Statistica Sinica **19** (2009), 377-390

A CHANGE-POINT MODEL FOR SURVIVAL DATA WITH LONG-TERM SURVIVORS

Xiaobing Zhao[1], Xianyi Wu[2] and Xian Zhou[3]

[1]*Jiang Nan University,* [2]*East China Normal University and* [3]*Macquarie University*

Abstract: Change-point hazard models have been extensively investigated by many authors, but the literature on change-point problems with survival data subject to censoring is rather small. In an earlier example provided by Matthews and Farewell (1982), a set of nonlymphoblastic leukemia data were fitted by using a change-point model. But for that data set, the Kaplan-Meier estimator of the distribution function levels off well below 1, which indicates the presence of "long-term survivors" in the data. In this paper, we propose a new change-point model for survival data that accounts for long-term survivors. Estimation methods for the proposed model are investigated, and large-sample properties of the estimators are established. A simulation study is carried out to evaluate the performance of the estimating methods. As an application, the nonlymphoblastic leukemia data are re-analyzed using the new model.

Key words and phrases: Change-point hazard model, long-term survivors, pseudo-maximum likelihood, random censoring.

1. Introduction

In order to analyze data concerning treatment of leukemia patients, a change-point model was introduced into the field of survival analysis by Matthews and Farewell (1982); they assumed the hazard function $\lambda(x)$ to be constant with the exception of a jump. More specifically, the hazard function took the form

$$\lambda(x) = \beta + \theta I_{\{x>\tau\}}, \tag{1.1}$$

where I_A is the indicator function of an event A, β and $\beta + \theta$ are the hazard rates before and after the change point, respectively, τ is the change-point for the hazard rate, and θ represents the jump size at the change-point. Here β, $\beta + \theta$ and τ are positive, but θ can be either positive or negative, reflecting an increase or a decrease in hazard rate at the change-point. Three different aspects of this model have been investigated in the literature. The first is model fitting by means of maximum likelihood methods, see Matthews and Farewell (1982), Nguyen, Rogers and Walker (1984) and Loader (1991). The second is testing a

constant hazard rate versus an alternative with a change-point, considered by Matthews, Farewell and Pyke (1985), Henderson (1990) and Loader (1991). The third aspect is the structural properties, explored by Chang, Chen and Hsiung (1994) and Gijbels and Gurler (2003). Their estimation procedures rely on a certain functional of the estimated cumulative hazard function. Apart from these investigations, the literature is rather small on change-point problems arising in survival/censored data. For example, see Sen (1993) and Pons (2003), each of whom considered a Cox model with change-point according to a threshold in a covariate, and Muller and Wang (1990), who proposed a kernel method to estimate the changes in hazard rate.

Although survival models with "long-term survivors" have been extensively studied for decades and many applications have been reported, see the monograph by Maller and Zhou (1996) for example, these models have not considered possible change-point phenomena. In reality, however, long-term survivors may well exist in change-point situations. For example, when nonlymphoblastic leukemia data are fitted with a change-point model, as in Matthews and Farewell (1982), the Kaplan-Meier estimator (KME) of the distribution function levels off below 1 (cf., Figure 1), indicating the presence of long-term survivors (those patients who will never suffer a relapse of the leukemia) in the data (cf., Maller and Zhou (1996)). This inspired us to pursue the change-point model (1.1) with possible presence of long-term survivors.

The paper is structured as follows. After formulating the problem in Section 2, we present the estimation approaches in Section 3. Asymptotic properties are investigated in Section 4. Some simulation results are reported in Section 5, and an application to a set of nonlymphoblastic leukemia data is demonstrated in Section 6. The proofs of theorems are provided in the Appendix.

2. Model Formulations

To motivate our approach, we first provide a brief review on the hazard rate function for survival data with long-term survivors. The failure time is assumed to be of the form

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

where $T^* < \infty$ denotes the failure time of a "susceptible" (who is not a long-term survivor) and η indicates, by the values 1 or 0, whether the sampled subject is a susceptible or long-term survivor, independently of T^*. Let $p = \Pr(\eta = 1)$ denote the proportion of susceptibles, $p \in (0, 1]$. Then the cumulative distribution function (cdf) of T is given by

$$F(x) = \Pr(T \le x) = \Pr(T \le x|\eta = 1)\Pr(\eta = 1) = \Pr(T^* \le x)p = pF_0(x), \quad (2.1)$$

where $F_0(x) = \Pr(T^* \le x)$ is a proper cdf (with $F_0(\infty) = 1$). Correspondingly, the hazard rate function, provided that T^* has a density function $f_0(x)$, is given by

$$\lambda(x) = \frac{F'(x)}{1 - F(x)} = \frac{pf_0(x)}{1 - pF_0(x)}. \tag{2.2}$$

Note that $\lambda(x) \to 0$ as $x \to \infty$ for $p < 1$, hence the hazard rate cannot remain constant when long-term survivors are present. A simple example is the exponential lifetime with long-term survivors, where T^* is exponentially distributed with a constant hazard rate ψ. Then the hazard rate of T is

$$\lambda(x) = \frac{p\psi \exp(-\psi x)}{1 - p + p\exp(-\psi x)},$$

which is no longer constant.

