_r b_{k-r}}{(r+1)(2r+1)}.$$

Thus, the first constants d_k's are: $d_1 = 1$, $d_3 = \frac{1}{6}$, $d_5 = \frac{7}{120}$, $d_7 = \frac{127}{7560}, \ldots$

Combining Equations (12) and (13), the expected value of any function of the PBS random variable Y, say $H(Y)$, can be determined by

$$\mathrm{E[H(Y)]} = \frac{\varphi}{(1-\mathrm{e}^{-\varphi})}\int_0^1 \mathrm{H}\left(\sum_{\mathrm{k}=0}^{\infty}\mathrm{m_k}\left(\mathrm{u}-\frac{1}{2}\right)^{\mathrm{k}}\right)\mathrm{e}^{-\varphi \mathrm{u}}\,\mathrm{du}. \tag{14}$$

By expanding the exponential function in Equation (11), we obtain another expansion for the PBS density function given by

$$f_{\mathrm{PBS}}(t) = \frac{\varphi}{(1-\mathrm{e}^{-\varphi})} f_{\mathrm{BS}}(t)\sum_{j=0}^{\infty}\frac{(-1)^j\varphi^j F_{\mathrm{BS}}(t)^j}{j!}. \tag{15}$$

Equations (12), (14) and (15) are the main results of this section and they provide accurate approximations with at most $r = 10$ terms and $k = 10$, respectively. They can be very useful to determine ordinary and incomplete moments, generating and characteristic functions, mean deviations and several other mathematical quantities for the PBS distribution.

4. Moments and generating function

First proposed by Greenwood et al.,[25] the probability-weighted moments (PWMs) are expectations of certain functions of a random variable whose mean exists. A general theory for PWMs covers the summarization and description of theoretical probability distributions and observed data samples, non-parametric estimation of the underlying distribution of an observed sample, estimation of parameters, quantiles of probability distributions and hypothesis tests. The PWM method can generally be used for estimating parameters of a distribution whose inverse form cannot be expressed explicitly. If p and r are positive integers, the PWMs of the BS distribution are formally defined by

$$\tau_{p,r} = \mathrm{E}[Y^p F_{\mathrm{BS}}(Y)^r] = \int_0^{\infty} t^p F_{\mathrm{BS}}(t)^r f_{\mathrm{BS}}(t)\,\mathrm{d}t.$$

Cordeiro and Lemonte [24] provided a closed-form expression for $\tau_{p,r}$. First, for $k = 0, 1, \ldots$, we define $a_k = (-1)^k 2^{(1-2k)/2}[\sqrt{\pi}(2k+1)k!]^{-1}$, $s_j = k_1 + \cdots + k_j$ and $A(k_1, \ldots, k_j) = \alpha^{-2s_j - j} a_{k_1} \ldots a_{k_j}$ and

$$I(p,\alpha) = \frac{K_{p+1/2}(\alpha^{-2}) + K_{p-1/2}(\alpha^{-2})}{2K_{1/2}(\alpha^{-2})}. \tag{16}$$

Here, the function $K_\nu(z)$ denotes the modified Bessel function of the third kind with ν representing its order and z the argument. Its integral representation is $K_\nu(z) = 0.5\int_{-\infty}^{\infty}\exp\{-z\cosh(t) - \nu t\}\,\mathrm{d}t$. They demonstrated that

$$\tau_{p,r} = \frac{\lambda^p}{2^r}\sum_{j=0}^{r}\binom{r}{j}\sum_{k_1,\ldots,k_j=0}^{\infty}A(k_1,\ldots,k_j)\sum_{m=0}^{2s_j+j}(-1)^m\binom{2s_j+j}{m}$$
$$\times\,\lambda^{-(2s_j+j-2m)/2} I\left(p+\frac{(2s_j+j-2m)}{2},\alpha\right), \tag{17}$$

where $I(p + (2s_j + j - 2m)/2, \alpha)$ can be determined from Equation (16).

The quantities $\tau_{p,r}$ can also be computed numerically from the baseline functions $F_{\mathrm{BS}}(t)$ and $f_{\mathrm{BS}}(t)$ using any statistical software with numerical facilities. However, establishing in-built routines for explicit expressions for $\tau_{p,r}$ can be more efficient than computing the PWMs by numerical integration. It can also be more accurate computationally to use these in-built routines. Other forms such as integral representations can be prone to rounding off errors among others.

Hence, the pth ordinary moment of Y can be expressed from Equation (15) as

$$\mu'_p = \mathrm{E}(Y^p) = \frac{\varphi}{(1-\mathrm{e}^{-\varphi})} \sum_{r=0}^{\infty} \frac{(-1)^r \varphi^r \tau_{p,r}}{r!}, \tag{18}$$

where $\tau_{p,r}$ is determined from Equation (17).

Another expansion for μ'_p can be immediately obtained from Equation (14) as

$$\mu'_p = \frac{\varphi}{(1-\mathrm{e}^{-\varphi})} \int_0^1 \left(\sum_{k=0}^{\infty} m_k \left(u - \frac{1}{2} \right)^k \right)^p \mathrm{e}^{-\varphi u}\, \mathrm{d}u.$$

Using an equation for a power series raised to a positive integer power [26, Section 0.314], we can write

$$\mu'_p = \frac{\varphi}{(1-\mathrm{e}^{-\varphi})} \sum_{k=0}^{\infty} c_{p,k} \int_0^1 \left(u - \frac{1}{2} \right)^k \mathrm{e}^{-\varphi u}\, \mathrm{d}u,$$

where the coefficients $c_{p,k}$ (for $k = 1, 2, \ldots$) can be determined from the recursive equation (with $c_{p,0} = m_0^p$)

$$c_{p,k} = (k m_0)^{-1} \sum_{r=1}^{k} [r(p+1) - k] m_r c_{p,k-r}. \tag{19}$$

The coefficient $c_{p,k}$ follows recursively from $c_{p,0}, \ldots, c_{p,k-1}$ and then from $m_0, \ldots, m_k$. Here, $c_{p,k}$ can be written explicitly in terms of the quantities m_r, although it is not necessary for programming numerically our expansions in any algebraic or numerical software.

By expanding the binomial term in the last expression for μ'_p, we can write

$$\mu'_p = \frac{\varphi}{(1-\mathrm{e}^{-\varphi})} \sum_{k=0}^{\infty} \sum_{r=0}^{k} \left(-\frac{1}{2} \right)^{k-r} \binom{k}{r} c_{p,k} \int_0^1 u^r\, \mathrm{e}^{-\varphi u}\, \mathrm{d}u.$$

The above integral can be expressed in terms of the confluent hypergeometric function [26, Section 9.1] using Maple. We obtain

$$\int_0^1 u^r\, \mathrm{e}^{-\varphi u}\, \mathrm{d}u = \frac{\mathrm{e}^{-\varphi}}{(r+1)} {}_1F_1(1, r+2; \varphi),$$

where

$${}_1F_1(p, q; y) = \sum_{j=0}^{\infty} \frac{(p)_j}{(q)_j} \frac{y^j}{j!}$$

is the confluent hypergeometric function and $(p)_j$ is the Pochhammer symbol defined as $(p)_j = p(p+1)\cdots(p+j-1)$. Hence, an alternative formula for the pth moment of Y is given by

$$\mu'_p = \frac{\varphi\, \mathrm{e}^{-\varphi}}{(1-\mathrm{e}^{-\varphi})} \sum_{k=0}^{\infty} \sum_{r=0}^{k} \frac{(-1/2)^{k-r} \binom{k}{r} c_{p,k}}{(r+1)} {}_1F_1(1, r+2; \varphi). \tag{20}$$

