# The Log-exponentiated-Weibull Regression Models with Cure Rate: Local Influence and Residual Analysis

Vicente G. Cancho, Edwin M. M. Ortega and Heleno Bolfarine
*University of São Paulo*

*Abstract*: In this paper the log-exponentiated-Weibull regression model is modified to allow the possibility that long term survivors are present in the data. The modification leads to a log-exponentiated-Weibull regression model with cure rate, encompassing as special cases the log-exponencial regression and log-Weibull regression models with cure rate typically used to model such data. The models attempt to estimate simultaneously the effects of covariates on the acceleration/deceleration of the timing of a given event and the surviving fraction; that is, the proportion of the population for which the event never occurs. Assuming censored data, we consider a classic analysis and Bayesian analysis for the parameters of the proposed model. The normal curvatures of local influence are derived under various perturbation schemes and two deviance-type residuals are proposed to assess departures from the log-exponentiated-Weibull error assumption as well as to detect outlying observations. Finally, a data set from the medical area is analyzed.

*Key words:* Cure rate models, exponentiated-Weibull distribution, influence diagnostic, survival data; residual analysis.

## 1. Introduction

Models for survival analysis typically assume that all units under study are susceptible to the event and will eventually experience this event if the follow-up is sufficiently long. However, there are situations where a fraction of individuals are not expected to experience the event of interest; that is, those individuals are cured or insusceptible. For example, researchers may be interested in analyzing the recurrence of a disease. Many individuals may never experience a recurrence; therefore, a cured fraction of the population exists. Cure rate models have been applied to estimate the possibility of a cured fraction. These models extend the understanding of time-to-event data by allowing the formulation of more accurate and informative conclusions. These conclusions are otherwise unobtainable

from an analysis which fails to account for a cured or insusceptible fraction of the population. If a cured component is not present, the analysis reduces to standard approaches of survival analysis. Cure rate models have been used for modeling time-to-event data for various types of cancer, including breast cancer, non-Hodgkin lymphoma, leukemia, prostate cancer and melanoma. Thus, a cure rate model is suitable for modeling data from cancer clinical trials. Berkson and Gage (1952) introduce the mixture cure rate models, Maller and Zhou (1996) give an extensive discussion of classic methods of inference for mixture cure rate models. An alternative formulation of the parametric cure rate models is discussed in Yakovlev and Tsodikov (1996). A Bayesian formulation of this model is given in Chen *et al.* (1999). Tsodikov *et al.* (2003) give an excellent review of such methods. Outside the applications in this area, Hoggart and Griffin (2001) focused on the time to a customer leaving a bank and Yamaguchi (1992) applied the mixture cure rate models to the analysis of permanent employment in Japan.

The literature presents many applications of the survival models with cure rate considering the Weibull family of distributions (see, Ibrahim *et al.*, 2001; Maller and Zhou , 1996). This family is suitable in situations where the failure rate function is constant or monotone. This paper examines the statistical inference aspects and the modeling of the presence of a cure rate of a given data set by using the log-exponentiated-Weibull regression model. The inferential part was carried out using the asymptotic distribution of the maximum likelihood estimators, which in situations when the sample is small, the normality is more difficult to be justified. As an alternative for classic analysis, we explore the use of the Bayesian methods via Markov Chain Monte Carlo.

Influence diagnostics is an important part in the analysis of a data set, as it provides us with an indication of bad model fitting or of influential observations. Cook (1986) proposed a diagnostic approach, named local influence, to assess the effect of small perturbations in the model and/or data on the parameter estimates. Several authors have applied the local influence methodology in regression models more general than the normal regression model (see, for example, Paula 1993, Galea *et al.*, 2000, Díaz-García *et al.*, 2004, and Le *et al.*, 2006). Moreover, some authors have investigated the assessment of local influence in survival analysis models: for instance, Pettitt and Bin Daud (1989) investigated local influence in proportional hazard regression models, Escobar and Meeker (1992) adapted local influence methods to regression analysis with censoring, Ortega *et al.* (2003) considered the problem of assessing local influence in log-exponentiated-Weibull regression models with censored observations and Ortega *et al.* (2008) investigated local influence in the Weibull mixture cure models.

In Section 2 we briefly describe the cure rate model. In Section 3 we suggest a log-exponentiated-Weibull regression model with cure fraction, in addition

with the maximum likelihood estimators and the Bayes estimator. In Section 4, we discuss the global and local influence method, the likelihood displacement is used to evaluate the influence of observations on the maximum likelihood estimators. We present residual from a fitted model using the martingale residual proposed by Therneau *et al.* (1990) and we proposed a modified deviance residual for the generalized log-gamma regression model with cure fraction for the log-exponentiated-Weibull regression model with cure rate in Section 5. In Section 6 presents the results of an analysis with a real data set and analysis residual some and conclusions appear in Section 7.

## 2. The Cure Rate Model

As in Yakovlev and Tsodikov (1996) and Chen *et al.* (1999), we formulate the model within a biological context. The promotion time for the $j$th tumor cell is denoted by $R_j$, $j = 1, \ldots, N$, where $N$ is random variable unobservable that denotes the number of competing causes that can produce a detectable cancer. If $N = 0$, we define $P(R_0 = \infty) = 1$ to represent a cure. Hence, the observable time to relapse of cancer or failure time is defined as $T = \min(R_j, \, 0 \le j \le N)$. If $N$ is a Poisson random variable with mean $\phi$ independent of the sequence $R_j$, $j = 1, 2, \ldots$, also assumed independent random variable with the same cumulative distribution function (c.d.f.) $F(.)$ and $S(.) = 1 - F(.)$, we have that the population survival function is given by

$$
\begin{aligned}
S_{pop}(t) &= P(N = 0) + P(R_1 > t, \ldots, R_N > t | N \ge 1) P(N \ge 1) \\
&= \exp(-\phi) + \sum_{k=1}^{\infty} \left[ S(t) \right]^k \frac{\phi^k \exp\{-\phi\}}{k!} \\
&= \exp\left\{ - \phi F(t) \right\}.
\end{aligned}
\tag{2.1}
$$

A corresponding cure fraction in model (2.1) is $\lim_{t \to \infty} S_{pop}(t) = \exp\{-\phi\} > 0$, that is not a proper model. As $\phi \to \infty$, the cure fraction tends to 0, whereas as $\phi \to 0$ the cure fraction tends to 1. Corresponding population density and hazard functions are $f_{pop}(t) = \phi f(t) \exp\{-\phi F(t)\}$ and $h_{pop}(t) = \phi f(t)$, respectively. They are not proper probability density function or hazard function. However, $h_{pop}(t)$ is multiplicative in $\phi$ and $f(t)$; thus, it has the proportional hazard structure. The population survival function (1) can be written as

$$
S_{pop}(t) = \exp\{-\phi\} + \left[ 1 - \exp\{-\phi\} \right] S^*(t),
$$

where

$$
S^*(t) = \frac{\exp\{-\phi F(t)\} - \exp\{-\phi\}}{1 - \exp\{-\phi\}},
$$

which is a proper survival function and

$$f^*(t) = \frac{\exp\{-\phi F(t)\}}{1 - \exp\{-\phi\}}\phi f(t),$$

that is a proper probability density function.

Suppose we have $n$ subjects and let $N_i$ latent variables that denote the number of competing causes that can produce a detectable cancer for the *ith* subject. Further, we assume that $N_i$'s are independent with Poisson distributions with means $\phi_i$, $i = 1, \ldots, n$. Further, suppose $R_{i1}, \ldots, R_{iN_i}$ are the promotion times for the $N_i$ competing causes for the $i$th subject, which are unobserved, and all have proper cumulative distribution function, $F(\cdot|\boldsymbol{\psi})$ and survival function $S(\cdot|\boldsymbol{\psi}) = 1 - F(\cdot|\boldsymbol{\psi})$, where $\boldsymbol{\psi}$ is the parameter vector.

As in Chen *et al.* (1999), we assume that the mean of $N_i$ is such that, $\phi(\mathbf{x}_i) = \exp(\mathbf{x}_i^T \boldsymbol{\gamma})$, where $\mathbf{x}_i^T = (x_{i1}, \ldots, x_{ip_1})$ denotes the $p_1 \times 1$ vector of covariate values for the *ith* subject, and $\boldsymbol{\gamma}^T = (\gamma_1, \ldots, \gamma_{p_1})$ denotes the corresponding vector of regression coefficients. Let $t_i$ denote the survival time for subject $i$, $t_i = \min(T_i, C_i)$, with $T_i = \min(R_{i0}, R_{i1}, \ldots, R_{iN_i})$ and $C_i$ is the censoring time whereas $\delta_i$ is the censoring indicator, assuming 1 if $t_i = T_i$ and 0 otherwise. The observed data is $\mathbf{D} = (n, \mathbf{t}, \delta, \mathbf{X})$, where $\mathbf{t} = (t_1, \ldots, t_n)^T$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^T$ and $\mathbf{X} = (\mathbf{x}_1^T, \ldots, \mathbf{x}_n^T)$, Also, let $\mathbf{N} = (N_1, \ldots, N_n)^T$. The complete data is given by $\mathbf{D}_c = (n, \mathbf{t}, \delta, \mathbf{X}, \mathbf{N})$, where $\mathbf{N}$ is an unobserved vector of latent variables. Chen *et al.*(1999) show that the likelihood function for the complete data is

$$\begin{aligned}
L(\boldsymbol{\gamma}, \boldsymbol{\psi}|\mathbf{D}_c) &= \left\{ \prod_{i=1}^n \Big[ N_i f(t_i|\boldsymbol{\psi}) \Big]^{\delta_i} \Big[ S(t_i|\boldsymbol{\psi}) \Big]^{N_i - \delta_i} \right\} \\
&\quad \times \exp\left\{ \sum_{i=1}^n \Big[ N_i \log\big[\phi(\mathbf{x}_i)\big] - \log(N_i!) - \phi(\mathbf{x}_i) \Big] \right\}. \quad (2.2)
\end{aligned}$$

Assuming out the unobserved latent variable $\mathbf{N}$ in (2.2) the marginal log-likelihood function for the observable data is given by

$$\begin{aligned}
L(\boldsymbol{\gamma}, \boldsymbol{\psi}|\mathbf{D}) &= \sum_{\mathbf{N}} L(\boldsymbol{\gamma}, \boldsymbol{\psi}|\mathbf{D}_c) \\
&= \prod_{i=1}^n \exp\big(\phi(\mathbf{x}_i) f(t_i|\boldsymbol{\psi})\big)^{\delta_i} \exp\big( -\phi(\mathbf{x}_i)\big[1 - S(t_i|\boldsymbol{\psi})\big]\big). \quad (2.3)
\end{aligned}$$

