

## Cure Rate Models Considering The Burr XII Distribution in Presence of Covariate

Emílio A. Coelho-Barros

*Universidade Tecnológica Federal do Paraná, Departamento Acadêmico de Matemática, DAMAT/UTFPR  
Cornélio Procópio, PR, Brazil  
eabarros@utfpr.edu.br*

Jorge Alberto Achcar

*Universidade de São Paulo, Departamento de Medicina Social, FMRP/USP  
Ribeirão Preto, SP - Brazil  
achcar@fmrp.usp.br*

Josmar Mazucheli

*Universidade Estadual de Maringá, Departamento de Estatística, DES/UEM  
Maringá, PR - Brazil  
jmazucheli@uem.br*

Received 5 May 2015

Accepted 5 September 2016

This paper presents estimates for the parameters included in long-term mixture and non-mixture lifetime models, applied to analyze survival data when some individuals may never experience the event of interest. We consider the case where the lifetime data have a three-parameter Burr XII distribution, which includes the popular Weibull mixture model as a special case. Classical and Bayesian procedures are used to get point estimates and confidence intervals for the unknown parameters. We consider a general survival model where the scale and shape parameters of the Burr XII distribution are dependent of some covariates. To illustrate the proposed methodology, we consider an application considering a leukaemia data set where the proposed model gives better fit for the data when compared to other existing models.

*Keywords:* Cure rate mixture model; Cure rate non-mixture model; Burr XII distribution.

### 1. Introduction

A long-term survivor mixture model, also known as standard cure rate model, assumes that the studied population is a mixture of susceptible individuals, who experience the event of interest and non-susceptible individuals that will never experience it. These individuals are not at risk with respect to the event of interest and are considered immune, non-susceptible or cured [1]. Different approaches, parametric and non-parametric, have been considered to model the proportion of immunes and interested readers can refer, for example, to Boag (1949), Berkson and Gage (1952), Haybittle (1965), Farewell (1982, 1986), Meeker (1987), Gamel *et al.* (1990), Ghitany and Maller (1992), Copas and Heydari (1997), Ng and McLachlan (1998), De Angelis *et al.* (1999), Peng

and Dear (2000), Sy and Taylor (2000), Balakrishnan and Pal (2013). Following Maller and Zhou (1996), the standard cure rate model assumes that a certain fraction  $p$  in the population is cured or never fail with respect to the specific cause of death or failure, while the remaining fraction  $(1 - p)$  of the individuals is not cured, leading to the survival function for the entire population written as:

$$S(t) = p + (1 - p)S_0(t), \quad (1.1)$$

where  $p \in (0, 1)$  is the mixing parameter and  $S_0(t)$  denotes a proper survival function for the non-cured group in the population. Considering a random sample of lifetimes  $(t_i, \delta_i, i = 1, \dots, n)$ , under the assumption of right censored lifetime, the contribution of the  $i^{th}$  individual for the likelihood function is:

$$L_i = [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i}, \quad (1.2)$$

where  $\delta_i$  is a censoring indicator variable, that is,  $\delta_i = 1$  for an observed lifetime and  $\delta_i = 0$  for a censored lifetime.

From the mixture survival function, (1.1), the probability density function is obtained from  $f(t_i) = -\frac{d}{dt}S(t_i)$  and given by:

$$f(t_i) = (1 - p)f_0(t_i), \quad (1.3)$$

where  $f_0(t_i)$  is the probability density function for the susceptible individuals. Substitution of the mixture density (1.3) and survival function (1.1) in the standard likelihood function (1.2) yields the likelihood for the long-term survivor mixture model:

$$L_i = [(1 - p)f_0(t_i)]^{\delta_i} [p + (1 - p)S_0(t_i)]^{1-\delta_i}. \quad (1.4)$$

Thus, the log-likelihood considering all observations is given by:

$$l = r \log(1 - p) + \sum_{i=1}^n \delta_i \log f_0(t_i) + \sum_{i=1}^n (1 - \delta_i) \log [p + (1 - p)S_0(t_i)], \quad (1.5)$$

where  $r = \sum_{i=1}^n \delta_i$  is the number of uncensored observations. Common choices for the survival function  $S_0(t)$ , in (1.1), are the exponential and Weibull distributions. Recently, the exponentiated exponential distributions was considered by [16] and [17]. Peng *et al.* (1998) investigated the use of a generalized Fisher-Snedecor distribution as baseline for  $S_0(t)$ . The generalized Fisher-Snedecor distribution is a supermodel that includes the most popular survival models as particular cases, such as the exponential, Weibull, log-normal, among others. Yamaguchi (1992) considered the generalized log-gamma distribution for the mixture cure rate model in the context of accelerated failure-time regression models. The Gompertz distribution was considered by Gieser *et al.* (1998), while the exponentiated Weibull and exponentiated exponential distributions were considered, respectively, by Bolfarine and Cancho (2001) and Kannan *et al.* (2010). The Conway-Maxwell Poisson cure rate model was proposed by Rodrigues *et al.* (2009a) as an alternative to the cure rate model discussed by Yin and Ibrahim (2005). Shao and Zhou (2004) proposed a mixture parametric model for survival data with long-term survivors considering the Burr XII distribution.

An alternative to a long-term survivor mixture model is the long-term survivor non-mixture model suggested by [26–28] which defines an asymptote for the cumulative hazard and hence for

the cure fraction. The survival function for a non-mixture cure rate model is defined as:

$$S(t) = p^{1-S_0(t)}, \tag{1.6}$$

where, like in (1.1),  $p \in (0, 1)$  is the mixing parameter and  $S_0(t)$  denotes a proper survival function for the non-cured group. Observe that if the probability of cure is large, then the intrinsic survival function  $S(t)$  is large —  $S_0(t)$  will be large which implies in  $F_0(t) = 1 - S_0(t)$  small. Larger values of  $F_0(t)$  at a fixed time  $t$  imply lower values of  $S(t)$ . This model was derived under the threshold model for tumor resistance (cancer research) where,  $F_0(t)$  refers to the distribution of division time for each cell in a homogeneous clone of cells. The non-mixture model (1.6) or the promotion time cure fraction has been used by Lambert *et al.* (2007, 2010) to estimate the probability of cure fraction in cancer lifetime data.

