The generalized exponential-power series cure rate model with covariates and latent activation schemes

José Flores Delgado[a,c] Vicente Garibay Cancho [b] and Jorge R. Chávez Fuentes[a]

[a]Departamento de Ciencias, Pontificia Universidad Católica Del Perú,
Apartado 1761, Lima, Perú

[b]Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo,
Avenida Trabalhador São-carlense, 400, 13566-590. São Carlos-SP, Brazil

[c]Programa de Pos-Graduação de Ciências de Computação e Matemática Computacional,
Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo,
Avenida Trabalhador São-carlense, 400, 13566-590. São Carlos-SP, Brazil

Abstract

This report presents a new family of cure rate models derived from a scenario with latent risk factors under different activation schemes that would lead to the occurrence of the event of interest. It is considered that the times of activation of the latent factors follow a generalized exponential distribution, while the number of this factors follows a power series series distribution. These models incorporate explanatory variables related with the cure rate through logistic regression. A sensitivity analysis including diagnostic measures based on case-deletion approaches and local influence is performed. To illustrate the results of this paper an example with real data is presented.

1 Introduction

The mixture cure rate model was proposed by Boag (1949) to considerer the possibility that a population may be have: immune (or cured) and susceptible (or no cured) individuals in regard to some event of interest. In this model the survival function, $S_p(t)$, for the entire population of individuals is given by

$$S_p(t) = p_0 + (1 - p_0)S(t), \tag{1}$$

where $S(t)$ is the survival function of the susceptible individuals and $p_0 = S_p(\infty)$ is the fraction of the immune individuals (cured rate). This model has been studied by several authors, for example, Farewell (1982) assumes that the cure rate for the ith individual in a sample of size n is given by a logistic regression between p_0 and the covariates as follows

$$p_{0i} = \frac{\exp(\boldsymbol{x}_i{}^\top \boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_i{}^\top \boldsymbol{\beta})}\,, \tag{2}$$

where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})^\top$ is a covariate vector and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^\top$ is a vector parameter.

For the survival function $S(t)$ the exponential, the gamma, and the Weibull distributions are commonly used.

Alternative cure rate models, known in the literature as The Bounded Cumulative Hazard Model or The Promotion Time Cure Model, were proposed by Yakovlev *et al.* (1993), Yakovlev & Tsodikov (1996), Tsodikov (1998) and Chen *et al.* (1999). These models introduced the well known concept of promotion or activation to explain the time to-event.

Extensions of the promotion model have been proposed in the literature, see e.g., Cancho & Bolfarine (2001), Rodrigues *et al.* (2009), Hashimoto *et al.* (2012) and Ortega *et al.* (2009).

The promotion time cure model was generalized by Cooner *et al.* (2007), developing a theory where the event of interest occurs due to latent risk factors according to different activation schemes. For example, the first activation scheme, that is, the event of interest occurs when the first risk factor is activated, and the last activation scheme occurs when all risk factors have been activated. The models presented in this work consider the first and last activation schemes and assume the logistic regression given in (2), for the cure rate, and the generalized exponential distribution, for the survival function $S(t)$.

The remaining of this paper is organized as follows. In Section 2 the theory of models with latent risk factors under activation schemes, introduced by Cooner *et al.* (2007), is described and new results based on it are given. Under the first and last activation schemes, two new cure rate models with covariates are derived in Section 3. These new models also can be derived from the approach developed by Cancho *et al.* (2011a) and will be called generalized exponential-power series cure rate models (EG-SP). Inference methods based on the likelihood is described in Section 3. In Section 4 a sensitivity analysis of the model is given. To illustrate the results of this paper an example with real data is presented in Section 6. Finally some conclusions are provided in Section 7.

2 Models with latent risk factors

The models with latent risk factors according to activation schemes introduced by Cooner *et al.* (2007) are defined by the following assumptions

(s_1) Let M be a discrete random variable taking values in the set of the natural numbers and probability function f_M such that $P(M = 0) < 1$. M represents the number of latent risk factors that lead to the occurrence of the event of interest. Let $p_0 = P(M = 0)$ denote the probability of absence of risk factors.

(s_2) Let K be a discrete random variable taking values in the set of the natural numbers and

such that $P(1 \leq K \leq M | M = m) = 1$, for $m = 1, 2, \ldots$ This variable represents the minimum number of factors that must be activated for the event to occur.

(s_3) Let $\{ Z_1, Z_2 \ldots \}$ be a sequence of independent, identically distributed, continuous random variables, independent of M, with common cumulative probability function F_a and survival function S_a. These random variables represent the times of activation of the latent factors.

(s_4) The time to-event of interest is defined by the extended random variable

$$T = \begin{cases} \infty, & \text{if } M = 0, \\ Z_{(K)}, & \text{if } M \geq 1, \end{cases} \tag{3}$$

where $Z_{(1)} \leq \cdots \leq Z_{(K)} \leq \cdots \leq Z_{(M)}$ are the ordered statistics of the variables $Z_1, \ldots, Z_M$. Thus when there aren't latent risk factors T is equal to infinity, otherwise it is equal to the time until K out of M risk factors are activated.

An expression for the survival function of T, S_p, is given in the following theorem.

Theorem 1. *Under the assumptions (s_1)-(s_4) the survival function of T is given by*

$$S_p(t) \;\; = \;\; p_0 + (1 - p_0) S^*(t), \;\; \textit{para } t > 0, \tag{4}$$

where

$$S^*(t) = \frac{1}{1 - p_0} \sum_{m=1}^{\infty} \sum_{k=1}^{m} \sum_{j=m-k+1}^{m} \binom{m}{j} S_a^j(t) [1 - S_a(t)]^{m-j} P(K = k | M = m) f_{_M}(m). \tag{5}$$

Furthermore the series in (5) *converges uniformly. Also S^* satisfies the usual properties of a survival function associated to a positive continuous random variable, that is,*

i) S^* is a continuous function,

ii) S^* is a decreasing function,

iii) $\lim_{t \to 0^+} S^*(t) = 1$ and $\lim_{t \to \infty} S^*(t) = 0$.

These properties are satisfied by S_p, except that in iii) $\lim_{t \to \infty} S(t) = p_0$.

Notice that S^* corresponds to the survival function of the susceptible individuals. Equation (5) is equivalent to equation (2) in Cooner *et al.* (2007). It is worth to mention that this equivalent form and the uniform convergence facilitate the study of the properties of $S(t)$ as well as it helps to generate new survival functions.

Proof. Since $S_p(t) = P(T > t)$, then

$$S_p(t) = P(T > t, M = 0) + P(T > t, M \geq 1) = P(M = 0) + P(T > t, M \geq 1). \tag{6}$$

On the other hand,

$$\begin{aligned} P(T > t, M \geq 1) &= P(Z_{(K)} > t, M \geq 1) \\ &= \sum_{m=1}^{\infty} \sum_{k=1}^{m} P(Z_{(k)} > t) P(K = k | M = m) P(M = m). \end{aligned} \tag{7}$$

By applying the known result of the order statistics

$$P\left(Z_{(k)} > t\right) = \sum_{j=m-k+1}^{m} \binom{m}{j} S_a^j(t)$$

in the last equation yields

$$P(T > t, M \geq 1) = \sum_{m=1}^{\infty} \sum_{k=1}^{m} \sum_{j=m-k+1}^{m} \binom{m}{j} S_a^j(t)[1 - S_a(t)]^{m-j} P(K = k | M = m) f_{M}(m). \tag{8}$$

Then, (5) follows from (6) and (8).

The uniform convergence in (5) follows from the Weierstrass criterion by observing that $|S^*(t)| \leq f_{M}(m)$ for all $t > 0$, where $f_{M}(m)$ is a probability function. Properties *i)* and *iii)* follow from the uniform convergence and the properties of the survival function S_{Z}. Property *ii)* follows immediately by observing that $S^*(t) = \frac{1}{1-p_0} P(T > t, M \geq 1)$. □

Del teorema anterior conviene dar las observaciones siguientes, de acuerdo con Cooner *et al.* (2007)

a) If $p_0 = 0$ a non cure rate model is obtained, where the survival function is S^* given by (5).

b) When $p_0 > 0$, the model with latent risk factors becomes a mixture cure rate model, where the cure rate is $p_0 = P(M = 0)$, and the survival function of the susceptible individuals is S^* given by (5).

c) The mixture cure rate model given in (1) becomes a model with laten risk factors when M follows a Bernoulli distribution with $P(M = 1) = 1 - p_0$ and $S_a(t) = S(t)$.

d) The promotion time cure rate model becomes a model with laten risk factors when $K = 1$, M follows a Poisson distribution and the times of activation follows an exponential distribution.

Thus, the model with latent risk factors by Cooner *et al.* (2007) describes how the time to-event is generated. The mixture cure rate and the promotion time models are particular cases of this model. Notice that this model also provides a general way to obtain survival distributions by specifying the distribution of M and K. When $P(M = 0) > 0$ a non-cure rate model is obtained.

Corollary 1 establishes that when the distribution of M is truncated in zero the survival function obtained is the one that corresponds to the susceptible individuals. This result is not given in Cooner *et al.* (2007).

Corollary 1. *Let $S_p(t) = p_0 + (1 - p_0)S^*(t)$ be the cure rate model obtained under assumptions of Theorem 1. When the distribution $f_{_M}$ is truncated in zero the resulting survival function is $S^*(t)$.*

Proof. Since the distribution of $f_{_M}$ truncated in zero is $\frac{1}{1-p_0} f_{_M}(m)$, the claim follows from 5. □

In subsections 2.1 and 2.2 examples are given to illustrate this result.