Now assume that the hazard rate of T^* is as in model (1.1) Then, its density and cdf are, respectively,

$$\begin{aligned} f_0(x) &= \lambda_0(x) \exp\left\{ -\int_0^x \lambda_0(t)dt \right\} \\ &= \begin{cases} \beta \exp(-\beta x), & \text{if } \ 0 \le x \le \tau, \\ (\beta + \theta)\exp\{-\beta x - \theta(x - \tau)\}, & \text{if } \ x > \tau, \end{cases} \\ F_0(x) &= \begin{cases} 1 - \exp(-\beta x), & 0 \le x \le \tau, \\ 1 - \exp\{-\beta x - \theta(x - \tau)\}, & x > \tau. \end{cases} \end{aligned}$$

Thus, by (2.2), the hazard rate of T is

$$\lambda(x) = \begin{cases} \dfrac{p\beta \exp(-\beta x)}{1 - p + p\exp(-\beta x)}, & 0 \le x \le \tau, \\ \dfrac{p(\beta + \theta)\exp\{-\beta x - \theta(x - \tau)\}}{1 - p + p\exp\{-\beta x - \theta(x - \tau)\}}, & x > \tau. \end{cases} \tag{2.3}$$

The hazard rate $\lambda(x)$ expressed in (2.3) also has a jump at τ of size

$$\frac{p\theta \exp(-\beta\tau)}{1 - p + p\exp(-\beta\tau)},$$

which is increasing in p if $\theta > 0$ and decreasing in p for $\theta < 0$, and it reaches its maximum (minimum) value θ at $p = 1$ for the case of $\theta > 0$ ($\theta < 0$). Obviously, the hazard rate does not follow the same mathematical form before and after point τ, and $\lambda(x) \le \lambda_0(x)$.

Recently, (1.1) has been extended by Wu, Zhao and Wu (2003) to the form

$$\lambda(x) = [\beta + \theta I_{\{x > \tau\}}]\lambda_0(x, \gamma), \tag{2.4}$$

where $\lambda_0(x,\gamma)$ is a continuous baseline hazard function dependent on unknown parameter vector γ. Model (2.4) covers many important models commonly used in survival analysis, such as exponential, Weibull, extreme, log-logistic, and generalized Weibull/gamma, etc. More recently, Dupuy (2006) further allowed $\lambda_0(x,\gamma)$ in (2.4) to accommodate a time-varying covariate $Z(t)$ among individuals, and specified the following hazards model

$$\lambda(x) = [\beta + \theta I_{\{x>\tau\}}] \exp\{(\gamma_1 + \gamma_2 I_{\{x>\tau\}})^\top Z(t)\}. \tag{2.5}$$

If long-term survivors are present, by incorporating (2.2), (2.4) or (2.5), the hazard rate of T becomes

$$\lambda(x) = \frac{-pdS_0(t)}{1-p+pS_0(t)}, \tag{2.6}$$

where $S_0(t) = \exp[(\beta+\theta)\Lambda_0(x,\gamma) - \theta\Lambda_0(\tau,\gamma)]$, and $\Lambda_0(x,\gamma) = \int_0^x \lambda_0(s,\gamma)ds$ with $\lambda_0(x,\gamma)$ defined in (2.4) or (2.5).

Wu et al. (2003) and Dupuy (2006) gave estimators of τ and β, θ, γ based on (2.4) and (2.5) by the modified estimation procedure proposed by Chang et al. (1994), which may be extended to models (2.3) or (2.6). In this paper, for ease of presentation, we only investigate (2.3) based on the estimated cumulative hazard function proposed by Chang et al. (1994), and further applied by Gijbels and Gurler (2003). The approach we employ, however, is nonparametric in nature, and can be applied to more general change-point models such as (2.6) as well.

As is common in change-point models, we suppose the existence of bounds τ_1 and τ_2 such that $0 < \tau_1 \le \tau \le \tau_2 < \infty$. In medical research, one often has to deal with censored survival times due to dropout of patients from the study or the termination of the observation period. In particular, a long-term survivor will always appear as being censored, but a censored observation is not necessarily from a long-term survivor. Following the usual formulation, we postulate a "true" survival time T_i for each individual i, which is only observed if it does not exceed individual i's censoring time C_i; otherwise, we observe C_i. Whether an individual i is censored or not is recorded in a censoring indicator δ_i, with $\delta_i = 1$ if T_i is an actual failure time (uncensored) and $\delta_i = 0$ if T_i is censored. The observable survival time Y_i, possibly censored, is given by $Y_i = T_i \wedge C_i = \min(T_i, C_i)$, $i = 1,\ldots,n$. It is further assumed that T_i is independent of C_i for each i, $(T_i, C_i), i = 1,\ldots,n$, are mutually independent pairs, and $C_1,\ldots,C_n$ have a common cdf G.