Recently De Castro *et al.* (2007) show that the likelihood function in (2.3) can be represented by the following expression:

$$L(\boldsymbol{\gamma}, \boldsymbol{\psi}|\mathbf{D}) = \prod_{i=1}^n \Big\{ f_{pop}(t_i|\boldsymbol{\psi}) \Big\}^{\delta_i} \Big\{ S_{pop}(t_i|\boldsymbol{\psi}) \Big\}^{1-\delta_i}. \quad (2.4)$$

This expression can be viewed as a generalization of the likelihood function found in models for censored data (Kalbfleish and Prentice, 2002), since the proper density and survival functions are replaced by their improper counterparts.
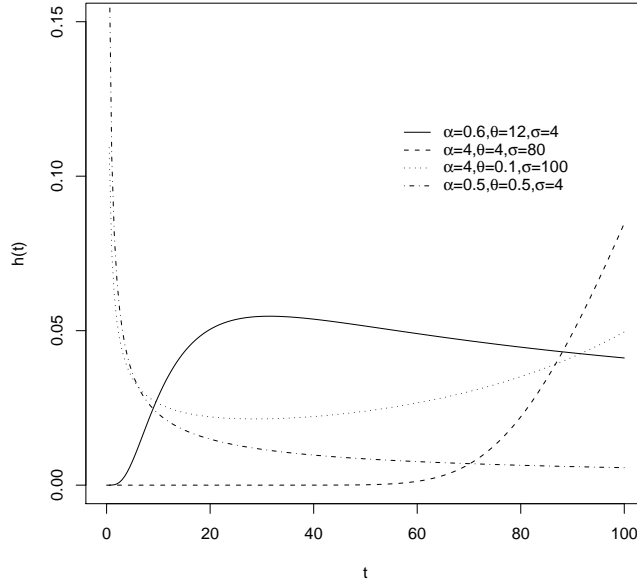


Figure 1: Some forms specials for hazard function for the exponetiated-Weibull family ($WE(\alpha, \theta, \sigma)$).

## 3. The Log-exponentiated-Weibull Regression Models with Cure Rate

The random variable T nonnegative has a exponentiated Weibull (WE) distribution is its probability density is of form

$$f(t) = \alpha\theta\lambda[1 - \exp(-(\lambda t)^\alpha)]^{\theta-1}\exp[-(\lambda t)^\alpha](\lambda t)^{\alpha-1}, \qquad (3.1)$$

and survival function is given by,

$$S(t) = 1 - [1 - \exp(-(\lambda t)^\alpha)]^\theta, \qquad (3.2)$$

where $t > 0$ and $\alpha > 0$, $\theta > 0$, are shape parameters and $\lambda > 0$ is scale parameter. It can be shown that the hazard function is given by

$$h(t) = \frac{\alpha\theta\lambda[1 - \exp(-(\lambda t))]^{\theta-1}\exp(-(\lambda t)^\alpha)(\lambda t)^{\alpha-1}}{1 - (1 - \exp(-(\lambda t)^\alpha))^\theta}, \qquad (3.3)$$

The great flexibility of this model to fit lifetime data, is given by the different forms that the hazard function (3.3) can take, that is, (i) if $\alpha \geq 1$ and $\alpha\theta \geq 1$, then the hazard function is monotonically increasing; (ii) if $\alpha \leq 1$ and $\alpha\theta \leq 1$, then the hazard function is monotonically decreasing; (iii) if $\alpha > 1$ and $\alpha\theta < 1$, then the hazard function is bathtub shaped and (iv) if $\alpha < 1$ and $\alpha\theta > 1$, then we have a unimodal hazard function. In Figure 1, we have some case specials for hazard function (3.3).

Applications of the WE distribution in reliability and survival studies were investigated by Mudolkar $et$ $al.$ (1995) and Cancho $et$ $al.$ (1999). Cancho and Bolfarine (2001) proposed the exponentiated-Weibull mixture model to modeling the presence of cure fraction in lifetime data. Some properties of this distribution have been studied more detailed in Mudholkar and Hutson (1996) and Nassar and Eissa (2003).

We assume that the promotion time in the cure rate model in (2.1) follows a exponentiated-Weibull distribution with density function given by (3.1). We consider the following log-exponentiated-Weibull regression models

$$Y_i = \mathbf{z}_i^\top \boldsymbol{\beta} + \sigma\varepsilon_i, \quad i = 1, \ldots, n, \tag{3.4}$$

where $Y_i$ is log survival time for subject $i$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{p_2})$ is a vector of unknown parameters to be estimated, $\mathbf{z}_i^\top = (z_{i1}, z_{i2}, \ldots, z_{ip_2})$ is the explanatory vector and $\sigma > 0$ is unknown parameter and $\varepsilon_i$ are random variables assumed to be identically distributed with common probability density function

$$f(\varepsilon) = \theta[1 - \exp\{-\exp(\varepsilon)\}]^{\theta-1} \exp\{\varepsilon - \exp(\varepsilon)\}, \infty < \varepsilon < \infty, \tag{3.5}$$

where $\theta > 0$ is unknown parameter. It is supposition implicates that the variable $Y_i$ has been log-exponentiated-Weibull distribution (see, Cancho $et$ $al.$,1999).

On the other hard, by considering the model with cure rate in (2.1) and by assuming that such $\log(R_{ij})$ follow log-exponentiated-Weibull distribution with $y_i = \min\{\log(T_i), \log(C_i)\}$ in which $\log(T_i) = \min\{\log(R_{i0}), \log(R_{i1}), \ldots, \log(R_{iN_i})\}$, $i = 1, 2, \ldots, n$ we can show that the population survival function is

$$S_{pop}(y_i) = \exp\left\{-\phi_i(\mathbf{x}_i)\left(\left\{1 - \exp\left[-\exp\left(\frac{y_i - \mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}^\theta\right)\right\}, \tag{3.6}$$

where $\phi_i(\mathbf{x}_i)) = \exp(\mathbf{x}_i^T\boldsymbol{\gamma})$. This model will be referred to as the log-exponentiated-Weibull regression model with cure rate (LEWR-CR). This model is an extension of an accelerated failure time model using the exponentiated-Weibull distribution and it allows to determine the effect of covariates both on the failure time and on the cure rate itself. The corresponding density function is given by

$$
f_{pop}(y_i) = \frac{\theta}{\sigma}\phi_i(\mathbf{x}_i)\exp\left\{-\phi_i(\mathbf{x}_i)\left(\left\{1-\exp\left[-\exp\left(\frac{y_i-\mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}^{\theta}\right)\right\}
$$
$$
\times\left[1-\exp\left\{-\exp\left(\frac{y_i-\mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right)\right\}\right]^{\theta-1}
$$
$$
\times\exp\left\{\frac{y_i-\mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}-\exp\left(\frac{y_i-\mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right)\right\}, \tag{3.7}
$$

where $f_{pop}(y) = -\frac{d}{dy}S_{pop}(y)$. Note that $f_{pop}(y)$ is not a proper probability density function since $S_{pop}(y)$ is not a proper survival function.

## 3.1 Likelihood inference for the LEWR-CR

Considering the result in (2.4) the log-likelihood function for $\boldsymbol{\alpha} = \left(\theta,\sigma,\boldsymbol{\gamma}^T,\boldsymbol{\beta}^T\right)^T$, corresponding to the observed data is given by

$$
l(\boldsymbol{\alpha}) = \sum_{i\in F}\log\left[\phi(\mathbf{x}_i)\right] + r\log(\theta) - r\log(\sigma) + \sum_{i\in F}\left(\frac{y_i-\mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right) -
$$
$$
\sum_{i\in F}\exp\left(\frac{y_i-\mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right) + (\theta-1)\sum_{i\in F}\log\left(1-\exp\left\{-\exp\left(\frac{y_i-\mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right)\right\}\right) -
$$
$$
\sum_{i=1}^{n}\phi(\mathbf{x}_i)\left(\left\{1-\exp\left[-\exp\left(\frac{y_i-\mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right)\right]\right\}^{\theta}\right), \tag{3.8}
$$

where F and C denote, respectively, the set of uncensored and censored individuals, and $r$ is number uncensored observations. This model allows to determine the effect of covariates both on the failure time and on the cure rate itself. The maximum likelihood estimates for the parameter vector $\boldsymbol{\alpha} = \left(\theta,\sigma,\boldsymbol{\gamma}^T,\boldsymbol{\beta}^T\right)^T$ can be obtained by maximizing the log-likelihood function. In this paper, the software Ox (MAXBFGS subroutine) (see Doornik, 2001) was used to compute maximum likelihood estimates (MLE). Covariance estimates for the maximum likelihood estimators $\widehat{\boldsymbol{\alpha}}$ can also be obtained using the Hessian matrix. Confidence intervals and hypothesis testing can be conducted by using the large sample distribution of the MLE which is a normal distribution with the covariance matrix as the inverse of the Fisher information since regularity conditions are satisfied. More specifically, the asymptotic covariance matrix is given by $\mathbf{I}^{-1}(\boldsymbol{\alpha})$ with $\mathbf{I}(\boldsymbol{\alpha}) = -E[\ddot{\mathbf{L}}(\boldsymbol{\alpha})]$ such that $\ddot{\mathbf{L}}(\boldsymbol{\alpha}) = \left\{\frac{\partial^2 l(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^T}\right\}$.

Since it is not possible to compute the Fisher information matrix $\mathbf{I}(\boldsymbol{\alpha})$ due to the censored observations (censoring is random and noninformative), it is possible

to use the matrix of second derivatives of the log likelihood, $-\ddot{\mathbf{L}}(\boldsymbol{\alpha})$, evaluated at the MLE $\boldsymbol{\alpha} = \widehat{\boldsymbol{\alpha}}$, which is a consistent estimator. The required second derivatives are computed numerically.