From (1.6), the survival and hazard function for the non-mixture cure rate model can be written, respectively, as:

$$S(t_i) = \exp[\log(p) F_0(t_i)] \tag{1.7}$$

and

$$h(t_i) = -\log(p) f_0(t_i). \tag{1.8}$$

Since  $f(t) = h(t)S(t)$ , the contribution of the  $i^{th}$  individual for the likelihood function is given by:

$$L_i = h(t_i)^{\delta_i} S(t_i) \tag{1.9}$$

that is:

$$L_i = [-\log(p) f_0(t_i)]^{\delta_i} \exp[\log(p) F_0(t_i)]. \tag{1.10}$$

Considering a random sample of lifetimes  $(t_i, \delta_i, i = 1, \dots, n)$  the log-likelihood is:

$$l = r \log[-\log(p)] + \sum_{i=1}^n \delta_i \log f_0(t_i) + \log(p) \sum_{i=1}^n [1 - S_0(t_i)], \tag{1.11}$$

where, as before,  $r = \sum_{i=1}^n \delta_i$ .

A Bayesian formulation of the non-mixture cure rate model is given in Chen *et al.* (1999). A model which includes a standard mixture model for cure rate was considered in Yin and Ibrahim (2005). Rodrigues *et al.* (2009b) extended the long-term survival model proposed by Chen *et al.* (1999).

In this paper, considering the Burr XII distribution, we compare the performance of the mixture and non-mixture cure fraction formulation when the scale and shape parameters are dependent of covariates. The Burr XII distribution provides more flexibility than the Weibull distribution which could be a special case of the Burr XII distribution if its parameters are extended to a limiting case. It is also important to point out that the Burr XII distribution is mathematically tractable with a closed form for its cumulative distribution function.

The paper is organized as follows: in Section 2, we introduce the likelihood function assuming the Burr XII distribution distribution for the susceptible individuals; in Section 3, we present a Bayesian analysis assuming the mixture and non-mixture models in presence or absence of covariates; in Section 4, we present an application with the leukaemia data of Kersey *et al.* (1999) in

various aspects of statistical inference, in particular, the comparison between the effects of the allogeneic and autologous treatments; finally in Section 5, we introduce some comments and remarks.

## 2. The Burr XII Distribution Cure Model

Burr (1942) suggested a number of cumulative distributions, where the most popular one is the so-called Burr XII distribution, whose three-parameter probability density function is given by:

$$f_0(t | \mu, \alpha, \lambda) = \frac{\alpha}{\mu^\alpha} t^{\alpha-1} \left[ 1 + \lambda \left( \frac{t}{\mu} \right)^\alpha \right]^{-\left(1 + \frac{1}{\lambda}\right)}, \quad (2.1)$$

where  $\mu > 0$  is the scale parameter;  $\alpha > 0$  and  $\lambda > 0$  are shape parameters. For  $\lambda \rightarrow +0$  we have the Weibull distribution as a particular case. The hazard function of a Burr XII distribution is decreasing if  $\alpha \leq 1$  and is unimodal with the mode at  $t = \frac{(\alpha-1)^{1/\alpha}}{\mu^{-1}\lambda^{1/\alpha}}$  when  $\alpha > 1$ . The three-parameter Burr XII distribution is much more flexible than the standard two-parameter Weibull distribution. Rodriguez (1977) showed that the possible combination of skewness and kurtosis covers a two-dimensional area in the skewness-kurtosis plane for the Burr XII distribution, but is restricted to a curve for the Weibull distribution. Some typical shapes of the three-parameter Burr XII distribution are shown in Figure 1.

From (2.1), the distribution and survival functions are written, respectively, by:

$$F_0(t | \mu, \alpha, \lambda) = 1 - \left[ 1 + \lambda \left( \frac{t}{\mu} \right)^\alpha \right]^{-\frac{1}{\lambda}} \quad (2.2)$$

$$S_0(t | \mu, \alpha, \lambda) = \left[ 1 + \lambda \left( \frac{t}{\mu} \right)^\alpha \right]^{-\frac{1}{\lambda}}$$

From (2.2), the Burr XII model in the presence of long-term survivors or immunes has a probability density function, a distribution function and a survival function given, respectively, as follows:

$$f(t | \theta) = (1-p) \frac{\alpha}{\mu^\alpha} t^{\alpha-1} \left[ 1 + \lambda \left( \frac{t}{\mu} \right)^\alpha \right]^{-\left(1 + \frac{1}{\lambda}\right)}, \quad (2.3)$$

$$F(t | \theta) = (1-p) \left\{ 1 - \left[ 1 + \lambda \left( \frac{t}{\mu} \right)^\alpha \right]^{-\frac{1}{\lambda}} \right\}, \quad (2.4)$$

$$S(t | \theta) = p + (1-p) \left[ 1 + \lambda \left( \frac{t}{\mu} \right)^\alpha \right]^{-\frac{1}{\lambda}}, \quad (2.5)$$

where  $\theta = (\mu, \alpha, \lambda, p)$ ,  $\mu$  is the scale parameter,  $\alpha$  and  $\lambda$  are shape parameters and  $p$  is the proportion of immunes or non-susceptible. Suppose the data are of the form  $(t_i, \delta_i)$ ,  $i = 1, \dots, n$ , where  $\delta_i = 1$  if  $t_i$  is uncensored and  $\delta_i = 0$  otherwise and that  $f(t_i)$  is given by (2.3). Under the assumption of right censored lifetime, the observed full likelihood function is:

$$L(\theta | \mathbf{t}, \delta) = L_1(\theta | \mathbf{t}, \delta) \times L_2(\theta | \mathbf{t}, \delta) \quad (2.6)$$