3. Pseudo-Maximum Likelihood Estimation

With the change-point model defined in Section 2, we now focus on estimating the change point τ, the susceptible proportion p, and the unknown parameter

vector (β, θ). From (2.3), the log-likelihood function can be written as

$$\log L(\tau, \beta, \theta, p) = \sum_{i=1}^{n} \Big\{ \delta_i \log(p f_0(Y_i)) + (1 - \delta_i) \log[1 - p F_0(Y_i)] \Big\} \tag{3.1}$$

$$= \sum_{i=1}^{n} l(\tau, \beta, \theta, p | Y_i, \delta_i), \quad \text{say,} \tag{3.2}$$

where $l(\tau, \beta, \theta, p|y, \delta)$ is the log-likelihood of a single observation (y, δ), and given by

$$\begin{aligned} l(\tau, \beta, \theta, p|y, \delta) = \delta \Big\{ & \log(p\beta) - \beta y + I(y > \tau)(\log(1 + \theta/\beta) - \theta(y - \tau)) \Big\} \\ & + \Big\{ (1 - \delta) I(y \le \tau) \log(1 - p + p \exp(-\beta y)) \Big\} \\ & + \Big\{ (1 - \delta) I(y > \tau) \log(1 - p + p \exp(-\beta y - \theta(y - \tau))) \Big\}. \end{aligned} \tag{3.3}$$

Due to the apparent irregularity of the likelihood function, the classical maximum likelihood method cannot be used (Nguyen, Rogers and Walker (1984)), while the estimating procedure proposed by Chang et al. (1994) cannot be directly employed with a nuisance parameter p. Thus we resort to the *pseudo-likelihood* approach, which overcomes such difficulties. It was proposed by Gong and Samaniego (1981) and further studied by others including Huang (1996) and Hu (1998). The key idea is to replace the true (but unknown) "nuisance" parameters p and τ in (3.1) by their consistent estimators and then treat the log-likelihood function $\log L(\hat{\tau}, \beta, \theta, \hat{p})$, now called the *pseudo log-likelihood function*, as a usual likelihood function of β and θ to generate the pseudo-MLE $(\hat{\beta}, \hat{\theta})$ of (β, θ).

The consistent estimators of τ and p can be obtained by nonparametric methods as follows. As in Maller and Zhou (1996), we take

$$\hat{p} = \hat{F}_n(Y_{(n)}), \tag{3.4}$$

where $\hat{F}_n(t)$ denotes the Kaplan-Meier estimator of the cdf of failure times, and $Y_{(n)}$ is the largest observation. Maller and Zhou (1996, Thms. 4.1−4.4), showed that $\hat{p}$ is consistent and asymptotically normal for p when $0 < p < 1$, provided that the follow-up is sufficient in a certain sense. To get a consistent estimator of τ, note from (2.3) that the cumulative hazard function of T is now given by

$$\Lambda(x) = \begin{cases} -\log(1 - p + p \exp(-\beta x)), & 0 \le x \le \tau, \\ -\log(1 - p + p \exp(-\beta x - \theta(x - \tau))), & \tau < x. \end{cases}$$

Let

$$\Lambda^*(x) = -\log\left\{\frac{1}{p}\Big[\exp(-\Lambda(x)) - 1 + p\Big]\right\} = \begin{cases} \beta x, & 0 \le x \le \tau, \\ \beta x + \theta(x-\tau), & \tau < x, \end{cases}$$

which is a piecewise linear function of x. Further define

$$X(t) = \left[\frac{\Lambda^*(D) - \Lambda^*(t)}{D-t} - \frac{\Lambda^*(t) - \Lambda^*(0)}{t}\right] g(t(D-t)) \tag{3.5}$$

for $0 < t < D$, where $D > \tau_2$ and $g(x) = x^q, 0 \le q \le 1$. Then we have

$$\begin{aligned} X(t) &= \frac{\Lambda^*(D) g(t(D-t))}{D-t} - \frac{\Lambda^*(t) g(t(D-t))}{(D-t)t} \\ &= \theta \frac{D-\tau}{D-t} g(t(D-t)) I_{\{t \le \tau\}} + \theta \frac{\tau}{t} g(t(D-t)) I_{\{t > \tau\}}, \end{aligned}$$

which is increasing (decreasing) on $[0, \tau]$ and decreasing (increasing) on $[\tau, D]$ for $\theta > 0$ ($\theta < 0$). Let $X_n(t)$ be the empirical version of (3.5) with unknown cumulative hazard function $\Lambda(\cdot)$ and p replaced, respectively, by the Nelson-Aalen type estimator

$$\hat{\Lambda}(t) = \sum_{i: Z_i \le t} \frac{\delta_i}{n-i+1} = \int_0^t \left\{\sum_{i=1}^n H_i(s)\right\}^{-1} d \sum_{i=1} N_i(s) \tag{3.6}$$

and the KME $\hat{p}$ in (3.4), where $Z_1 < \cdots < Z_n$ are the order statistics of $Y_1, \ldots, Y_n$, $N_i(t) = 1_{[Y_i, \infty)}(t \wedge C_i)$, and $H_i(t) = 1_{(0, Y_i]}(t)$. Then similar to Chang et al. (1994), an estimator of τ is given by

$$\hat{\tau}_n = \begin{cases} \inf\{t \in [\tau_1, \tau_2] : X_n(t\pm) = \sup_{u \in [\tau_1, \tau_2]} X_n(u)\} & \text{if } \theta > 0, \\ \inf\{t \in [\tau_1, \tau_2] : X_n(t\pm) = \inf_{u \in [\tau_1, \tau_2]} X_n(u)\} & \text{if } \theta < 0. \end{cases} \tag{3.7}$$

Chang et al. (1994) showed, in the absence of long-term survivors, that $\hat{\tau}_n$ is a consistent estimator of τ, with $\hat{\tau}_n - \tau = O_p(n^{-1})$, and is asymptotically normal under some mild conditions. Using the asymptotic properties of $\hat{p}$, and the Nelson-Aalen estimator of $\Lambda(\cdot)$, we establish the asymptotic properties of $\hat{\tau}_n$ in Theorem 1 (its proof is given in the Appendix).