For testing the adequacy of log-Weibull regression model with cure rate, that is, $H_0 : \theta = 1$, we can consider the likelihood ratio statistics, which in this case is given by

$$\Lambda_n = -2\Big\{\ell(\tilde{\boldsymbol{\alpha}}) - \ell(\hat{\boldsymbol{\alpha}})\Big\} \tag{3.9}$$

where $\hat{\boldsymbol{\alpha}}$ is the maximum likelihood estimator that follows by maximizing the log-likelihood in (3.8) and $\tilde{\boldsymbol{\alpha}}$ the restricted maximum likelihood estimator computer under $H_0$, that is, with $\theta = 1$. For large samples, $\Lambda_n$ is distributed approximately like the chi-square distribution with one degree of freedom. For testing the adequacy of the log-exponential regression model with cure rate, that is, $H_0 : (\sigma, \theta)^\top = (1, 1)^\top$, the likelihood deviance $\Lambda_n$ is as given in equation (3.9) but $\tilde{\boldsymbol{\alpha}}$ being the restricted maximum likelihood ratio estimator computed under $H_0$, that is, with $\sigma = 1$ and $\theta = 1$. In this case $\Lambda_n$ is distributed in large samples approximately like the chi-square distribution with two degrees of freedom.

## 3.2 Bayesian inference for the LEWR-CR

The use of the Bayesian method besides being an alternative analysis, allows the incorporation of previous knowledge on the parameters through informative priori densities. When there is not this information one can consider noninformative priors. In the Bayesian approach, the relevant information to the model parameters is obtained through posterior marginal distributions. As such, two difficulties arise. The first refers to the attainment of the required marginal posterior distributions and the second to the calculation of the moments of interest. Both cases require solving integrals that many times do not present analytical solution. In this paper we have used the simulation method of Markov Chain Monte Carlo (MCMC), such as the Gibbs sampler and Metropolis-Hasting algorithm to encompass such difficulties.

We consider the joint prior density for $\boldsymbol{\alpha} = \left(\theta, \sigma, \boldsymbol{\gamma}^T, \boldsymbol{\beta}^T\right)^T$ of the form

$$\pi(\boldsymbol{\alpha}) = \left(\prod_{i=1}^{p_2} \pi(\beta_i)\right)\left(\prod_{k=1}^{p_1} \pi(\gamma_k)\right)\pi(\sigma)\pi(\theta), \tag{3.10}$$

where $\beta_i \sim N(a_i, b_i)$, $i = 1, \ldots, p_2$, $\gamma_i \sim N(a_{1i}, b_{1i})$, $i = 1, \ldots, p_1$, $\sigma \sim IG(c, d)$ $\theta \sim G(e, f)$, with $N(a, b)$ denoting the Normal distribution, $IG(a, b)$ denoting the Inverse Gamma distribution with shape parameter $a > 0$ and scale parameter $b > 0$ and $G(e, f)$ denoting the Gamma distribution with mean $e/f$ and variance $a/f^2$. We assume that the hyperparameters are specified.

Combining the likelihood function, $L(\boldsymbol{\alpha}) = \exp(\ell(\boldsymbol{\alpha}))$, with $\ell(\boldsymbol{\alpha})$ in (3.8) and prior specification (3.10), the joint posterior distribution for $\boldsymbol{\alpha}$ is given by,

$$\pi(\boldsymbol{\alpha}|\mathcal{D}) \quad \propto \quad \left(\frac{\theta}{\sigma}\right)^r \exp\left\{\sum_{i \in F} \log\left[\phi(\mathbf{x}_i)\right] + \sum_{i \in F}\left(\frac{y_i - \mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right) - \sum_{i \in F}\exp\left(\frac{y_i - \mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right) + \right.$$

$$\left. (\theta - 1)\sum_{i \in F}\log\left(G_i(\boldsymbol{\beta},\sigma)\right) - \sum_{i=1}^{n}\phi(\mathbf{x}_i)\left(G_i(\boldsymbol{\beta},\sigma)\right)^\theta\right\}\pi(\boldsymbol{\alpha}), \qquad (3.11)$$

where, $\mathcal{D}$ denotes the observed data, $r$ denotes the number of uncensored observations, $G_i(\boldsymbol{\beta},\sigma) = 1 - \exp\left\{-\exp\left(\frac{y_i - \mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right)\right\}$ and $\phi_i(\mathbf{x}_i)) = \exp\left(\mathbf{x}_i^T\boldsymbol{\gamma}\right)$.

To implement the MCMC methodology, the full conditionals of the parameters are given by

$$\pi(\boldsymbol{\beta}|\boldsymbol{\gamma},\sigma,\theta) \quad \propto \quad \exp\left\{\sum_{i \in F}\left(\frac{y_i - \mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right) - \sum_{i \in F}\exp\left(\frac{y_i - \mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right) + \right. \qquad (3.12)$$

$$\left. (\theta - 1)\sum_{i \in F}\log\left(G_i(\boldsymbol{\beta},\sigma)\right) - \sum_{i=1}^{n}\phi(\mathbf{x}_i)\left(G_i(\boldsymbol{\beta},\sigma)\right)^\theta\right\}\pi(\boldsymbol{\beta})$$

$$\pi(\boldsymbol{\gamma}|\boldsymbol{\beta},\sigma,\theta) \quad \propto \quad \exp\left\{\sum_{i \in F}\log\left[\phi(\mathbf{x}_i)\right] - \sum_{i=1}^{n}\phi(\mathbf{x}_i)\left(G_i(\boldsymbol{\beta},\sigma)\right)^\theta\right\}\pi(\boldsymbol{\gamma}) \qquad (3.13)$$

$$\pi(\sigma|\boldsymbol{\beta},\boldsymbol{\gamma},\theta) \quad \propto \quad \sigma^{-r}\exp\left\{\sum_{i \in F}\left(\frac{y_i - \mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right) - \sum_{i \in F}\exp\left(\frac{y_i - \mathbf{z}_i^T\boldsymbol{\beta}}{\sigma}\right) + \right. \qquad (3.14)$$

$$\left. (\theta - 1)\sum_{i \in F}\log\left(G_i(\boldsymbol{\beta},\sigma)\right) - \sum_{i=1}^{n}\phi(\mathbf{x}_i)\left(G_i(\boldsymbol{\beta},\sigma)\right)^\theta\right\}\pi(\sigma),$$

$$\pi(\theta|\boldsymbol{\beta},\boldsymbol{\gamma},\sigma) \quad \propto \quad \theta^r\exp\left\{\theta\sum_{i \in F}\log\left(G_i(\boldsymbol{\beta},\sigma)\right) - \sum_{i=1}^{n}\phi(\mathbf{x}_i)\left(G_i(\boldsymbol{\beta},\sigma)\right)^\theta\right\}\pi(\theta). \qquad (3.15)$$

Since the above conditional posteriors do not present standard forms, the use of the Metropolis-Hasting sampler is required.

## 4. Influence Diagnostics

Local influence calculation can be carried out for model (6). If likelihood displacement $LD(\boldsymbol{\omega}) = 2\{l(\hat{\boldsymbol{\alpha}}) - l(\hat{\boldsymbol{\alpha}}_{\boldsymbol{\omega}})\}$ is used, where $\hat{\boldsymbol{\alpha}}_{\boldsymbol{\omega}}$ denotes the MLE under the perturbed model, the normal curvature for $\boldsymbol{\alpha}$ at direction $\mathbf{d}, \|\mathbf{d}\| = 1$, is given by $C_{\mathbf{d}}(\boldsymbol{\alpha}) = 2|\mathbf{d}^T\boldsymbol{\Delta}^T\left[\ddot{\mathbf{L}}(\boldsymbol{\alpha})\right]^{-1}\boldsymbol{\Delta}\mathbf{d}|$, where $\boldsymbol{\Delta}$ is a $(p_1 + p_2 + 4)$ matrix that depends on the perturbation scheme and whose elements are given by $\Delta_{vi} = \partial^2 l(\boldsymbol{\alpha}|\boldsymbol{\omega})/\partial\theta_v\partial\omega_i$, $i = 1, 2, \ldots, n$ and $v = 1, 2, \ldots, p_1 + p_2 + 4$ evaluated at $\hat{\boldsymbol{\alpha}}$

and $\boldsymbol{\omega}_0$, where $\boldsymbol{\omega}_0$ is the no perturbation vector (see Cook, 1986). For the log-exponentiated-Weibull regression models with cure rate, the elements of $\ddot{\mathbf{L}}(\boldsymbol{\alpha})$ are given in appendix A. We can also calculate normal curvatures $C_{\mathbf{d}}(\theta), C_{\mathbf{d}}(\sigma)$, $C_{\mathbf{d}}(\boldsymbol{\gamma})$ and $C_{\mathbf{d}}(\boldsymbol{\beta})$ to perform various index plots, for instance, the index plot of $\mathbf{d}_{max}$, the eigenvector corresponding to $C_{\mathbf{d}_{max}}$, the largest eigenvalue of the matrix $\mathbf{B} = -\boldsymbol{\Delta}^T [\ddot{\mathbf{L}}(\boldsymbol{\alpha})]^{-1} \boldsymbol{\Delta}$ and the index plots of $C_{\mathbf{d}_i}(\theta)$, $C_{\mathbf{d}_i}(\sigma)$ $C_{\mathbf{d}_i}(\boldsymbol{\gamma})$ and $C_{\mathbf{d}_i}(\boldsymbol{\beta})$ named total local influence (see, for example, Lesaffre and Verbeke, 1998), where $\mathbf{d}_i$ denotes an $n=$ vector of zeros with one at the $i-th$ position. Thus, the curvature at direction $\mathbf{d}_i$ assumes the form $C_i = 2|\boldsymbol{\Delta}_i^T [\ddot{\mathbf{L}}(\boldsymbol{\alpha})]^{-1} \boldsymbol{\Delta}_i|$ where $\boldsymbol{\Delta}_i^T$ denotes the $ith$ row of $\boldsymbol{\Delta}$. It is usual to point out those cases such that

$$C_i \geq 2\bar{C}, \quad \bar{C} = \frac{1}{n} \sum_{i=1}^{n} C_i. \tag{4.1}$$

## 4.1 Curvature calculations

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

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{vi})_{\left[(p_1+1)+(p_2+1)+2\times n\right]} = \left( \frac{\partial^2 l(\boldsymbol{\alpha}|\boldsymbol{\omega})}{\partial\theta_v\boldsymbol{\omega}_i} \right)_{\left[(p_1+1)+(p_2+1)+2\times n\right]}, \tag{4.2}$$

where $v = 1, 2, \ldots, p_1 + p_2 + 4$ and $i = 1, 2, \ldots, n$.