The remaining of this section is dedicated to show three specifications for the distribution of K called by Cooner *et al.* (2007) as activation schemes.

2.1 The first-activation scheme

In this case $K = 1$ that is, the first activation makes the event of interest to appears. Thus the time to-event, the survival function of the susceptible individuals, $S^*_F(t)$, and the survival function of the entire population, $S_{pF}(t)$, where F denotes the first activation, become

$$T = \begin{cases} \infty, & \text{if } M = 0, \\ Z_{(1)} = min\{ Z_1, \ldots, Z_M \}, & \text{if } M \geq 1, \end{cases} \tag{9}$$

$$S^*_F(t) = \frac{1}{1 - p_0} \sum_{m=1}^{\infty} S^m_a(t) f_{_M}(m) \tag{10}$$

and

$$S_{pF}(t) = \sum_{m=0}^{\infty} S^m_a(t) f_{_M}(m) = g_{_M}(S_a(t)), \tag{11}$$

where $g_{_M}$ is the probability generating function of M.

The promotion model is obtained from this scheme when M follows a Poisson distribution. Recently new cure rate models of this activation scheme have appeared in the literature. For instance in Cancho *et al.* (2011a) and Rodrigues *et al.* (2009) the Weibull distribution is considered for the time of activation and for the distribution of M the power series is considered in Cancho *et al.* (2011a), while Rodrigues *et al.* (2009) considers the COM-Poisson.

On the other hand, non-cure rate models obtained without using the theory of latent risk factors can also be considered as particular cases of the first activation scheme by taking for the number of risk factor, the geometric, the Poisson, the logarithmic and the power series distributions truncated in zero (see e.g., Adamidis & Loukas (1998), Kus (2007), Tahmasbi & Rezaei (2008) and Chahkandi & Ganjali (2009), Morais & Barreto-Souza (2011)).

2.2 The last-activation scheme

In this case $K = M$; that is, the last activation leads to observed the event of interest. Thus the time to-event, the survival function of the susceptible individuals, $S_L^*(t)$, and the survival function of the entire population, $S_{pL}(t)$, where L denotes the last activation, become

$$T = \begin{cases} \infty, & \text{if } M = 0, \\ Z_{(M)} = max\{ Z_1, \ldots, Z_M \}, & \text{if } M \geq 1, \end{cases} \tag{12}$$

$$S_L^*(t) = 1 - \frac{1}{1 - p_0} \sum_{m=1}^{\infty} F_a^m(t) f_{_M}(m). \tag{13}$$

and

$$S_{pL}(t) = 1 + P(M = 0) - g_{_M}(F_a(t)). \tag{14}$$

Recently new cure rate models of this activation scheme have appeared in the literature. For instance in Cancho *et al.* (2011a) the Weibull distribution is considered for the times of activation and the power series distribution for M. On the other hand, non-cure rate models obtained without using the theory of latent risk factors can also be considered as particular cases of the first activation scheme by taking the exponential and the Weibull distribution for the times of activation and for the number of risk factor, the geometric, the Poisson, and the power series distributions truncated in zero (see e.g., Adamidis *et al.* (2005), Cancho *et al.* (2011b) and Flores *et al.* (2011)).

2.3 The uniform random-scheme

In this case, the conditional distribution of K given $M = m$ is uniform on $\{1, \ldots, m\}$, that is, the kth activation leads to observe the event of interest with probability $1/m$ for $k = 1, \ldots, m$. Thus the survival function of the susceptible individuals, $S_R^*(t)$, and the survival function of the entire population, $S_{pR}(t)$, where R denotes the uniform random activation, become

$$S_R^*(t) = S_a(t) \tag{15}$$

and

$$S_{pR}(t) = p_0 + (1 - p_0)S_a(t), \tag{16}$$

Notice that, in this case, the distribution of M only affects the determination of the cure rate $p_0 = P(M = 0)$.

Corollary 2 establishes the stochastic order between the survival functions of these activation scheme. This result is not given in Cooner *et al.* (2007).

Corollary 2. *Inequalities* (17) *and* (18) *hold for every distribution of* M

$$S_F^*(t) \leq S_a(t), \quad S_a(t) \leq S_L^*(t) \tag{17}$$

and

$$S_{pF}^*(t) \leq S_a(t), \quad S_a(t) \leq S_{pL}^*(t). \tag{18}$$

Proof. The inequalities given in (17) follow from (10) and (13) and (18) follows immediately from (17). □

A similar result is given in Kim *et al.* (2011), where K and M are independents and the time to-event T is equal to infinity, if $M < K$, and is $Z_{(K)}$ if $M \geq K$.

3 Derivation of the model

Assume that in the model with latent risk factors, developed in Section 2, M follows a power series distribution, that is, its probability function is given by

$$f_{_M}(m) = P(M = m) = \frac{a_m \theta^m}{C(\theta)},\ m = 0,\ 1,\ 2,\ \ldots\ ;\ \theta \in \Theta, \tag{19}$$

where $a_0,\ a_1,\ a_2,\ \ldots$ is a sequence of non-negative real numbers, where at leats one of them is strictly positive, $\Theta = (0,\ s)$, where s is a positive number no greater than the ratio of convergence of the power series $\sum_{m=0}^{\infty} a_m \theta^m$, and $C(\theta) = \sum_{m=0}^{\infty} a_m \theta^m,\ \forall\, \theta \in (0,\ s)$. Notice, in particular, that C is positive and infinitely many differentiable. The cure rate is given by $p_0 = P(M = 0) = \frac{a_0}{C(\theta)}$ and the probability generating function is $g(s) = \frac{C(\theta\, s)}{C(\theta)},$ if $s\theta \in \Theta$. The Poisson, geometric and logarithmic distributions, among others, are particular cases of this distribution. These cases will be considered in this paper and are showed in Table 1.

Table 1: Particular cases of the power series distribution

Distribution	$f_M(m)$	a_m	$C(\theta)$	Θ	p_0
Poisson	$\frac{e^{-\theta}\theta^m}{m!}, m = 0, 1, 2, \ldots$	$\frac{1}{m!}$	e^{θ}	$(0,\ \infty)$	$e^{-\theta}$
Geometric	$\theta^m(1-\theta)$, $m = 0,\ 1,\ 2, \ldots$	1	$(1-\theta)^{-1}$	$(0,\ 1)$	$1-\theta$
Logarithmic	$\frac{\theta^{m+1}}{-(m+1)\,log(1-\theta)}$, $m = 0,\ 1, 2, \ldots$	$\frac{1}{m}$	$-log(1-\theta)$	$(0,\ 1)$	$-\theta/log(1-\theta)$

Observe that in each case the power series converges to $\theta \in \Theta$.

Moreover, it will be assumed that the times of activations follow the two-parameter generalized exponential (GE) distribution with parameters $\lambda > 0$ and $\alpha > 0$, introduced by Gupta & Kundu (1999). The failure rate function of this distribution is similar to the Weibull distributions, its cumulative probability, survival and density functions are given for all $t > 0$ by

$$F_a(t) = (1 - e^{-\lambda t})^{\alpha}, \tag{20}$$

$$S_a(t) = 1 - (1 - e^{-\lambda t})^{\alpha} \tag{21}$$

and

$$f_a(t) = \alpha\lambda e^{-\lambda t}(1 - e^{-\lambda t})^{\alpha-1}, \tag{22}$$

respectively. With these assumptions the cure rate models derived with the first, the last and the uniform-random schemes will be denoted by GEPSF, GEPSL and GEPSR, respectively. The survival functions of these models are obtained by (11) (14) and (16) and are given by

$$S_{pF}(t) = \frac{C(\theta S_a(t))}{C(\theta)}, \tag{23}$$

$$S_{pL}(t) = 1 + \frac{a_0}{C(\theta)} - \frac{C(\theta(1 - S_a(t))}{C(\theta)} \tag{24}$$

and

$$S_{pR}(t) = \frac{a_0}{C(\theta)} + (1 - \frac{a_0}{C(\theta)})S_a(t), \tag{25}$$

respectively. Notice that $\theta S_a(t) < \theta < s$ and $\theta(1 - S_a(t)) < \theta < s$.

When the Poisson distribution is considered the cure rate models derived with the first, the last and the uniform-random schemes are denoted by GEPF, GEPL and GEPR, respectively.

Similarly when the geometric distribution is considered the cure rate models derived with the first, the last and the uniform-random schemes are denoted by GEGF, GEGL and GEGR, respectively. Finally when the logarithmic distribution is considered the cure rate models for these schemes are denoted GELF, GELL and GELR, respectively. The survival function of these models are obtained by replacing the associate $C(\theta)$ and a_0, given in Table 2, in the equations (23), (24) and (25). Now, as in Cancho *et al.* (2011a), the cure rate p_0 is incorporated in these functions by the reparametrization $\theta = C^{-1}(a_0/p_0)$. Thus, for the Poisson, geometric and logarithmic distributions θ becomes $-\log(p_0)$, $1-p_0$ and $1+p_0W(-p_0e^{-1/p_0})$, respectively, where $W(\cdot)$ stands for the Lambert W function (Corless *et al.*, 1996). With this reparametrization the survival and density extended function of these models are given in Table 2, as reported in Cancho *et al.* (2011a).