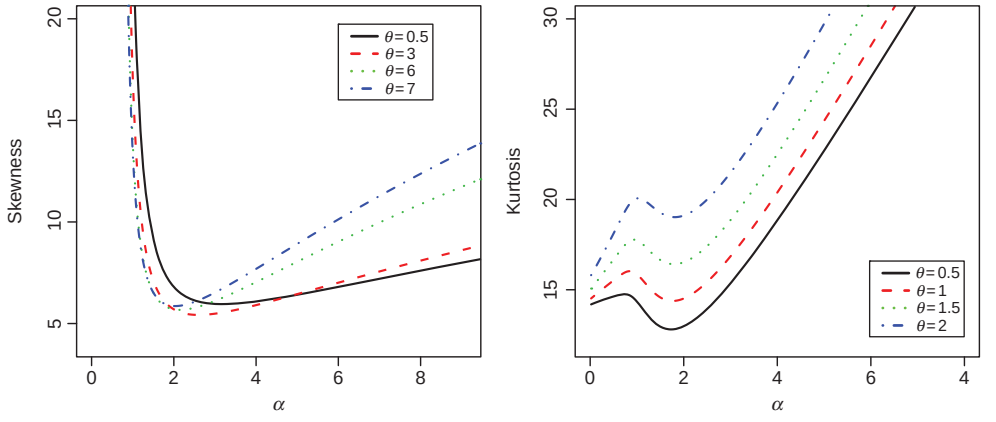

Figure 3. Skewness and kurtosis of the PBS distribution as functions of α for some values of φ. (left panel) For $\lambda = 3.5$. (right panel). For $\lambda = 0.5$.

When φ approaches zero, Equation (20) gives the pth moment of the BS distribution

$$\mu'_p = \sum_{k=0}^{\infty} \sum_{r=0}^{k} \frac{(-1/2)^{k-r} c_{p,k}}{(r+1)} \binom{k}{r}.$$

Equations (18) and (20) are the main results of this section and can provide accurate approximations with at most $r = 10$ terms and $k = 10$, respectively.

The central moments (μ_s) and cumulants (κ_s) of Y can be determined from Equation (18) as

$$\mu_s = \sum_{k=0}^{p} \binom{s}{k} (-1)^k \mu_1'^s \mu'_{s-k} \quad \text{and} \quad \kappa_s = \mu'_s - \sum_{k=1}^{s-1} \binom{s-1}{k-1} \kappa_k \mu'_{s-k},$$

respectively, where $\kappa_1 = \mu'_1$. Thus, $\kappa_2 = \mu'_2 - \mu'^2_1$, $\kappa_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1$, $\kappa_4 = \mu'_4 - 4\mu'_3\mu'_1 - 3\mu'^2_2 + 12\mu'_2\mu'^2_1 - 6\mu'^4_1$, etc.

The skewness and kurtosis measures of Y can be obtained from the third and fourth standardized cumulants given by $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ and $\gamma_2 = \kappa_4/\kappa_2^2$, respectively. Plots of these quantities as functions of α for selected values of φ, by fixing $\lambda = 3.5$ and $\lambda = 0.5$, are displayed in Figure 3. These plots indicate that the skewness decreases and increases when α increases (for fixed φ) and that the kurtosis first increases, decreases and then increases again when α increases (for fixed λ).

The nth descending factorial moment of Y can be also determined from Equation (18) as $\mu'_{(n)} = \mathrm{E}[Y^{(n)}] = \mathrm{E}\{Y(Y-1)(Y-2)\cdots(Y-n+1)\} = \sum_{r=0}^{n} s(n,r)\mu'_r$, where $s(n,r) = (r!)^{-1}[D^r x^{(n)}]_{x=0}$ are the Stirling numbers of the first kind. They count the number of ways to permute a list of n items into r cycles.

The moment-generating function of Y, say $M(s) = \mathrm{E}(e^{sY})$, can follow from Equation (18) or Equation (20) by $M(t) = \sum_{p=0}^{\infty} (\mu'_p s^p)/p!$.

5. Mean residual lifetime

The MRL function, also called expected remaining life function or mean excess function, has been widely studied in lifetime context. It plays an important role in many fields such as biomedical

science, industrial reliability, life insurance and demography, among others. The MRL function $e(y)$ for the PBS distribution is given by

$$e(x) = E(Y - x \mid Y > x) = \frac{1}{S_{\mathrm{PBS}}(x)} \int_x^{\infty} (t - x) f_{\mathrm{PBS}}(t)\, \mathrm{d}t,$$

where $S_{\mathrm{PBS}}(x)$ comes from Equation (8) in terms of the BS survival function $S_{\mathrm{BS}}(x)$ and φ. From Equations (12) and (13), we can write

$$e(x) = \frac{\varphi}{(1 - \mathrm{e}^{-\varphi}) S_{\mathrm{PBS}}(x)} \left[\sum_{k=0}^{\infty} m_k \sum_{j=0}^{k} \left(-\frac{1}{2} \right)^{k-j} \binom{k}{j} T_j(F_{\mathrm{BS}}(x)) - x T_0(F_{\mathrm{BS}}(x)) \right], \tag{21}$$

where $T_j(p) = \int_p^1 u^j \mathrm{e}^{-\varphi u}\, \mathrm{d}u$.

Further, we derive a simple formula for $T_j(p)$ based on the complementary incomplete gamma function defined by $\Gamma(\alpha, a) = \int_a^{\infty} z^{\alpha-1} \mathrm{e}^{-z}\, \mathrm{d}z$. After a very simple algebra, we obtain

$$T_j(p) = \varphi^{-(j+1)} [\Gamma(j + 1, p\varphi) - \Gamma(j + 1, \varphi)]. \tag{22}$$

Equations (21) and (22) give the MRL function for the PBS distribution, i.e. the expected time for the uncured individuals at time x.

6. Inference

Consider the situation where the time to the event is not completely observed and is subjected to right censoring. Let C_i denote the censoring time. We then observe $t_i = \min\{T_i, C_i\}$ and $\delta_i = \mathrm{I}(T_i \leq C_i)$, where $\delta_i = 1$ if T_i is the observed time to the event defined before and $\delta_i = 0$ if it is right censored, for $i = 1, \ldots, n$. Let $\boldsymbol{\gamma} = (\alpha, \lambda)^{\mathrm{T}}$ denote the parameter vector of the distribution function of the time-to-event $F(t)$. From n pairs of times and censoring indicators $(t_1, \delta_1), \ldots, (t_n, \delta_n)$, the observed full likelihood function under non-informative censoring reduces to

$$L(\varphi, \boldsymbol{\gamma}) \propto \prod_{i=1}^{n} [f_{\mathrm{pop}}(t_i; \varphi, \boldsymbol{\gamma})]^{\delta_i} [S_{\mathrm{pop}}(t_i; \varphi, \boldsymbol{\gamma})]^{1-\delta_i}, \tag{23}$$

where $S_{\mathrm{pop}}(t_i)$ and $f_{\mathrm{pop}}(t_i)$ are given in Equations (4) and (5), respectively.