Considering the model defined in (12) and its log-likelihood function given by (13).

## Case-weights perturbation

Consider the vector of weights $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)^T$.
In this case the log-likelihood function takes the form

$$\begin{aligned}
l(\boldsymbol{\alpha}|\boldsymbol{\omega}) &= \sum_{i\in F} \omega_i \log\left[\phi(\mathbf{x}_i)\right] + \log(\theta) \sum_{i\in F} \omega_i - \log(\sigma) \sum_{i\in F} \omega_i + \sum_{i\in F} \omega_i h_i - \\
&\quad \sum_{i\in F} \omega_i \exp\{h_i\} + (\theta-1) \sum_{i\in F} \omega_i \log\left[1 - \exp\left\{-\exp(h_i)\right\}\right] - \\
&\quad \sum_{i\in F} \omega_i \phi(\mathbf{x}_i)\left[1 - \exp\left\{-\exp(h_i)\right\}\right]^\theta - \sum_{i\in C} \omega_i \phi(\mathbf{x}_i)\left[1 - \exp\left\{-\exp(h_i)\right\}\right]^\theta, \tag{4.3}
\end{aligned}$$

where $0 \leq \omega_i \leq 1$ and $\boldsymbol{\omega} = (1, \ldots, 1)^T$. Let us denote $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\theta, \boldsymbol{\Delta}_\sigma, \boldsymbol{\Delta}_\gamma, \boldsymbol{\Delta}_{\boldsymbol{\beta}})^T$.
Then the elements of vector $\boldsymbol{\Delta}_\theta$ take the form

$$\Delta_i = \begin{cases} \hat{\theta}1 + \log(\hat{g}_i)\left[1 - \hat{\phi}(\mathbf{x}_i)\hat{g}_i^{\hat{\theta}}\right] & \text{if} \quad i\epsilon F, \\ -\hat{\phi}(\mathbf{x}_i)\log(\hat{g}_i)\hat{g}_i^{\hat{\theta}} & \text{if} \quad i\epsilon C. \end{cases}$$

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

$$\Delta_i = \begin{cases} \hat{h}_i \hat{\sigma}^{-1}\left\{ \hat{h}_i^{-1} - 1 + \exp\{\hat{h}_i\} + \exp\{\hat{h}_i\}(1-\hat{g}_i)\hat{g}_i^{-1}\left[1-\hat{\theta}\left(-1+\hat{\phi}(\mathbf{x}_i)\hat{g}_i^{\hat{\theta}}\right)\right]\right\} & \text{if } i\epsilon F, \\ \hat{h}_i \hat{\sigma}^{-1}\hat{\theta}\exp\{\hat{h}_i\}(1-\hat{g}_i)\hat{\phi}(\mathbf{x}_i)\hat{g}_i^{(\hat{\theta}-1)} & \text{if } i\epsilon C. \end{cases}$$

The matrix $\boldsymbol{\Delta}_\gamma$ $(p_1 \times n)$ is expressed as

$$\Delta_{ki} = \begin{cases} x_{ik}\left[1 - \exp\{\mathbf{x}_i^T\hat{\boldsymbol{\beta}}\}\hat{g}_i^{\hat{\theta}}\right] & \text{if } i\epsilon F, \\ -x_{ik}\exp\{\mathbf{x}_i^T\hat{\boldsymbol{\beta}}\}\hat{g}_i^{\hat{\theta}} & \text{if } i\epsilon C. \end{cases}$$

The matrix $\boldsymbol{\Delta}_\beta$ $(p_2 \times n)$ is expressed as

$$\Delta_{ji} = \begin{cases} z_{ij}\hat{\sigma}^{-1}\exp\{\hat{h}_i\}\left\{-\exp\{\hat{h}_i\}^{-1} - 1 + (1-\hat{g}_i)\hat{g}_i^{-1}\left[-\hat{\theta}+1+\hat{\theta}\hat{\phi}(\mathbf{x}_i)\hat{g}_i^{\hat{\theta}}\right]\right\} & \text{if } i\epsilon F, \\ z_{ij}\hat{\theta}\hat{\sigma}^{-1}\hat{\phi}(\mathbf{x}_i)\exp\{\hat{h}_i\}(1-\hat{g}_i)\hat{g}_i^{(\hat{\theta}-1)} & \text{if } i\epsilon C. \end{cases}$$

## Response perturbation

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

$$\begin{aligned} l(\boldsymbol{\alpha}|\boldsymbol{\omega}) = & \sum_{i\in F}\log\left[\phi(\mathbf{x}_i)\right] + r\log(\theta) - r\log(\sigma) + \sum_{i\in F}h_i^* - \\ & \sum_{i\in F}\exp\{h_i^*\} + (\theta-1)\sum_{i\in F}\log\left[1-\exp\left\{-\exp(h_i^*)\right\}\right] - \\ & \sum_{i\in F},\phi(\mathbf{x}_i)\left[1-\exp\left\{-\exp(h_i^*)\right\}\right]^\theta - \sum_{i\in C}\phi(\mathbf{x}_i)\left[1-\exp\left\{-\exp(h_i^*)\right\}\right]^\theta, \quad (4.4) \end{aligned}$$

where $h_i^* = \sigma^{-1}(y_i + \omega_i S_y - \mathbf{z}_i^T\boldsymbol{\beta})$.
In addition, the elements of the vector $\boldsymbol{\Delta}_\theta$ take form

$$\Delta_i = \begin{cases} S_y\hat{\sigma}^{-1}\exp\{\hat{h}_i\}(1-\hat{g}_i)\hat{g}_i^{-1}\left\{1 - \hat{\phi}(\mathbf{x}_i)\hat{g}_i^{\hat{\theta}}\left[\hat{\theta}\log(\hat{g}_i)+1\right]\right\} & \text{if } i\epsilon F, \\ S_y\hat{\sigma}^{-1}\hat{\phi}(\mathbf{x}_i)\exp\{\hat{h}_i\}(1-\hat{g}_i)\hat{g}_i^{(\hat{\theta}-1)}\left[\hat{\theta}\log(\hat{g}_i)+1\right] & \text{if } i\epsilon C. \end{cases}$$

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

by

$$
\Delta_i = \begin{cases}
S_x\hat{\sigma}^{-2}\exp\{\hat{h}_i\}\Big\{-\exp\{-\hat{h}_i\}+\hat{h}_i+1+(1-\hat{g}_i)\hat{g}_i^{(\hat{\theta}-1)} \\
\quad\Big[-(\hat{\theta}-1)\hat{g}_i^{(\hat{\theta}-1)}\Big\{\hat{h}_i(1-\exp\{\hat{h}_i\})-(1-\hat{g}_i)(1+\hat{h}_i)+1\Big\}+ \\
\quad\hat{\theta}\hat{\phi}(\mathbf{x}_i)\hat{h}_i\Big\{\exp\{\hat{h}_i\}\big[(\hat{\theta}-1)(1-\hat{g}_i^{-1})\hat{g}_i^{-1}-1\big]+\hat{h}_i^{-1}+1\Big\}\Big]\Big\} & \text{if} \quad i\epsilon F, \\
S_x\hat{\sigma}^{-2}h_i\exp\{\hat{h}_i\}\hat{\phi}(\mathbf{x}_i)\hat{\theta}(1-\hat{g}_i)\hat{g}_i^{(\hat{\theta}-1)} \\
\quad\Big\{\exp\{\hat{h}_i\}\big[(\hat{\theta}-1)(1-\hat{g}_i^{-1})\hat{g}_i^{-1}-1\big]+\hat{h}_i^{-1}+1\Big\} & \text{if} \quad i\epsilon C.
\end{cases}
$$

The entries of the matrix $\boldsymbol{\Delta_\gamma}$ $(p_1\times n)$ can be expressed as

$$
\Delta_{ki} = \begin{cases}
-x_{ik}S_y\hat{\sigma}^{-1}\hat{\theta}\exp\{\hat{h}_i\}\hat{\phi}(\mathbf{x}_i)(1-\hat{g}_i)\hat{g}_i^{(\hat{\theta}-1)} & \text{if} \quad i\epsilon F, \\
-x_{ik}S_y\hat{\sigma}^{-1}\hat{\theta}\exp\{\hat{h}_i\}\hat{\phi}(\mathbf{x}_i)(1-\hat{g}_i)\hat{g}_i^{(\hat{\theta}-1)} & \text{if} \quad i\epsilon C.
\end{cases}
$$

Furthermore, the elements the matrix $\boldsymbol{\Delta_\beta}$, $(p_2\times n)$ can be expressed as

$$
\Delta_{ji} = \begin{cases}
z_{ij}S_x\hat{\sigma}^{-2}\exp\{\hat{h}_i\}\Big\{1-(1-\hat{g}_i)\hat{g}_i^{\hat{\theta}-1}\Big[(\hat{\theta}-1)(\hat{g}_i+\exp\{\hat{h}_i\})\hat{g}_i^{(\hat{\theta}-1)}+ \\
\quad\hat{\theta}\hat{\phi}(\mathbf{x}_i)\Big\{\exp\{\hat{h}_i\}\big[(\hat{\theta}-1)\hat{g}_i^{-1}(1-\hat{g}_i)-1\big]+1\Big\}\Big]\Big\} & \text{if} \quad i\epsilon F, \\
z_{ij}S_x\hat{\sigma}^{-2}\hat{\theta}\hat{\phi}(\mathbf{x}_i)\exp\{\hat{h}_i\}(1-\hat{g}_i)\hat{g}_i^{\hat{\theta}-1}\Big\{\exp\{\hat{h}_i\}\big[(\hat{\theta}-1)\hat{g}_i^{-1}(1-\hat{g}_i)-1\big]+1\Big\} & \text{if} \quad i\epsilon C.
\end{cases}
$$

## Explanatory variable perturbation (Cure rate)

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

In this case the log-likelihood function takes the form

$$
\begin{aligned}
l(\boldsymbol{\alpha}|\boldsymbol{\omega}) &= \sum_{i\in F}\log\big[\phi^*(\mathbf{x}_i)\big]+r\log(\theta)-r\log(\sigma)+\sum_{i\in F}h_i- \\
&\quad \sum_{i\in F}\exp\{h_i\}+(\theta-1)\sum_{i\in F}\log\Big[1-\exp\big\{-\exp(h_i)\big\}\Big]- \\
&\quad \sum_{i\in F},\phi^*(\mathbf{x}_i)\Big[1-\exp\big\{-\exp(h_i)\big\}\Big]^\theta-\sum_{i\in C}\phi^*(\mathbf{x}_i)\Big[1-\exp\big\{-\exp(h_i)\big\}\Big]^\theta, \quad (4.5)
\end{aligned}
$$

where $\phi^*(\mathbf{x}_i)=\exp\{\mathbf{x}_i^{*T}\}=\gamma_0+\gamma_1 x_{i1}+\ldots+\gamma_t(x_{it}+\omega_i S_x)+\ldots+\gamma_{p_1}x_{ip_1}$.