Theorem 1. *Assume the i.i.d. censoring model with $0 < p \le 1$, and that F is continuous at τ_H in case $\tau_H < \infty$. Further assume that $\tau_{F_0} > D$, where τ_H and τ_{F_0} are the right extremes of $H = 1 - (1-F)(1-G)$ and F_0, respectively. Then the estimator $\hat{\tau}_n$ of τ defined at* (3.7) *is consistent.*

Now we can estimate parameters β and θ as follows. First replace the parameters p and τ in (3.2) by their consistent estimators $\hat{p}$ defined in (3.4), and $\hat{\tau}_n$ defined in (3.7), respectively. Then treat $\log L(\beta, \theta | \hat{\tau}_n, \hat{p})$ as a likelihood function of

β and θ only. The estimators $\hat{\beta}$ and $\hat{\theta}$ are obtained by applying the usual method of maximum likelihood, but to the pseudo-likelihood function $\log L(\beta, \theta|\hat{\tau}_n, \hat{p})$.

4. Asymptotic Properties of Parameter Estimators

4.1. Asymptotic properties of parameter estimators

We first introduce some notations that are convenient in the theory of empirical process (cf., Huang (1996, pp.553−557)). For ease of presentation, we consider the case $\theta > 0$ only. Let $\phi = (\phi_1, \phi_2, \phi_3, \phi_4) = (\tau, \beta, \theta, p) \in \Theta = [\tau_1, \tau_2] \times (0, \infty) \times (0, \infty) \times (0, 1)$, $\mu = (\beta, \theta) \in \Theta_1 = (0, \infty) \times (0, \infty)$, and $\nu = (\tau, p) \in \Theta_2 = [\tau_1, \tau_2] \times (0, 1)$. Write P for the probability measure of $X = (Y, \delta)$ and $PM = \int M dP$ for any function $M(x)$. Let P_n be an empirical version of P from $\{X_i = (Y_i, \delta_i), i = 1, \ldots, n\}$. Then $P_n M = \int M dP_n = n^{-1} \sum_{i=1}^n M(X_i)$. For the log-likelihood function $l(\phi|X)$ of a single observation, the score function is denoted by $\dot{l}_{\phi_i}(\phi|X) = \partial l(\phi|X)/\partial \phi_i$, $i = 1, 2, 3, 4$. The true value of $\phi = (\tau, \beta, \theta, p)$ is denoted by $\phi_0 = (\tau_0, \beta_0, \theta_0, p_0)$ and the true distribution by $P_0(x)$. In addition, we temporarily assume the existence of $P_0 \dot{l}_{\phi_i}(\phi|X)$.

Based on independent observations $X_1, \ldots, X_n$, the log-likelihood function can be written as $\log L(\phi|X) = nP_n l(\phi|\cdot) = \sum_{i=1}^n l(\phi|X_i)$. For some small numbers $A_1, A_2 > 0$ and $\eta_1, \eta_2 > 0$, define the parameter spaces for μ and ν as

$$\begin{cases} C_0 = \{\mu = (\beta, \theta) : \beta \geq A_1, \theta \geq A_2\}, \\ C_\eta = \{\nu = (\tau, p) : |\tau - \tau_0| \leq \eta_1, |p - p_0| \leq \eta_2\}. \end{cases} \tag{4.1}$$

From (3.3), the log-likelihood of a single observation $X = (Y, \delta)$ is given by

$$\begin{aligned} l(\mu, \nu|X) = \delta \Big[\log(p\beta) - \beta Y + \Big(\log\Big(1 + \frac{\theta}{\beta}\Big) - \theta(Y - \tau)\Big) I(Y > \tau) \Big] \\ + \Big\{ (1 - \delta) I(Y \leq \tau) \log(1 - p + p \exp(-\beta Y)) \Big\} \\ + \Big\{ (1 - \delta) I(Y > \tau) \log(1 - p + p \exp(-\beta Y - \theta(Y - \tau)) \Big\}, \end{aligned}$$

and the first partial derivatives of $l(\mu, \nu|x)$ with respect to $\mu = (\beta, \theta)$ are

$$\begin{aligned} \dot{l}_\beta(\mu, \nu|X) = -Y + \delta \Big[\frac{1}{\beta} - \frac{\theta}{\beta(\beta + \theta)} I(Y > \tau) \Big] + (1 - \delta)(1 - p) Y \\ \times \Big[\frac{I(Y \leq \tau)}{1 - p + p \exp(-\beta Y))} + \frac{I(Y > \tau)}{1 - p + p \exp(-\beta Y - \theta(Y - \tau))} \Big], \end{aligned} \tag{4.2}$$

$$\dot{l}_\theta(\mu, \nu|X) = \frac{\delta I(Y > \tau)}{\beta + \theta} - \frac{(1 - \delta) p (Y - \tau) \exp(-\beta Y - \theta(Y - \tau)) I(Y > \tau)}{1 - p + p \exp(-\beta Y - \theta(Y - \tau))}. \tag{4.3}$$

From (4.2) and (4.3), it is easy to check that $|\dot{l}_\theta(\mu,\nu|X)| < (\beta+\theta)^{-1} + (1-\delta)(Y-\tau)I(Y>\tau)$. Thus we have

$$P_0\dot{l}_\mu(\mu,\nu|X) < \infty, \quad P_0\dot{l}^2_\mu(\mu,\nu|X) < \infty, \tag{4.4}$$

provided the censoring distribution has a finite variance. Note that μ_0 is the unique point such that $P_0\dot{l}_\mu(\mu,\nu_0|X) = 0$, hence we obtain $\hat{\mu}$ by solving $P_n\dot{l}_\mu(\mu,\nu_0|X) = 0$.