In many medical problems, the lifetimes are affected by explanatory variables such as the cholesterol level, blood pressure, weight and many others. Parametric models to estimate univariate survival functions for censored data regression problems are widely used. The parameter φ in Equation (4) is now linked to a vector $\mathbf{x}_i$ of explanatory variables by $\varphi_i = \exp(\mathbf{x}_i^{\top} \boldsymbol{\beta})$, for $i = 1, \ldots, n$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\top}$ denotes the vector of regression coefficients. Denote by $\boldsymbol{\theta} = (\boldsymbol{\gamma}^{\top}, \boldsymbol{\beta}^{\top})^{\top}$ the vector of model parameters. By substituting Equations (4) and (5) into Equation (23), the log-likelihood function can be reduced to

$$\begin{aligned} l(\boldsymbol{\theta}) = {} & \sum_{i=1}^{n} \delta_i \left\{ \log[f_{\mathrm{BS}}(t_i)] + \mathbf{x}_i^{\top} \boldsymbol{\beta} - \exp(\mathbf{x}_i^{\top} \boldsymbol{\beta}) \Phi \left[\alpha^{-1} \left(\sqrt{\frac{t_i}{\lambda}} - \sqrt{\frac{\lambda}{t_i}} \right) \right] \right\} \\ & - \sum_{i=1}^{n} (1 - \delta_i) \exp(\mathbf{x}_i^{\top} \boldsymbol{\beta}) \Phi \left[\alpha^{-1} \left(\sqrt{\frac{t_i}{\lambda}} - \sqrt{\frac{\lambda}{t_i}} \right) \right]. \end{aligned} \tag{24}$$

The score functions for the parameters are given by

$$U_\alpha(\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \left\{ \frac{[\dot{f}_{\mathrm{BS}}(t_i)]_\alpha}{f_{\mathrm{BS}}(t_i)} + \alpha^{-2} \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) q(\lambda, t_i) \phi[\alpha^{-1} q(\lambda, t_i)] \right\}$$
$$- \alpha^{-2} \sum_{i=1}^{n} (1 - \delta_i) \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) q(\lambda, t_i) \phi[\alpha^{-1} q(\lambda, t_i)]\},$$

$$U_\lambda(\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \left\{ \frac{[\dot{f}_{\mathrm{BS}}(t_i)]_\lambda}{f_{\mathrm{BS}}(t_i)} + \alpha^{-1} \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \left(\frac{t_i}{2\lambda^2}\sqrt{\frac{\lambda}{t_i}} + \frac{1}{2t_i}\sqrt{\frac{t_i}{\lambda}} \right) \phi[\alpha^{-1} q(\lambda, t_i)] \right\}$$
$$- \sum_{i=1}^{n} (1 - \delta_i) \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \phi[\alpha^{-1} q(\lambda, t_i)] \left(\frac{t_i}{2\lambda^2}\sqrt{\frac{\lambda}{t_i}} + \frac{1}{2t_i}\sqrt{\frac{t_i}{\lambda}} \right),$$

$$U_{\beta_j}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \{x_{ij}[1 - \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})] \Phi[\alpha^{-1} q(\lambda, t_i)]\} - \sum_{i=1}^{n} (1 - \delta_i) x_{ij} \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \Phi\{\alpha^{-1} q(\lambda, t_i)\}.$$

Here, $\phi(\cdot)$ is the standard normal density function, $q(\lambda, t_i) = \sqrt{t_i/\lambda} - \sqrt{\lambda/t_i}$, $[\dot{f}_{\mathrm{BS}}(t_i)]_\alpha = \partial f_{\mathrm{BS}}(t_i)/\partial\alpha$, $[\dot{f}_{\mathrm{BS}}(t_i)]_\lambda = \partial f_{\mathrm{BS}}(t_i)/\partial\lambda$ and $j = 1, \ldots, p$. The maximum-likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained by solving the nonlinear equations $U_\alpha(\boldsymbol{\theta}) = 0$, $U_\lambda(\boldsymbol{\theta}) = 0$ and $U_{\beta_j}(\boldsymbol{\theta}) = 0$. These equations cannot be solved analytically and statistical software can be used to solve them numerically. We can use iterative techniques such as a Newton–Raphson-type algorithm to calculate the estimate $\hat{\boldsymbol{\theta}}$. We use the software Ox (MAXBFGS subroutine) [27] to compute $\hat{\boldsymbol{\theta}}$.

The inference procedures for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ can be based on the normal approximation

$$(\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\gamma}}^\top)^\top \sim \mathrm{N}_{(p+2)}\{(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top, -\ddot{\mathbf{L}}^{-1}(\boldsymbol{\theta})\},$$

where $-\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \{\partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\boldsymbol{\theta}^\top\}$ is the $(p+2) \times (p+2)$ observed information matrix

$$\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{L}_{\beta_j\beta_{j'}} & \mathbf{L}_{\beta_j\gamma_k} \\ \mathbf{L}_{\gamma_j\gamma_k} & \mathbf{L}_{\gamma_k\gamma_{k'}} \end{pmatrix},$$

whose sub-matrices are given in the appendix.

7. Bootstrap re-sampling method

The bootstrap re-sampling method was proposed by Efron [23] by treating the observed sample as if it represents the population. From the information obtained from such a sample, B bootstrap samples of a similar size to that of the observed sample are generated, from which it is possible to estimate various characteristics of the population, such as the mean, variance, percentiles and so on. According to the literature, the re-sampling method may be non-parametric or parametric. In this study, the non-parametric bootstrap method is addressed, according to which the distribution function F can be estimated by the empirical distribution $\hat{F}$.

Let $\mathbf{T} = (T_1, \ldots, T_n)$ be an observed random sample and $\hat{F}$ be the empirical distribution of $\mathbf{T}$. Thus, a bootstrap sample $\mathbf{T}^*$ is constructed by re-sampling with replacement n elements of the sample $\mathbf{T}$. For the B bootstrap samples generated, say $T_1^*, \ldots, T_B^*$, the bootstrap replication of the parameter of interest for the bth sample is given by a general function of the type [23]:

$$\hat{\boldsymbol{\theta}}_b^* = s(T_b^*),$$

that is, the value of $\hat{\boldsymbol{\theta}}$ for sample T_b^*, $b = 1, \ldots, B$.

The bootstrap estimator of the standard error [28] is the standard deviation (SD) of these bootstrap samples given by

$$\widehat{\mathrm{EP}}_B = \left[\frac{1}{(B-1)} \sum_{b=1}^{B} (\hat{\theta}_b^* - \bar{\theta}_B)^2 \right]^{1/2},$$

where B is the number of bootstrap samples generated and $\bar{\theta}_B = (1/B) \sum_{b=1}^{B} \hat{\theta}_b^*$. According to Efron and Tibshirani,[28] assuming $B \geq 200$, it is generally sufficient to present good results to determine the bootstrap estimate. However, to achieve greater accuracy, a reasonably high B value must be considered. In this study, we consider $B = 3000$ bootstrap samples. We describe the bias corrected (BC) and accelerated (BCa) methods for constructing approximated confidence intervals based on the bootstrap re-sampling method. For further details on bootstrap intervals, see, for example [28–30].

7.1. *BCa bootstrap interval*

The bootstrap interval based on the BCa method considers that the percentiles used in delimitating the bootstrap confidence intervals depend on the corrections for tendency $\hat{a}$ and acceleration $\hat{z}_0$. The bias correction value $\hat{z}_0$ is generated based on the proportion of estimates of the bootstrap samples that are smaller than the original estimate $\hat{\boldsymbol{\theta}}$. The expression for $\hat{z}_0$ is given by

$$\hat{z}_0 = \Phi^{-1} \left(\frac{\sharp(\hat{\boldsymbol{\theta}}_b^* < \hat{\boldsymbol{\theta}})}{B} \right), \quad b = 1, \ldots, B,$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution, B is the number of generated bootstrap samples, $\hat{\boldsymbol{\theta}}$ is the MLE of the observed sample and $\hat{\boldsymbol{\theta}}_b^*$ is the MLE of the bth bootstrap sample. Let $\hat{\boldsymbol{\theta}}_{(i)}$ be the MLE of the sample without the ith observation. We define

$$\hat{a} = \frac{\sum_{i=1}^{n} [\hat{\boldsymbol{\theta}}_{(\cdot)} - \hat{\boldsymbol{\theta}}_{(i)}]^3}{6\{\sum_{i=1}^{n} [\hat{\boldsymbol{\theta}}_{(\cdot)} - \hat{\boldsymbol{\theta}}_{(i)}]^2\}^{3/2}},$$

where n is the sample size and $\hat{\boldsymbol{\theta}}_{(\cdot)} = \sum_{i=1}^{n} \hat{\boldsymbol{\theta}}_{(i)}/n$.