In addition, the elements of the vector $\boldsymbol{\Delta}_\theta$ are expressed as

$$\Delta_i = \left\{ \begin{array}{ll} -\hat{\gamma}_t S_x \hat{g}_i^{\hat{\theta}} \log(\hat{g}_i) & \text{if } i = 1, 2, \ldots, n. \end{array} \right.$$

The elements of vector $\boldsymbol{\Delta}_\sigma$ are expressed as

$$\Delta_i = \left\{ \begin{array}{ll} \hat{\gamma}_t S_x \hat{\theta} \hat{\sigma}^{-1} \hat{h}_i \exp\{\hat{h}_i\}(1 - \hat{g}_i) \hat{g}_i^{\hat{\theta}} & \text{if } i = 1, 2, \ldots, n. \end{array} \right.$$

The elements the matrix $\boldsymbol{\Delta}_\gamma$ may be expressed when $k \neq t$

$$\Delta_{ki} = \left\{ \begin{array}{ll} -x_{ik} \hat{\gamma}_t S_x \left[ \hat{\phi}(\mathbf{x}_i) \right]^{-2} & \text{if } i \epsilon F, \\ 0 & \text{if } i \epsilon C. \end{array} \right.$$

The elements of the vector $\boldsymbol{\Delta}\gamma$, when $k = t$ are given by

$$\Delta_{ti} = \left\{ \begin{array}{ll} S_x \left\{ \hat{\phi}^{-1}(\mathbf{x}_i) \left[ -x_{it} \hat{\gamma}_t \hat{\phi}^{-1}(\mathbf{x}_i) + 1 \right] - \hat{g}_i \right\} & \text{if } i \epsilon F, \\ -S_x \hat{g}_i^{\hat{\theta}} & \text{if } i \epsilon C. \end{array} \right.$$

The elements the matrix $\boldsymbol{\Delta}_\beta$ can be expressed

$$\Delta_{ji} = \left\{ \begin{array}{ll} z_{ij} \hat{\gamma}_t S_x \hat{\theta} \hat{\sigma}^{-1} \exp\{\hat{h}_i\}(1 - \hat{g}_i) \hat{g}_i^{(\hat{\theta}-1)} & \text{if } i = 1, 2, \ldots, n. \end{array} \right.$$

## Explanatory variable perturbation (Failure Time $T$)

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

$$
\begin{aligned}
l(\boldsymbol{\alpha}|\boldsymbol{\omega}) = & \sum_{i \in F} \log \left[ \phi(\mathbf{x}_i) \right] + r \log(\theta) - r \log(\sigma) + \sum_{i \in F} h_i^* - \\
& \sum_{i \in F} \exp\{h_i^*\} + (\theta - 1) \sum_{i \in F} \log \left[ 1 - \exp\left\{ -\exp(h_i^*) \right\} \right] - \\
& \sum_{i \in F} \phi(\mathbf{x}_i) \left[ 1 - \exp\left\{ -\exp(h_i^*) \right\} \right]^\theta - \sum_{i \in C} \phi(\mathbf{x}_i) \left[ 1 - \exp\left\{ -\exp(h_i^*) \right\} \right]^\theta \quad (4.6)
\end{aligned}
$$

where $h_i^* = \sigma^{-1}(y_i - \mathbf{z}_i^{*T} \boldsymbol{\beta})$ and $\mathbf{z}_i^{*T} = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \cdots + \beta_t(z_{it} + \omega_i S_t) + \cdots + \beta_{p_2} x_{ip_2}$.

In addition, the elements of the vector $\boldsymbol{\Delta}_\theta$ are expressed as

$$\Delta_i = \left\{ \begin{array}{ll} -\hat{\beta}_t S_x \hat{\sigma}^{-1} \exp\{\hat{h}_i\}(1 - \hat{g}_i) \hat{g}_i^{(\hat{\theta}-1)} \left[ \hat{g}_i^{-\hat{\theta}} + \hat{\theta} \log(\hat{g}_i) + 1 \right] & \text{if } i \epsilon F, \\ -\hat{\beta}_t S_x \hat{\sigma}^{-1} \exp\{\hat{h}_i\}(1 - \hat{g}_i) \hat{g}_i^{(\hat{\theta}-1)} \left[ \hat{\theta} \log(\hat{g}_i) + 1 \right] & \text{if } i \epsilon C. \end{array} \right.$$

The elements of vector $\mathbf{\Delta}_\sigma$ are expressed as

$$
\Delta_i = \begin{cases}
\beta_t S_x \hat{\sigma}^{-2} \Big\{ 1 - \exp\{\hat{h}_i\}(1 + \hat{h}_i) + (1 - \hat{g}_i)\hat{g}_i^{(\hat{\theta}-2)} \exp\{\hat{h}_i\} \Big[(\hat{\hat{\theta}} - 1)\hat{h}_i \hat{g}_i^{-\theta} \big\{ \exp\{\hat{h}_i\} + \\
\hat{g}_i(1 + \hat{h}_i^{-1}) \big\} - \hat{\theta}\hat{\phi}(\mathbf{x}_i)\big\{ \hat{h}_i \exp\{\hat{h}_i\}(\hat{\theta} - 1)(1 - \hat{g}_i) + \hat{g}_i(\hat{h}_i \exp\{\hat{h}_i\} + \hat{h}_i + 1)\big\}\Big]\Big\} & \text{if}\quad i\epsilon F, \\[2mm]
-\hat{\beta}_t \hat{\theta} S_x \hat{\sigma}^{-2} \exp\{\hat{h}_i\}\hat{\phi}(\mathbf{x}_i)(1 - \hat{g}_i)\hat{g}_i^{\hat{\theta}-2} \\
\Big\{(\hat{\theta} - 1)h_i \exp\{\hat{h}_i\}(1 - \hat{g}_i) + \hat{g}_i\Big[\hat{h}_i \exp\{\hat{h}_i\} + \hat{h}_i + 1\Big]\Big\} & \text{if}\quad i\epsilon C.
\end{cases}
$$

The elements the matrix $\mathbf{\Delta}_\beta$ may be expressed when $j \neq t$,

$$
\Delta_{ji} = \begin{cases}
z_{ij}\hat{\beta}_t S_x \hat{\sigma}^{-2} \exp\{\hat{h}_i\}(1 - \hat{g}_i)\Big\{(1 - \hat{g}_i)^{-1} + (\hat{\theta} - 1)\big[\hat{g}_i + \exp\{\hat{h}_i\}\big] - \\
\hat{\theta}\hat{\phi}(\mathbf{x}_i)\hat{g}_i\Big[\hat{\theta}\exp\{\hat{h}_i\}(1 - \hat{g}_i) + \hat{g}_i\big(1 - \exp\{\hat{h}_i\}\big)\Big]\Big\} & \text{if}\quad i\epsilon F, \\[2mm]
-z_{ij}\hat{\beta}_t S_x \hat{\theta}\hat{\sigma}^{-2}\hat{\phi}(\mathbf{x}_i)\exp\{\hat{h}_i\}(1 - \hat{g}_i)\hat{g}_i\Big[\hat{\theta}\exp\{\hat{h}_i\}(1 - \hat{g}_i) + \hat{g}_i\big(1 - \exp\{\hat{h}_i\}\big)\Big] & \text{if}\quad i\epsilon C.
\end{cases}
$$

The elements of the vector $\mathbf{\Delta}_\beta$, when $j = t$ are given by

$$
\Delta_{ti} = \begin{cases}
S_x \hat{\sigma}^{-1}\Big\{ -1 - \exp\{\hat{h}_i\}\Big[\hat{z}_{it}\hat{\beta}_t + \hat{\theta}\hat{\phi}(\mathbf{x}_i)(1 - \hat{g}_i)\hat{g}_i^{(\hat{\theta}-1)} + 1\Big] - \\
\hat{h}_i(\hat{\theta} - 1)(1 - \hat{g}_i)\hat{g}_i^{-1}\Big\} + z_{it}\hat{\beta}_t S_x \hat{\sigma}^{-2}\exp\{\hat{h}_i\}(1 - \hat{g}_i) \\
\Big\{(\hat{\theta} - 1)\big[\hat{g}_i + \exp\{\hat{h}_i\}\big] + \hat{\theta}\hat{\phi}(\mathbf{x}_i)\hat{g}_i\Big[\exp\{\hat{h}_i\}\big[\hat{\theta}(1 - \hat{g}_i) - \hat{g}_i\big] + \hat{g}_i\Big]\Big\} & \text{if}\quad i\epsilon F, \\[2mm]
S_x \hat{\theta}\hat{\sigma}^{-1}\hat{\phi}(\mathbf{x}_i)\exp\{\hat{h}_i\}\hat{g}_i^{(\hat{\theta}-1)}(1 - \hat{g}_i) & \text{if}\quad i\epsilon C.
\end{cases}
$$

The elements the matrix $\mathbf{\Delta}_\gamma$ may be expressed

$$
\Delta_{ki} = \Big\{ \; x_{ik} S_x \hat{\beta}_t \hat{\theta}\hat{\sigma}^{-1}\hat{\phi}(\mathbf{x}_i)\exp\{\hat{h}_i\}(1 - \hat{g}_i)\hat{g}_i^{\hat{\theta}-1} \quad \text{if}\quad i = 1, 2, \ldots, n,
$$

where $\hat{g}_i = 1 - \exp\{-\exp(\hat{h}_i)\}$, $\hat{h}_i = \frac{y_i - \mathbf{z}_i^T \hat{\boldsymbol{\beta}}}{\hat{\sigma}}$, $k = 0, 1, \ldots, p_1$, $j = 0, 1, \ldots, p_2$, and $i = 1, 2, \ldots, n$.

## 5. Residual Analysis

In order to study departures from the error assumptions as the well as presence of outlying observations, we will consider two kinds of residuals: deviance component residual (see, for instance, McCullagh and Nelder, 1989) and martingale-type residual (see for instance, Barlow and Prentice, 1988; Therneau *et al.*, 1990). More details can be found in Ortega *et al.* (2003).

