



Published in final edited form as:  
*Stat Sin.* 2010 ; 20: 661–674.

## EFFICIENT ESTIMATION FOR AN ACCELERATED FAILURE TIME MODEL WITH A CURE FRACTION

**Wenbin Lu**

Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A

Wenbin Lu: lu@stat.ncsu.edu

### Abstract

We study the accelerated failure time model with a cure fraction via kernel-based nonparametric maximum likelihood estimation. An EM algorithm is developed to calculate the estimates for both the regression parameters and the unknown error density, in which a kernel-smoothed conditional profile likelihood is maximized in the M-step. We show that with a proper choice of the kernel bandwidth parameter, the resulting estimates are consistent and asymptotically normal. The asymptotic covariance matrix can be consistently estimated by inverting the empirical Fisher information matrix obtained from the profile likelihood using the EM algorithm. Numerical examples are used to illustrate the finite-sample performance of the proposed estimates.

### Key words and phrases

Cure model; EM algorithm; kernel smoothing; profile likelihood; survival data

### 1. Introduction

In some medical studies, of cancer or AIDS for example, it is often observed that a substantial proportion of study subjects never experience the event of interest and are thus treated as *cured* or *nonsusceptible* subjects. A number of survival models with a cure rate have been proposed in the literature for analyzing such data. One commonly used modeling approach considers a two-component mixture model that assumes that the underlying population is a mixture of susceptible and nonsusceptible subjects. Various parametric mixture cure models have been studied. For example, Berkson and Gage (1952) considered the exponential-logistic mixture, and Farewell (1982, 1986) considered the Weibull-logistic mixture for survival data with a cure fraction.

More recently, semiparametric mixture cure models have attracted much attention. Kuk and Chen (1992) proposed the so-called proportional hazards cure model in which the proportional hazards model (Cox (1972)) is used for survival times of susceptible subjects, while the logistic regression is used for the cure fraction. They developed a Monte Carlo simulation-based algorithm for conducting maximum marginal likelihood estimation. The proportional hazards cure model was further studied by Peng and Dear (2000) and Sy and Taylor (2000) using a semiparametric EM algorithm. In addition, Fang, Li and Sun (2005) and Lu (2008) studied nonparametric maximum likelihood estimation for the proportional hazards cure model and derived the asymptotic properties of the resulting estimates.

Other semiparametric mixture cure models have been studied in the literature. For example, Lu and Ying (2004) proposed a general class of transformation cure models where the linear transformation model is used for failure times of susceptible subjects. The authors developed a set of martingale representation-based asymptotic unbiased estimating equations for

parameter estimation, and derived the large sample properties of the resulting estimators; however, the proposed algorithm for solving the equations may fail to converge. Moreover, the resulting estimators for the regression parameters are not efficient. In standard survival data analysis when there is no cure fraction, the accelerated failure time model (Kalbfleisch and Prentice (1980), Cox and Oakes (1984)) is a useful alternative to the proportional hazards model due to its direct physical interpretation (Reid (1994)). In the presence of a nonsusceptible population, Li and Taylor (2002) and Zhang and Peng (2007) considered the accelerated failure time mixture cure model and proposed an EM-type algorithm for parameter estimation. Instead of directly maximizing the conditional likelihood in the M-step, they employed different estimation methods. Specifically, Li and Taylor (2002) used an M-estimator of Ritov (1990), while Zhang and Peng (2007) considered a modified Gehan-type weighted log-rank estimation. The theoretical properties of the proposed estimates have not been studied. In addition, since both estimates do not maximize the observed likelihood function they are not efficient, and classical likelihood based methods cannot be applied here to obtain the variance of the proposed estimates. They all use the bootstrap method to obtain the variance estimates.

In this paper, we propose a kernel-based nonparametric maximum likelihood estimation method for the accelerated failure time mixture cure model. An EM algorithm is developed to implement the estimation. As opposed to the methods of Li and Taylor (2002) and Zhang and Peng (2007), we maximize a kernel-smoothed conditional profile likelihood in the M-step. The proposed kernel estimation method is motivated by a recent work of Zeng and Lin (2007) in efficient estimation for the accelerated failure time model without cure fraction. We show that with a proper choice of the kernel bandwidth parameter, the resulting estimates are consistent, asymptotically normal and efficient. In addition, we propose an EM-aided numerical differentiation method to compute individual profile likelihood scores, then estimate the limiting covariance matrix by inverting the empirical Fisher information matrix obtained from them.

## 2. Model and Estimation

Under the mixture modelling approach, a decomposition of the event time is given by

$$T = \eta T^* + (1 - \eta)\infty, \quad (2.1)$$

where  $T^* < \infty$  denotes the failure time of a susceptible subject and  $\eta$  indicates, by the value 1 or 0, whether the study subject is susceptible or not. The accelerated failure time mixture cure model is specified by the following two terms:

$$\log(T^*) = \beta' Z + \varepsilon, \quad (2.2)$$

$$P(\eta=1|X, Z) = \frac{\exp(\gamma' X)}{1 + \exp(\gamma' X)}, \quad (2.3)$$

where  $\beta$ ,  $p$ -dimensional, and  $\gamma$ ,  $q$ -dimensional, are unknown regression parameter vectors of primary interest, and  $\varepsilon$  is the error term with a completely unspecified continuous density function. The baseline covariates  $Z$  and  $X$  may share some common components and  $X$  includes 1 so that  $\gamma$  contains the intercept term. Furthermore, we assume that the censoring

time  $C$  is independent of  $T^*$  and  $\eta$ , conditional on  $Z$  and  $X$ . Define  $\tilde{T} = \min(T, C)$  and  $\delta = I(T \leq C)$ . Then the observations consist of  $(\tilde{T}_i, \delta_i, Z_i, X_i)$ ,  $i = 1, \dots, n$ , independent copies of  $(\tilde{T}, \delta, Z, X)$ .