The main results on the asymptotic properties of the estimators in Section 3 are presented in the next three theorems; their proofs are given in the Appendix.

Theorem 2. *Suppose that the parameter spaces are listed in* (4.1), *and that* $\hat{p}$ *and* $\hat{\tau}$ *are given by* (3.4) *and* (3.7), *respectively. Then* $P_n\dot{l}_\mu(\hat{\mu},\hat{\nu}) = o_{p^*}(n^{-1/2})$ *almost surely, where* $\dot{l}_\mu(\mu,\nu)$ *is defined in* (4.2)–(4.3), *and* $\hat{\mu}$ *converges in outer probability to* μ_0.

Theorem 3. *Under the conditions of Theorem* 2, $\sqrt{n}(\hat{\mu}-\mu_0) = O_{p^*}(1)$.

Theorem 4. *Under the conditions of Theorem* 2, $\sqrt{n}(\hat{\mu}-\mu_0)$ *is asymptotically normal with mean 0 and variance* $\{P_0\ddot{l}_{\mu\mu}(\mu_0,\nu_0)\}^{-2}V$, *where* $V = Var(\Lambda_1 + P_0\ddot{l}_{\mu\nu}(\mu_0,\nu_0)\Lambda_2$.

Remark 1. A precise representation of V, in the asymptotic variance of $\sqrt{n}(\hat{\mu}-\mu_0)$ in Theorem 4, can be found in Corollary 3.1.4 of Hu (1998) for our i.i.d. setup. In this case there exists $\alpha(X,\mu_0,\nu_0) \neq 0$ such that $\sqrt{n}P_0\ddot{l}_{\mu\mu}(\mu_0,\nu_0)(\hat{\nu}-\nu_0) = \sqrt{n}P_n\alpha(\cdot,\mu_0,\nu_0) + o_p(1)$, which gives $V = Var[\dot{l}_\mu(\mu_0,\nu_0|X)] + Var[\alpha(X,\mu_0,\nu_0)]$, where $\alpha(\cdot,\mu_0,\nu_0)$ is defined by (3.1.21) in Hu (1998). Without such an $\alpha(X,\mu_0,\nu_0)$, a closed form of V is not available, but we can estimate the variance by the bootstrap method as discussed below.

4.2. Bootstrap method for standard errors of estimators

In Section 4.1, we discussed the asymptotic properties of estimators. However, since the asymptotic variances of $\hat{\mu}$ and $\hat{\nu}$ are often intractable, we resort to the bootstrap method.

The bootstrap technique was introduced by Efron (1979), originally as a tool for "estimating" ad-hoc-estimators that could not be calculated explicitly (see also Efron (1982) and Efron and Tibshirani (1986)). Many authors, including Singh (1981) and Bickel and Freedman (1981), showed that the bootstrap, as an estimator for the distribution of a (standardized) estimator, often gives a better approximation to the true distribution than its limiting distribution.

For the bootstrap method on censored data, Efron and Tibshirani (1986) resampled from the pairs (Y_i,δ_i), $i = 1,\ldots,n$, ignoring the special structure provided by its parametric model. This plan is likely to be "model-robust" as

it is less sensitive to departures from parametric model (Hjort (1992)). Following Efron and Tibshirani (1986), we can obtain our bootstrap estimators. The simulation algorithm proceeds in three steps.

1. Re-sample the pairs $\{(Y_1, \delta_1), \ldots, (Y_n, \delta_n)\}$ with probability $1/n$ at each pair (Y_i, δ_i). Denote the re-sampled data by $d^*(1), \ldots, d^*(B)$, for some positive integer B.
2. For each set of bootstrap data $d^*(b)$, evaluate the estimates of interest.
3. Calculate the sample means and standard deviations of the statistics of interest.

We use the above procedure in the analysis on a set of nonlymphoblastic leukemia data in Section 6 below.

5. Simulation Results

We now investigate the performance of the proposed method to estimate τ, β, θ, p via simulation. We simulated data from the change-point model with hazard function defined at (2.3), as $\tau = 1$, $\beta = 1$, $\theta = 1$, and $p = 0.9$. The distribution function corresponding to (2.3) is

$$F(x) = \begin{cases} 1 - p + p\exp\{-\beta x\}, & x \in [0, \tau], \\ 1 - p + p\exp\{-(\beta + \theta)x + \theta\tau\}, & x > \tau. \end{cases}$$

Moreover, we took a uniform censoring distribution on the interval $[1, 2]$, which results in a censoring proportion of about 30%. Finally, we let $q = 0$ for $g(t)$ in (3.5), so that $g(t) \equiv 1$.

Since the method involves some choices of intervals and/or other parameters, we simulated data for a number of settings. The results presented are based on 1,000 repetitions with various sample sizes. Table 1 summarizes them and reports on the average estimates denoted by $m(\cdot)$, and standard deviations denoted by $s(\cdot)$, of the four parameters. The choices of the intervals $[\tau_1, \tau_2]$ and the upper limit D are also listed in Table 1.