Hence, the BCa bootstrap interval of coverage $100(1-2\alpha)\%$ becomes

$$[\hat{\boldsymbol{\theta}}^*_{(B\alpha_1)}, \hat{\boldsymbol{\theta}}^*_{(B\alpha_2)}],$$

where

$$\alpha_1 = \Phi \left\{ \hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-1}(\alpha)}{1 - \hat{a}[\hat{z}_0 + \Phi^{-1}(\alpha)]} \right\} \quad \text{and} \quad \alpha_2 = \Phi \left\{ \hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-1}(1-\alpha)}{1 - \hat{a}[\hat{z}_0 + \Phi^{-1}(1-\alpha)]} \right\}.$$

Note that α_1 and α_2 are corrections to the bootstrap percentiles.

8. Simulation study

We conduct a simulation study to evaluate the performance of the parameter estimation procedure for the proposed model. We assume the BS distribution for the time to the event (Z) under the parameter values $\alpha = 2$ and $\lambda = 2$. For $i = 1, \ldots, n$, the number of causes (N_i) of the event of

Table 1. Averages of the MLEs, SDs of the averages and RMSEs of the cured fractions $p_0^{(0)}$ and $p_0^{(1)}$ for simulated data from the BS cure rate model.

n	Average		SD		RMSE	
50	0.248	0.570	0.074	0.116	0.093	0.118
100	0.252	0.571	0.056	0.078	0.082	0.082
200	0.249	0.570	0.041	0.054	0.070	0.059
400	0.249	0.571	0.028	0.037	0.063	0.045
800	0.252	0.574	0.020	0.026	0.063	0.039

interest for the ith individual is generated from the Poisson distribution with parameter $\varphi_i = \exp(\beta_0 + \beta_1 x_i)$. In the simulation study, we have one binary covariate x with values drawn from a Bernoulli distribution with parameter 0.5. We take $\beta_0 = 0.5$ and $\beta_1 = -1$, so that the cured fraction for the two levels of x are $p_0^{(0)} = 0.192$ and $p_0^{(1)} = 0.545$, respectively. The censoring times are sampled from the uniform distribution on the interval $(0, \tau)$, where τ was set in order to control the proportion of censored observations. This proportion is taken on average approximately equal to 50%.

We consider sample sizes $n = 50, 100, 200, 400$ and $n = 800$. For each configuration, we perform 1000 simulations and calculate the averages of the MLEs of the cured fractions ($p_0^{(0)}$ and $p_0^{(1)}$), SDs of these averages and the square root of the mean square errors (RMSEs) of the estimates. The results are listed in Table 1 for the simulated data from the BS cure rate model. We note that the averages of the MLEs are close to the true parameter values and that, as expected, the SDs and RMSEs decrease as the sample size increases.

9. Diagnostic analysis

In order to assess the sensitivity of the MLEs, the global influence and local influence under three perturbation schemes are carried out.

9.1. *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 observation from the data set. The case-deletion model for Equations (4) and (5) is given by

$$S_{\text{pop}}(t_{(i)}) = \exp\left\{-\varphi_{(i)}\Phi\left[\alpha^{-1}\left(\sqrt{\frac{t_{(i)}}{\lambda}} - \sqrt{\frac{\lambda}{t_{(i)}}}\right)\right]\right\}$$

and

$$f_{\text{pop}}(t_{(i)}) = \varphi_{(i)} f_{\text{BS}}(t_{(i)}) \exp\left\{-\varphi_{(i)}\Phi\left[\alpha^{-1}\left(\sqrt{\frac{t_{(i)}}{\lambda}} - \sqrt{\frac{\lambda}{t_{(i)}}}\right)\right]\right\},$$

where $i = 1, \ldots, n$, $\varphi_{(i)} = \exp(\mathbf{x}_{(i)}^\top \boldsymbol{\beta})$ and the quantity with subscript '(i)' means the original quantity with the ith case deleted. For model (5), the log-likelihood function for $\boldsymbol{\theta}$ is denoted by $l_{(i)}(\boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}_{(i)} = (\hat{\boldsymbol{\beta}}_{(i)}^\top, \hat{\boldsymbol{\gamma}}_{(i)}^\top)^\top$ be the MLE of $\boldsymbol{\theta}$ calculated by maximizing $l_{(i)}(\boldsymbol{\theta})$. To assess the influence of the ith case on the MLE $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\gamma}}^\top)^\top$, the basic idea is to compare the difference between $\hat{\boldsymbol{\theta}}_{(i)}$ and $\hat{\boldsymbol{\theta}}$. If deletion of an observation 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 observation is regarded as an

influential observation. A first measure of global influence is defined as the standardized norm of $\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}$ (generalized Cook distance)

$$\mathrm{GD}_i(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^\top [-\ddot{\mathbf{L}}(\boldsymbol{\theta})](\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}).$$

Another alternative is to calculate the statistics $\mathrm{GD}_i(\boldsymbol{\beta})$ or $\mathrm{GD}_i(\boldsymbol{\gamma})$, whose values reveal the impact of the ith case on the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. Another popular measure of the difference between $\hat{\boldsymbol{\theta}}_{(i)}$ and $\hat{\boldsymbol{\theta}}$ is the likelihood distance

$$\mathrm{LD}_i(\boldsymbol{\theta}) = 2\{l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{(i)})\}.$$

Further, we can compute $\hat{\beta}_j - \hat{\beta}_{j(i)} (j = 1, \ldots, p)$ to assess the difference between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{(i)}$. Alternative global influence measures are possible. We can also consider the behaviour of some test statistics, such as the Wald test for explanatory variables or censoring effect, under a case-deletion scheme.

As $\hat{\boldsymbol{\theta}}_{(i)}$ is required for every case, a quite heavy total computational burden may be involved. In this case, the following one-step approximation for $\hat{\boldsymbol{\theta}}_{(i)}$ can be used to reduce the burden

$$\hat{\boldsymbol{\theta}}_{(i)} \cong \hat{\boldsymbol{\theta}} + \ddot{\mathbf{L}}(\hat{\boldsymbol{\theta}})^{-1} \dot{l}_{(i)}(\hat{\boldsymbol{\theta}}),$$

where $\dot{l}_{(i)}(\boldsymbol{\theta}) = \partial l_{(i)}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ is evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