The observed likelihood function can be written as

$$L_n^o(\theta, f) = \prod_{i=1}^n \left( \left[ \pi(\gamma' X_i) e^{-\beta' Z_i} f\{e^{R_i(\beta)}\} \right]^{\delta_i} \times \left[ 1 - \pi(\gamma' X_i) + \pi(\gamma' X_i) S\{e^{R_i(\beta)}\} \right]^{1-\delta_i} \right), \quad (2.4)$$

where  $\theta = (\beta', \gamma)'$ ,  $\pi(a) = \exp(a) / \{1 + \exp(a)\}$ ,  $R_i(\beta) = \log(\tilde{T}_i) - \beta' Z_i$ , and  $f$  and  $S$  are, respectively, the density and survival functions of  $e^e$ . The direct maximization of (2.4) with respect to  $\theta$  and  $f$  is quite intractable due to the presence of a cure fraction. In addition, even when there is no cure fraction (i.e.  $\pi \equiv 1$ ), as discussed by Zeng and Lin (2007), the maximum of (2.4) does not exist. Instead, they proposed to maximize a kernel-smoothed profile likelihood function to obtain the estimates.

Here, we develop an EM algorithm to approximately maximize the observed likelihood function  $L_n^o(\theta, f)$ , in which a kernel-smoothed conditional profile likelihood is used in the M-step. To be specific, we first introduce the complete likelihood

$$L_n^c(\theta, f) = \prod_{i=1}^n \left\{ \left[ \pi(\gamma' X_i) e^{-\beta' Z_i} f\{e^{R_i(\beta)}\} \right]^{\delta_i \eta_i} \times \left( \left\{ 1 - \pi(\gamma' X_i) \right\}^{1-\eta_i} \left[ \pi(\gamma' X_i) S\{e^{R_i(\beta)}\} \right]^{\eta_i} \right)^{1-\delta_i} \right\}. \quad (2.5)$$

Write  $l_n^o(\theta, f) = (1/n) \log\{L_n^o(\theta, f)\}$  and  $l_n^c(\theta, f) = (1/n) \log\{L_n^c(\theta, f)\}$ . Note that  $\delta_i = 1$  implies  $\eta_i = 1$ . Then we have  $l_n^c(\theta, f) = l_{n,1}^c(\gamma) + l_{n,2}^c(\beta, \lambda)$ , where

$$l_{n,1}^c(\gamma) = \frac{1}{n} \sum_{i=1}^n [\eta_i \log\{\pi(\gamma' X_i)\} + (1 - \eta_i) \log\{1 - \pi(\gamma' X_i)\}], \quad (2.6)$$

$$l_{n,2}^c(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^n \left( -\delta_i \beta' Z_i + \delta_i \log[\lambda\{e^{R_i(\beta)}\}] - \eta_i \Lambda\{e^{R_i(\beta)}\} \right), \quad (2.7)$$

with  $\lambda$  and  $\Lambda$  being the hazard and cumulative functions of  $e^e$ , respectively.

In the E-step of the EM algorithm, we compute the conditional expectations of  $l_{n,1}^c$  and  $l_{n,2}^c$  given the observed data and current parameter estimates. Let  $\mathcal{O}_i$  denote the observed data of the  $i$ th study subject and  $\Omega^{[k]} = (\beta^{[k]}, \theta^{[k]}, \hat{\lambda}^{[k]})$  be the parameter estimates at the  $k$ th iteration. Then

$$\begin{aligned} \tilde{l}_{n,1}^c(\gamma) &= E\{l_{n,1}^c(\gamma) | \mathcal{O}, \widehat{\Omega}^{[k]}\} \\ &= \frac{1}{n} \sum_{i=1}^n [w_i^{[k]} \log\{\pi(\gamma' X_i)\} + (1 - w_i^{[k]}) \log\{1 - \pi(\gamma' X_i)\}], \end{aligned} \quad (2.8)$$

$$\begin{aligned} \tilde{l}_{n,2}^c(\beta, \lambda) &= E\{l_{n,2}^c(\beta, \lambda) | \mathcal{O}, \widehat{\Omega}^{[k]}\} \\ &= \frac{1}{n} \sum_{i=1}^n \left( -\delta_i \beta' Z_i + \delta_i \log[\lambda \{e^{R_i(\beta)}\}] - w_i^{[k]} \Lambda\{e^{R_i(\beta)}\} \right), \end{aligned} \quad (2.9)$$

where  $\mathcal{O} = \{\mathcal{O}_i: i = 1, \dots, n\}$  and

$$\begin{aligned} w_i^{[k]} &= P(\eta_i = 1 | \mathcal{O}_i, \widehat{\Omega}^{[k]}) \\ &= \delta_i + (1 - \delta_i) \frac{\pi(\widehat{\gamma}^{[k]} | X_i) \widehat{S}^{[k]}(e^{R_i(\widehat{\beta}^{[k]})})}{1 - \pi(\widehat{\gamma}^{[k]} | X_i) + \pi(\widehat{\gamma}^{[k]} | X_i) \widehat{S}^{[k]}(e^{R_i(\widehat{\beta}^{[k]})})}. \end{aligned} \quad (2.10)$$