Table 2: Survival function ($S_{\rm p}$) and density function ($f_{\rm p}$) for some models

Model	$S_{\rm p}(t)$	$f_{\rm p}(t)$
GEPF	$p_0^{F_a(t)}$	$-\log(p_0)p_0^{F_a(t)}f_a(t)$
GEPL	$1+p_0-p_0^{S_a(t)}$	$-\log(p_0)p_0^{S_a(t)}f_a(t)$
GEPR	$p_0+(1-p_0)S_a(t)$	$(1-p_0)f_a(t)$
GEGF	$\{1+(p_0^{-1}-1)F_a(t)\}^{-1}$	$\dfrac{p_0^{-1}-1}{\{1+(p_0^{-1}-1)F_a(t)\}^2}f_a(t)$
GEGL	$1+p_0-\{1+(p_0^{-1}-1)S(y)\}^{-1}$	$\dfrac{p_0^{-1}-1}{\{1+(p_0^{-1}-1)S_a(t)\}^2}f_a(t)$
GEGR	$p_0+(1-p_0)S_a(t)$	$(1-p_0)f_a(t)$
GELF	$-\dfrac{\log(1-W_0S_a(t))}{W_0S_a(t)}p_0$	$\dfrac{W_0S_a(t)+\{1-W_0S_a(t)\}\log(1-W_0S_a(t))}{\{1-W_0S_a(t)\}W_0S_a(t)^2}p_0f_a(t)$
GELL	$1+p_0+\dfrac{\log(1-W_0F_a(t))}{W_0F_a(t)}p_0$	$\dfrac{W_0F_a(t)+\{1-W_0F_a(t)\}\log(1-W_0F_a(t))}{\{1-W_0F_a(t)\}W_0F_a(t)^2}p_0f_a(t)$
GELR	$p_0+(1-p_0)S_a(t)$	$(1-p_0)f_a(t)$

Remark. $W_0 = 1+p_0\,W(-e^{-1/p_0}/p_0)$, where $W(\cdot)$ is the Lambert W function (Corless *et al.*, 1996).

The functions $f_p(t)$ and $S_p(t)$ have three parameters, α, λ and p_0. Thus for all $t > 0$, $\alpha > 0$ $\lambda > 0$ and $p_0 \in (0, 1)$, will be defined

$$\begin{aligned} f_p(t,\boldsymbol{\gamma},p_0) &= f_p(t,\alpha,\lambda,p_0) = f_p(t) \\ S_p(t,\boldsymbol{\gamma},p_0) &= S_p(t,\alpha,\lambda,p_0) = S_p(t), \end{aligned} \tag{26}$$

with $\boldsymbol{\gamma} = (\alpha,\lambda)^\top$.

As mentioned in Section 1, the cure rate for the ith individual in a sample of size n is given by a logistic regression between p_0 and the covariates as given in equation (2). Under this link function the models are *identifiable* in the sense of Li *et al.* (2001). In this case the models generated by the uniform-random scheme are equals to the exponential cure rate model proposed by Kannan *et al.* (2010), that is, GEPR=GEGR=GELR=GE.

4 Inference

Let us consider that for the i-th individual, in a sample of size n, it is observed $t_i = \min\{T_i, C_i\}$ and $\delta_i = \mathrm{I}(T_i \leq C_i)$, where T_i is the time to-event, C_i is the censoring time and $\mathrm{I}(T_i \leq C_i)$ is the indicator function.

Let $\boldsymbol{\psi}^\top = (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top)$ be the model vector parameters and $\Psi \subset \mathbb{R}^{2+} \times \mathbb{R}^k$ the parametric space. The likelihood associated with $(t_1, \delta_1, \boldsymbol{x}_1), \ldots, (t_n, \delta_n, \boldsymbol{x}_n)$ can be written by

$$\mathcal{L}(\boldsymbol{\psi}; \boldsymbol{D}) = \prod_{i=1}^{n} f_p(t_i, \boldsymbol{\gamma}, p_{0_i})^{\delta_i} \, S_p(t_i, \boldsymbol{\gamma}, p_{0_i})^{1-\delta_i}, \tag{27}$$

where $\boldsymbol{D} = (\boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{X})$, $\boldsymbol{t} = (t_1, \ldots, t_n)^\top$, $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$, $\boldsymbol{X} = (\boldsymbol{x}_\mathbf{1}^\top, \ldots, \boldsymbol{x}_\boldsymbol{n}^\top)$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^\top$, $f_p(\,\cdot\,,\,\cdot\,,\,\cdot\,)$ and $S_p(\,\cdot\,,\,\cdot\,,\,\cdot\,)$ are the extended density and survival functions in equations (26).

The log-likelihood associated with $\boldsymbol{D}$ can be written as

$$\ell(\boldsymbol{\psi}) = \sum_{i=1}^{n} [\, \delta_i log(f_p(t_i; \boldsymbol{\gamma}, p_{0_i})) + (1 - \delta_i) log(S_p(t_i; \boldsymbol{\gamma}, p_{0_i}))\,], \tag{28}$$

The maximum likelihood estimation is obtained by direct maximization of (28)) via the BBoptim function of the R program (R Development Core Team, 2011). The Lambert W function in Table 2 can be found in the R package emdbook. Under suitable regularity conditions it can be shown that the asymptotic distribution of the maximum likelihood estimator $\widehat{\boldsymbol{\psi}}$ is multivariate normal (see details in Lawless (2003)) with mean vector $\boldsymbol{\psi}$ and covariance matrix $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\psi}})$, which can be estimated by $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\psi}}) = -\ddot{\mathrm{L}}^{-1}(\hat{\boldsymbol{\psi}})$, the observed information matriz, that is,

$$\ddot{\mathrm{L}}(\boldsymbol{\psi}) = \frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^\top}\,. \tag{29}$$

5 Sensitivity analysis

This section outlines a methodology to perform a sensitivity analysis.

5.1 Global influence

To analyse the influence of observations on the parameters estimates it is common to use the case-deletion approach (see Cook (1977)) which measures the effect of removing the ith sample case.

Let $\ell_{(i)}$ be the log-likelihood function when the ith sample case is removed. Then, by (28)

$$\ell_{(i)}(\boldsymbol{\psi}) = \sum_{j\in\{1,\ldots,n\}-i} [\,\delta_i log(f_p(t_i;\boldsymbol{\gamma},p_{0_i})) + (1-\delta_i) log(S_p(t_i;\boldsymbol{\gamma},p_{0_i}))\,], \tag{30}$$

where subscript "(i)" means the original quantity with the ith case deleted. Let $\hat{\boldsymbol{\psi}}_{(i)} = (\hat{\boldsymbol{\gamma}}_{(i)}^\top, \hat{\boldsymbol{\beta}}_{(i)}^\top)^\top$ denote the maximum estimate of $\ell_{(i)}(\boldsymbol{\psi})$. Thus, the ith case is regarded as an influential observation if the difference between $\hat{\boldsymbol{\psi}}_{(i)}$ and $\hat{\boldsymbol{\psi}}$ is large.

Since $\hat{\boldsymbol{\psi}}_{(i)}$ must be calculated for the n cases, the following approximation, given in Cook & Weisberg (1982), is used to simplify the computational time:

$$\hat{\boldsymbol{\psi}}_{(i)} \approx \hat{\boldsymbol{\psi}} - \ddot{\mathrm{L}}^{-1}(\hat{\boldsymbol{\psi}})\dot{\ell}_{(i)}(\hat{\boldsymbol{\psi}}), \tag{31}$$

where $\dot{\ell}_{(i)}(\hat{\boldsymbol{\psi}})$ is the derivative $\frac{\partial \ell_{(i)}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}$ evaluated at $\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}$.

By substituting in (31) the decompositions

$$\dot{\ell}_{(i)}(\hat{\boldsymbol{\psi}}) = \begin{bmatrix} \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\gamma}} \\ \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\beta}}\,) \end{bmatrix} \tag{32}$$

and

$$-\ddot{\mathrm{L}}^{-1}(\hat{\boldsymbol{\psi}}) = \begin{bmatrix} A_{\boldsymbol{\gamma}} & A_{\boldsymbol{\gamma\beta}} \\ A_{\boldsymbol{\gamma\beta}}^\top & A_{\boldsymbol{\beta}} \end{bmatrix} \tag{33}$$

the following approximations are obtained:

$$\hat{\boldsymbol{\gamma}}_{(i)} \approx \boldsymbol{\gamma} - \big(A_{\boldsymbol{\gamma}} \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\gamma}} + A_{\boldsymbol{\gamma\beta}} \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\beta}} \big) \tag{34}$$

$$\hat{\boldsymbol{\beta}}_{(i)} \approx \boldsymbol{\beta} - \big(A_{\boldsymbol{\beta}} \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\beta}} + A_{\boldsymbol{\gamma\beta}}^\top \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\gamma}} \big) \tag{35}$$

Usual measures for the difference between $\hat{\boldsymbol{\psi}}_{(i)}$ and $\hat{\boldsymbol{\psi}}$ are detailed in the following subsection.