From Table 1 we can see that the choice of the interval $[\tau_1, \tau_2]$ as well as the upper limit D affects the estimation more for smaller sample sizes, but the impact diminishes with larger samples. The estimations of τ and of θ are sensitive to the choice of D and τ_1, τ_2 with sample size 50 or 100, while the estimation of β is less sensitive, and the sensitivity is reduced with increasing sample size. The estimated variances of $\hat{\theta}$ appear quite large, but that can be overcome by increasing sample size and/or choosing a larger τ_1, D and a small τ_2, which is supported by simulations not reported here, such as with $\tau_1 = 0.75$, $\tau_2 = 1.25$ and $D = 4.5$. The estimator of θ performs well even for a small sample size $n = 50$. The estimation of the susceptible proportion p is quite accurate and

Table 1. Performance of the estimators for the true values of the parameters $\tau = 1$, $\beta = 1$, $\theta = 1$, $p = 0.9$.

$[\tau_1, \tau_2]$	D	n	$m(\hat{p})$	$s(\hat{p})$	$m(\hat{\tau})$	$s(\hat{\tau})$	$m(\hat{\beta})$	$s(\hat{\beta})$	$m(\hat{\theta})$	$s(\hat{\theta})$
[0.5, 1.5]	2.5	50	0.889	0.061	1.131	0.258	0.990	0.210	1.600	1.054
		100	0.892	0.047	1.127	0.244	1.025	0.150	1.533	0.831
		400	0.892	0.046	1.221	0.214	1.005	0.099	1.148	0.501
		800	0.894	0.019	1.190	0.234	1.005	0.081	1.029	0.316
[0.5, 1.5]	3	50	0.893	0.062	0.916	0.244	0.961	0.213	1.276	0.768
		100	0.891	0.044	0.965	0.207	0.977	0.154	1.123	0.581
		400	0.894	0.024	1.061	0.015	1.016	0.078	1.167	0.309
		800	0.894	0.017	1.011	0.018	1.013	0.064	1.118	0.242
[0.5, 1.5]	3.5	50	0.889	0.061	0.847	0.240	0.952	0.209	1.175	0.786
		100	0.891	0.046	0.862	0.211	0.969	0.154	1.041	0.523
		400	0.894	0.022	0.965	0.111	0.996	0.075	1.043	0.288
		800	0.894	0.018	0.994	0.013	1.001	0.057	1.051	0.051
[0.25, 1.75]	3.5	50	0.888	0.059	0.689	0.360	0.874	0.415	1.056	0.765
		100	0.890	0.043	0.730	0.320	0.919	0.167	0.961	0.504
		400	0.893	0.024	0.948	0.177	0.991	0.083	1.034	0.283
		800	0.894	0.017	0.995	0.073	1.002	0.053	1.056	0.192

Table 2. Performance of bootstrap estimation for the standard deviations.

$\hat{s}(\hat{p})$	$\hat{s}(\hat{\tau})$	$\hat{s}(\hat{\beta})$	$\hat{s}(\hat{\theta})$
0.0415 (0.0081)	0.3046 (0.0861)	0.2009 (0.0508)	0.5166 (0.0708)

stable throughout all settings, but is slightly biased downwards, which is expected as so is the KME.

We also checked the performance of bootstrap estimation for the standard deviations of the parameter estimators. Table 2 below reports the simulation results for $\tau_1 = 0.25$, $\tau_2 = 1.75$, $D = 3.5$, and sample size $n = 100$. In Table 2, $\hat{s}(\cdot)$ represents the average of 200 bootstrap estimates (with $B = 1,000$) of the standard deviations of the relevant parameter estimator, and the standard errors of these 200 estimates are also reported in parentheses.

Table 2 shows that the performance of the bootstrap estimation is quite satisfactory. The results for other cases are similar.

6. An Application

We consider the application of our model and estimation procedure on a set of nonlymphoblastic leukemia data that were analyzed in Matthews and Farewell (1982) by change-point models without long-term survivors. The data consist of survival times (in days), defined as the time from remission induction to relapse for 84 patients with acute nonlymphoblastic leukemia. Of these, 51 are uncensored (relapse observed) and 33 are censored (relapse not observed). The data

Table 3. Ordered remission durations for 84 patients with acute nonlymphoblastic leukemia.

24	46	57	57	64	65	82	89	90	90	111
117	128	143	148	152	166	171	186	191	197	209
223	230	247	249	254	258	264	269	270	273	284
294	304	304	332	341	393	395	487	510	516	518
518	534	608	642	697	955	1160	68^*	119^*	182^{*24}	583^*
1310^*	1538^*	1634^*	1908^*	1996^*	2057^*					

are listed in Table 3 below, where * indicates censored observations, and 182^{*24} indicates 24 observations censored at 182 days.