In the M-step, we maximize  $\tilde{l}_{n,1}^c(\gamma)$  with respect to  $\gamma$ , and maximize  $\tilde{l}_{n,2}^c(\beta, \lambda)$  with respect to  $\beta$  and  $\lambda$ , respectively. The maximization of  $\tilde{l}_{n,1}^c(\gamma)$  can be easily done using the Newton-Raphson method. But for  $\tilde{l}_{n,2}^c(\beta, \lambda)$ , following the discussion of Zeng and Lin (2007), we can show that it cannot achieve its maximum for finite  $\beta$  due to the lack of smoothness in the estimation of  $\lambda$ . To overcome this difficulty, a smoothed estimate for  $\lambda$  needs to be introduced. As in Zeng and Lin (2007), we start with a piece-wise constant hazard function and then study its limits using a kernel smoother. To be specific, for all possible  $\beta$ 's in a bounded set, we consider a compact interval  $[0, M]$  containing all  $e^{R_i(\beta)}$ 's and partition this interval into  $J_n$  equally spaced subintervals,  $0 \equiv x_0 < x_1 < \dots < x_{J_n} \equiv M$ . Then the piece-wise constant hazard function is written as

$$\lambda(x) = \sum_{j=1}^{J_n} c_j I(x \in [x_{j-1}, x_j)), \quad 0 \leq x < M.$$

Thus, for any  $x$ , the cumulative hazard function can be represented as

$$\Lambda(x) = \sum_{j=1}^{J_n} c_j (x - x_{j-1}) I(x_{j-1} \leq x < x_j) + \frac{M}{J_n} \sum_{j=1}^{J_n} c_j I(x \geq x_j).$$

Plugging these functions into  $\tilde{l}_{n,2}^c(\beta, \lambda)$ , for a fixed  $\beta$ , we maximize the resulting likelihood function with respect to the  $c_j$ 's. The solution of  $c_j$  can be obtained from the score equation and has the closed form

$$\widehat{c}_j = \frac{\sum_{i=1}^n \delta_i I(x_{j-1} \leq e^{R_i(\beta)} < x_j)}{\sum_{i=1}^n w_i^{[k]} [ \{e^{R_i(\beta)} - x_{j-1}\} I(x_{j-1} \leq e^{R_i(\beta)} < x_j) + I(e^{R_i(\beta)} \geq x_j) M / J_n ]}.$$

After plugging the  $\widehat{c}_j$ 's into  $\tilde{l}_{n,2}^c(\beta, \lambda)$  and discarding some constants, we obtain the conditional profile log-likelihood function

$$\begin{aligned}
\tilde{l}_{n,2}^p(\beta) &= -\frac{1}{n} \sum_{i=1}^n \delta_i \beta' Z_i \\
&+ \sum_{j=1}^{J_n} \left\{ \frac{1}{n} \sum_{i=1}^n \delta_i I(e^{R_i(\beta)} \in [x_{j-1}, x_j]) \right\} \\
&\times \log \left\{ \frac{J_n}{nM} \sum_{i=1}^n \delta_i I(e^{R_i(\beta)} \in [x_{j-1}, x_j]) \right\} \\
&- \sum_{j=1}^{J_n} \left\{ \frac{1}{n} \sum_{i=1}^n \delta_i I(e^{R_i(\beta)} \in [x_{j-1}, x_j]) \right\} \\
&\times \log \left\{ \frac{J_n}{nM} \sum_{i=1}^n w_i^{[k]} \{e^{R_i(\beta)} - x_{j-1}\} I(e^{R_i(\beta)} \in [x_{j-1}, x_j]) + \frac{1}{n} \sum_{i=1}^n w_i^{[k]} I(e^{R_i(\beta)} \geq x_j) \right\}.
\end{aligned}$$

Applying similar techniques as used by Zeng and Lin (2007), we can show that as  $n \rightarrow \infty$ ,  $J_n \rightarrow \infty$  and  $J_n/n \rightarrow 0$ ,  $\tilde{l}_{n,2}^p(\beta)$  is asymptotically equivalent to the kernel-smoothed conditional profile log-likelihood function

$$\tilde{l}_{n,2}^s(\beta) = -\frac{1}{n} \sum_{i=1}^n \delta_i \log(\tilde{T}_i) + \frac{1}{n} \sum_{i=1}^n \delta_i \log \left[ \frac{1}{n} \sum_{j=1}^n \delta_j K_h \{R_j(\beta) - R_i(\beta)\} \right] - \frac{1}{n} \sum_{i=1}^n \delta_i \log \left\{ \frac{1}{n} \sum_{j=1}^n w_j^{[k]} \int_{-\infty}^{R_j(\beta) - R_i(\beta)} K_h(u) du \right\}, \quad (2.11)$$

where  $K_h(x) = K(x/h)/h$  is a kernel function with bandwidth  $h$  obtained from a symmetric probability density function  $K(x)$ . Since  $\tilde{l}_{n,2}^s(\beta)$  is a smooth function of  $\beta$ , it can be easily maximized using the Newton-Raphson or other gradient-based algorithms. Let  $\hat{\beta}^{[k+1]}$  denote the maximizer of  $\tilde{l}_{n,2}^s(\beta)$ . Then given  $\hat{\beta}^{[k+1]}$ , we estimate  $\lambda(x)$  by

$$\hat{\lambda}^{[k+1]}(x) = \frac{x^{-1} \sum_{j=1}^n \delta_j K_h \{R_j(\hat{\beta}^{[k+1]}) - \log x\}}{\sum_{j=1}^n w_j^{[k]} \int_{-\infty}^{R_j(\hat{\beta}^{[k+1]}) - \log x} K_h(u) du}, \quad x > 0. \quad (2.12)$$