9.2. *Local influence*

Another approach is suggested by Cook,[18] where instead of removing observations, some weights are given to them. Local influence calculation can be carried out for model (4). If likelihood displacement $\mathrm{LD}(\boldsymbol{\omega}) = 2\{l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}$ is used, where $\boldsymbol{\omega}$ is the perturbation vector such that $\boldsymbol{\omega} \in \Omega \subseteq \mathbf{R}^q$, Ω is an open set and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denotes the MLE under the perturbed model, i.e. the MLE based on $l(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})$. The normal curvature for $\boldsymbol{\theta}$ at the direction $\mathbf{d}, \| \mathbf{d} \| = 1$, is given by $C_{\mathbf{d}}(\boldsymbol{\theta}) = 2|\mathbf{d}^\top \boldsymbol{\Delta}^\top [\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1} \boldsymbol{\Delta} \mathbf{d}|$, where $\boldsymbol{\Delta}$ is a $(p+2) \times n$ matrix that depends on the perturbation scheme, whose elements are given by $\Delta_{vi} = \partial^2 l(\boldsymbol{\theta} \mid \boldsymbol{\omega})/\partial \phi_v \partial \omega_i$, for $i = 1, \ldots, n$ and $v = 1, \ldots, p+2$, evaluated at $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega}_0$, where $\boldsymbol{\omega}_0$ is the no perturbation vector.[18] For the PBS regression model with cure rate, the elements of $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ are given in the appendix. We can also calculate normal curvatures $C_{\mathbf{d}}(\boldsymbol{\beta})$ and $C_{\mathbf{d}}(\boldsymbol{\gamma})$ 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 $\mathbf{B} = -\boldsymbol{\Delta}^\top [\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1} \boldsymbol{\Delta}$, and the index plots of $C_{\mathbf{d}_i}(\boldsymbol{\beta})$ and $C_{\mathbf{d}_i}(\boldsymbol{\gamma})$, called the total local influence, where $\mathbf{d}_i$ denotes an $n \times 1$ vector of zeros with one at the ith position. Thus, the curvature at direction $\mathbf{d}_i$ takes the form $C_i = 2|\boldsymbol{\Delta}_i^\top [\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1} \boldsymbol{\Delta}_i|$, where $\boldsymbol{\Delta}_i^\top$ denotes the ith row of $\boldsymbol{\Delta}$. It is usual to point out those cases such that $C_i \geq 2\bar{C}$, where $\bar{C} = (1/n) \sum_{i=1}^n C_i$. Another influence measure for the ith observation is $U_i = \sum_{k=1}^{n_1} \theta_k e_{ki}^2$, where $\{(\theta_k, \mathbf{e}_k) \mid k = 1, \ldots, n\}$ are the eigenvalue–eigenvector pairs of $\mathbf{B}$ with $\theta_1 \geq \cdots \geq \theta_{n_1} \geq \theta_{n_1+1} = \cdots = \theta_n = 0$ and $\{\mathbf{e}_k = (e_{k1}, \ldots, e_{kn})^\top\}$ is the associated orthonormal basis. Zhu and Zhang [31] studied the influence measure u_i systematically under a case-weighted perturbation. Hence, this influence measure expresses local sensitivity to the log-likelihood of the perturbations.

9.3. *Curvature calculations*

For three perturbation schemes, we calculate the matrix

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{vi})_{[(p+2)\times n]} = \left(\frac{\partial^2 l(\boldsymbol{\theta} \mid \boldsymbol{\omega})}{\partial \boldsymbol{\theta}_v \boldsymbol{\omega}_i} \right)_{[(p+2)\times n]},$$

where $v = 1, \ldots, p+2$ and $i = 1, \ldots, n$. We shall consider the model defined in Equations (4) and (5) and its log-likelihood function given by Equation (24).

9.3.1. *Case-weight perturbation*

First, we consider a case-weight perturbation which modifies the weight given to each subject in the log-likelihood. Consider the vector of weights $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$. In this case, the log-likelihood function takes the following form:

$$l(\boldsymbol{\theta} \mid \boldsymbol{\omega}) = \sum_{i=1}^{n} \delta_i \omega_i \left\{ \log[f_{\mathrm{BS}}(t_i)] + \mathbf{x}_i^\top \boldsymbol{\beta} - \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \Phi \left[\alpha^{-1} \left(\sqrt{\frac{t_i}{\lambda}} - \sqrt{\frac{\lambda}{t_i}} \right) \right] \right\}$$
$$- \sum_{i=1}^{n} (1 - \delta_i) \omega_i \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \Phi \left[\alpha^{-1} \left(\sqrt{\frac{t_i}{\lambda}} - \sqrt{\frac{\lambda}{t_i}} \right) \right],$$

where $0 \leq \omega_i \leq 1$ and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. The elements of the matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{\boldsymbol{\beta}}^\top, \boldsymbol{\Delta}_{\boldsymbol{\gamma}}^\top)^\top$ are calculated numerically.

9.3.2. *Response perturbation*

Since t_i values have different variances, they require a scaling of the perturbation vector $\boldsymbol{\omega}$ by an estimator of the SD of t_i. Here, we shall consider that each t_i is perturbed as $t_{iw} = t_i + \omega_i S_t$, where S_t is a scale factor that may be estimated by the SD of t and $\omega_i \in \mathbf{R}$. Then, the perturbed log-likelihood function can be expressed as

$$l(\boldsymbol{\theta} \mid \boldsymbol{\omega}) = \sum_{i=1}^{n} \delta_i \left\{ \log[f_{\mathrm{BS}}(t_i^*)] + \mathbf{x}_i^\top \boldsymbol{\beta} - \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \Phi \left[\alpha^{-1} \left(\sqrt{\frac{t_i^*}{\lambda}} - \sqrt{\frac{\lambda}{t_i^*}} \right) \right] \right\}$$
$$- \sum_{i=1}^{n} (1 - \delta_i) \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \Phi \left[\alpha^{-1} \left(\sqrt{\frac{t_i^*}{\lambda}} - \sqrt{\frac{\lambda}{t_i^*}} \right) \right],$$

where $t_i^* = (t_i + \omega_i S_t)$ and $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$. The elements of the matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{\boldsymbol{\beta}}^\top, \boldsymbol{\Delta}_{\boldsymbol{\gamma}}^\top)^\top$ are calculated numerically.

9.3.3. *Explanatory variable perturbation*

Cook [18] described a general scheme for perturbing the whole design matrix X in linear regression models. Some authors have studied the perturbation of explanatory variables. This perturbation has a more complicated impact on the estimates. The errors-in-variable model considers the errors of the explanatory variables and so the local influence under the perturbation of these variables may be in connection with the model. We take an additive perturbation on a particular continuous explanatory variable, say x_t, by setting $x_{it\omega} = x_{it} + \omega_i S_x$, where S_x is a scaled factor, $\omega_i \in \mathbf{R}$. This perturbation scheme leads to the perturbed log-likelihood function

$$l(\boldsymbol{\theta} \mid \boldsymbol{\omega}) = \sum_{i=1}^{n} \delta_i \left\{ \log[f_{\mathrm{BS}}(t_i)] + \mathbf{x}_i^{*\top} \boldsymbol{\beta} - \exp(\mathbf{x}_i^{*\top} \boldsymbol{\beta}) \Phi \left[\alpha^{-1} \left(\sqrt{\frac{t_i}{\lambda}} - \sqrt{\frac{\lambda}{t_i}} \right) \right] \right\}$$
$$- \sum_{i=1}^{n} (1 - \delta_i) \exp(\mathbf{x}_i^{*\top} \boldsymbol{\beta}) \Phi \left[\alpha^{-1} \left(\sqrt{\frac{t_i}{\lambda}} - \sqrt{\frac{\lambda}{t_i}} \right) \right],$$

where $(\mathbf{x}_i^{*\top}\boldsymbol{\beta}) = \beta_1 x_{i1} + \cdots + \beta_t(x_{it} + \omega_i S_x) + \cdots + \beta_p x_{ip}$ and $\boldsymbol{\omega}_0 = (0, \ldots, 0)^{\mathrm{T}}$. The elements of the matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{\boldsymbol{\beta}}^{\top}, \boldsymbol{\Delta}_{\boldsymbol{\gamma}}^{\top})^{\top}$ are calculated numerically.

10. Application

In this section, we discuss an application of the local influence methodology to a real data set on cancer recurrence. The data are of a study designed to determine if female breast cancer patients, originally classified as lymph node negative by standard light microscopy (SLM), could be more accurately classified by immunohistochemical (IH) examination of their lymph nodes with an anticytokeratin monoclonal antibody cocktail. Identical sections of lymph nodes were sequentially examined by SLM and IH. The data are collected by Sedmak et al. [32] and 45 female breast cancer patients with negative axillary lymph nodes and a minimum 10-year follow-up were selected from the Ohio State University Hospital Cancer Registry.[33]

The following variables are associated with each patient ($i = 1, \ldots, 45$), t_i: observed time (in months) and x_{i1}: IH response (0 = negative, 1 = positive). First, we consider the PBS model (4) with a regressor variable to define the cured fraction by

$$\log\{-\log(p_{0i})\} = \beta_0 + \beta_1 x_{i1}.$$

We use the MAXBFGS subroutine in Ox to obtain the MLEs of the parameters. We report in Table 2 the non-parametric bootstrap estimates and the BCa confidence intervals. At the 5% level, there is no difference of the groups under treatment. We note that $\widehat{\beta}_1$ is negative, which implies that the patients in the group with positive IH response ($x_1 = 1$) should have estimated survival probabilities lower than those probabilities for the patients in the group with negative IH response ($x_1 = 0$).