5.1.1 With The generalized Cook distance

A measure for the difference between $\hat{\boldsymbol{\psi}}_{(i)}$ and $\hat{\boldsymbol{\psi}}$, denoted by GD_i, is the following generalized Cook distance

$$GD_i = (\hat{\boldsymbol{\psi}}_{(i)} - \hat{\boldsymbol{\psi}})^\top \big(\text{-}\ddot{\mathrm{L}}(\hat{\boldsymbol{\psi}})\big)(\hat{\boldsymbol{\psi}}_{(i)} - \hat{\boldsymbol{\psi}}). \tag{36}$$

Since $\hat{\boldsymbol{\psi}}^{\top}$ maximises the log-likelihood function, -$\ddot{\mathrm{L}}$ is positive definite and GD_i is, indeed, a mathematical distance. Thus, the distance between $\hat{\boldsymbol{\psi}}_{(i)}$ and $\hat{\boldsymbol{\psi}}$ is a measure of the influence of the ith sample case on the vector of parameters estimates. Equation (31) gives the following approximation for this distance

$$GD_i \approx \dot{\ell}_{(i)}(\hat{\boldsymbol{\psi}})^{\top}\big(\text{-}\ddot{\mathrm{L}}^{-1}(\hat{\boldsymbol{\psi}})\big)\dot{\ell}_{(i)}(\hat{\boldsymbol{\psi}}). \tag{37}$$

Now, following in Li *et al.* (2012), some measures of the influence of the ith sample case on sub vector of parameters estimates $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}$ are derived. By substituting (34) and (35) in (37) is follows

$$GD_i \approx \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\gamma}}^{\top} A_{\boldsymbol{\gamma}} \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\gamma}} + 2\frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\gamma}}^{\top} A_{\boldsymbol{\gamma\beta}} \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\beta}} + \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\beta}}^{\top} A_{\boldsymbol{\beta}} \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\beta}}.$$

From this approximation, the following quantities

$$GD_i(\boldsymbol{\gamma}) \approx \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\gamma}}^{\top} A_{\boldsymbol{\gamma}} \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\gamma}} \tag{38}$$

and

$$GD_i(\boldsymbol{\beta}) \approx \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\beta}}^{\top} A_{\boldsymbol{\beta}} \frac{\partial \ell_{(i)}(\hat{\boldsymbol{\psi}})}{\partial \boldsymbol{\beta}}, \tag{39}$$

are considered measures of the influence of the ith sample case on the estimates $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}$, respectively.

5.1.2 With the likelihood displacement

The likelihood displacement of $\hat{\boldsymbol{\psi}}_{(i)}$ with respect to $\hat{\boldsymbol{\psi}}$ is given by

$$LD_i = 2\{\ell(\hat{\boldsymbol{\psi}}) - \ell(\hat{\boldsymbol{\psi}}_{(i)})\}. \tag{40}$$

Since $\hat{\boldsymbol{\psi}}^{\top}$ maximises the log-likelihood function LD_i is non-negative but it is not a mathematical distance. As emphasized in Cook (1986), this measure is extensively used because of its large sample properties, for instance one of these establishes that

$$\{\, \boldsymbol{\psi} : 2[\, \ell(\hat{\boldsymbol{\psi}}) - \ell(\boldsymbol{\psi})\,] < \chi^2 \,\},$$

is a confidence region for $\boldsymbol{\psi}$, where χ^2 is a value of a chi-squared distribution with $k+2$ degrees of freedom. Since

$$LD(\boldsymbol{\psi}) \approx (\boldsymbol{\psi} - \hat{\boldsymbol{\psi}})^{\top}\big(\text{-}\ddot{\mathrm{L}}(\hat{\boldsymbol{\psi}})\big)(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}),$$

in a neighborhood of $\hat{\boldsymbol{\psi}}$, then

$$LD_i \approx GD_i, \tag{41}$$

whenever $\hat{\boldsymbol{\psi}}_{(i)}$ is closed to $\hat{\boldsymbol{\psi}}$, as observed in Li *et al.* (2012).

5.2 Local influence

A different approach to perform a sensitivity analysis considers the effect of small perturbations in the model or in the data according to a particular scheme. This approach, which is called Local influence, was proposed by Cook (1986) for the linear regression model and later it was extended to more general model as described in Escobar & Meeker (1992). Let $\boldsymbol{w} = (w_1, \ldots, w_n)$ be the vector of numbers that quantifies the perturbation and $\ell W(\boldsymbol{\psi}, \boldsymbol{w})$ the associated log-likelihood function, where $\boldsymbol{\psi} \in \Psi$ and $\boldsymbol{w} \in \Omega$, an open subset of $\mathbb{R}^n$. It is assumed that there exists $\boldsymbol{w}_0$ such that no perturbation occurs when $\boldsymbol{w} = \boldsymbol{w}_0$. Thus, $\ell W(\boldsymbol{\psi}, \boldsymbol{w}_0) = \ell(\boldsymbol{\psi})$ and $\hat{\boldsymbol{\psi}}$ maximizes $\ell W(\,\cdot\,, \boldsymbol{w}_0)$.

The log-likelihood displacement function, LDW, is defined by

$$LDW(\boldsymbol{w}) = 2\{\ell(\hat{\boldsymbol{\psi}}) - \ell(\hat{\boldsymbol{\psi}}_w)\}, \tag{42}$$

where $\hat{\boldsymbol{\psi}}_w$ maximizes $\ell W(\,\cdot\,, \boldsymbol{w})$. As pointed out in Cook & Weisberg (1982), large values of LDW indicate that $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\psi}}_W$ differ considerably relative to the contours of the unperturbed log-likelihood $\ell(\boldsymbol{\psi})$. Cook (1986) states that the graph of the surface $(\boldsymbol{w}, LDW(\boldsymbol{w}))$ contains essential information on the influence of the perturbation scheme in question. In order to measure this influence he uses curvatures as follows.

The curvature of LDW, in the direction of the unit vector $\boldsymbol{h}$, evaluated in $\boldsymbol{w}_0$ is given by

$$C_{\boldsymbol{h}} = \frac{\partial^2 LD(\boldsymbol{w}_0 + ah)}{\partial a^2} = 2\boldsymbol{h}^\top \ddot{A}\boldsymbol{h}, \tag{43}$$

where

$$\ddot{A} = -\frac{\partial^2 L(\hat{\boldsymbol{\psi}}_w,\, \boldsymbol{w}_0)}{\partial \boldsymbol{w} \partial \boldsymbol{w}^\top}. \tag{44}$$

From (43) the maximum curvature, C_{max}, is obtained when $\boldsymbol{h}$ is the eigenvector, $\boldsymbol{h}_{max}$, associated with the largest eigenvalue of $\ddot{A}$. Thus, $\boldsymbol{h}_{max} = (\boldsymbol{h}_{max_1}, \ldots, \boldsymbol{h}_{max_n})^\top$ indicates how to perturb the model or data to obtain the greatest local change in LDW. According to Cook (1986), the plot of $(i,\, \boldsymbol{h}_{max_i})$, $i = 1, \ldots, n$, identifies potential influential cases and $\boldsymbol{h}_{max_i} > 2$ can be used as a benchmark to indicate that the ith observation is influential.

Other benchmarks from which an observation can be considered influential are given next. Let $\boldsymbol{u}_i \in \mathbb{R}^n$ be the vector of zeros whit a single 1 in the ith position, $C_i = 2A_{ii}$ the curvature associated with $\boldsymbol{u}_i$ and $M(0)_i = C_i / \sum_{j=1}^{k+2} \lambda_j$, $i = 1, \ldots, n$, where $\lambda_1 \geq \cdots \geq \lambda_{k+2}$ be the $k+2$ positive eigenvalues of $2\ddot{A}$. Let $\bar{M}(0)$ and $S_{M(0)}$ denote the sample mean and the standard deviation of $M(0)_1, \ldots, M(0)_n$, respectively. Notice that $\bar{M}(0) = 1/n$, because $\sum_{j=1}^{k+2} \lambda_j = trace(2\ddot{A})$. The plot of $(i, M(0)_i)$, $i = 1, \ldots, n$, is another graphical way to identify possible influential cases. Poon & Poon (1999), Zhu & Lee (2001) and Escobar & Meeker (1992)

proposed the benchmarks $M(0)_i > 2/n$, $M(0)_i > 1/n + 2S_{M(0)}$ and $M(0)_i > \chi^2_{0,5;k+2}/\sum_{j=1}^{k+2}\lambda_j$, respectively, to indicate that the ith observation could be considered influential.

Thus, the matrix $\ddot{A}$ is key to carry out the analysis of local influence. The following identity, given by Cook (1986), is used to facilitate the calculation of $\ddot{A}$:

$$\ddot{A} = \Delta^\top(-\ddot{L}(\hat{\boldsymbol{\psi}}))^{-1}\Delta, \tag{45}$$

where

$$\Delta = \frac{\partial^2 \ell W(\hat{\boldsymbol{\psi}}, \boldsymbol{w}_0)}{\partial \boldsymbol{\psi} \partial \boldsymbol{w}^\top}. \tag{46}$$

Hence it is important to calculate the matriz Δ.

5.2.1 Perturbation schemes

The perturbation schemes for the model and data, used in this paper, are described bellow. Some identities to obtain the matrix Δ are derived.

a) **Weight perturbation**. In this case, the vector $\boldsymbol{w} = (w_1, \ldots, w_n)$, used to quantify the perturbation, affects the log-likelihood function as follows

$$\ell W(\boldsymbol{\psi}, \boldsymbol{w}) = \sum_{l=1}^{n}[\, w_l \delta_l log(f_p(t_l, \boldsymbol{\gamma}, p_{0_l})) + w_l(1-\delta_l) log(S_p(t_l, \boldsymbol{\gamma}, p_{0_l}))\,], \tag{47}$$

where $0 \leq w_i \leq 1$. The non-perturbation vector is $\boldsymbol{w}_0 = (1, \ldots, 1)^\top$.