In summary, the EM-algorithm is given as follows: *Step 0.* Set  $\Omega \equiv (\theta, \lambda)$  at its initial estimate  $\hat{\Omega}^{[0]}$ ; *Step 1.* At the  $k$ th iteration, compute  $w_i^{[k]}$ ,  $\tilde{l}_{n,1}^c(\gamma)$ , and  $\tilde{l}_{n,2}^s(\beta)$  based on the current estimates  $\hat{\Omega}^{[k]}$ ; *Step 2.* Compute updated estimates  $\hat{\gamma}^{[k+1]}$  and  $\hat{\beta}^{[k+1]}$  by minimizing  $-\tilde{l}_{n,1}^c(\gamma)$  and  $-\tilde{l}_{n,2}^s(\beta)$  obtained from Step 1, respectively, then compute the estimator  $\hat{\lambda}^{[k+1]}(x)$  using (2.12); *Step 3.* Set  $k = k + 1$ . Repeat Steps 1 and 2 until convergence.

For computational convenience, we choose the standard normal density function for the kernel  $K$ . In order to calculate  $w_i^{[k]}$ , we also need  $\hat{S}^{[k]}$ , or equivalently  $\hat{\Lambda}^{[k]}$  which can be obtained from  $\hat{\lambda}^{[k]}$  by numerical integration. For the stability of the EM algorithm, we set  $\hat{S}^{[k]}(x) = 0$  for  $x > e^{R_{(L)}(\hat{\beta}^{[k]})}$ , where  $R_{(L)}(\hat{\beta}^{[k]})$  is the largest uncensored residual of  $n$  study subjects. Such a constraint has been widely used in the estimation of semiparametric mixture cure models (Peng and Dear (2000), Lu and Ying (2004), Zhang and Peng (2007), among

others). The choice of initial estimates  $\hat{\Omega}^{[0]}$  and the kernel bandwidth  $h$  is discussed in detail in Section 4.

### 3. Asymptotic Properties and Variance Estimation

Let  $\theta_0 = (\beta_0', \gamma_0')$  denote the true value of  $\theta$  and  $\lambda_0(x)$  be the true value of  $\lambda(x)$ . In addition, let  $\hat{\theta}$  and  $\hat{\lambda}(x)$  denote the estimates at convergence obtained from the EM algorithm. We derive here the asymptotic properties of the estimates  $(\hat{\theta}, \hat{\lambda})$  and propose a EM-aided numerical differentiation method for computing the variance of  $\hat{\theta}$  based on profile likelihood scores.

#### Theorem 3.1

Suppose that the regularity conditions (C1)–(C5) hold and that  $h \rightarrow 0$  and  $nh^2 \rightarrow \infty$ . Then

$$\sup_{x \in [0, \tau]} |\hat{\Lambda}_n(x) - \Lambda_0(x)| \rightarrow 0 \text{ a.s. and } \left| \hat{\theta}_n - \theta_0 \right| \rightarrow 0 \text{ a.s.}$$

#### Theorem 3.2

Suppose that the regularity conditions (C1)–(C5) hold and that  $nh^{2m} \rightarrow 0$  and  $nh^6 \rightarrow \infty$ . Then, as  $n \rightarrow \infty$ ,  $n^{1/2}(\hat{\theta} - \theta_0)$  converges in distribution to a mean-zero normal random vector with covariance matrix achieving the semiparametric efficiency bound of  $\theta_0$ .

The proofs of the above theorems, along with their irregularity conditions, are given in the Appendix, available at <http://www.stat.sinica.edu.tw/statistica>.

Next, we derive the variance estimate of  $\hat{\theta}$  obtained from the proposed EM algorithm. The variance formula of Louis (1982) for parametric EM algorithms is not really feasible here due to the infinite dimensional parameter  $\lambda(x)$  and the kernel-smoothed conditional profile likelihood used in the M-step. An alternative way to compute variance estimates is to invert the empirical Fisher information matrix of the profile likelihood; this has been widely done in nonparametric maximum likelihood estimation (Nielsen, Gill, Andersen and Sørensen (1992), Murphy, Rossini and van der Vaart (1997), Zeng, Cai, and Shen (2006), among others). Theoretical properties of the profile likelihood estimation have been rigorously studied in Murphy and van der Vaart (2000). The empirical Fisher information matrix of the profile likelihood usually does not have an analytical form. Chen and Little (1999) proposed an EM-aided numerical differentiation method for computing the second derivative of the log profile likelihood at the maximum; the validity of the method has been established by the authors. Here we adopt a modification of the method. To be specific, we used the EM-aided numerical differentiation method to calculate the individual profile likelihood scores at the maximum, then obtain the empirical Fisher information matrix of the profile likelihood based on these scores. This modified approach can ensure that the resulting information matrix is positively definite.