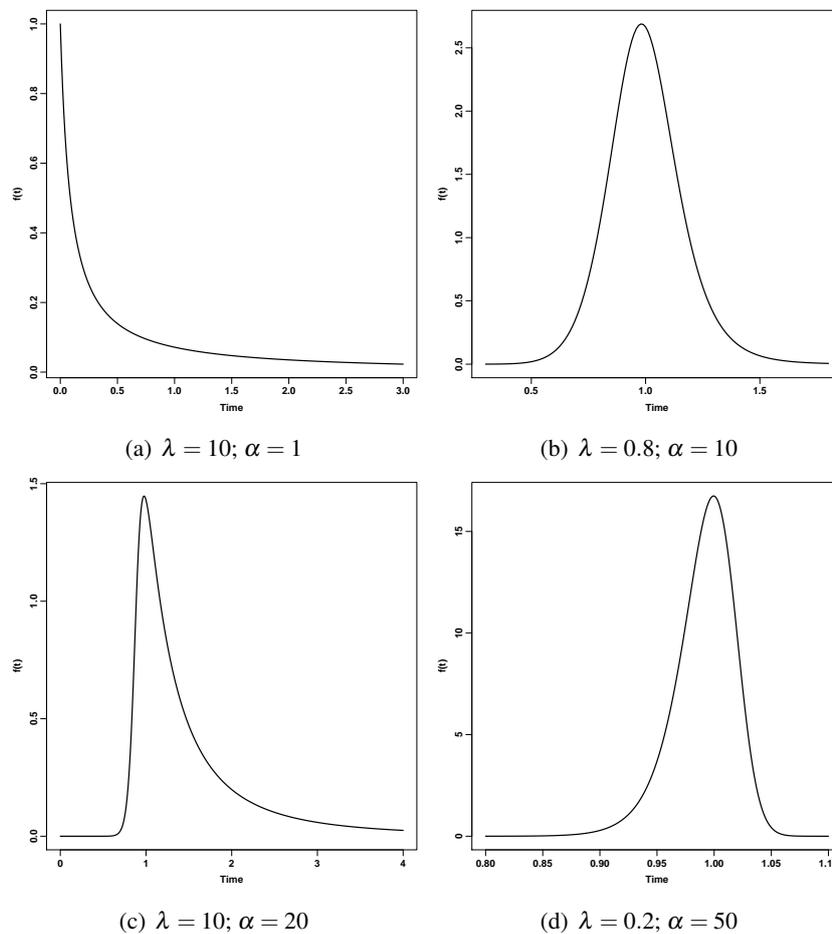


Fig. 1. Behavior of probability density functions considering different shapes values of the three-parameter Burr XII distribution ( $\mu = 1$ ).

such that the log-likelihood terms,  $l_j(\theta | \mathbf{t}, \delta) = \log[L_j(\theta | \mathbf{t}, \delta)]$ ,  $j = 1, 2$ , are given, respectively, by:

$$l_1(\theta | \mathbf{t}, \delta) = r \log(1 - p) + r \log(\alpha) - r\alpha \log(\mu) + (\alpha - 1)\tilde{t} - \left(1 + \frac{1}{\lambda}\right) \sum_{i=1}^n \delta_i \log(A_i) \tag{2.7}$$

and:

$$l_2(\theta | \mathbf{t}, \delta) = \sum_{i=1}^n (1 - \delta_i) \log \left\{ p + (1 - p) A_i^{-\frac{1}{\lambda}} \right\}, \tag{2.8}$$

where  $r = \sum_{i=1}^n \delta_i$ ,  $\tilde{t} = \sum_{i=1}^n \delta_i \log(t_i)$ ,  $A_i = 1 + B_i$  and  $B_i = \lambda \left(\frac{t_i}{\mu}\right)^\alpha$ .

Given the observed lifetime data,  $(t_i, \delta_i)$ ,  $i = 1, \dots, n$ , and defining  $l(\theta | \mathbf{t}, \delta) = \log L(\theta | \mathbf{t}, \delta)$ , the maximum likelihood estimates for  $\theta = (\mu, \alpha, \lambda, p)$ , denoted by  $\hat{\theta} = (\hat{\mu}, \hat{\alpha}, \hat{\lambda}, \hat{p})$ , are obtained

by solving, for example using the Newton-Raphson method, the following likelihood equations:

$$\frac{\partial}{\partial \mu} l(\theta | \mathbf{t}, \delta) = -\frac{r\alpha}{\mu} + \frac{\alpha(1 + \frac{1}{\lambda})}{\mu} \sum_{i=1}^n \frac{\delta_i B_i}{A_i^\alpha} + \frac{(1-p)\alpha}{\lambda \mu} \sum_{i=1}^n \frac{A_i^{-(1+\frac{1}{\lambda})} B_i}{p + (1-p)A_i^{-\frac{1}{\lambda}}} = 0 \quad (2.9)$$

$$\frac{\partial}{\partial \alpha} l(\theta | \mathbf{t}, \delta) = \frac{r}{\alpha} - r \log(\mu) + \tilde{r} - \left(1 + \frac{1}{\lambda}\right) \sum_{i=1}^n \frac{\delta_i B_i \log\left(\frac{t_i}{\mu}\right)}{A_i} - \frac{(1-p)}{\lambda} \sum_{i=1}^n \frac{A_i^{-(1+\frac{1}{\lambda})} B_i \log\left(\frac{t_i}{\mu}\right)}{p + (1-p)A_i^{-\frac{1}{\lambda}}} = 0 \quad (2.10)$$

$$\frac{\partial}{\partial \lambda} l(\theta | \mathbf{t}, \delta) = \frac{1}{\lambda^2} \sum_{i=1}^n \delta_i \log(A_i) - \frac{\lambda+1}{\lambda^2} \sum_{i=1}^n \frac{\delta_i B_i}{A_i} + \frac{(1-p)}{\lambda^2} \sum_{i=1}^n \frac{A_i^{-\frac{1}{\lambda}} [\log(A_i) - B_i A_i^{-1}]}{p + (1-p)A_i^{-\frac{1}{\lambda}}} = 0 \quad (2.11)$$

$$\frac{\partial}{\partial p} l(\theta | \mathbf{t}, \delta) = -\frac{r}{1-p} + \sum_{i=1}^n \frac{1 - A_i^{-\frac{1}{\lambda}}}{p + (1-p)A_i^{-\frac{1}{\lambda}}} = 0 \quad (2.12)$$

Under the non-mixture formulation and using (2.2), the probability density function, the distribution function and the survival function are given respectively by:

$$f(t | \theta) = -\log(p) \frac{\alpha}{\mu^\alpha} t^{\alpha-1} \left[1 + \lambda \left(\frac{t}{\mu}\right)^\alpha\right]^{-(1+\frac{1}{\lambda})} p \left\{1 - [1 + \lambda \left(\frac{t}{\mu}\right)^\alpha]^{-\frac{1}{\lambda}}\right\} \quad (2.13)$$

$$F(t | \theta) = 1 - p \left\{1 - [1 + \lambda \left(\frac{t}{\mu}\right)^\alpha]^{-\frac{1}{\lambda}}\right\}, \quad (2.14)$$

$$S(t | \theta) = p \left\{1 - [1 + \lambda \left(\frac{t}{\mu}\right)^\alpha]^{-\frac{1}{\lambda}}\right\}. \quad (2.15)$$

In this case, the log-likelihood function for the non-mixture Burr XII cure model can be written as:

$$l(\theta | \mathbf{t}, \delta) = r \log[-\log(p)] + r \log(\alpha) - r\alpha \log(\mu) + (\alpha - 1)\tilde{r} - \left(1 + \frac{1}{\lambda}\right) \sum_{i=1}^n \delta_i \log(A_i) + \log(p) \sum_{i=1}^n \left(1 - A_i^{-\frac{1}{\lambda}}\right) \quad (2.16)$$

where, again,  $r = \sum_{i=1}^n \delta_i$ ,  $\tilde{r} = \sum_{i=1}^n \delta_i \log(t_i)$ ,  $A_i = 1 + B_i$  and  $B_i = \lambda \left(\frac{t_i}{\mu}\right)^\alpha$ .