## 5.1 Martingale-type and deviance modified components residuals

This martingale-type modified residual was introduced in counting processes and can be written in log-exponentiated-Weibull regression models with rate cure as

$$
r_{M_i} = \begin{cases} 1 - \log\left[\exp\{-\hat{\phi}(\mathbf{x}_i)\hat{g}_i^{\hat{\theta}}\}\right] + \log\left[1 - \exp\{-\hat{\phi}(\mathbf{x}_i)\}\right] & \text{if} \quad i\epsilon F, \\ \log\left[1 - \exp\{-\hat{\phi}(\mathbf{x}_i)\}\right] - \log\left[\exp\{-\hat{\phi}(\mathbf{x}_i)\hat{g}_i^{\hat{\theta}}\}\right] & \text{if} \quad i\epsilon C. \end{cases} \tag{5.1}
$$

More details about counting processes can be found, for instance, in Fleming and Harrington (1991) and Ortega (2001). These authors show that the distribution of the deviance component residual based on the martingale residual has very close asymptotic distribution to the normal distribution.

Therefore, we have that the deviance component residual for log-exponentiated-Weibull regression models with rate cure becomes

$$
r_{D_i} = \begin{cases} \text{sgn}(r_{M_i})\sqrt{2}\{r_{M_i} + \log(1 - r_{M_i})\}^{\frac{1}{2}} & \text{if} \quad i\epsilon F, \\ \text{sgn}(r_{M_i})\sqrt{2}\{r_{M_i}\}^{\frac{1}{2}} & \text{if} \quad i\epsilon C. \end{cases} \tag{5.2}
$$

where $\hat{g}_i = 1 - \exp\{-\exp(\hat{h}_i)\}$, $\hat{h}_i = \hat{\sigma}^{-1}(y_i - \mathbf{z}_i^T\hat{\boldsymbol{\beta}})$, $\hat{\phi}(\mathbf{x}_i) = \exp\{\mathbf{x}_i^T\hat{\boldsymbol{\gamma}}\}$.

## 5. Residual Analysis

In order to study departures from the error assumptions as the well as presence of outlying observations, we will consider two kinds of residuals: deviance component residual (see, for instance, McCullagh and Nelder, 1989) and martingale-type residual (see for instance, Barlow and Prentice, 1988; Therneau *et al.*, 1990). More details can be found in Ortega *et al.* (2003).

## 5.1 Martingale-type and deviance modified components residuals

This martingale-type modified residual was introduced in counting processes and can be written in log-exponentiated-Weibull regression models with rate cure as

$$
r_{M_i} = \begin{cases} 1 - \log\left[\exp\{-\hat{\phi}(\mathbf{x}_i)\hat{g}_i^{\hat{\theta}}\}\right] + \log\left[1 - \exp\{-\hat{\phi}(\mathbf{x}_i)\}\right] & \text{if} \quad i\epsilon F, \\ \log\left[1 - \exp\{-\hat{\phi}(\mathbf{x}_i)\}\right] - \log\left[\exp\{-\hat{\phi}(\mathbf{x}_i)\hat{g}_i^{\hat{\theta}}\}\right] & \text{if} \quad i\epsilon C. \end{cases} \tag{5.1}
$$

More details about counting processes can be found, for instance, in Fleming and Harrington (1991) and Ortega (2001). These authors show that the distribution of the deviance component residual based on the martingale residual has very close asymptotic distribution to the normal distribution.

Therefore, we have that the deviance component residual for log-exponentiated-Weibull regression models with rate cure becomes

$$r_{D_i} = \begin{cases} \operatorname{sgn}(r_{M_i})\sqrt{2}\{r_{M_i} + \log(1 - r_{M_i})\}^{\frac{1}{2}} & \text{if} \quad i\epsilon F, \\ \operatorname{sgn}(r_{M_i})\sqrt{2}\{r_{M_i}\}^{\frac{1}{2}} & \text{if} \quad i\epsilon C. \end{cases} \quad (5.2)$$

where $\hat{g}_i = 1 - \exp\{-\exp(\hat{h}_i)\}$, $\hat{h}_i = \hat{\sigma}^{-1}(y_i - \mathbf{z}_i^T\hat{\boldsymbol{\beta}})$, $\hat{\phi}(\mathbf{x}_i) = \exp\{\mathbf{x}_i^T\hat{\boldsymbol{\gamma}}\}$.

## 6. Application

In this section, the application of the local influence theory to a set of real data on cancer recurrence is discussed. The data are part of an assay on cutaneous melanoma (a type of malignant cancer) for the evaluation of postoperative treatment performance with a high dose of a certain drug (interferon alfa-2b) in order to prevent recurrence. Patients were included in the study from 1991 to 1995, and follow-up was conducted until 1998. The data were collected by Ibrahim *et al.* (2001b). This data set has recently been analyzed by Mizoi (2004), using a Weibull model with cure fraction. The data present the survival times, T, as the time until the patient's death. The original size of the sample was $n = 427$ patients, 10 of which did not present a value for covariate tumor thickness. When such cases were removed, a sample of size $n = 417$ patients was retained. The percentage of censored observations is 56%. The following variables were associated with participant $i$, $i = 1, 2, \ldots, 417$.

- $t_i$: observed time (in years);

- $x_{i1}$: treatment (0= observation, 1=interferon);

- $x_{i2}$: age (in years);

- $x_{i3}$: nodule (nodule category: 1 to 4);

- $x_{i4}$: sex (0=male, 1=female);

- $x_{i5}$: p.s. (performance status-patient's functional capacity scale as regards his daily activities: 0=fully active, 1=other);

- $x_{i6}$: tumor (tumor thickness in mm.).

The survival function graph, Kaplan-Meier estimate, is presented in Figure 2, from where a significant fraction of survivors can be observed.
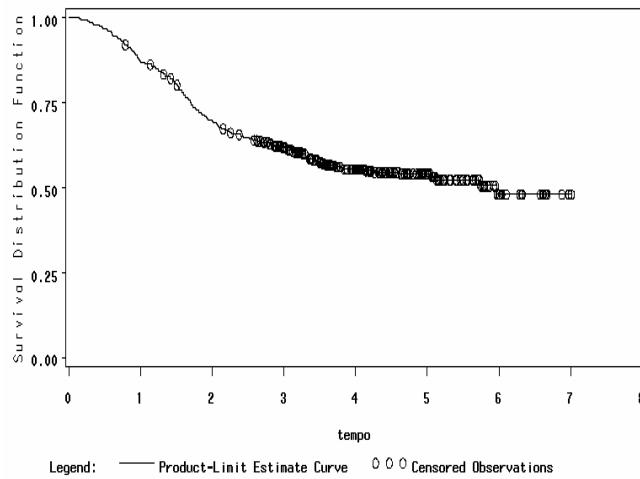
Figure 2: Plot of the Survivor Function for the melanoma data

Firstly, we consider the following regression model:

$$y_i = \log(t_i) = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \sigma \varepsilon_i, \ i = 1, \ldots, 417,$$

where $\varepsilon_i$ are independent random variable with common probability density function given by (3.5) and $\phi(\mathbf{x}_i) = \exp(\gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_3 x_{i3} + \gamma_4 x_{i4} + \gamma_5 x_{i5} + \gamma_6 x_{i6})$, where $y_i$ denotes the lifetime logarithm.

## 6.1 Maximum likelihood results

To obtain the maximum likelihood estimates (MLEs) for the parameters in the log-exponentiate-Weibull regression model we use the subroutine MAXBFGS in Ox, whose results are given in the following Table 1.

We note the covariate *treatment* is significative (at 1%) in the log of time $T$, the predictor *nodule* is significative (at 5%) for both in the log of time and cure fraction, also the predictor *sex* is significative (at 10%).

The value of the likelihood ratio to test the null hypothesis $H_{01} : \theta = 1$ in (3.9) provides $\Lambda_n = 18.772$, which clearly is significante at the 5%, with critical value $\chi^2_{1,0.05} = 3.841$. Clearly, the LER-CR model is also not adequate, since for testing $H_0 : \theta = 1, \sigma = 1$, the observed value of likelihood ratio statistics is $\Lambda_n = 49.104$ (2 degrees of freedom) with p-value $\approx 0$. Thus, the likelihood ratio test indicates that the LWER-CR model presents a much better fit that the LWR-CR to the data set under considerations.

Table 1: Maximum likelihood estimates for the log-exponentiated-Weibull regression model with cure rate

| Parameter | Estimate | SE | p-value |
|-----------|----------|------|---------|
| $\sigma$ | 1.5567 | 0.2745 | - |
| $\theta$ | 6.2222 | 2.030 | - |
| $\beta_{treatment}$ | 0.4088 | 0.1586 | 0.0099 |
| $\beta_{age}$ | -0.0031 | 0.0057 | 0.5833 |
| $\beta_{nodule}$ | -0.1868 | 0.0778 | 0.0164 |
| $\beta_{sex}$ | -0.2674 | 0.1611 | 0.0970 |
| $\beta_{p.s}$ | -0.0729 | 0.2159 | 0.7355 |
| $\beta_{tumor}$ | 0.0164 | 0.0227 | 0.4697 |
| $\gamma_0$ | -1.4227 | 0.4868 | 0.0035 |
| $\gamma_{treatment}$ | 0.3393 | 0.1892 | 0.0730 |
| $\gamma_{age}$ | 0.0076 | 0.0071 | 0.2837 |
| $\gamma_{nodule}$ | 0.2523 | 0.0898 | 0.0049 |
| $\gamma_{sex}$ | -0.3196 | 0.1890 | 0.0908 |
| $\gamma_{p.s}$ | 0.0940 | 0.2523 | 0.7094 |
| $\gamma_{tumor}$ | 0.0356 | 0.0276 | 0.1970 |

Table 2: Posterior summaries for the log-exponentiated-Weibull regression model with cure rate.