Write  $l_{n,1}^c(\gamma) = \sum_{i=1}^n l_{n,1i}^c(\gamma)$  and  $l_{n,2}^c(\beta, \lambda) = \sum_{i=1}^n l_{n,2i}^c(\beta, \lambda)$ . Define  $l_{n,i}^c(\theta, \lambda) = l_{n,1i}^c(\gamma) + l_{n,2i}^c(\beta, \lambda)$ ,  $i = 1, \dots, n$ . Perturb the  $j$ th component of  $\theta$  by a small amount  $d$ , denoted as  $\theta_j^*$ . Fix the  $j$ th component  $\theta_j$  of  $\theta$  at  $\theta_j^*$  and run the proposed EM algorithm to compute the estimates for all other parameters. Let  $\theta_{[-j]}^*$  denote the resulting estimate for the parameters in  $\theta$  except for  $\theta_j$ , and  $\lambda_{[-j]}^*$  denote the resulting estimate for  $\lambda$ . Following the suggestion of Chen and Little (1999), we use two-sided perturbation, i.e.,  $\theta_j^{*1} = \hat{\theta}_j + d$  and  $\theta_j^{*2} = \hat{\theta}_j - d$ . Correspondingly, we

have the estimates  $(\theta_{[-j]}^{*1}, \lambda_{[-j]}^{*1})$  and  $(\theta_{[-j]}^{*2}, \lambda_{[-j]}^{*2})$ . Then the  $j$ th component of the profile likelihood score for the  $i$ th study subject is

$$S_{ij}^{pl} = [E_{\Omega=\Omega_j^*1} \{l_{n,i}^c(\Omega_j^*1) | \mathcal{O}_i\} - E_{\Omega=\Omega_j^*2} \{l_{n,i}^c(\Omega_j^*2) | \mathcal{O}_i\}] (2d)^{-1},$$

where the two expectations are taken with respect to  $\eta_i$  given the observed data  $\mathcal{O}_i$  and the estimates  $\Omega_j^*1 = (\theta_j^{*1}, \theta_{[-j]}^{*1}, \lambda_{[-j]}^{*1})$  and  $\Omega_j^*2 = (\theta_j^{*2}, \theta_{[-j]}^{*2}, \lambda_{[-j]}^{*2})$ , respectively. Define

$$S_i^{pl} = (S_{i1}^{pl}, \dots, S_{i,p+q}^{pl})'. \text{ The empirical Fisher information matrix of the profile likelihood can be obtained as } \widehat{I} = \sum_{i=1}^n S_i^{pl} (S_i^{pl})'.$$

## 4. Numerical Studies

### 4.1. Simulations

We examine here the finite sample performance of the proposed estimates. Event times  $T$  were generated from the accelerated failure time cure model defined in (2.2) and (2.3). A binary covariate  $Z$  was generated from a Bernoulli distribution with success probability 0.5. We set  $X = (1, Z)$ . The error  $\varepsilon$  in (2.2) was given by  $a_0 + a_1 V$ , where  $a_0, a_1$  were two constants and  $V$  was a random variable generated from three scenarios: the extreme value distribution, the standard logistic distribution, and the standard normal distribution. The censoring time  $C$  was generated from a uniform distribution on  $[0, a_2]$ , with  $a_2$  a constant. The parameters were set as  $\beta_0 = 1.0$  and  $\gamma_0 = (0.5, -0.5)$  or  $(1.0, -0.5)$ , which give approximately 43.9% and 32.3% overall cure fractions, respectively. The constants  $a_0, a_1$  and  $a_2$  were chosen to obtain the desired censoring proportions. For example, when the error was from the extreme value distribution, we chose  $a = (a_0, a_1, a_2) = (-0.5, 0.5, 8)$ , which gives approximately 50.5% censoring proportion for the 43.9% cure fraction and 40.6% censoring proportion for the 32.3% cure fraction. In each scenario, we conducted 500 runs of simulations with the sample size  $n = 100$ .

For the bandwidth parameter  $h$  in the kernel-smoothed conditional profile likelihood, we followed the suggestions of Zeng and Lin (2007) and used the optimal bandwidths (Jones (1990), Jones and Sheather (1991)),  $(8\sqrt{2}/3)^{1/5} \sigma_1 n^{-1/5}$  and  $4^{1/3} \sigma_2 n^{-1/3}$ , where  $\sigma_1$  is the sample standard deviation of the  $(\log \tilde{T} - \tilde{\beta}_{1,ls}' Z)$  for uncensored data with  $\hat{\beta}_{1,ls}$  the least square estimate of  $\beta$  using only uncensored data, while  $\sigma_2$  was the sample standard deviation of the  $(\log \tilde{T} - \tilde{\beta}_{2,ls}' Z)$  for all the data with  $\hat{\beta}_{2,ls}$  the corresponding least square estimate of  $\beta$ . We considered both bandwidths in the simulations and found the results comparable. Thus, we only present the results using the first bandwidth here.

In the proposed EM algorithm, we also need to obtain the initial estimate  $\hat{\Omega}^{[0]}$ . Here we chose the initial estimate for  $\beta$  as  $\hat{\beta}^{[0]} = \hat{\beta}_{1,ls}$ . For computing  $\hat{\gamma}^{[0]}$ , we considered a logistic regression of  $\delta$  on  $X$ , i.e., we treated all the censored subjects as cured at the initial step. For  $\hat{S}^{[0]}(\cdot)$ , we used the Kaplan-Meier estimate based on the  $(\tilde{T}e^{-(\hat{\beta}^{[0]})'Z}, \delta)$ . In addition, we set  $\hat{S}^{[0]}(x) = 0$  for all the  $x$  greater than the largest uncensored transformed time  $\tilde{T}e^{-(\hat{\beta}^{[0]})'Z}$ . We found that the proposed initial estimates combined with the chosen bandwidth parameter worked quite well in all our simulations and that the proposed EM algorithm usually converged within 10 iterations. To compute the variance estimate of  $\hat{\theta}$ , we used the proposed EM-aided numerical differentiation method discussed at the end of the previous section.