Differentiating (2.16) with respect to  $\mu$ ,  $\alpha$ ,  $\lambda$  and  $p$  setting the results equal to zero we have:

$$\frac{\partial}{\partial \mu} l(\theta | \mathbf{t}, \delta) = -\frac{r\alpha}{\mu} + \frac{\alpha(1 + \frac{1}{\lambda})}{\mu} \sum_{i=1}^n \frac{\delta_i B_i}{A_i} - \frac{\alpha \log(p)}{\lambda \mu} \sum_{i=1}^n A_i^{-(1+\frac{1}{\lambda})} B_i = 0 \quad (2.17)$$

$$\frac{\partial}{\partial \alpha} l(\theta | \mathbf{t}, \delta) = \frac{r}{\alpha} - r \log(\mu) + \tilde{r} - \left(1 + \frac{1}{\lambda}\right) \sum_{i=1}^n \frac{\delta_i B_i \log\left(\frac{t_i}{\mu}\right)}{1 + B_i} + \frac{\log(p)}{\lambda} \sum_{i=1}^n \frac{A_i^{-\frac{1}{\lambda}} B_i \log\left(\frac{t_i}{\mu}\right)}{(1 + B_i)} = 0 \quad (2.18)$$

$$\frac{\partial}{\partial \lambda} l(\theta | \mathbf{t}, \delta) = \frac{1}{\lambda^2} \sum_{i=1}^n \delta_i \log(A_i) - \frac{\lambda+1}{\lambda^2} \sum_{i=1}^n \frac{\delta_i B_i}{1 + B_i} - \frac{\log(p)}{\lambda^2} \sum_{i=1}^n A_i^{-\frac{1}{\lambda}} \left[\log(A_i) - \frac{B_i}{A_i}\right] = 0 \quad (2.19)$$

$$\frac{\partial}{\partial p} l(\theta | \mathbf{t}, \delta) = \frac{r}{p \log(p)} + \frac{1}{p} - \frac{1}{p} \sum_{i=1}^n A_i^{-\frac{1}{\lambda}} = 0 \quad (2.20)$$

The equation (2.20) can be solved algebraically for  $p$ , giving:

$$\widehat{p}(\mu, \alpha, \lambda) = \exp \left[ \sum_{i=1}^n \frac{\delta_i}{A_i^{-\frac{1}{\lambda}} - 1} \right]. \quad (2.21)$$

To obtain  $\widehat{\mu}$ ,  $\widehat{\alpha}$  and  $\widehat{\lambda}$ , we substitute  $\widehat{p}(\mu, \alpha, \lambda)$  into the equations (2.17), (2.18) and (2.19) and solve for  $\mu$ ,  $\alpha$  and  $\lambda$  using numerical methods. The  $100 \times (1 - \psi)$  % confidence intervals for  $\mu$ ,  $\alpha$ ,  $\lambda$  and  $p$  can be obtained from the usual asymptotic normality of the maximum likelihood estimators with  $Var(\widehat{\mu})$ ,  $Var(\widehat{\alpha})$ ,  $Var(\widehat{\lambda})$  and  $Var(\widehat{p})$  estimated from the inverse of the observed Fisher information matrix, that is, the inverse of the matrix of negative second derivatives of the log-likelihood function locally at  $\widehat{\mu}$ ,  $\widehat{\alpha}$ ,  $\widehat{\lambda}$  and  $\widehat{p}$ .

In the presence of one covariate  $x_i$ ,  $i = 1, \dots, n$ , we can assume a link function for  $\mu$ ,  $\alpha$ ,  $\lambda$  and  $p$ , that is,  $\log(\mu_i) = \beta_0 + \beta_1 x_i$ ,  $\log(\alpha_i) = \alpha_0 + \alpha_1 x_i$ ,  $\log(\lambda_i) = \gamma_0 + \gamma_1 x_i$  and  $\log\left(\frac{p_i}{1-p_i}\right) = \eta_0 + \eta_1 x_i$ , where  $x_i$ , for example, taking the value 0 if individual  $i$  is in the treatment group 1 or the value 1 if individual  $i$  is in the treatment group 2. In this way, we can have interest in test the following hypothesis:  $H_0 : \beta_1 = 0$  (no treatment effect in the susceptible patients),  $H_0 : \alpha_1 = 0$  (no treatment effect in the shape of the lifetime distribution),  $H_0 : \gamma_1 = 0$  (no treatment effect in the shape of the lifetime distribution) or  $H_0 : \eta_1 = 0$  (no treatment effect in the proportion of cured individuals).

### 3. A Bayesian Analysis

For a Bayesian analysis of the mixture and non-mixture models introduced in Section 1, we assume an prior uniform distribution defined in the interval  $(0, 1)$ ,  $U(0, 1)$ , for the probability of cure  $p$  and  $Gamma(0.001, 0.001)$  prior distributions for the scale parameter  $\mu$  and shape parameters  $\alpha$  and  $\lambda$ , where  $Gamma(a, b)$  denotes a gamma distribution with mean  $a/b$  and variance  $a/b^2$ . We further assume prior independence among  $p$ ,  $\mu$ ,  $\alpha$  and  $\lambda$ . Observe that we are using approximately non-informative priors for the parameters of the models.

In the presence of a covariate vector  $\mathbf{x} = (x_1, \dots, x_k)'$  affecting the parameters  $\mu$  and  $\alpha$ , but not affecting the shape parameter  $\lambda$ , let us assume the following regression model,

$$\mu_i = \beta_0 \exp(\beta_1 x_{1i} + \dots + \beta_k x_{ki}) \text{ and } \alpha_i = \alpha_0 \exp(\alpha_1 x_{1i} + \dots + \alpha_k x_{ki}). \quad (3.1)$$

Assuming the mixture and non-mixture models introduced in Section 1, let us consider a gamma prior distribution  $Gamma(0.001, 0.001)$  for the regression parameters  $\beta_0$  and  $\alpha_0$  and a normal prior distribution  $N(0, 100)$  for the regression parameters  $\beta_l$  and  $\alpha_l$ ,  $l = 1, \dots, k$ , where  $N(\mu, \sigma^2)$  denotes a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . We also assume prior independence among the parameters.

Posterior summaries of interest are obtained from simulated samples for the joint posterior distribution using standard Markov Chain Monte Carlo (MCMC) methods as the Gibbs sampling algorithm [35] or the Metropolis-Hastings algorithm [36].

### 4. An Application

In this section we analyze a leukaemia data set consisting of 90 observations introduced by Kersey *et al.* (1987) and reproduced by Maller and Zhou (1996). In this data 46 patients were treated by allogeneic transplant (Group I) and the other 44 by autologous transplant (Group II). The survival

time refers to the number of days to recurrence of leukaemia for patients after one of the two treatments. The medical problems of interest include: the existence of “cured” patients (who will never suffer a recurrence of leukaemia) and the estimation of their proportion; the failure distributions of susceptible patients; and comparison between the effects of the two treatments.