Since $\frac{\partial^2 \ell W(\boldsymbol{\psi}, \boldsymbol{w})}{\partial \psi_j \partial w_i} = \frac{\partial}{\partial \psi_j} GW_i(\boldsymbol{\psi})$, with

$$GW_i(\boldsymbol{\psi}) = \delta_i log(f_p(t_i, \boldsymbol{\gamma}, p_{0_i})) + (1-\delta_i) log(S_p(t_i, \boldsymbol{\gamma}, p_{0_i})). \tag{48}$$

The computational calculation of Δ is directly from the jacobian of $GW(\boldsymbol{\psi}) = (GW_1(\boldsymbol{\psi}), \ldots, GW_n(\boldsymbol{\psi}))$, that must be evaluated in $\hat{\boldsymbol{\psi}}$.

b) **Perturbation of responses**. In this case the perturbation is originated in the responses, ti, from $t_i + w_i s_T$, $i = 1, \ldots, n$, where $\boldsymbol{w} = (w_1, \ldots, w_n)$ is the vector used to quantify the perturbation and s_T is a scale factor estimated by the standard deviation of responses. The non-perturbation vector is $\boldsymbol{w}_0 = (0, \ldots, 0)^\top$.

The perturbed log-likelihood function is given by

$$\ell W(\boldsymbol{\psi}, \boldsymbol{w}) = \sum_{l=1}^{n}[\, \delta_l log(f_p(t_l + w_l s_t, \boldsymbol{\gamma}, p_{0_l})) + (1-\delta_l) log(S_p(t_l + w_l s_t, \boldsymbol{\gamma}, p_{0_l}))\,] \tag{49}$$

From this equation it follows that $\frac{\partial}{\partial w_i}\ell W(\boldsymbol{\psi}, \boldsymbol{w}_0) = GR_i(\boldsymbol{\psi})$, where

$$GR_i(\boldsymbol{\psi}) = s_T\Big\{ \frac{\delta_i \frac{\partial}{\partial t} f_p(t_i + s_T, \boldsymbol{\gamma}, p_{0_i})}{f_p(t_i + s_T, \boldsymbol{\gamma}, p_{0_i})} - \frac{(1-\delta_i) f_p(t_i + s_T, \boldsymbol{\gamma}, p_{0_i})}{S_p(t_i + s_T; \boldsymbol{\gamma}, p_{0_i})} \Big\}. \tag{50}$$

The computational calculation of Δ is directly from the jacobian of $GR(\boldsymbol{\psi}) = (GR_1(\boldsymbol{\psi}), \ldots, GR_n(\boldsymbol{\psi}))$, that must be evaluated in $\hat{\boldsymbol{\psi}}$.

c) **Perturbation of the explanatory variables**. In this case, the perturbation is originated in one of the covariates. If $(x_{1k_o}, \ldots, x_{nk_o})$ is the vector of the values of $\boldsymbol{X}_{k_o}$ in the sample, then the perturbed vector of values is $(x_{1k_o} + w_1 s_{\boldsymbol{X}_{ko}}, \ldots, x_{nk_o} + w_n s_{\boldsymbol{X}_{ko}})$, where $s_{\boldsymbol{X}_{ko}}$ is the standard deviation of the observed values of this covariate. The vector of no-perturbación is $\boldsymbol{w}_0 = (0, \ldots, 0)^\top$. Let $\boldsymbol{xw}_i$ be the vector obtained when the k_oth entry in $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})^\top$ is substituted by $x_{ik_o} + w_i s_{\boldsymbol{X}_{ko}}$, and $pw_{0i} = \frac{\exp(\boldsymbol{xw}_i{}^\top \boldsymbol{\beta})}{1+\exp(\boldsymbol{xw}_i{}^\top \boldsymbol{\beta})}$, for $i = 1, \ldots, n$. Thus, the perturbed log-likelihood function is given by

$$\ell W(\boldsymbol{\psi}, \boldsymbol{w}) = \sum_{l=1}^{n} [\, \delta_l log(f_p(t_l, \boldsymbol{\gamma}, pw_{0l})) + (1-\delta_l) log(S_p(t_l, \boldsymbol{\gamma}, pw_{0l}))\,]. \tag{51}$$

Since

$$pw_{0i} = \frac{\exp(\boldsymbol{x}_i{}^\top \boldsymbol{\beta}) \exp(w_i s_{\boldsymbol{X}_{ko}} \beta_{ko})}{1 + \exp(\boldsymbol{x}_i{}^\top \boldsymbol{\beta}) \exp(w_i s_{\boldsymbol{X}_{ko}} \beta_{ko})}\,, \tag{52}$$

then

$$\frac{\partial}{\partial w_i}\ell W(\boldsymbol{\psi}, \boldsymbol{w}) = \frac{\partial}{\partial w_i}\{\, \delta_i log(f_p(t_i, \boldsymbol{\gamma}, pw_{0i})) + (1-\delta_i) log(S_p(t_i, \boldsymbol{\gamma}, pw_{0i}))\,\} \tag{53}$$

and, by the chain rule,

$$\frac{\partial}{\partial w_i}\ell W(\boldsymbol{\psi}, \boldsymbol{w}_0) = GE_i(\boldsymbol{\psi}), \tag{54}$$

where

$$GE_i(\boldsymbol{\psi}) = \frac{s_{\boldsymbol{X}_{ko}} \beta_{ko} \exp(-\boldsymbol{x}_i{}^\top \boldsymbol{\beta})}{\big(1 + \exp(-\boldsymbol{x}_i{}^\top \boldsymbol{\beta})\big)^2} \Big\{ \frac{\delta_i \frac{\partial}{\partial p} f_p(t_i, \boldsymbol{\gamma}, p_{0_i})}{f_p(t_i, \boldsymbol{\gamma}, p_{0_i})} + \frac{(1-\delta_i) \frac{\partial}{\partial p} S_p(t_i, \boldsymbol{\gamma}, p_{0_i})}{S_p(t_i, \boldsymbol{\gamma}, p_{0_i})} \Big\}. \tag{55}$$

The computational calculation of Δ is directly from the jacobian of $GE(\boldsymbol{\psi}) = (GE_1(\boldsymbol{\psi}), \ldots, GE_n(\boldsymbol{\psi}))$, that must be evaluated in $\hat{\boldsymbol{\psi}}$.

6 Application

To illustrate the results of this paper the data set considered by Cancho *et al.* (2011a) is reanalysed. The data set includes 205 patients observed after operation for removal of malignant melanoma in the period 1962–77. These data are available in the timereg package in R (Scheike,

2009). The observed time (T) ranges from 0.0274 to 15.25 years and refers to the time until the patient's death or the censoring time. Patients dead from other causes, as well as patients still alive at the end of the study are censored observations (72%). The covariates considered are $\boldsymbol{X}_2$ = ulceration status (present=1, $n = 90$; absent=0, $n = 115$), $\boldsymbol{X}_3$ = tumor thickness (in mm, mean = 2.92 and standard deviation = 2.96), $\boldsymbol{X}_4$ = age (in years, mean= 52,46 and standard deviation = 16,67) and $\boldsymbol{X}_5$ = sex (male=1, female=0), with coefficients β_2, β_3, β_4 and β_5, respectively, and β_1 denotes the intercept. The Kaplan-Meier estimated of the surviving function (see Figure 1) levels off above 0.6 and the presence of a plateau indicates that the models that ignore the possibility of cure does not fit for these data. Therefore, the dataset may be fitted by a cure rate model with covariates, in particular, the GE-PS cure rate models GEPF, GEPL, GEGF, GEGL, GELF, GELL and GE, given in Table 2, are used.

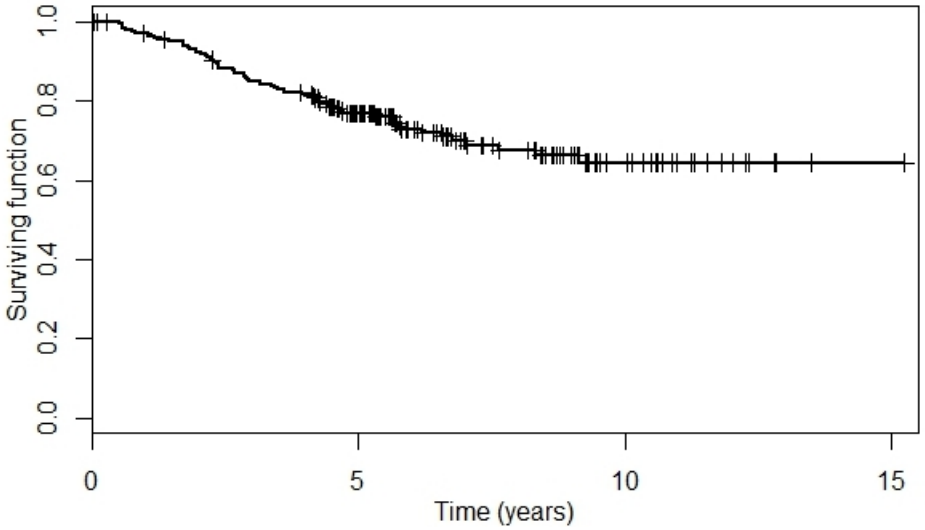

Figure 1: Kaplan-Meier estimate of the surviving function.

6.1 Fitting models and parameter estimation

The maximum likelihood estimation for the vector parameters, $\boldsymbol{\psi} = (\alpha, \lambda, \beta_1, \beta_2, \ldots, \beta_5)$, is obtained by direct maximization of (27) or (28) via the BBoptim function of the R program (R Development Core Team, 2011). Table 3 presents the maximum values of the log-likelihood function ($l(\cdot)$), the estimated AIC and BIC criteria considering the all fitted distributions.

Table 3: AIC and SBC for the fitted models.