Following the suggestion of Chen and Little (1999), we chose the perturbation  $\delta = \alpha/n$ , with  $\alpha$  a positive constant. We tried different values of  $\alpha$  in simulations, and found  $\alpha = 2$  gave reasonable variance estimates for all the scenarios. The proposed EM algorithm and the EM-aided numerical differentiation method were implemented in R, the code is available from the author upon request. The simulation results are summarized in Table 2. For comparison, we also report the rank regression-based estimator of Zhang and Peng (2007), denoted as ZP. We did not evaluate the variance of the ZP estimators because of the heavy computational burden. The relative efficiency (RE) of the ZP estimator compared with the proposed estimator was computed as the ratio of sample variances of the two estimators with the ZP estimator being a reference.

It is clear that the proposed estimates were unbiased in all the scenarios, and the proposed variance estimates based on the EM-aided numerical differentiation method matched the sample standard deviations of the parameter estimates reasonably well. Furthermore, the Wald-type 95% confidence intervals had proper coverage probabilities. Under the extreme value error, the proposed estimators were slightly less efficient than the ZP estimators. Under other error distributions, the proposed estimators were generally more efficient than the ZP estimators, especially for the short-term parameters  $\beta$ .

#### 4.2. Application to breast cancer data

We applied the proposed method to a data set obtained from a breast cancer study of 139 breast cancer patients who were randomly assigned to three treatment groups (control, treatment A and treatment B). The endpoint of interest is time to relapse or death. There were 95 censored and 44 uncensored among 139 patients. Besides the treatment assignment, four other covariates, namely clinical stage I, pathological stage, histological stage, and number of lymph nodes were recorded. The data set was first analyzed by Farewell (1986) using a Weibull-logistic cure model. Kuk and Chen (1992) and Peng and Dear (2000) further studied a subset of the data with three covariates: treatment assignment, clinical stage I, and number of lymph nodes, using the proportional hazards cure model. The number of lymph nodes was converted to a binary covariate indicating whether more than four lymph nodes had disease involvement.

To check the proportional hazards assumption for the survival distribution of susceptible subjects, we used the method of Zhang and Peng (2007). To be specific, we plot in Figure 1 the logarithm of the estimated cumulative hazard functions for the uncensored patients in the three treatment groups, respectively, based on the Kaplan-Meier estimators of survival functions. Figure 1 shows that the logarithm of cumulative hazard functions of the three treatment groups clearly cross each other and thus the proportional hazards assumption is not appropriate for this data set. Here, instead, we considered the accelerated failure time cure model for the same subset of the data and applied the proposed kernel-smoothing-based profile likelihood method for parameters estimation. The results are summarized in Table 1. Based on the results, we observe that compared with the control group, treatment A has a significant beneficial effect on the short-term survival of susceptible subjects while treatment B has a significant beneficial effect on long-term survival, i.e. the cured fraction. In addition, clinical stage I has significant effects on both the short-term and the long-term survivals while the number of lymph nodes is not significant for either of them. Our findings are generally in agreement with those obtained by Peng and Dear (2000) using the proportional hazards cure model. Note that the interpretations of short-term parameters ( $\beta$ 's) in the proportional hazards cure model and accelerated failure time cure model are different. In general, they show opposite signs since one is for the hazard ratio and the other directly describes the log survival time.



## 5. Concluding Remarks

In this paper, we have developed a kernel-smoothing-based EM algorithm for efficient estimation in the accelerated failure time cure model, and derived the asymptotic properties for the resulting estimates. A convenient EM-aided numerical differentiation method was also proposed for computing the variance estimates. The mixture modeling approach is one of the commonly used methods for formulating cure models. Another widely used approach is to consider bounded cumulative hazard models (see Tsodikov (1998, 2001), Tsodikov, Ibrahim and Yakovlev (2003), Zeng, Ying and Ibrahim (2006)). Such cure models may have nice biological interpretations, but the short-term and long-term effects cannot be naturally separated as in mixture cure models. It is of great interest to develop some diagnostic tools for various types of cure models.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