| Parameter | Mean | Median | S.D. | 2,5% | 97.5% | $\hat{R}$ |
|-----------|------|--------|------|------|-------|-----------|
| $\sigma$ | 1.632 | 1.613 | 0.2998 | 1.109 | 2.271 | 1.001 |
| $\theta$ | 5.936 | 5.725 | 1.869 | 2.917 | 10.14 | 1.008 |
| $\beta_{treatment}$ | 0.4627 | 0.4623 | 0.1853 | 0.0952 | 0.8342 | 1.002 |
| $\beta_{age}$ | -0.0018 | -0.0018 | 0.0062 | -0.0138 | 0.0107 | 1.007 |
| $\beta_{nodule}$ | -0.1868 | -0.1844 | 0.0827 | -0.3543 | -0.0317 | 1.004 |
| $\beta_{sex}$ | -0.3008 | -0.2961 | 0.1836 | -0.6857 | 0.0456 | 1.006 |
| $\beta_{p.s}$ | -0.109 | -0.1056 | 0.1815 | -0.4726 | 0.2467 | 1.000 |
| $\beta_{tumor}$ | 0.0206 | 0.0199 | 0.0263 | -0.0291 | 0.0742 | 1.001 |
| $\gamma_0$ | -1.456 | -1.489 | 0.4593 | -2.292 | -0.5273 | 1.019 |
| $\gamma_{treatment}$ | 0.415 | 0.415 | 0.2124 | 0.0082 | 0.8523 | 1.001 |
| $\gamma_{age}$ | 0.0094 | 0.0095 | 0.0072 | -0.0047 | 0.02397 | 1.001 |
| $\gamma_{nodule}$ | 0.246 | 0.2477 | 0.09103 | 0.05907 | 0.4167 | 1.011 |
| $\gamma_{sex}$ | -0.3494 | -0.3432 | 0.1994 | -0.7444 | 0.03491 | 1.003 |
| $\gamma_{p.s}$ | 0.04082 | 0.04001 | 0.03024 | -0.01716 | 0.1022 | 1.012 |
| $\gamma_{tumor}$ | 0.02779 | 0.05206 | 3.154 | -6.206 | 6.223 | 1.071 |

## 6.2 Bayesian analysis

We consider now a Bayesian analysis for the data set, using minimal prior information. The following independent priors were considered to perform the Gibbs sampler: $\beta_i \sim N(0, 100)$, $i = 1, \ldots, 6$   $\gamma_i \sim N(0, 100)$, $i = 0, 1, \ldots, 6$

$\sigma \sim IG(1, 0.01)$, $\theta \sim G(1, 0.01)$, so that we have a vague prior distribution. Considering these prior densities we generated two parallel independent runs of the Gibbs sampler chain with size 50000 for each parameter, disregarding the first 10000 iterations to eliminate the effect of the initial values and to avoid correlation problems, we considered a spacing of size 10, obtaining a sample of size 4000 from each chain. To monitor the convergence of the Gibbs samples we used the between and within sequence information, following the approach developed in Gelman and Rubin (1992) to obtain the potential scale reduction, $\hat{R}$. In all cases, these values were close to one, indicating chain convergence. In Table 2 we report posterior summaries for the parameters of the log-exponentiated-Weibull regression model with cure rate.

To compare the LEWR-CR model and LWR-CR model fits by inspecting the estimated of the expected value of Akaike's Information Criterion (EAIC), the expected value of Bayesian Information Criterion (EBIC) and Deviance information criterion (DIC) (see, Spiegelhalter *et al.*, 2002), all put together in Table 3. According to EAIC, EBIC and DIC, the log-exponentiated-Weibull regression (LEWR-CR) model improves the corresponding log-Weibull regression (LWR-CR) model. Similar conclusions can be made analyzing the credibility interval for the parameter $\theta$ under LEWR-CR model since it does not contain $\theta = 1$.

Table 3: EAIC, EBIC and DIC criteria

| Model | EAIC | EBIC | DIC |
|---|---|---|---|
| LEWR-CR | 903.900 | 964.396 | 888.578 |
| LWR-CR | 920.600 | 977.063 | 904.946 |

## 6.3 Local influence analysis

In this section, we will make an analysis of local influence for the cancer data.

### Case-weights perturbation

By applying the local influence theory developed in Section 4, where case-weight perturbation is used, value $C_{\mathbf{d}_{max}}(\boldsymbol{\alpha}) = 1.5853$, $C_{\mathbf{d}_{max}}(\boldsymbol{\gamma}) = 1.2163$ and $C_{\mathbf{d}_{max}}(\boldsymbol{\beta}) = 1.4495$ was obtained as maximum curvature. In Figure 3, the graph for the eigenvector corresponding to $|\mathbf{d}_{max}(\boldsymbol{\alpha})|$, $|\mathbf{d}_{max}(\boldsymbol{\gamma})|$ and $|\mathbf{d}_{max}(\boldsymbol{\beta})|$ for all points is presented. Clearly, the most influential is observation 341 on $\boldsymbol{\alpha}$ (See Figure 3(a)). But marginally we noticed that the observation 47 can be influential on $\boldsymbol{\gamma}$ and the observation 341 on $\boldsymbol{\beta}$. We also mentioned that the observation 47 introduce one of the largest lifetimes and the observation 341 presents the smallest lifetime.
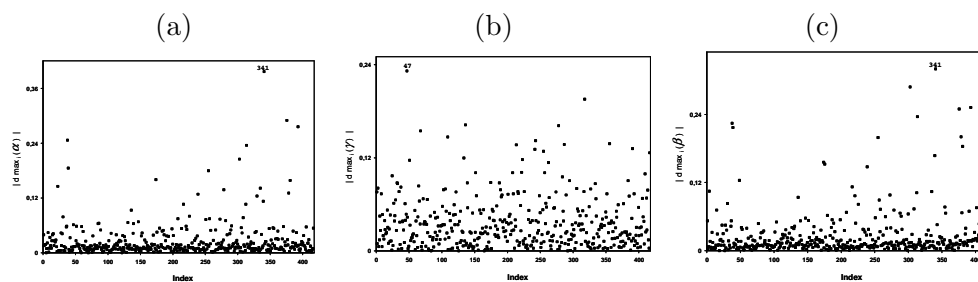
(a) (b) (c)



Figure 3: Case-weights perturbation. (a) Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\alpha}$. (b) Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\gamma}$. (c) Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\beta}$.

(d) (e) (f)
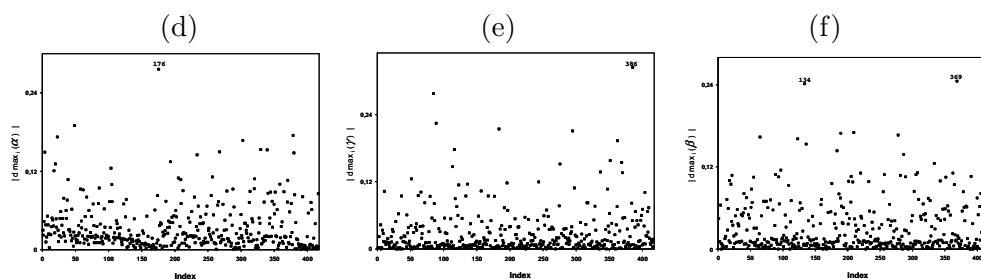


Figure 4: Response perturbation. (d) Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\alpha}$. (e) Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\gamma}$. (f) Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\beta}$.

**Influence using response variable perturbation**

Next, the influence of perturbations on the observed survival times will be analyzed. The value for the maximum curvature calculated was $C_{\mathbf{d}_{max}}(\boldsymbol{\alpha}) = 19.303$, $C_{\mathbf{d}_{max}}(\boldsymbol{\gamma}) = 1.3153$ and $C_{\mathbf{d}_{max}}(\boldsymbol{\beta}) = 6.6636$. Figure 4, containing the graph for $|\mathbf{d}_{max}(\boldsymbol{\gamma})|$, $|\mathbf{d}_{max}(\boldsymbol{\alpha})|$ and $|\mathbf{d}_{max}(\boldsymbol{\beta})|$ for all points. Results in Figure 4(d) suggest that the observation 176 is the most influential on $\boldsymbol{\alpha}$. But marginally we noticed that the observation 386 can be influential on $\boldsymbol{\gamma}$ and the observations 134 and 369 on $\boldsymbol{\beta}$.

**Influence using explanatory variable perturbation**

The perturbation of the covariables age($x_2$) and tumor($x_6$) is investigated here. After perturbation of covariable age, value $C_{\mathbf{d}_{max}}(\boldsymbol{\alpha}) = 1.1243$, $C_{\mathbf{d}_{max}}(\boldsymbol{\gamma}) = 1.1012$ and $C_{\mathbf{d}_{max}}(\boldsymbol{\beta}) = 0.4667$ was obtained as maximum curvature, and after perturbation of covariable tumor, values $C_{\mathbf{d}_{max}}(\boldsymbol{\alpha}) = 1.0351$, $C_{\mathbf{d}_{max}}(\boldsymbol{\gamma}) = 0.9418$ and $C_{\mathbf{d}_{max}}(\boldsymbol{\beta}) = 0.8326$ were achieved. The respective index plot of $|\mathbf{d}_{max}(\boldsymbol{\alpha})|$, $|\mathbf{d}_{max}(\boldsymbol{\gamma})|$ and $|\mathbf{d}_{max}(\boldsymbol{\beta})|$ against the observation index are shown in Figures 5 and 6. Results in Figura 5(g) and 6(j) suggest that the observations 176 and 351

are the most influential in $\boldsymbol{\alpha}$. But marginally we noticed that the observation 47, 196 and 351 can be influential on $\gamma$ and the observation 176 on $\beta$
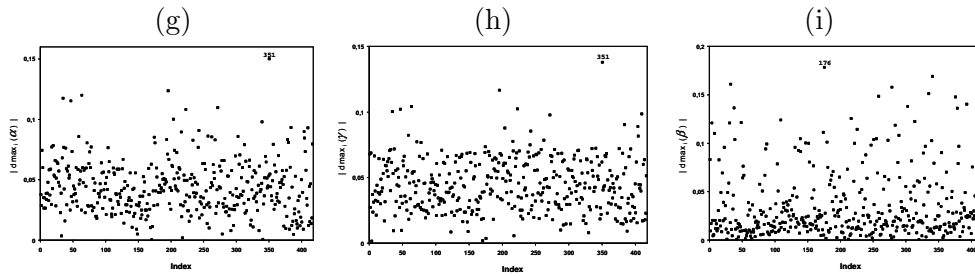


Figure 5: Age explanatory variable perturbation. (g) Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\alpha}$. (h) Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\gamma}$. (i) Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\beta}$.
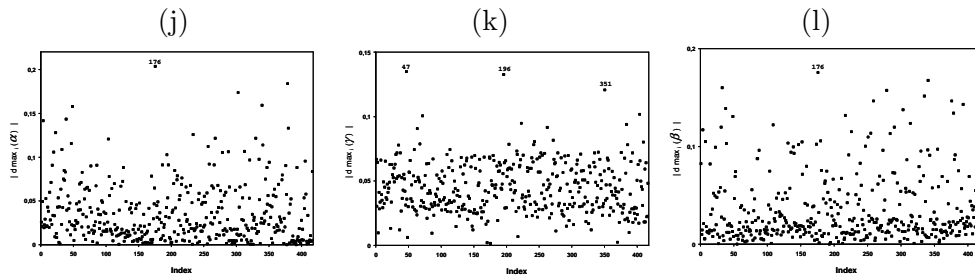


Figure 6: Tumor explanatory variable perturbation. (j) Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\alpha}$. (k) Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\gamma}$. (l) Index plot of $\mathbf{d}_{max}$ for $\boldsymbol{\beta}$.