Model	$l(\cdot)$	AIC	SBC
GEPF	-205,3233	424,6466	421,29262
GEGF	-203,0158	420,0316	416,6776
GELF	-200,9897	415.9794	412,6254
GEPL	-212,0622	438,1244	434,77042
GEGL	-214,9450	443,8900	440,53602
GELL	-217,0028	448,0056	444,65162
GE	-208,1857	430,3714	427,01742

According to the AIC and SBC criteria, the GELF cure rate model outperforms its concurrent distributions in both considered criteria. Table 4 shows the maximum likelihood estimates (MLE), the standard errors estimates (SEE) by the observed Fisher information matrix and the 95% confidence intervals for the parameters of the all fitted distributions. All the covariates, except the age, have a significant effect on the determination of the cured fraction.

Table 4: MLE and SEE of the parameters for the GELF model.

Parameter	Estimate (est)	Standard error (se)	|est| / se	95% IC
α	2,9685	0,6079	4,8835	(1,7771; 4,1598)
λ	0,2227	0,0846	2,633	(0,0569; 0,3885)
β_1	1,8318	0,625	2,9311	(0,6069; 3,0567)
β_2	-1,3875	0,3147	4,4091	(-2,0043; -0,7707)
β_3	-0,1141	0,0343	3,3254	(-0,1814; -0,0469)
β_4	-0,0041	0,0077	0,5302	(-0,0192; 0,011)
β_5	-0,6083	0,2735	2,2245	(-1,1443; -0,0723)

6.2 Sensitivity analysis

In this subsection a sensitivity analysis is made for the GELF proposed model, according to Section 5. The measures of global and local influence ware calculated.

6.2.1 Global influence

To measure the influence of the ith sample case on the vector of parameters estimates it is calculated the case-deletion measures GD_i and LD_i described in Subsection 5.1. Theses measures are similar, as shown in Figure 2. This similarity suggests that the removing of the ith sample

case does not affect significantly the estimation of the parameters, as pointed out in (41). Notice from the figure that since the cases 5, 6, 7, 10, 11, 13, 15 and 29 are above of the benchmark they are the most influential.

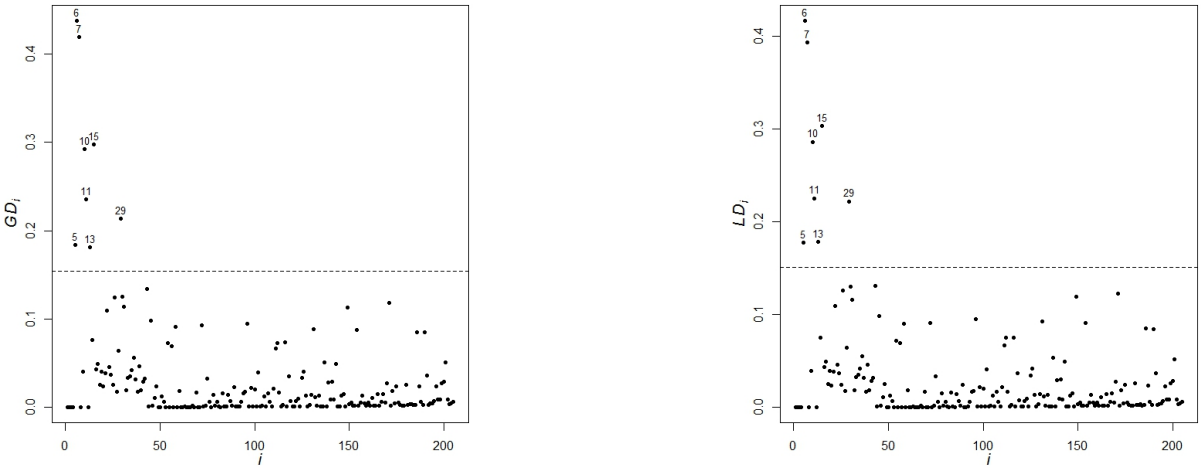

Figure 2: Index plot of case-deletion measures. Left panel: the generalized Cook's distance (GC_i). Right Panel: The likelihood distance (LD_i).

Influence measures on the estimation of sub-vectors $\boldsymbol{\gamma} = (\alpha, \lambda)$, the parameters of time activation, and $\boldsymbol{\beta}$, the coefficients of regression, were also calculated and illustrated in Figure 3. Notice from the figure that since the cases 5, 6, 7, 10, 11, 26, 29 and 31 are above of the benchmark they are the most influential.

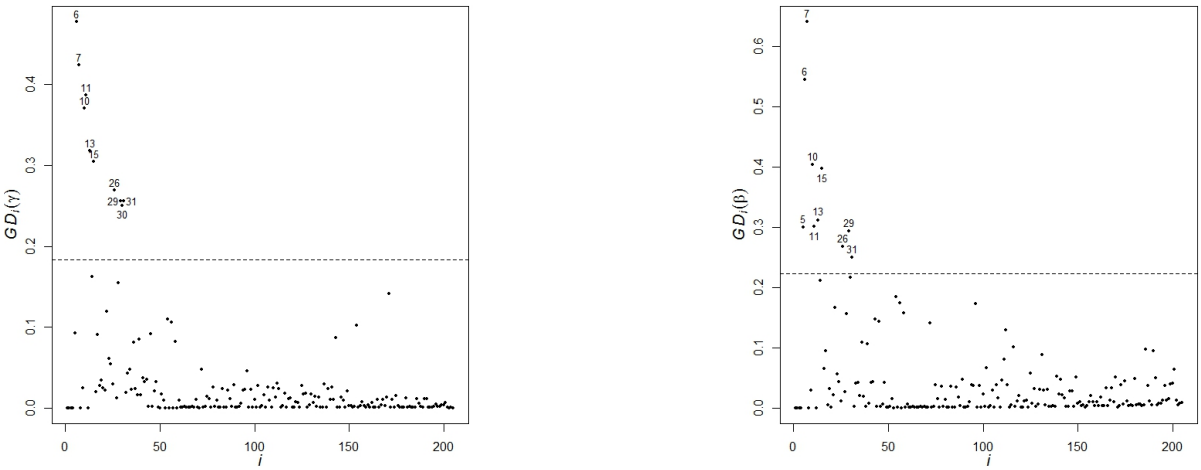

Figure 3: Index plot of case-deletion measures for the subvectors parameters. Left panel: the generalized Cook's distance $GC_i(\boldsymbol{\gamma})$. Right Panel: the generalized Cook's distance $GC_i(\boldsymbol{\beta})$.

6.2.2 Local influence

The influence measures given in subsection 5.2 will be calculated under the perturbation schemes.

Case weight perturbation. The value $C_{max} = 2,8160$ is a maximum curvature. The possible influential cases are indicated by the plots in Figure 4.

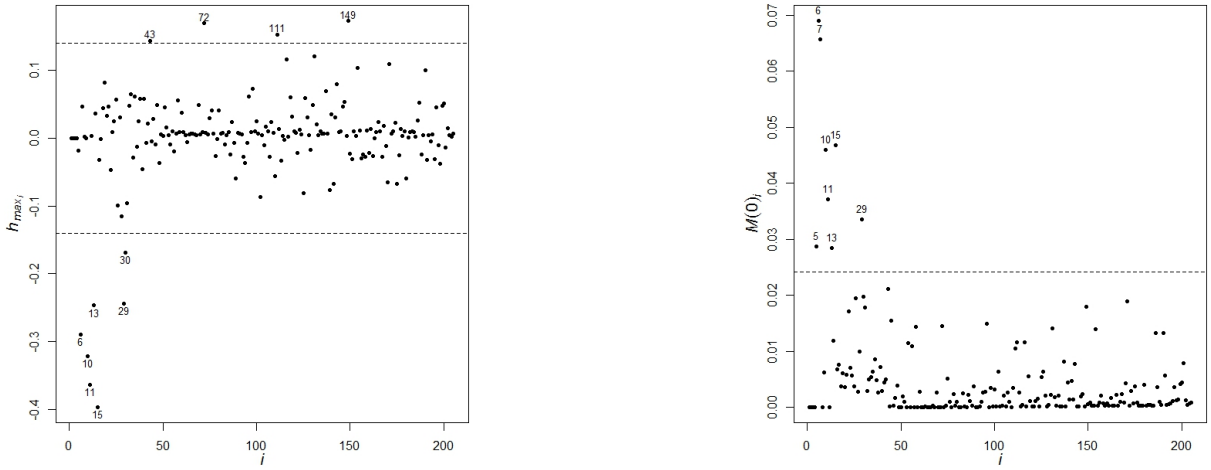

Figure 4: Index plot of local influence measures with the weight perturbation scheme. Left panel: Direction of maximum curvature, $\boldsymbol{h}_{max}$. Right Panel: $M(0)_i$.

Perturbation of responses. The value $C_{max} = 54,7037$ is a maximum curvature. As shown in Figure 5, the cases 5, 6, 7, 9 and 10 would be the most influential.