The Kaplan-Meier estimator (KME) of the distribution function for the data is plotted in Figure 1, together with its pointwise 95% confidence intervals (cf., Maller and Zhou (1996, p.11)). The KME levels off below 1 and the upper limits of the confidence intervals are below 1 as well, indicating strong evidence for the presence of long-term survivors. Hence our proposed model (2.3) is suitable to analyze the data. Based on the change-point model (1.1), the MLE of the parameters were calculated to be $(\hat{\beta}, \hat{\theta}, \hat{\tau}) = (0.00204, -0.0016, 697.0)$ in Matthews and Farewell (1982), which implies that the hazard function has a change-point of around 697.0 days with a jump of -0.0016. Furthermore, Qin and Sun (1997)

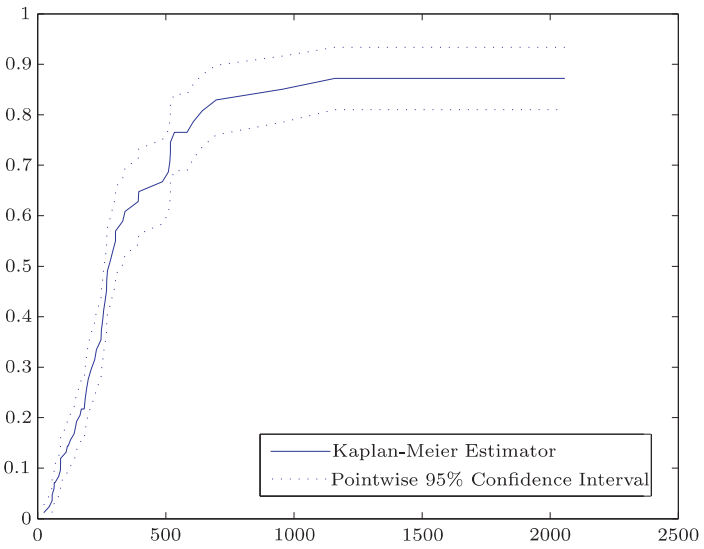

Figure 1. Kaplan-Meier Estimator of the Distribution Function (DF) for the Nonlymphoblastic Leukemia Data.

Table 4. Bootstrap estimates for nonlymphoblastic leukemia data with $B = 1,000$.

$\hat{p}$	$s(\hat{p})$	$\hat{\tau}$	$s(\hat{\tau})$	$\hat{\beta}$	$s(\hat{\beta})$	$\hat{\theta}$	$s(\hat{\theta})$
0.8712	0.0181	697.94	0.0640	0.0026	0.0020	-0.0017	0.0012

found strong evidence against the null hypothesis of constant hazard rate (no change-point). Using (2.3) with $[\tau_1, \tau_2] = [10, 800]$ and $D = 1,600$, and the bootstrap procedure described in Section 4.2, our estimates for (τ, β, θ, p), together with their standard errors (denoted by $s(\cdot)$), are shown in Table 4 below.

The results in Table 4 give an estimate of p at around 0.87, indicating a proportion of 13% for the long-term survivors. The change-point τ is estimated as 698 days, with an estimated jump of about -0.001 for the hazard rate. Compared with the results of Matthews and Farewell (1982), we see that the impact of allowing long-term survivors is rather small on the estimates of the parameters τ and θ, but much greater on β (30%) and the jump size (60%). Furthermore, the bootstrap method allowed us to easily estimate the standard errors of the estimators, which are quite small for $\hat{p}$ and $\hat{\tau}$, indicating high precision of estimation. The standard errors of the other estimators ($\hat{\beta}$ and $\hat{\theta}$) are also reasonable.

Matthews and Farewell (1982) analyzed these data and gave a test for a change-point in the hazard function, but they did not consider the possibility of long-term survivors. In this section, we fit the data using the survival cure model in (2.3), which is supported by the Kaplan Meier estimator of the failure distribution (cf., Figure 1). But is there really a change-point present in the hazard function for the data? This is an important question that determines which model is more appropriate to describe the data. It can be answered by testing the null hypothesis $\mathcal{H}_0$: there is no change point in the survival distribution or $\tau = +\infty$ (cf., Qin and Sun (1997)). Following the results of Qin and Sun (1997), together with our model (2.3), the modified deviance test statistic is given by

$$R_{MF}(\tau) = \sup_{\tau_1 \le \tau \le \tau_2} 2\Big\{ \log L(\tau, \hat{\beta}_\tau, \hat{\theta}_\tau, \hat{p}_\tau) - \log L(\infty, \hat{\beta}_0, \hat{\theta}_0, \hat{p}_0) \Big\},$$

where $\log L$ is defined in (3.2). likelihood estimator of the parameters corresponding to model (2.3), with and without change point are $\hat{\beta}_\tau$, $\hat{\theta}_\tau$, $\hat{p}_\tau$ and $\hat{\beta}_0$, $\hat{\theta}_0$, $\hat{p}_0$, respectively. This gives a deviance $R_{MF}(\tau) = 13.7032$. Hence, the p-value of the deviance test is $\Pr(R_{MF}(\tau) > 13.7032) = \Pr(\chi^2_3 > 13.7032) = 0.0022$, significantly less than 0.01. Hence there is overwhelming evidence to reject the null hypothesis $\mathcal{H}_0$ and conclude that a change point does exist for the data.