From expressions (2.7), (2.8) and (2.16), we get the estimates presented in Tables 1 and 2 — the maximum likelihood estimates for mixture and non-mixture models, respectively. In Tables 3 and 4, we have the inference results considering the Bayesian approach for mixture and non-mixture models, respectively.

We also have in Tables 1 and 4, the AIC (Akaike information criterion) and the Monte Carlo estimates of DIC (Deviance Information Criterion) used as a discrimination criterion for different models. Smaller values of AIC and DIC indicates better models. A brief introduction to AIC and DIC is presented at Appendix A.

The maximum likelihood estimates were obtained in *SAS/NLMIXED* procedure, [37], by applying the Newton-Raphson algorithm. To obtain the Bayesian estimates we have used MCMC (Markov Chain Monte Carlo) methods available in *SAS* software 9.2, *SAS/MCMC* [38]. A single chain has been used in the simulation of samples for each parameter of both models considering a “burn-in-sample” of size 15,000 to eliminate the possible effect of the initial values. After this “burn-in” period, we simulated other 200,000 Gibbs samples taking every 100<sup>th</sup> sample, to get approximated uncorrelated values which result in a final chain of size 2,000. Usual existing convergence diagnostics available in the literature for a single chain using the *SAS/MCMC* procedure indicated convergence for all parameters.

Table 1. Maximum likelihood (standard error) estimates for  $\mu$ ,  $\alpha$ ,  $\lambda$  and  $p$  in each group — mixture model.

| Group | $\hat{\mu}$         | $\hat{\alpha}$     | $\hat{\lambda}$    | $\hat{p}$           | AIC   |
|-------|---------------------|--------------------|--------------------|---------------------|-------|
| I     | 171.15<br>(112.77)  | 1.3824<br>(0.6862) | 1.4245<br>(3.0362) | 0.2050<br>(0.1894)  | 497.2 |
| II    | 110.33<br>(19.7609) | 3.2248<br>(1.0359) | 1.5414<br>(1.0503) | 0.2006<br>(0.06142) | 457.2 |

Table 2. Maximum likelihood (standard error) estimates for  $\mu$ ,  $\alpha$ ,  $\lambda$  and  $p$  in each group — non-mixture model.

| Group | $\hat{\mu}$         | $\hat{\alpha}$     | $\hat{\lambda}$    | $\hat{p}$           | AIC   |
|-------|---------------------|--------------------|--------------------|---------------------|-------|
| I     | 321.81<br>(178.20)  | 1.2938<br>(0.6086) | 1.3991<br>(5.5520) | 0.2119<br>(0.2522)  | 497.3 |
| II    | 146.08<br>(31.8094) | 3.0038<br>(0.8740) | 1.6820<br>(1.4796) | 0.2002<br>(0.06260) | 457.4 |

Table 3. Posterior means (standard deviation) for  $\mu$ ,  $\alpha$ ,  $\lambda$  and  $p$  in each group — mixture model.

| Group | $\hat{\mu}$        | $\hat{\alpha}$     | $\hat{\lambda}$    | $\hat{p}$          | DIC   |
|-------|--------------------|--------------------|--------------------|--------------------|-------|
| I     | 170.2<br>(15.5727) | 1.3224<br>(0.3386) | 1.5235<br>(1.2280) | 0.2046<br>(0.0984) | 495.3 |
| II    | 114.4<br>(22.5142) | 3.2585<br>(1.1278) | 1.8328<br>(1.3489) | 0.2073<br>(0.0622) | 457.3 |

Table 4. Posterior means (standard deviation) for  $\mu$ ,  $\alpha$ ,  $\lambda$  and  $p$  in each group — non-mixture model.

| Group | $\hat{\mu}$        | $\hat{\alpha}$     | $\hat{\lambda}$    | $\hat{p}$          | DIC   |
|-------|--------------------|--------------------|--------------------|--------------------|-------|
| I     | 302.0<br>(60.1777) | 1.3200<br>(0.2091) | 1.1538<br>(0.5350) | 0.2497<br>(0.0673) | 494.0 |
| II    | 158.4<br>(25.4148) | 2.7506<br>(0.5098) | 1.3057<br>(0.4480) | 0.2141<br>(0.0603) | 455.8 |

In Figure 2, we have the plots of the estimated survival functions based on the maximum likelihood estimates considering mixture and non-mixture models in presence of cure fraction and the plot of the non-parametric Kaplan-Meier estimate for the survival function [39]. We also have in Figure 2, the plot of the estimated survival function based on the Weibull and Burr XII distributions not considering the cure fraction modeling.

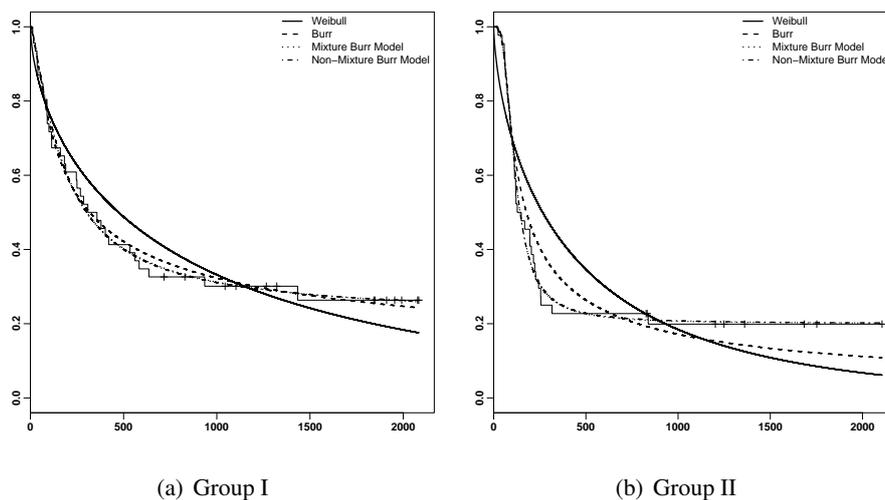


Fig. 2. Fitted models for the data.

From the fitted survival models (see, Figure 2), we conclude that the survival times are very well fitted by the mixture and non-mixture cure fraction models. Also from Figure 2, we conclude that the Burr XII distributions not considering the cure fraction modeling is better fitted than the

Weibull distributions not considering the cure fraction modeling, this shows that Burr XII distributions with three parameters is more flexible than the Weibull distribution. From the results of Tables 1 — 4, using classical or Bayesian inference approaches give similar results; the obtained DIC discrimination values from both models also give similar results.

We can also consider a binary variable related to the different groups where  $x_{1i} = 1$  for Group II and 0 for the Group I. Then we assume the regression models given in (3.1). For these regression models we consider three cases: model without covariates (Model 1), regression model for  $\mu$  (Model 2) and regression model for  $\mu$  and  $\alpha$  (Model 3).