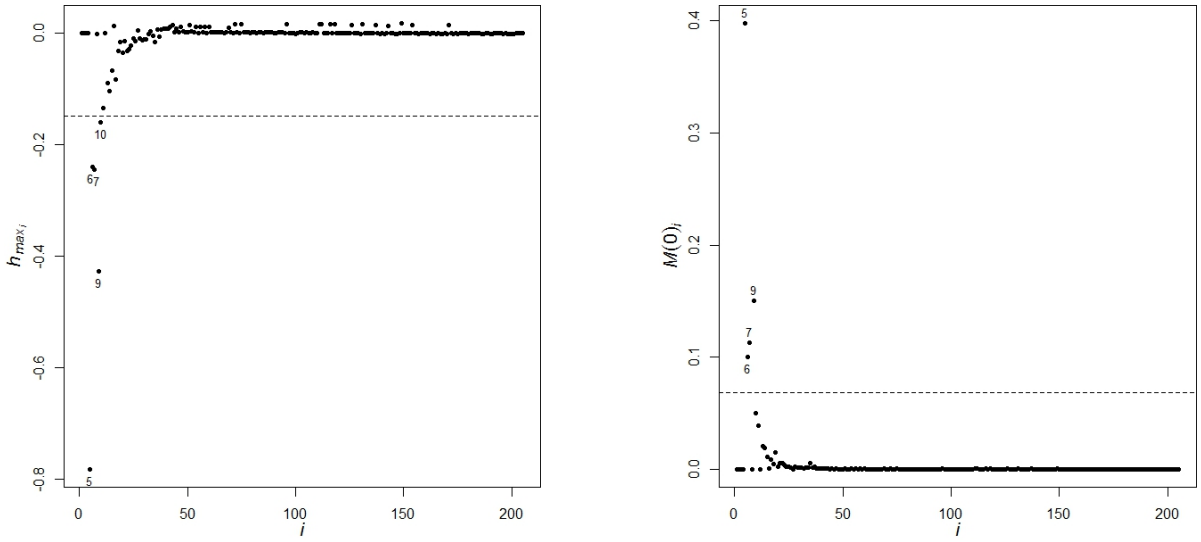

Figure 5: Index plot of local influence measures with the weight perturbation scheme for responses. Left panel: Direction of maximum curvature, $\boldsymbol{h}_{max}$. Right Panel: $M(0)_i$.

Perturbation of explanatory variable thickness. The value $C_{max} = 81,7252$ is a maximum curvature. The plots in Figure 6 suggests that the cases 5, 6, 7, 9, 10, 19, 21 and 35 are the most influential.

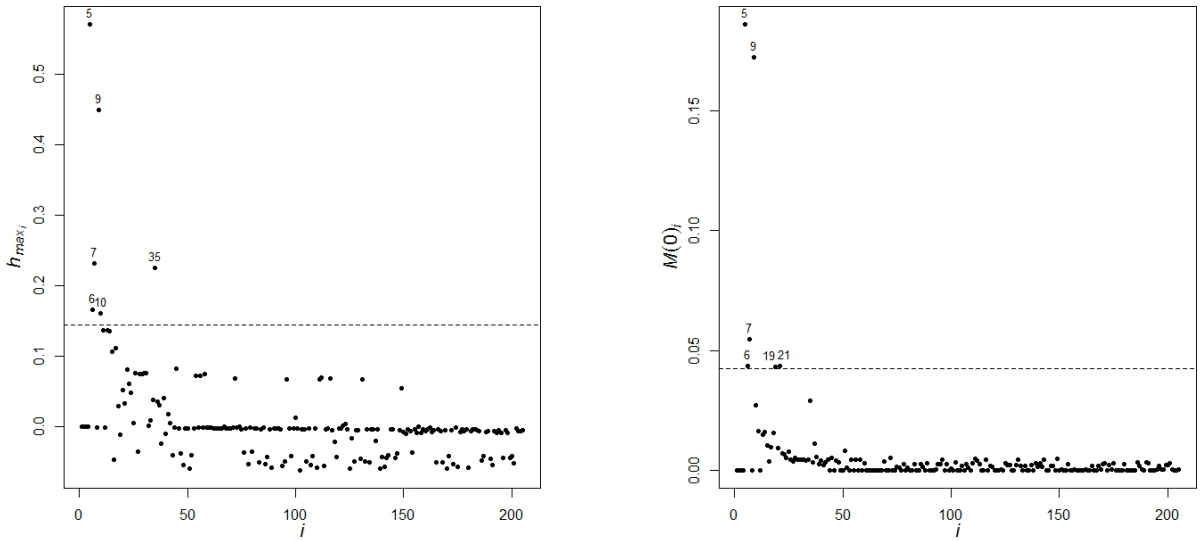

Figure 6: Index plot of local influence measures with the perturbation scheme for the covariate thickness. Left panel: Direction of maximum curvature, $\boldsymbol{h}_{max}$. Right Panel: $M(0)_i$.

Perturbation of explanatory variable age. The value $C_{max} = 1,0305$ is a maximum curvature. The plots in Figure 7 suggests which cases are the most influential.

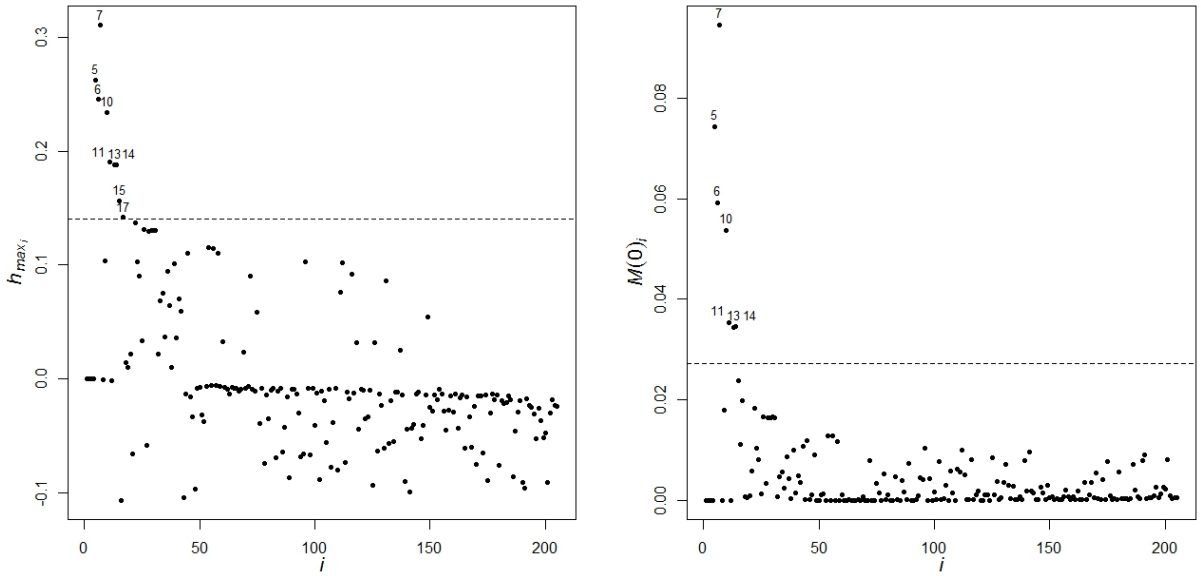

Figure 7: Index plot of local influence measures with the weight perturbation scheme. Left panel: Direction of maximum curvature, $\boldsymbol{h}_{max}$. Right Panel: $M(0)_i$.

In conclusion, according to the sensitivity analysis performed, the subset of potentially influential cases is $I = \{5, 6, 7, 9, 10, 11, 13, 14, 15, 17, 19, 21, 26, 29, 35, 43, 72, 111, 149\}$.

6.2.3 Impact of the detected influential observations

Each one of the cases in I is removed and then proposed model is refitted. To measure the impact of this removing on the estimation of ψ_i let us consider the following relative rate changes (rce):

$$rce(\psi_i) = 100(\hat{\psi}_i - \hat{\psi}_{(i)})/\hat{\psi}_i,$$

and the standard error estimate (rcse)

$$rcse(\psi_i) = 100(\hat{se}_{\psi_i} - \hat{se}_{\psi_{(i)}})/\hat{se}_{\psi_i},$$

These quantities are shown in Table 5. Considerables rates changes only occur in the estimates of the regression coefficient β_4, associated with the covariate age. Since this covariate is removed from the model because is not significant, then the proposed model seems not to be sensitive to the influential observations.

Table 5: Relative changes in the estimates (rce) and in the standard error estimates (rcse) of parameters

	α		λ		β_1		β_2		β_3		β_4		β_5	
Case	rce	rcse	rce	rcse	rce	rcse	rce	rcse	rce	rcse	rce	rcse	rce	rcse
5	0,5	1,1	0,5	2,1	1	0,3	0,7	0,6	10,8	4,5	27,4	1,7	4,6	0,5
6	7,9	12,2	6,5	1,1	10,3	1,3	1,1	0,1	5,3	1,7	76,6	0,4	4,1	1,9
7	8,2	13,4	9,6	1,4	6,1	4,3	0,9	0,5	9,1	0,6	77,8	1,6	6,2	0,8
9	2,8	2,1	4,6	1,8	1,7	3,8	0,9	0,1	6	10,3	13,7	2,4	2,7	0,4
10	5,8	9,3	3,5	0,4	0,8	1,0	1,0	0,1	1,0	1,5	26,3	0,9	16,8	0,6
11	6,3	9,3	3,2	0,5	2,8	0,0	1,4	0,1	2,2	1,0	5,7	0,6	13,6	0,2
13	3,7	5,8	0,8	0,3	1,8	0,4	2,0	0,1	2,1	0,8	25,6	0,6	11,4	0,1
14	0,3	1,5	1,3	0,5	3,6	0,6	2,3	0,0	2,5	1,8	18,7	1,2	6,8	0,4
15	1,8	2,7	2,6	0,3	8,0	3,6	2,8	0,4	0,1	0,0	73,7	1,5	9,2	0,9
17	2,3	1,6	4,1	0,4	0,8	2,5	2,2	0,1	0,9	1,4	15,6	0,5	5,4	0,4
19	2,6	1,6	3,8	1,7	5,5	5,5	1,4	0,3	0,3	2,5	30	11,7	3,8	0,8
21	2,8	1,7	4,9	2,2	0,6	5,6	0,8	0,1	0,8	6,5	18,3	4,5	6,2	5,2
26	0,5	0,6	2,9	0,1	0,0	3,0	4,3	1,9	0,7	0,2	21,1	0,3	4,5	0,7
29	0,8	1,2	5,3	0,7	7,6	6,7	4,2	1,8	4,3	0,3	57,3	0,9	5,6	0,8
35	3,3	1,9	5,0	3,2	2,2	3,6	2,0	6,5	5,4	8,3	4,5	0,5	2,7	0,1
43	0,4	1,7	2,2	1,9	1,4	0,1	0,2	0,0	9,5	4,3	1,4	2,1	10,2	1,8
72	1,2	0,3	0,4	2,0	1,4	0,0	3,5	2,4	2,4	1,6	37,0	2,5	5,8	1,0
111	0,4	0,8	3,3	2,0	1,8	1,3	3,1	2,5	2,5	1,4	20,0	2,3	4,1	1,3
149	2,1	3,7	9,8	2,3	4,4	4,0	3,0	2,8	4,8	1,4	23,8	4,1	1,9	2,0