10.1. *Global influence analysis*

We use Ox to compute the case-deletion measure $\mathrm{GD}_i(\boldsymbol{\theta})$ in Section 9.1. The index plot of this measure is displayed in Figure 4. The cases 1, 37 and 45 are possible influential observations.

10.2. *Local influence analysis*

In this section, we analyse the local influence for the cancer data.

10.2.1. *Case-weight perturbation*

By applying the local influence methodology developed in Section 9, where case-weight perturbation is used, we obtain the value $C_{\mathbf{d}_{\max}}(\boldsymbol{\theta}) = 1.57$ for the maximum curvature. In Figure 5, the

Table 2. MLEs and non-parametric bootstrap estimates for the parameters of the PBScr model fitted to the breast cancer data.

	MLEs			Non-parametric bootstrap estimates		
Parameter	Estimate	SE	p-Value	Estimate	SE	95% CI (BCa)
α	0.545	0.105	–	0.527	0.074	(0.385 ; 0.666)
λ	58.035	8.669	–	58.025	0.294	(57.514 ; 58.531)
β_0	0.560	0.378	0.138	0.568	0.165	(0.261 ; 0.872)
β_1	−1.032	0.137	0.018	−1.043	0.208	(−1.412 ; −0.649)

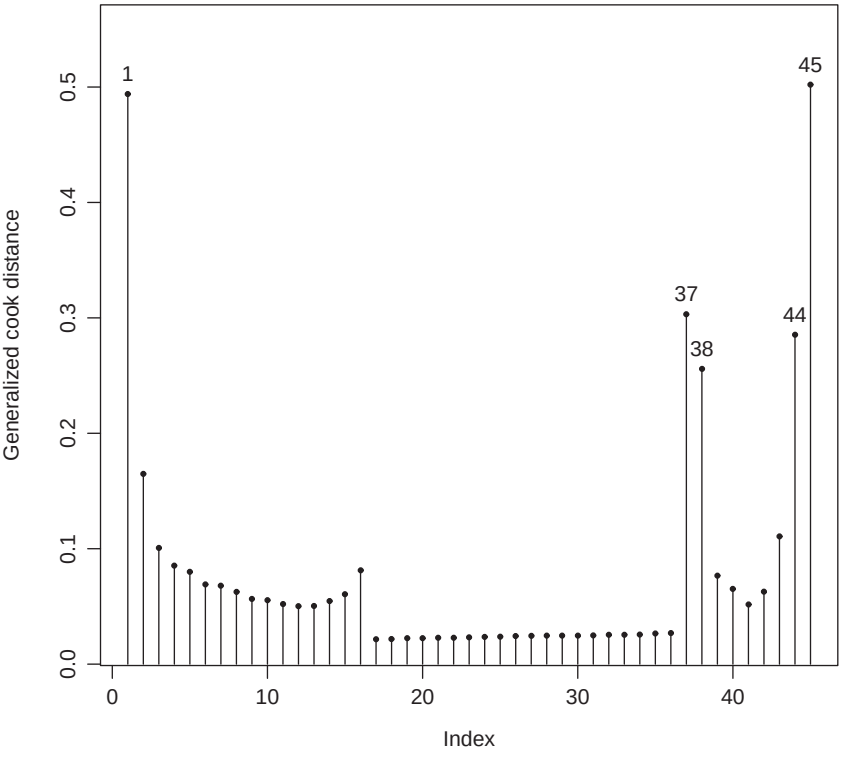

Figure 4. Index plot of $GD_i(\boldsymbol{\theta})$.

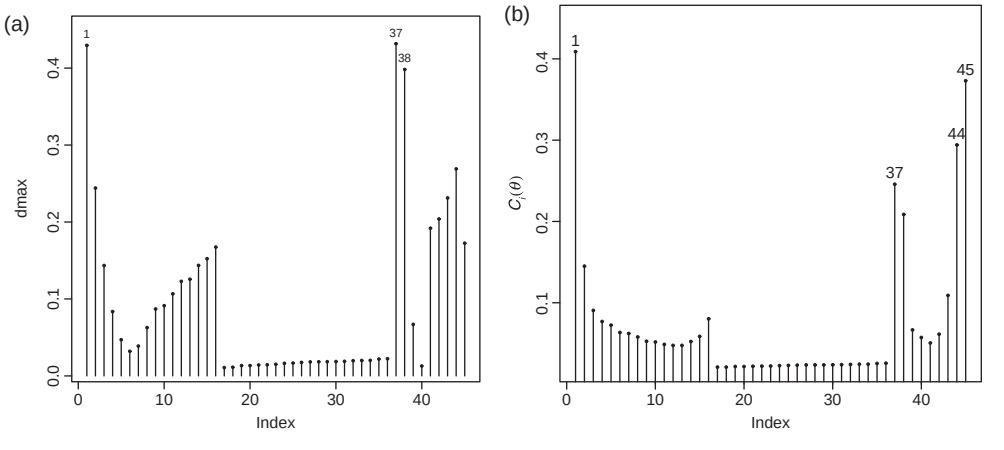

Figure 5. Index plots of $\mathbf{d}_{\max}$ (Figure 5(a)) and $C_i(\boldsymbol{\theta})$ (Figure 5(b)) under case-weight perturbation scheme.

index plots of $\mathbf{d}_{\max}(\boldsymbol{\theta})$ and C_i for all points are presented. Clearly, the most influential observations on $\hat{\boldsymbol{\theta}}$ are 1, 37, 38 and 45.

10.2.2. *Influence using response variable perturbation*

Next, we examine the influence of perturbations on the observed survival times. The value of the maximum curvature is $C_{\mathbf{d}_{\max}}(\boldsymbol{\theta}) = 9.60$. In Figure 6, we display the plots for $\mathbf{d}_{\max}(\boldsymbol{\theta})$ and C_i for

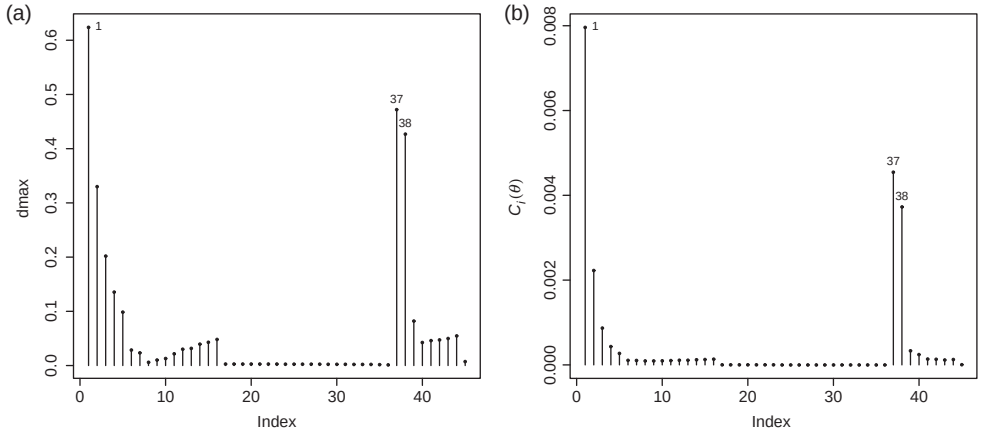

Figure 6. Index plots of $\mathbf{d}_{\max}$ (Figure 6(a)) and $C_i(\boldsymbol{\theta})$ (Figure 6(b)) under the response perturbation scheme.

all points. The plots in Figure 6(a) and 6(b) indicate that the observations 1, 37 and 38 are the most influential cases.