## 6.4 Residual analysis

In order to detect possible outlying observations as well as departures from the assumptions of the log-exponentiated-Weibull regression models with rate cure, we present in Figure 7, the graphs of $r_{Di}$ against the order observations.

By analyzing the residuals deviance graph, a random behavior is observed for the datas. As we can observe through the local influence analysis and analysis and the residual analysis it doesn't exist jointly influential points. Thus, the final model becomes the one given by

$$y_i = \beta_1 x_{i1} + \beta_3 x_{i3} + \beta_4 x_{i4} + \sigma\varepsilon_i, \quad \phi(\mathbf{x}_i) = \exp\{\gamma_0 + \gamma_1 x_{i1} + \gamma_3 x_{i3} + \gamma_4 x_{i4}\} \ (6.1)$$
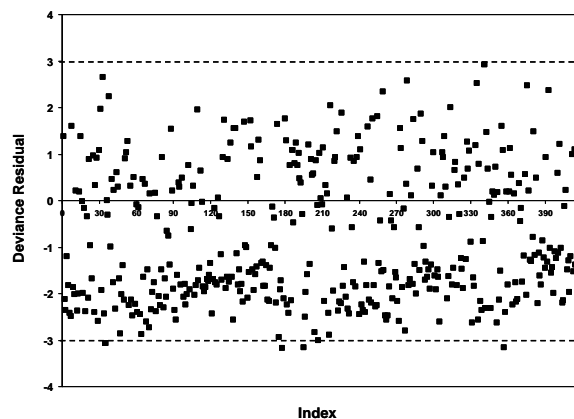
Figure 7: Index plot of the deviance

The parameters MLEs are reported in the Table 4.

Table 4: Maximum likelihood estimates for the log-exponentiated-Weibull regression model with cure rate

| Parameter | Estimate | SE | p-value |
|-----------|----------|-----|---------|
| $\sigma$ | 1.4990 | 0.1887 | - |
| $\theta$ | 5.7294 | 1.1124 | - |
| $\beta_{treatment}$ | 0.4247 | 0.1509 | 0.0049 |
| $\beta_{nodule}$ | -0.1937 | 0.0704 | 0.0059 |
| $\beta_{sex}$ | -0.2698 | 0.1593 | 0.0905 |
| $\gamma_0$ | -0.8691 | 0.2929 | 0.0031 |
| $\gamma_{treatment}$ | 0.3032 | 0.1843 | 0.0998 |
| $\gamma_{nodule}$ | 0.2459 | 0.0871 | 0.0047 |
| $\gamma_{sex}$ | -0.3514 | 0.1863 | 0.0592 |

We note the covariate *treatment* and *nodule* is significative (at 5%) in the log of time $T$, the predictor *nodule* and *sexo* is significative (at 5%) in the cure fraction. The covariate *nodule* is significative for both in the log of time and cure fraction, also the predictor *sex* and *treatment* is significative (at 10%) in the log of time and cure fraction respectively.

We may interpret the estimated coefficients of the final model as following. The predictor nodule decelerates the lifetime of individual's and alters the cured proportion significantly, and that significant difference exists among the levels of treatments (observation and interferon) in relation log of time.

## 7. Concluding Remarks

In this study, the log-exponentiated-Weibull regression (LEWR) model was modified in order to include long-term individuals. In the proposal under consideration, log-linear parametric modeling was taken as a basis for survival time. The model attempts to estimate simultaneously the covariates effects on the accelaration/decelaration of the timing of a given event and surviving fraction, that is, the proportion of the population for which the event never occurs.

Continuing with modeling investigation, we applied local influence theory (Cook (1986) and Thomas and Cook (1990)) and conducted a study based on martingale and deviance residuals in a survival model with a cure fraction. The necessary matrices for application of the technique were obtained by taking into account various types of perturbations to the data elements and to the model. By applying such results to a data set, indication was found of which observations or set of observations would sensitively influence the analysis results. This fact is illustrated in application (Section 6). By means of a real data set, it was observed that, for some perturbation schemes, the presence of certain observations could considerably change the levels of significance of certain variables. The results of the applications indicate that the use of the local influence technique in models with a cure fraction may be rather useful in the detection of possibly influential points by admitting two types of estimation methodology: maximum likelihood and Bayesian. In order to measure the quality of fitting, martingale and deviance residuals were used, which showed that the model fitting was correct. Finally we can observe that the log-exponentiated-Weibull regression models with rate considered is a robust model.

### Acknowledgment

### References

Barlow, W. E. and Prentice, R. L. (1988). Residuals for relative risk regression. *Biometrika*, **75**, 65-74.

Berkson, J. and Gage, R. P. (1952). Survival cure for cancer patients following treatment. *Journal of the American Statistical Association*, **47**, 501-515.

Cancho, V. G. and Bolfarine, H. (2001). Modeling the presence of immunes by using the exponentiated-Weibull model. *J. Appl. Stat.*, **28**, 659-671.

Cancho, V. G., Bolfarine, H., and Achcar, J. A. (1999). A Bayesian analysis for the exponentiated-Weibull distribution. *J. Appl. Stat. Sci.*, **8**, 227-242.

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

Cook, R. D. (1986). Assessment of local influence. *J. Roy. Statist. Soc. Ser. B*, **48**, 133-169. With discussion.

De Castro, A. F. M., Cancho, V. G., and Rodrigues, J. (2007). A flexible model for survival data with a surviving fraction. Technical Report 245, Des-Universidade de Federal de Sã Carlos.

Díaz-García, J. A., Galea Rojas, M., and Leiva-Sánchez, V. (2003). Influence diagnostics for elliptical multivariate linear regression models. *Commun. Stat., Theory Methods*, **32**, 625-641.

Doornik, J. A. (2001). *Ox 3.0 : an object-oriented matrix programming language*, 4th edition. Timberlake Consultants Ltd.

Escobar, L. A. and Meeker, W. Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics*, **48**, 507-528.

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

Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis.* John Wiley & Sons.

Galea, M., Riquelme, M., and Paula, G. A. (2000). Diagnostic methods in elliptical linear regression models. *Braz. J. Probab. Stat.*, **14**, 167-184.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457-511.

Greenhouse, J. B. and Wolfe, R. A. (1984). A competing risks derivation of a mixture model for the analysis of survival data. *Commun. Stat., Theory Methods*, **13**, 3133-3154.

Hoggart, C. and Griffin, J. E. (2001). A bayesian partition model for customer attrition. In *Bayesian Methods with Applications to Science, Policy, and Official Statistics (Select Papers from ISBA 2000)* (Edited by E. I. George), 61-70. Creta, International Society for Bayesian Analysis.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). *Bayesian survival analysis.* Springer-Verlag.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data*, second edition. John Wiley & Sons.

Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570-582.

Maller, R. A. and Zhou, X. (1996). *Survival analysis with long-term survivors*. John Wiley & Sons.

McCullagh, P. and Nelder, J. A. (1983). *Generalized linear models*. Chapman & Hall.

Mizoi, M. F. (2001). *Influência local em modelos de sobrevivência com fracão de cura*. Doctor thesis, IME-University of São Paulo, São Paulo-Brazil. (in Portuguese).

Mudholkar, G. S. and Hutson, A. D. (1996). The exponentiated Weibull family: some properties and a flood data application. *Comm. Statist. Theory Methods*, **25**, 3059-3083.

Mudholkar, G. S., Srivastava, D. K., and Freimer, M. (1995). The exponentiated Weibull family: A reanalysis of the bus-motor-failure data. *Technometrics*, **37**, 436-445.

Nassar, M. M. and Eissa, F. H. (2003). On the exponentiated Weibull distribution. *Comm. Statist. Theory Methods*, **32**, 1317-1336.

Ortega, E. M. M. (2001). *Influence Analysis in Generalized Log-Gamma Regression Models*. Doctor thesis, IME-University of SãoPaulo, São Paulo-Brazil. (in Portuguese).

Ortega, E. M. M., Bolfarine, H., and Paula, G. A. (2003). Influence diagnostics in generalized log-gamma regression models. *Comput. Statist. Data Anal.*, **42**, 165-186.

Ortega, E. M. M., Cancho, V. G., and Lachos, V. H. (2008). Influence diagnostics in the weibull mixutre model with covariates. *Statistics and Operations Research Transactions*, In press.

Pettitt, A. N. and Bin Duad, L. (1898). Case-weighted measures of influence for proportional hazards regression. *Applied Statistics*, **38**, 51-67.

Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, **56**, 227-236.

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

Tsodikov, A. D., Ibrahim, J. G., and Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *J. Amer. Statist. Assoc.*, **98**, 1063-1078.

Yakovlev, A. and Tsodikov, A. D. (1996). *Stochastic models of tumor latency and their biostatistical applications*, volume 1 of *Mathematical Biology and Medicine*. World Scientific, New Yersey.

Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of "permanent employment" in Japan. *Journal of the American Statistical Association*, **87**, 284-292.

Vicente, G. Cancho
Instituto de Ciências Matemáticas e de Computacão
USP - Caixa Postal 668 , 13560-970
São Carlos-SP-Brazil
garibay@icmc.usp.br

Edwin, M. M. Ortega
Departamento de Ciências Exatas - LCE - ESALQ - USP
Caixa-Postal 9, 13418-900
Piracicaba-SP-Brazil
edwin@esalq.usp.br

Heleno Bolfarine
Departamento de Estatística - IME - USP
Caixa Postal 66281, 05315-970 São Paulo-SP-Brazil
hbolfar@ime.usp.br