6.3 Final model and goodness of fit

In Table 6 are listed the estimates of the parameters of GELF cure rate model, with their estimates standard errors and the 95% confidence intervals. The covariates ulceration status, $\boldsymbol{X}_2$, tumor thickness, $\boldsymbol{X}_3$, and sex, $\boldsymbol{X}_5$, have a significant effect on the cured fraction. The proportion of patients cured is lower for patients with ulceration than for those without it. The covariate tumor thickness has a significant effect on the reduction of the cured fraction. The proportion of people cured is greater for the woman than for the men. The estimated cure rate for the ith individual is given by

$$\hat{p}_{0i} = \frac{exp(1,6351 - 1,3947x_{i2} - 0,1174x_{i3} - 0,6014x_{i5})}{1 + exp(1,6351 - 1,3947x_{i2} - 0,1174x_{i3} - 0,6014x_{i5})},$$

where x_{i1}, x_{i2} and x_{i5} are the associated values of covariates ulceration, thickness and sex, respectively. An estimation for the cure rate in the entire population of individuals is given by

$$\hat{p}_0 = \frac{1}{205}\sum_{i=1}^{205}\hat{p}_{0i} = 0,5948.$$

Table 6: Maximum likelihood estimates of the parameters for the GELF model.

Parameter	Estimate (est)	Standard error (se)	95% IC
α	2,9844	0,6133	(1,7824; 4,1864)
λ	0,2262	0,0851	(0,0595; 0,3930)
β_1	1,6351	0,4849	(0,6847; 2,5855)
β_2	-1,3947	0,3152	(-2,0125; -0,7770)
β_3	-0,1174	0,0337	(-0,1835; -0,0514)
β_5	-0,6014	0,2725	(-1,1356; -0,0673)

Figure 8 shows the fitted survival function superimposed to the empirical survival function. It can be seen that the fitness is good.

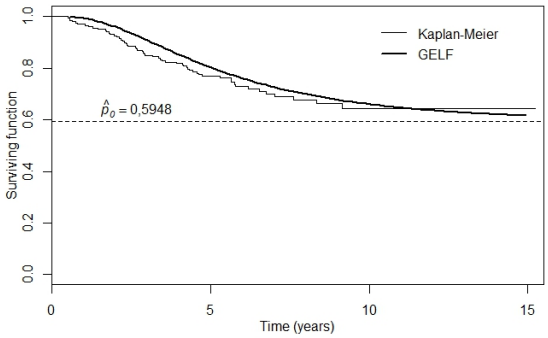

Figure 8: Kaplan-Meier estimate of the surviving function.

7 Concluding remarks

In this report the theory of models with latent risk factors under activation schemes, introduced by Cooner *et al.* (2007), was described and new results based on it were given. A new family of cure rate models with covariates called the generalized exponential-power series was derived by applying this theory. In addition the procedure to perform a sensitivity analysis for general models was summarized. Analytical expressions were provided to facilitate computational required to accomplish the analysis of local influence. To show the flexibility and potential of this family as a cure rate model, particular cases of this family to a real data set were fitted. For the best fitted model to the data, a sensitivity analysis was performed including diagnostic measures based on case-deletion and local influence approaches. This sensitivity analysis showed the robustness of the model against influential observations.

References

Adamidis, K. & Loukas, S. (1998). A lifetime distribution with decreasing failure rate. *Statistics and Probability Letters*, **39**(1), 35–42.

Adamidis, K., Dimitrakopoulou, T. & Loukas, S. (2005). On an extension of the exponential-geometric distribution. *Statistics and Probability Letters*, **73**(3), 259 – 269.

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society B*, **11**(1), 15–53.

Cancho, V. G. & Bolfarine, H. (2001). Modeling the presence of immunes by using the exponentiated-Weibull model. *Journal of Applied Statistics*, **28**(6), 659–671.

Cancho, V. G., de Castro, M. & Dey, D. K. (2011a). Long-term survival models with latent activation under a flexible family of distributions. Accepted for publication in *Brazilian Journal of Probability and Statistics*. Available at http://www.redeabe.org.br/bjps/future_papers.htm.

Cancho, V. G., Louzada-Neto, F. & Barriga, G. D. C. (2011b). The Poisson-exponential lifetime distribution. *Computational Statistics and Data Analysis*, **55**, 677–686.

Chahkandi, M. & Ganjali, M. (2009). On some lifetime distributions with decreasing failure rate. *Computational Statistics and Data Analysis*, **53**(12), 4433–4440.

Chen, M.-H., Ibrahim, J. G. & Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94**(447), 909–919.

Cook, R. D. (1977). Detection of infuential observations in linear regression. *Technometrics*, **19**, 15–18.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B*, **48**, 133–169.

Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall/CRC, Boca Raton, FL.

Cooner, F., Banerjee, S., Carlin, B. P. & Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*, **102**(478), 560–572.

Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J. & Knuth, D. E. (1996). On the lambert W function. *Advances in Computational Mathematics*, **5**(1), 329–359.

Escobar, L. A. & Meeker, William Q., J. (1992). Assessing influence in regression analysis with censored data. *Biometrics*, **48**(2), pp. 507–528.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, **38**(4), pp. 1041–1046.

Flores, D. J., Borges, P., Cancho, V. G. & Louzada-Neto, F. (2011). The complementary exponential power series distribution. Accepted for publication in *Brazilian Journal of Probability and Statistics*. Available at http://www.redeabe.org.br/bjps/future_papers.htm.

Gupta, R. D. & Kundu, D. (1999). Theory & methods: Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, **41**(2), 173–188.

Hashimoto, E., Cordeiro, G. & Ortega, E. (2012). The new Neyman type A beta Weibull model with long-term survivors. Accepted for publication in *Computational Statistics*. Available at http://dx.doi.org/10.1007/s00180-012-0338-9.

Kannan, N., Kundu, D., Nair, P. & Tripathi, R. C. (2010). The generalized exponential cure rate model with covariates. *Journal of Applied Statistics*, **37**(10), 1625–1636.

Kim, S., Chen, M.-H. & Dey, D. (2011). A new threshold regression model for survival data with a cure fraction. *Lifetime Data Analysis*, **17**, 101–122. 10.1007/s10985-010-9166-9.

Kus, C. (2007). A new lifetime distribution. *Computational Statistics and Data Analysis*, **51**(9), 4497–4509.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley, New York, NY, first edition.

Li, A.-P., Chen, Z.-X. & Xie, F.-C. (2012). Diagnostic analysis for heterogeneous log-Birnbaum-Saunders regression models. *Statistics and Probability Letters*, **82**(9), 1690–1698.

Li, C.-S., Taylor, J. M. & Sy, J. P. (2001). Identifiability of cure models. *Statistics and Probability Letters*, **54**, 389–395.

Morais, A. L. & Barreto-Souza, W. (2011). A compound class of weibull and power series distributions. *Computational Statistics and Data Analysis*, **55**(3), 1410–1425.

Ortega, E., Cancho, V. & Paula, G. (2009). Generalized log-gamma regression models with cure fraction. *Lifetime Data Analysis*, **15**, 79–106. 10.1007/s10985-008-9096-y.

Poon, W.-Y. & Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61**(4), 51–61. 10.1111/1467-9868.00162.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rodrigues, J., de Castro, M., Cancho, V. G. & Balakrishnan, N. (2009). Com-poisson cure rate survival models and an applications to a cutaneous melanoma data. *Journal of Statistical Planning and Inference*, **139**, 3605–3611.

Scheike, T. (2009). *timereg package*. With contributions from Torben Martinussen and Jeremy Silver. R package version 1.1-6.

Tahmasbi, R. & Rezaei, S. (2008). A two-parameter lifetime distribution with decreasing failure rate. *Computational Statistics and Data Analysis*, **52**, 3889–3901.

Tsodikov, A. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics*, **54**(1), 1508–1516.

Yakovlev, A., Yu, A. B., Bardou, V.-J., Fourquet, A., Hoang, T., Rochefodiere, A. & Tsodikov, A. D. (1993). A simple stochastics model of tumor recurrence an its aplications to data on premenopausal breast cancer. In S. F. de Biométrie, editor, *Biometrie et Analyse de Donnes Spatio-Temporelles No 12, B*, pages 33–82, France.

Yakovlev, A. Y. & Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, Singapore.

Zhu, H.-T. & Lee, S.-Y. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **63**(1), pp. 111–126.