Taking into account the influence analysis results, we note that the MLEs from the PBScr model are not highly sensitive under deletion of the outstanding observations. In general, the significance of the parameter estimates does not change (at the 5% level) after removing the set of observations which is the most influential. Therefore, we do not have inferential changes after removing the observations handed out in the diagnostic plots.

We estimate the proportion of cured individuals using the invariance property of the MLEs, that is,

$$\hat{p}_{0i} = \exp[-\exp(0.560 - 1.032x_{i1})] \quad \text{and} \quad \hat{p}_0 = \frac{\sum_{i=1}^{45} \hat{p}_{0i}}{45} = 0.463.$$

Thus, the estimated mean cure fraction is $\hat{p}_0 = 0.463$. The estimated survival function and the Kaplan–Meier estimate are given in Figure 7(a). The model provides a good fit for the cured fraction. Figure 7(b) displays the estimated survival function for the non-cured population stratified by IH response.

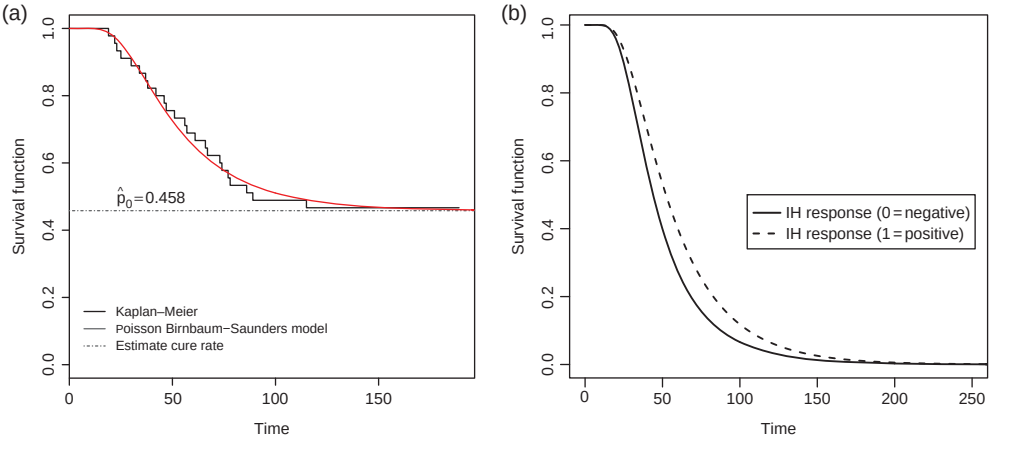

Figure 7. (a) Curve Kaplan–Meier and estimated survival function for the PBScr model. (b) Estimated survival function for the non-cured population (PBS model) for the breast cancer data set.

Finally, we turn out our attention to the role of the covariates on the cured rate, p_0. The estimates of the cure rate patients stratified by IH response category are $\hat{p}_{0j} = \exp[-\exp(\hat{\beta}_0 + j\hat{\beta}_1)]$ for $j = 0, 1$. Estimates of the surviving fraction of patients stratified by IH from 0 to 1 (and standard errors) are 0.174(0.0148) and 0.535(0.088), respectively. This result indicates that the covariate (IH) has a significant effect on the increase of the cured fraction.

11. Concluding remarks

We propose the PBScr model for fitting survival data with cure fraction. The PBScr model is conceived inside a latent competing causes scenario with cure fraction. Further, we introduce a tree-parameter continuous model called the PBS distribution, which extends the BS distribution. We provide a mathematical treatment of the new distribution including the density function and explicit expressions for the moments. We use the quasi-Newton method to obtain the maximum-likelihood estimates and perform asymptotic tests for the parameters. We provide applications of influence diagnostics (global, local and total influence) for the PBScr model. The required matrices for the applications of the techniques are obtained by taking into account some usual perturbations in the model/data. Finally, the PBScr model can be an interesting option to explain/predict survival time for long-term individuals.

References

[1] Birnbaum ZW, Saunders SC. A new family of life distributions. J Appl Probab. 1969;6:319–327.
[2] Johnson N, Kotz S, Balakrishnan N. Continuous univariate distributions. Vol. 2. New York: John Wiley; 1995.
[3] Balakrishnan N, Leiva V, Lopez J. Acceptance sampling plans from truncated life tests based on the generalized Birnbaum–Saunders distribution. Comm Statist Simulation Comput. 2007;36:643–656.
[4] Leiva V, Barros M, Paula GA, Sanhueza A. Generalized Birnbaum–Saunders distributions applied to air pollutant concentration. Environmetrics. 2008a;19:235–250.
[5] Leiva V, Riquelme M, Balakrishnan N, Sanhueza A. Lifetime analysis based on the generalized Birnbaum–Saunders distribution. Comput Statist Data Anal. 2008b;52:2079–2097.
[6] Cancho VG, Ortega E, Paula GA. On estimation and influence diagnostics for log-Birnbaum–Saunders Student-*t* regression models: full Bayesian analysis. J Statist Plann Inference. 2010;140:2486–2496.
[7] Díaz-García JA, Leiva-Sánchez V. A new family of life distributions based on the elliptically contoured distributions. J Statist Plann Inference. 2005;128:445–457.
[8] Sanhueza A, Leiva V, Balakrishnan N. The generalized Birnbaum–Saunders distribution and its theory, methodology, and application. Comm Statist Theory Methods. 2008;37:645–670.
[9] Gómez HW, Olivares-Pacheco JF, Bolfarine H. An extension of the generalized Birnbaum–Saunders distribution. Statist Probab Lett. 2009;79:331–338.
[10] Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. J R Stat Soc Ser B. 1949;11:15–53.
[11] Berkson J, Gage RP. Survival cure for cancer patients following treatment. J Amer Statist Assoc. 1952;47:501–515.
[12] Yakovlev AY, Tsodikov AD. Stochastic models of tumor latency and their biostatistical applications. Singapore: World Scientific; 1996.
[13] Ibrahim JG, Chen M-H, Sinha D. Bayesian survival analysis. New York: Springer; 2001.
[14] Tsodikov AD, Ibrahim JG, Yakovlev AY. Estimating cure rates from survival data: an alternative to two-component mixture models. J Amer Statist Assoc. 2003;98:1063–1078.
[15] Cooner F, Banerjee S, Carlin BP, Sinha D. Flexible cure rate modeling under latent activation schemes. J Amer Statist Assoc. 2007;102:560–572.
[16] Ortega EMM, Cancho VG, Paula GA. Generalized log-gamma regression models with cure fraction. Lifetime Data Anal. 2009;15:79–106.
[17] Cancho VG, Edwin MO, Bolfarine H. The log-exponentiated-Weibull regression models with cure rate: local influence and residual analysis. J Data Sci. 2009;7:433–458.
[18] Cook RD. Assessment of local influence. J Roy Statist Soc Ser B. 1986;48:133–169.
[19] Ortega EMM., Bolfarine H, Paula GA. Influence diagnostics in generalized log-gamma regression models. Comput Statist Data Analy. 2003;42:165–186.
[20] Silva GO, Ortega EMM, Cancho VG, Barreto ML. Log-Burr XII regression models with censored data. Comput Statist Data Anal. 2008;52:3820–3842.