In Tables 5 and 6 we have the maximum likelihood estimates for regression models considering mixture and non-mixture models, respectively. In Tables 7 and 8, we have the inference results considering the Bayesian approach for regression models considering mixture and non-mixture models, respectively.

Table 5. Maximum Likelihood Estimates (MLE) and Standard Error estimates (SE) for regression models — mixture model.

| Model   | Parameter        | <i>MLE</i> | <i>SE</i> | 95% Confidence Interval |
|---------|------------------|------------|-----------|-------------------------|
| Model 1 | $\hat{\mu}$      | 120.94     | 23.9370   | (73.3850; 168.49)       |
|         | $\hat{\alpha}$   | 2.1418     | 0.4924    | (1.1635; 3.1201)        |
|         | $\hat{\lambda}$  | 1.7671     | 0.9948    | (-0.2092; 3.7433)       |
|         | $\hat{p}$        | 0.2135     | 0.05406   | (0.1061; 0.3210)        |
| Model 2 | $\hat{\alpha}$   | 1.7451     | 0.4172    | (0.9162; 2.5741)        |
|         | $\hat{\lambda}$  | 0.8798     | 0.7466    | (-0.6035; 2.3630)       |
|         | $\hat{p}$        | 0.2306     | 0.04861   | (0.1340; 0.3272)        |
|         | $\hat{\beta}_0$  | 182.19     | 66.5977   | (49.8834; 314.50)       |
|         | $\hat{\beta}_1$  | -0.3782    | 0.3185    | (-1.0109; 0.2546)       |
| Model 3 | $\hat{\lambda}$  | 1.5246     | 0.8963    | (-0.2560; 3.3051)       |
|         | $\hat{p}$        | 0.2003     | 0.05370   | (0.09359; 0.3070)       |
|         | $\hat{\beta}_0$  | 167.73     | 57.7851   | (52.9296; 282.53)       |
|         | $\hat{\beta}_1$  | -0.4165    | 0.3060    | (-1.0245; 0.1915)       |
|         | $\hat{\alpha}_0$ | 1.4043     | 0.3375    | (0.7338; 2.0747)        |
|         | $\hat{\alpha}_1$ | 0.8268     | 0.2451    | (0.3399; 1.3136)        |

Table 6. Maximum Likelihood Estimates (MLE) and Standard Error estimates (SE) for regression models — non-mixture model.

| Model   | Parameter        | MLE     | SE      | 95% Confidence Interval |
|---------|------------------|---------|---------|-------------------------|
| Model 1 | $\hat{\mu}$      | 177.16  | 41.2033 | (95.3025; 259.02)       |
|         | $\hat{\alpha}$   | 2.0748  | 0.4643  | (1.1523; 2.9972)        |
|         | $\hat{\lambda}$  | 2.3927  | 1.9924  | (-1.5656; 6.3509)       |
|         | $\hat{\rho}$     | 0.2024  | 0.06668 | (0.06989; 0.3348)       |
| Model 2 | $\hat{\alpha}$   | 1.7262  | 0.3497  | (1.0314; 2.4209)        |
|         | $\hat{\lambda}$  | 0.9231  | 0.9830  | (-1.0298; 2.8761)       |
|         | $\hat{\rho}$     | 0.2292  | 0.05031 | (0.1293; 0.3292)        |
|         | $\hat{\beta}_0$  | 270.27  | 97.1221 | (77.3201; 463.22)       |
|         | $\hat{\beta}_1$  | -0.3783 | 0.3037  | (-0.9816; 0.2250)       |
| Model 3 | $\hat{\lambda}$  | 1.6699  | 1.3020  | (-0.9167; 4.2565)       |
|         | $\hat{\rho}$     | 0.2003  | 0.05569 | (0.08969; 0.3110)       |
|         | $\hat{\beta}_0$  | 316.98  | 125.46  | (67.7352; 566.22)       |
|         | $\hat{\beta}_1$  | -0.7733 | 0.3558  | (-1.4802; -0.06655)     |
|         | $\hat{\alpha}_0$ | 1.3190  | 0.2892  | (0.7444; 1.8935)        |
|         | $\hat{\alpha}_1$ | 0.8211  | 0.2480  | (0.3284; 1.3138)        |

Table 7. Posterior Means (PM) and Standard Deviation (SD) for regression models — mixture model.

| Model   | Parameter        | PM      | SD      | Credible Interval |
|---------|------------------|---------|---------|-------------------|
| Model 1 | $\hat{\mu}$      | 122.6   | 28.7453 | (77.2546; 194.9)  |
|         | $\hat{\alpha}$   | 2.1934  | 0.5451  | (1.2602; 3.4459)  |
|         | $\hat{\lambda}$  | 2.2280  | 1.3879  | (0.3819; 5.9824)  |
|         | $\hat{\rho}$     | 0.2009  | 0.0632  | (0.0646; 0.3176)  |
| Model 2 | $\hat{\alpha}$   | 1.8272  | 0.4815  | (1.0778; 2.9824)  |
|         | $\hat{\lambda}$  | 1.3434  | 1.1575  | (0.0332; 4.5494)  |
|         | $\hat{\rho}$     | 0.2173  | 0.0593  | (0.0851; 0.3261)  |
|         | $\hat{\beta}_0$  | 174.9   | 52.0182 | (81.5222; 277.9)  |
|         | $\hat{\beta}_1$  | -0.3018 | 0.2916  | (-0.8292; 0.3199) |
| Model 3 | $\hat{\lambda}$  | 1.7100  | 1.0094  | (0.3734; 4.1672)  |
|         | $\hat{\rho}$     | 0.2005  | 0.0528  | (0.0997; 0.3093)  |
|         | $\hat{\beta}_0$  | 175.8   | 52.2661 | (88.4162; 298.2)  |
|         | $\hat{\beta}_1$  | -0.4070 | 0.2806  | (-0.9384; 0.1674) |
|         | $\hat{\alpha}_0$ | 1.4214  | 0.3461  | (0.8879; 2.2550)  |
|         | $\hat{\alpha}_1$ | 0.8004  | 0.2449  | (0.3295; 1.2986)  |

Table 8. Posterior Means (PM) and Standard Deviation (SD) for regression models — non-mixture model.