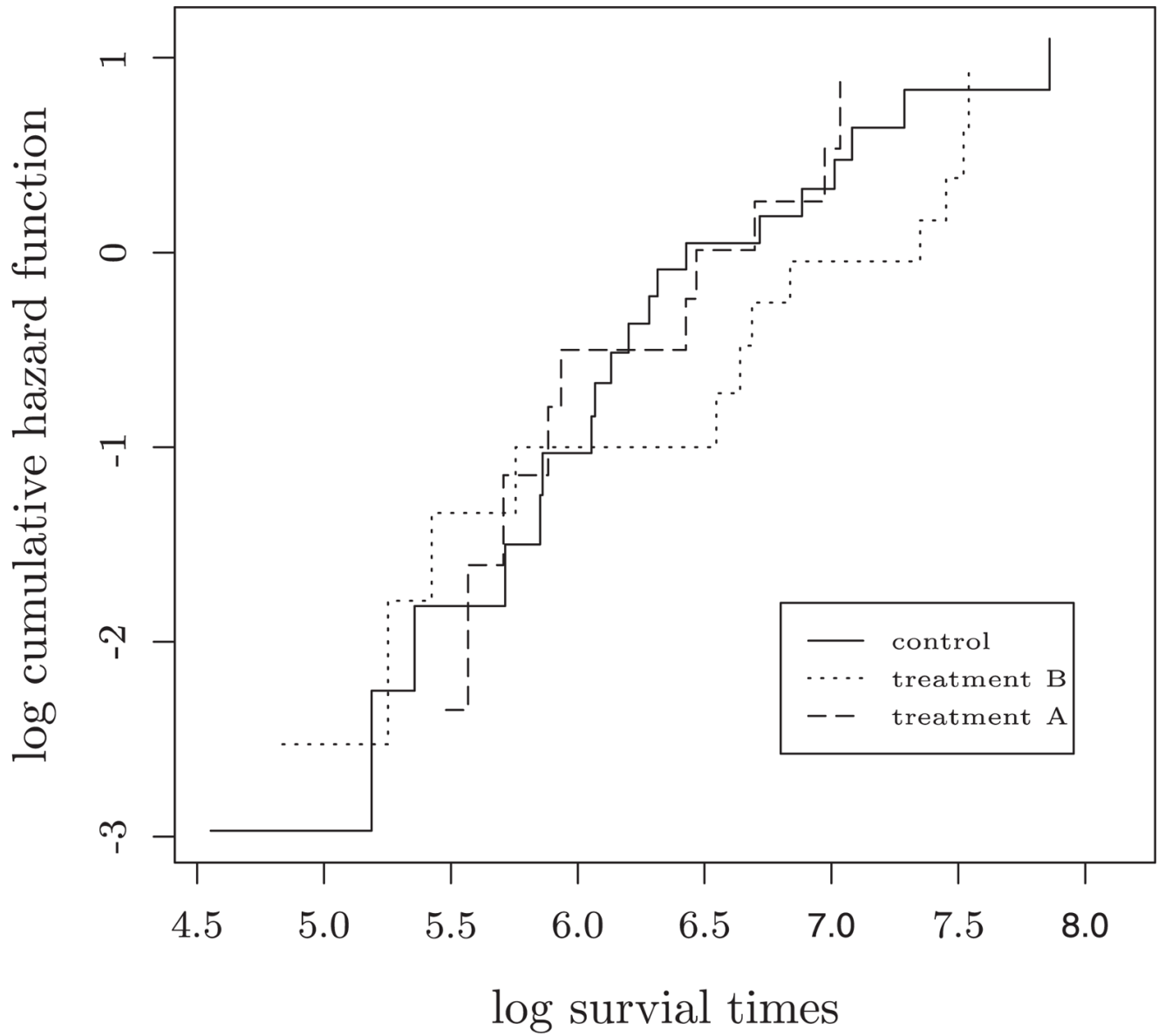
## Acknowledgments

The author is grateful to an associate editor and an anonymous referee for helpful comments and suggestions. This research was supported in part by the grant from the National Institutes of Health.

## References

- Ash, RB. Real Analysis and Probability. Academic Press; San Diego: 1972.
- Berkson J, Gage RP. Survival curve for cancer patients following treatment. *J Amer Statist Assoc* 1952;47:501–515.
- Bickel, PJ.; Klaassen, CAJ.; Ritov, Y.; Wellner, JA. Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press; Baltimore: 1993.
- Chen HY, Little RJA. Proportional hazards regression with missing covariates. *J Amer Statist Assoc* 1999;94:896–908.
- Cox DR. Regression models and life tables (with Discussion). *J Roy Statist Soc Ser B* 1972;34:187–220.
- Cox, DR.; Oakes, D. Analysis of Survival Data. Chapman and Hall; New York: 1984.
- Fang HB, Li G, Sun J. Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scand J Statist* 2005;32:59–75.
- Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* 1982;38:1041–1046. [PubMed: 7168793]
- Farewell VT. Mixture models in survival analysis: are they worth the risk? *Canad J Statist* 1986;14:257–262.
- Jones MC. The performance of kernel density functions in kernel distribution function estimation. *Statist Probab Lett* 1990;9:129–132.
- Jones MC, Sheather SJ. Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statist Probab Lett* 1991;11:511–514.
- Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. 1. Wiley; New York: 1980.
- Kress, R. Linear Integral Equations. Springer-Verlag; New York: 1989.
- Kuk AYC, Chen CH. A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 1992;79:531–541.
- Li CS, Taylor JMG. A semi-parametric accelerated failure time cure model. *Statist Medicine* 2002;21:3235–3247.
- Louis TA. Finding the observed information when using the EM algorithm. *J Roy Statist Soc Ser B* 1982;44:226–233.