We also performed a goodness-of-fit test for the fitted change-point model (2.3) using the method developed by Li and Sun (2000) under the assumption of $\tau = 697.94$, as follows. First generate L standard normal random samples

$\xi_1^{(l)}, \ldots, \xi_n^{(l)}$, $l = 1, \ldots, L$. Then calculate

$$Z_{nl}^*(t) = \sqrt{n}(1-F_{\hat{\omega}}(t)) \int_0^t \frac{\sum_{i=1}^n \xi_i^{(l)} dN_i(s)}{\sum_{i=1}^n H_i(s)} - \frac{F'_{\hat{\omega}}(t)^\top \hat{I}^{-1}}{\sqrt{n}} \int_0^\infty \frac{h'_{\hat{\omega}}(t)}{h_{\hat{\omega}}(t)} \sum_{i=1}^n \xi_i^{(l)} dN_i(s)$$

for $l = 1, \ldots, L$, where $N_i(s)$ and $H_i(s)$ are defined in (3.6), $\omega = (\beta, \theta, p)$ with maximum likelihood estimator $\hat{\omega}$, $F_w(t)$ and $h_w(t)$ are the failure distribution and hazard function in model (2.3) with partial derivative vectors $F'_w(t)$ and $h'_w(t)$, respectively, and $\hat{I}$ is the estimated information matrix of ω. Denote by $c_n(\alpha)$ the $100(1-\alpha)$th (sample) percentile of $\sup_{0<t<\tau_0} |Z_{nl}^*(t)|, \ldots, l = 1, \ldots, L$. According to the method of Li and Sun (2000), we reject (2.3) if $\sup_{0<t<\tau_0} |Z_n(t)| > c_n(\alpha)$, where $Z_n(t) = n^{1/2}(F_n(t) - F_{\hat{\omega}}(t))$ and $F_n(t)$ is the Kaplan-Meier estimator. For $\alpha = 0.05$, we calculated $\sup_{0<t<\tau_0} |Z_n(t)| = 0.3892$ and $c_n(\alpha) = 0.5578$ with $L = 200$ and $\tau_0 = 2,100$. Thus we accept the hypothesis that the data are drawn from the change-point model (2.3).

Remark 2. Many commonly used parametric distributions have hazard functions with very flexible shapes, particularly the generalized gamma (GG) distribution. We fitted the data by using the GG distribution, and performed a goodness-of-fit test using the above method. The test result, however, rejected the null hypothesis that the data are drawn from the GG distribution. This further supports the need to consider the change-point model for the data.

Acknowledgement

We are grateful to an associate editor and two referees for their valuable comments and suggestions that helped improve this paper.

References

Anderson, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer, New York.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196-1217.

Chang, I. S., Chen, C. H. and Hsiung, C. A. (1994). Estimation in change-point hazard rate models with random censorship. Change-point Problems, 78-92. *IMS Lecture Notes Monograph Ser.* 23, Inst. Math. Statist., Hayward, CA.

Dupuy, J. F. (2006). Estimation in a change-point hazard regression model. *Statist. Probab. Lett.* **76**, 182-190.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.* **1**, 54-77.

Gijbels, I. and Gurler, U. (2003). Estimation of a change point in a hazard function based on censored data. *Lifetime Data Anal.* **9**, 395-411.

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: theory and applications. *Ann. Statist.* **9**, 861-869.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

Hjort, N. L. (1992). On inference in parametric survival data. *Internat. Statist. Rev.* **60**, 355-387.

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24**, 540-568.

Hu, H. L. (1998). *Pseudo maximum likelihood estimation for semiparametric models.* Ph.D. Thesis, University of Washington.

Henderson, R. (1990). A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika* **77**, 835-843.

Li, G. and Sun, Y. Q. (2000). A simulation-based goodness-of-fit test for survival data. *Statist. Probab. Lett.* **47**, 403-410.

Loader, C. R. (1991). Inference for a hazard rate change point. *Biometrika* **78**, 749–757.

Matthews, D. E. and Farewell, V. T. (1982). On testing for constant hazard against a change-point alternative. *Biometrics* **38**, 463-468.

Matthews, D. E., Farewell, V. T. and Pyke, R. (1985). Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. *Ann. Statist.* **13**, 583-591.

Maller, R. A. and Zhou, X. (1996). *Survival Analysis with Long-term Survivors.* Wiley, New York.

Muller, H. G. and Wang, J. L. (1990). Nonparametric analysis of changes in hazard rates for censored survival data: an alternative to change-point models. *Biometrika* **77**, 305-314.

Nguyen, H. T., Rogers, G. S. and Walker, E. A. (1984). Estimation in change-point hazard rate models. *Biometrika* **71**, 299-304.

Pons, O. (2003). Estimation in a Cox regression model with a change-point according to a threshold in a covariate. *Ann. Statist.* **31**, 442-463.

Qin, J. and Sun, J. (1997). Statistical analysis of right-censored failure-time data with partially specified hazard rates. *Canad. J. Statist.* **25**, 442-463.

Sen, P. K. (1993). Some change-point problems in survival analysis: relevance of nonparametrics in applications. *Applied Change Point Problems in Statistics*, 325-336, Baltimore, MD.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9**, 1187-1195.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

Wang, J. G. (1987). A note on the uniform consistency of the Kaplan-Meier estimator, *Ann. Statist.* **13**, 1313-1316.

Wu, C. Q., Zhao, L. C. and Wu, Y. H. (2003). Estimation in change-point hazard function mdels. *Statist. Probab. Lett.* **63**, 41-48.

School of Science, Jiang Nan University, Wuxi, Jiangsu Province, China.
E-mail: maxbzhao@hotmail.com

Department of Statistics and Actuarial, East China Normal University, Shanghai, China.
E-mail: xywu@stat.ecnu.edu.cn

Department of Actuarial Studies, Macquarie University, NSW, Sydney, Australia.
E-mail: xzhou@efs.mq.edu.au

(Received December 2006; accepted August 2007)