| Model   | Parameter        | PM      | SD      | Credible Interval |
|---------|------------------|---------|---------|-------------------|
| Model 1 | $\hat{\mu}$      | 191.3   | 47.1324 | (118.4;309.2)     |
|         | $\hat{\alpha}$   | 2.0741  | 0.4720  | (1.2757;3.0983)   |
|         | $\hat{\lambda}$  | 3.4496  | 3.0633  | (0.2589;12.2087)  |
|         | $\hat{p}$        | 0.1879  | 0.0668  | (0.0432;0.3062)   |
| Model 2 | $\hat{\alpha}$   | 1.6350  | 0.2545  | (1.1474;2.1417)   |
|         | $\hat{\lambda}$  | 1.3797  | 1.8359  | (0.00255;7.8565)  |
|         | $\hat{p}$        | 0.2170  | 0.0674  | (0.0482;0.3320)   |
|         | $\hat{\beta}_0$  | 295.4   | 9.5409  | (276.6;313.4)     |
|         | $\hat{\beta}_1$  | -0.3886 | 0.1929  | (-0.7404;0.0111)  |
| Model 3 | $\hat{\lambda}$  | 1.6695  | 1.1204  | (0.1919;4.6356)   |
|         | $\hat{p}$        | 0.2044  | 0.0499  | (0.1049;0.3055)   |
|         | $\hat{\beta}_0$  | 338.9   | 15.2548 | (308.1;356.9)     |
|         | $\hat{\beta}_1$  | -0.7833 | 0.1830  | (-1.1088;-0.4185) |
|         | $\hat{\alpha}_0$ | 1.2762  | 0.2040  | (0.9006;1.7011)   |
|         | $\hat{\alpha}_1$ | 0.7752  | 0.2432  | (0.2887;1.2575)   |

Different model selection methods to choose the most adequate model could be adopted under the Bayesian paradigm [40]. We consider the Deviance Information Criterion (DIC) which is specifically useful for selecting models under the Bayesian approach, where samples of the posterior distribution for the model parameters are obtained by using MCMC methods (see, Appendix Appendix A).

In Bayesian context using MCMC methods, we have used the DIC (Appendix A) given automatically by the SAS software, in place of the usual AIC criteria used in classical approach (see, Table 9). Other proposal has been indicated in the literature [41–43].

Table 9. Discrimination criterion.

| Model   | Mixture Model |       | Non-Mixture Model |       |
|---------|---------------|-------|-------------------|-------|
|         | DIC           | AIC   | DIC               | AIC   |
| Model 1 | 959.5         | 959.7 | 958.2             | 959.7 |
| Model 2 | 959.8         | 960.5 | 958.7             | 960.5 |
| Model 3 | 949.5         | 950.4 | 948.7             | 950.6 |

From the results of Table 9, we conclude that Model 3 (regression model for  $\mu$  and  $\alpha$ ) is better fitted by the data. Since DIC is a little bit smaller considering the non-mixture Model 3 when compared to the other models, we use this model to get our final inferences of interest. From Table 8 and using the non-mixture Model 3, we conclude that the parameters  $\beta_1$  and  $\alpha_1$  have significant treatment effect in the ratio of susceptible patients.

## 5. Concluding Remarks

Usually in the analysis of lifetime data we could have the presence of a cure fraction, where a proportion of the patients will never experiment the event of interest, in many cases, death of the patient. To analyze this kind of data, we could use different existing parametric formulations, as mixture and non-mixture models. These formulations, usually assume standard existing distributions as the Weibull, log-normal or exponential distributions for the susceptible individuals. The use of a Burr XII distribution with mixture or non-mixture cure rate could be of great interest in applications since this model has a great flexibility of fit as compared to other standard lifetime distributions. Computationally, especially using the Bayesian paradigm, the obtained results are very similar as observed in the application introduced in Section 4. The great advantage of the mixture model is related to the simple interpretations, especially for medical researchers, where we have the proportion of cured and non-cured individuals given directly in the survival function expression.

## References

- [1] R. A. Maller and X. Zhou, *Survival analysis with long-term survivors* Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley Series in Probability and Statistics: Applied Probability and Statistics (John Wiley & Sons Ltd., Chichester, 1996).
- [2] J. Boag, Maximum likelihood estimation of the proportion of patients cured by Cancer therapy, *Journal of the Royal Statistical Society, B* **11** (1949) 15–53.
- [3] J. Berkson and R. P. Gage, Survival curve for cancer patients following treatment, *Journal of the American Statistical Association* **47** (1952) 501–515.
- [4] J. L. Haybittle, A two parameter model for the survival curve of treated cancer patients, *Journal of the American Statistical Association* **53** (1965) 16–26.
- [5] V. T. Farewell, The use of mixture models for the analysis of survival data with long-term survivors, *Biometrics* **38** (1982) 1041–1046.
- [6] V. T. Farewell, Mixture models in survival analysis: Are they worth the risk?, *The Canadian Journal of Statistics* **14**(3) (1986) 257–262.
- [7] W. Q. Meeker, Limited failure population life tests: Application to integrated circuit reliability, *Technometrics* **29**(1) (1987) 51–65.
- [8] J. W. Gamel, I. W. McLean and S. H. Rosenberg, Proportion cured and mean log survival time as functions of tumor size, *Statistics in Medicine* **9** (1990) 999–1006.
- [9] M. E. Ghitany and R. A. Maller, Asymptotic results for exponential mixture models with long term survivors, *Statistics* **23** (1992) 321–336.
- [10] J. B. Copas and F. Heydari, Estimating the risk of reoffending by using exponential mixture models, *Journal of the Royal Statistical Society, A* **160**(2) (1997) 237–252.
- [11] S. K. Ng and G. J. McLachlan, On modifications to the long-term survival mixture model in the presence of competing risks, *Journal of Statistical Computation and Simulation* **61** (1998) 77–96.
- [12] R. De Angelis, R. Capocaccia, T. Hakulinen, B. Soderman and A. Verdecchia, Mixture models for cancer survival analysis: application to population-based data with covariates, *Statistics in Medicine* **18**(4) (1999) 441–454.
- [13] Y. Peng and K. B. G. Dear, A nonparametric mixture model for cure rate estimation, *Biometrics* **56** (2000) 237–243.
- [14] J. P. Sy and J. M. G. Taylor, Estimation in a Cox proportional hazards cure model, *Biometrics* **56** (2000) 227–236.
- [15] N. Balakrishnan and S. Pal, Expectation maximization-based likelihood inference for flexible cure rate models with weibull lifetimes., *Statistical Methods in Medical Research* (2013).
- [16] J. Mazucheli, E. A. Coelho-Barros and J. A. Achcar, The exponentiated exponential mixture and non-mixture cure rate model in the presence of covariates, *Computer Methods and Programs in Biomedicine* **112**(1) (2013) 114–124.