- Lu W. Maximum likelihood estimation in the proportional hazards cure model. *Ann Inst Statist Math* 2008;60:545–574.
- Lu W, Ying Z. On semiparametric transformation cure models. *Biometrika* 2004;91:331–343.
- Luenberger, DG. *Optimization by Vector Space Methods*. John Wiley; New York: 1969.
- Murphy SA, Rossini AJ, van der Vaart AW. Maximum likelihood estimation in the proportional odds model. *J Amer Statist Assoc* 1997;92:968–976.
- Murphy SA, van der Vaart AW. On profile likelihood. *J Amer Statist Assoc* 2000;95:45–485.
- Nielsen GG, Gill RD, Andersen PK, Sørensen TIA. A counting process approach to maximum likelihood estimation in frailty models. *Scand J Statist* 1992;19:25–43.
- Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics* 2000;56:237–243. [PubMed: 10783801]
- Reid N. A conversation with Sir David Cox. *Statist Sci* 1994;9:439–455.
- Ritov Y. Estimation in a linear regression model with censored data. *Ann Statist* 1990;18:303–328.
- Scharfsten DO, Tsiatis AA, Gilbert PB. Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Analysis* 1998;4:355–391. [PubMed: 9880995]
- Schuster EF. Estimation of a probability density function and its derivatives. *Ann Math Statist* 1969;40:1187–1195.
- Serfling, RJ. *Approximation Theorems of Mathematical Statistics*. John Wiley; New York: 1980.
- Shorack, GR.; Wellner, JA. *Empirical Processes with Applications to Statistics*. John Wiley; New York: 1986.
- Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics* 2000;56:227–236. [PubMed: 10783800]
- Tsodikov AD. A proportional hazards model taking account of long-term survivors. *Biometrics* 1998;54:138–146.
- Tsodikov AD. Estimation of survival based on proportional hazards when cure is a possibility. *Mathematical and Computer Modelling* 2001;33:1227–1236.
- Tsodikov AD, Ibrahim JG, Yakovlev AY. Estimating cure rates from survival data: an alternative to two-component mixture models. *J Amer Statist Assoc* 2003;98:1063–1078.
- van der Vaart AW. Efficiency of infinite dimensional M-estimators. *Statist Neerlandica* 1995;49:9–30.
- van der Vaart, AW.; Wellner, JA. *Weak Convergence and Empirical Processes*. Springer-Verlag; New York: 1996.
- Zeng D, Cai J, Shen Y. Semiparametric additive risks model for interval-censored data. *Statist Sinica* 2006;16:287–302.
- Zeng D, Lin DY. Efficient estimation for the accelerated failure time model. *J Amer Statist Assoc* 2007;102:1387–1396.
- Zeng D, Ying G, Ibrahim JG. Semiparametric transformation models for survival data with a cure fraction. *J Amer Statist Assoc* 2006;101:670–684.
- Zhang J, Peng Y. A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statist Medicine* 2007;26:3157–3171.



**Figure 1.** Logarithm of the cumulative hazard functions for uncensored subjects.

Table 1

Analysis results for breast cancer data.

	$\beta$	SE	p-value	$\hat{\gamma}$	SE	p-value
intercept				0.210	0.491	0.669
treatment A	1.113	0.434	0.010	-0.067	0.133	0.612
treatment B	-0.107	0.421	0.800	-1.170	0.583	0.045
clinical stage I	1.350	0.562	0.016	-0.338	0.116	0.004
Lymph nodes	-0.552	0.485	0.255	1.152	0.763	0.131

<sup>†</sup>SE, estimated standard error

Table 2

Simulation results for accelerated failure time cure model.

Case I ( $\gamma_0 = 0.5$ )		Proposed					ZP	
Parameters	Bias	SD	SE	CP	Bias	SD	RE	
extreme value error								
$\beta_0$	0.017	0.220	0.220	0.938	0.015	0.197	1.247	
$\gamma_0$	0.020	0.322	0.335	0.960	0.012	0.305	1.112	
$\gamma_2$	0.026	0.466	0.479	0.942	0.030	0.452	1.064	
logistic error								
$\beta_0$	0.018	0.138	0.158	0.964	0.013	0.154	0.801	
$\gamma_0$	0.012	0.314	0.322	0.960	0.016	0.316	0.988	
$\gamma_2$	0.023	0.462	0.465	0.950	0.021	0.470	0.965	
normal error								
$\beta_0$	0.031	0.176	0.182	0.948	0.004	0.190	0.859	
$\gamma_0$	0.028	0.321	0.317	0.958	0.009	0.321	0.999	
$\gamma_2$	0.003	0.483	0.457	0.940	0.032	0.471	1.052	
Case II ( $\gamma_0 = 1.0$ )		Proposed					ZP	
extreme value error								
$\beta_0$	0.018	0.178	0.202	0.962	0.005	0.179	0.992	
$\gamma_0$	0.035	0.375	0.369	0.944	0.031	0.341	1.208	
$\gamma_2$	0.014	0.542	0.518	0.938	-0.008	0.503	1.163	
logistic error								
$\beta_0$	0.015	0.127	0.142	0.970	0.007	0.143	0.790	
$\gamma_0$	0.035	0.344	0.356	0.966	0.037	0.349	0.971	
$\gamma_2$	-0.002	0.521	0.498	0.934	-0.008	0.536	0.946	

Parameters	Proposed					ZP		
	Bias	SD	SE	CP	RE	Bias	SD	RE
normal error								
$\beta_0$	0.012	0.145	0.162	0.962	0.017	0.160	0.820	
$\gamma_{01}$	0.041	0.366	0.355	0.958	0.036	0.370	0.980	
$\gamma_{02}$	0.040	0.534	0.502	0.940	0.040	0.529	1.018	

<sup>†</sup>SD, sample standard deviation of parameter estimates; SE, mean of estimated standard errors; CP, coverage probability; RE, relative efficiency of Zhang and Peng's estimator (ZP) compared with the proposed estimator (Proposed).