[21] Fachini JB, Ortega EMM, Louzada-Neto F. Influence diagnostics for polyhazard models in the presence of covariates. Stat Methods Appl. 2008;17:413–433.
[22] Hashimoto EM, Ortega EMM, Cancho VG, Cordeiro GM. The log-exponentiated Weibull regression model for interval-censored data. Comput Statist Data Anal. 2010;54:1017–1035.
[23] Efron B. Bootstrap methods: another look at the jackknife. Ann Statist. 1979;7:1–26.
[24] Cordeiro GM, Lemonte AJ. The β-Birnbaum–Saunders distribution: an improved distribution for fatigue life modeling. Comput Statist Data Anal. 2011;55:1445–1461.
[25] Greenwood JA, Landwehr JM, Matalas NC, Wallis JR. Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. Water Resour Res. 1979;15:1049–1054.
[26] Gradshteyn IS, Ryzhik IM. Table of integrals, series, and products. 6th ed. New York: Academic Press; 2000.
[27] Doornik JA. Object-oriented matrix programming using ox. London: Timberlake Consultants Press and Oxford; 2007.
[28] Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman & Hill; 1993.
[29] DiCiccio T, Efron B. Bootstrap confidence intervals. Statist Sci. 1996;11:189–212.
[30] Davison A, Hinkley D. Bootstrap methods and their application. Cambridge: Cambridge University Press; 1997.
[31] Zhu H, Zhang H. A diagnostic procedure based on local influence. Biometrika. 2004;91:579–589.
[32] Sedmak DD, Meineke TA, Knechtges DS, Anderson J. Prognostic significance of cytokeratin-positive breast cancer metastases. Mod Pathol. 1989;2:516–520.
[33] Klein JP, Moeschberger ML. Survival analysis techniques for censored and truncated data. New York: Springer; 2003.

Appendix

The elements of the observed information matrix $L(\boldsymbol{\theta})$ for the parameters $(\alpha, \lambda, \beta_j)$ are:

$$\mathbf{L}_{\alpha,\alpha} = \sum_{i=1}^{n} \delta_i \left\{ \frac{[\ddot{f}_{\mathrm{BS}}(t_i)]_{\alpha\alpha} f_{\mathrm{BS}}(t_i) - [\dot{f}_{\mathrm{BS}}(t_i)]_{\alpha}^2}{[f_{\mathrm{BS}}(t_i)]^2} \right\}$$
$$+ \sum_{i=1}^{n} \delta_i \{\alpha^{-3} \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) q(\lambda, t_i) \phi[\alpha^{-1} q(\lambda, t_i)][-2 + \alpha^{-2} q^2(\lambda, t_i)]\}$$
$$- \alpha^{-3} \sum_{i=1}^{n} (1 - \delta_i) \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) q(\lambda, t_i) \phi\{\alpha^{-1} q(\lambda, t_i)\}[-2 + \alpha^{-2} q^2(\lambda, t_i)],$$

$$\mathbf{L}_{\alpha,\lambda} = \sum_{i=1}^{n} \delta_i \left\{ \frac{[\ddot{f}_{\mathrm{BS}}(t_i)]_{\alpha\lambda} f_{\mathrm{BS}}(t_i) - [\dot{f}_{\mathrm{BS}}(t_i)]_{\alpha} [\dot{f}_{\mathrm{BS}}(t_i)]_{\lambda}}{[f_{\mathrm{BS}}(t_i)]^2} \right\}$$
$$+ \alpha^{-2} \sum_{i=1}^{n} \delta_i \left\{ \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \left(\frac{t_i}{2\lambda^2} \sqrt{\frac{\lambda}{t_i}} + \frac{1}{2t_i} \sqrt{\frac{t_i}{\lambda}} \right) \phi[\alpha^{-1} q(\lambda, t_i)]\{\alpha^{-1} q^2(\lambda, t_i) - 1\} \right\}$$
$$+ \alpha^{-2} \sum_{i=1}^{n} (1 - \delta_i) \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \phi\{\alpha^{-1} q(\lambda, t_i)\} \left(\frac{t_i}{2\lambda^2} \sqrt{\frac{\lambda}{t_i}} + \frac{1}{2t_i} \sqrt{\frac{t_i}{\lambda}} \right) [1 - \alpha^{-2} q(\lambda, t_i)],$$

$$\mathbf{L}_{\alpha,\beta_j} = \alpha^{-2} \sum_{i=1}^{n} \delta_i \{x_{ij} \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) q(\lambda, t_i) \phi[\alpha^{-1} q(\lambda, t_i)]\}$$
$$- \alpha^{-2} \sum_{i=1}^{n} (1 - \delta_i) x_{ij} \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) q(\lambda, t_i) \phi[\alpha^{-1} q(\lambda, t_i)],$$

$$\mathbf{L}_{\lambda,\lambda} = \sum_{i=1}^{n} \delta_i \left\{ \frac{[\ddot{f}_{\mathrm{BS}}(t_i)]_{\lambda\lambda} f_{\mathrm{BS}}(t_i) - [\dot{f}_{\mathrm{BS}}(t_i)]_{\lambda}^2}{[f_{\mathrm{BS}}(t_i)]^2} \right\} + \alpha^{-1} \sum_{i=1}^{n} \delta_i \left\{ \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \phi[\alpha^{-1} q(\lambda, t_i)] \right.$$

$$\times \left[-\frac{t_i}{\lambda^3}\sqrt{\frac{\lambda}{t_i}} + \frac{1}{4\lambda^2}q(\lambda, t_i) + \alpha^{-1}q(\lambda, t_i)\left(\frac{t_i}{2\lambda^2}\sqrt{\frac{\lambda}{t_i}} + \frac{1}{2t_i}\sqrt{\frac{t_i}{\lambda}}\right)^2 \right] \Bigg\}$$

$$+ \sum_{i=1}^{n}(1-\delta_i)\exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\phi[\alpha^{-1}q(\lambda, t_i)]$$

$$\times \left\{ -\alpha^{-2}q(\lambda, t_i)\left[\frac{t_i}{2\lambda^2}\sqrt{\frac{\lambda}{t_i}} + \frac{1}{2t_i}\sqrt{\frac{t_i}{\lambda}}\right]^2 + \left(\frac{t_i}{\lambda^3} + \frac{1}{4\lambda^2}\right)\sqrt{\frac{\lambda}{t_i}} - \frac{1}{2\lambda t_i}\sqrt{\frac{t_i}{\lambda}} \right\},$$

$$\mathbf{L}_{\lambda,\beta_j} = \alpha^{-1}\sum_{i=1}^{n}\delta_i \left\{ x_{ij}\exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\left(\frac{t_i}{2\lambda^2}\sqrt{\frac{\lambda}{t_i}} + \frac{1}{2t_i}\sqrt{\frac{t_i}{\lambda}}\right)\phi[\alpha^{-1}q(\lambda, t_i)] \right\}$$

$$- \sum_{i=1}^{n}(1-\delta_i)x_{ij}\exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\phi[\alpha^{-1}q(\lambda, t_i)]\left[\frac{t_i}{2\lambda^2}\sqrt{\frac{\lambda}{t_i}} + \frac{1}{2t_i}\sqrt{\frac{t_i}{\lambda}}\right],$$

$$\mathbf{L}_{\beta_j,\beta_s} = -\sum_{i=1}^{n}\delta_i\{x_{ij}x_{is}\exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\Phi[\alpha^{-1}q(\lambda, t_i)]\}$$

$$- \sum_{i=1}^{n}(1-\delta_i)x_{ij}x_{is}\exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\Phi\{\alpha^{-1}q(\lambda, t_i)\},$$

where $\phi(\cdot)$ is the standard normal density function, $q(\lambda, t_i) = \sqrt{t_i/\lambda} - \sqrt{\lambda/t_i}$, $[\ddot{f}_{\mathrm{BS}}(t_i)]_{\alpha\alpha} = \partial^2 f_{\mathrm{BS}}(t_i)/\partial\alpha^2$, $[\ddot{f}_{\mathrm{BS}}(t_i)]_{\alpha\lambda} = \partial^2 f_{\mathrm{BS}}(t_i)/\partial\alpha\,\partial\lambda$, $[\ddot{f}_{\mathrm{BS}}(t_i)]_{\lambda\lambda} = \partial^2 f_{\mathrm{BS}}(t_i)/\partial\lambda^2$, and $j, s = 1, \ldots, p$.