- [17] R. D. Gupta and D. Kundu, Exponentiated exponential family: an alternative to gamma and Weibull distributions, *Biom. J.* **43**(1) (2001) 117–130.
- [18] Y. Peng, K. B. G. Dear and J. W. Denham, A generalized F mixture model for cure rate estimation, *Statistics in Medicine* **17**(8) (1998) 813–830.
- [19] K. Yamaguchi, Accelerated failure-time regression model with a regression model for the surviving fraction: an application to the analysis of permanent employment in japan, *Journal of the American Statistical Association* **87** (1992) 284–292.
- [20] P. W. Gieser, M. N. Chang, P. V. Rao, J. J. Shuster and J. Pullen, Modelling cure rates using the gompertz model with covariate information., *Statistics in Medicine* (17) (1998) 831–839.
- [21] V. G. Cancho and H. Bolfarine, Modeling the presence of immunes by using the exponentiated-Weibull model, *Journal of Applied Statistics* **28**(6) (2001) 659–671.
- [22] N. Kannan, D. Kundu, P. Nair and R. C. Tripathi, The generalized exponential cure rate model with covariates, *Journal of Applied Statistics* **37**(9-10) (2010) 1625–1636.
- [23] J. Rodrigues, M. de Castro, V. G. Cancho and N. Balakrishnan, COM-Poisson cure rate survival models and an application to a cutaneous melanoma data, *J. Statist. Plann. Inference* **139**(10) (2009) 3605–3611.
- [24] G. Yin and J. G. Ibrahim, Cure rate models: a unified approach, *The Canadian Journal of Statistics* **33**(4) (2005) 559–570.
- [25] Q. Shao and X. Zhou, A new parametric model for survival data with long-term survivors., *Stat Med* **23**(22) (2004) 3525–43.
- [26] A. Y. Yakovlev, A. D. Tsodikov and B. Asselain, *Stochastic models of tumor latency and their biostatistical applications* (World Scientific, Singapore, 1996).
- [27] A. D. Tsodikov, J. G. Ibrahim and A. Y. Yakovlev, Estimating cure rates from survival data: an alternative to two-component mixture models, *Journal of the American Statistical Association* **98**(464) (2003) 1063–1078.
- [28] P. C. Lambert, P. W. Dickman, C. L. Weston and J. R. Thompson, Estimating the cure fraction in population-based cancer studies by using finite mixture models, *Journal of the Royal Statistical Society. Series C. Applied Statistics* **59**(1) (2010) 35–55.
- [29] P. C. Lambert, J. R. Thompson, C. L. Weston and P. W. Dickman, Estimating and modeling the cure fraction in population-based cancer survival analysis, *Biostatistics* **8**(3) (2007) 576–594.
- [30] M.-H. Chen, J. G. Ibrahim and D. Sinha, A new Bayesian model for survival data with a surviving fraction, *Journal of the American Statistical Association* **94**(447) (1999) 909–919.
- [31] J. Rodrigues, V. G. Cancho, M. de Castro and F. Louzada-Neto, On the unification of long-term survival models, *Statistics & Probability Letters* **79**(6) (2009) 753–759.
- [32] J. H. Kersey, D. Weisdorf, M. E. Nesbit, T. W. LeBien, W. G. Woods, P. B. McGlave, T. Kim, D. A. Vallera, A. I. Goldman and B. Bostrom, Comparison of autologous and allogeneic bone marrow transplantation for treatment of high-risk refractory acute lymphoblastic leukemia., *The New England Journal of Medicine* **317**(8) (1987) 461–467.
- [33] I. W. Burr, Cumulative frequency functions, *Annals of Mathematical Statistics* **13** (1942) 215–232.
- [34] R. N. Rodriguez, A guide to the Burr type XII distributions, *Biometrika* **64**(1) (1977) 129–134.
- [35] A. E. Gelfand and A. F. M. Smith, Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association* **85**(410) (1990) 398–409.
- [36] S. Chib and E. Greenberg, Understanding the metropolis-hastings algorithm, *The American Statistician* **49**(4) (1995) 327–335.
- [37] SAS, *The NLMIXED Procedure, SAS/STAT® User's Guide, Version 9.22*, 2010).
- [38] SAS, *The MCMC Procedure, SAS/STAT® User's Guide, Version 9.22*, 2010).
- [39] E. L. Kaplan and P. Meier, Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* **53** (1958) 457–481.
- [40] A. Berg, R. Meyer and J. Yu, Deviance information criterion for comparing stochastic volatility models, *Journal of Business and Economic Statistics* **22** (2004) 107–120.
- [41] S. P. Brooks, Discussion on the paper by spiegelhalter, best, carlin and van de linde (2002), *Journal of the Royal Statistical Society, Series B* **64**(3) (2002) 616–618.

- [42] G. Celeux, F. Forbes, C. P. Robert and D. M. Titterton, Deviance information criteria for missing data models, *Bayesian Analysis* **1** (2005) 651–674.
- [43] A. E. Gelfand, D. K. Dey and H. Chang, Model determination using predictive distributions with implementation via sampling-based methods (with discussion), in *Bayesian Statistics*, (Oxford University Press, New York, 1992) pp. 147–167.
- [44] D. Spiegelhalter, N. Best, B. Carlin and A. van der Linde, Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society, Series B, Methodological* **64** (2002) 583–639.
- [45] H. Akaike, *Information theory and an extension of the maximum likelihood principle* (N. Petrov and F. Caski (eds), Budapest: Adadémiai Kiadó, 1973). Proceedings of the 2<sup>nd</sup> International Symposium on Information Theory.
- [46] H. Akaike, A new look at statistical model identification, *IEEE Transactions Automatic Control* **19** (1974) 716–722.

### Appendix A. Deviance Information Criterion

The deviance can be expressed as,

$$D(\theta) = -2\log L(\theta | \mathbf{y}) + c, \quad (\text{A.1})$$

where  $L(\theta | \mathbf{y})$  is the likelihood function for the unknown parameters in  $\theta$  given the observed data  $\mathbf{y}$  and  $c$  is a constant not required for comparing models.

Spiegelhalter *et al.* (2002), defined the DIC criterion by,

$$DIC = D(\hat{\theta}) + 2n_D, \quad (\text{A.2})$$

where  $D(\hat{\theta})$  is the deviance evaluated at the posterior mean  $\hat{\theta}$  and  $n_D$  is the effective number of parameters in the model, namely  $n_D = \bar{D} - D(\hat{\theta})$ , where  $\bar{D} = E[D(\theta)]$  is the posterior deviance measuring the quality of the goodness-of-fit of the current model to the data. Smaller values of DIC indicate better models. Note that these values could be negative.

Another commonly used measure of goodness-of-fit is the Akaike information criterion (AIC) [45, 46] given by,

$$AIC = -2\log L(\hat{\theta} | \mathbf{y}) + 2p, \quad (\text{A.3})$$

where  $L(\hat{\theta} | \mathbf{y})$  is the maximized likelihood value and  $p$  is the number of parameters in the model. Smaller values of AIC indicate better models.