On the Accelerated Failure Time Model for Current Status and Interval Censored Data

Lu Tian*

Tianxi Cai†

*Harvard University, ltian@hsph.harvard.edu

†Harvard University, tcai@hsph.harvard.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/harvardbiostat/paper14>

Copyright ©2004 by the authors.

On the Accelerated Failure Time Model for Current Status and Interval Censored Data

Lu Tian and Tianxi Cai

Abstract

This paper introduces a novel approach to making inference about the regression parameters in the accelerated failure time (AFT) model for current status and interval censored data. The estimator is constructed by inverting a Wald type test for testing a null proportional hazards model. A numerically efficient Markov chain Monte Carlo (MCMC) based resampling method is proposed to simultaneously obtain the point estimator and a consistent estimator of its variance-covariance matrix. We illustrate our approach with interval censored data sets from two clinical studies. Extensive numerical studies are conducted to evaluate the finite sample performance of the new estimators.

On the Accelerated Failure Time Model for Current Status and Interval Censored Data

LU TIAN

*Department of Preventive Medicine, Northwestern University,
680 N. Lake Shore Drive, Suite 1102, Chicago, Illinois 60611*

and

TIANXI CAI

*Department of Biostatistics, Harvard University,
655 Huntington Avenue, Boston, Massachusetts 02115.*

Abstract

This paper introduces a novel approach to making inference about the regression parameters in the accelerated failure time (AFT) model for current status and interval censored data. The estimator is constructed by inverting a Wald type test for testing a null proportional hazards model. A numerically efficient Markov chain Monte Carlo (MCMC) based resampling method is proposed to simultaneously obtain the point estimator and a consistent estimator of its variance-covariance matrix. We illustrate our approach with interval censored data sets from two clinical studies. Extensive numerical studies are conducted to evaluate the finite sample performance of the new estimators.

Key words: *accelerated failure time model, current status data, interval censoring, nonparametric maximum likelihood estimator (NPMLE), MCMC.*

1 Introduction

Interval censored data arise frequently when subjects in a study are not continuously observed. Instead, subjects undergo periodic examinations, and the event defining failure is only known to whether or not have occurred between exam times. Current status data arises when it is only feasible to have one examination or observation time to see if the failure time T has occurred or has not occurred by the examination time C . Specifically, one observes (C, Δ) , where $\Delta = I_{\{T \geq C\}}$, and $I_{\{\cdot\}}$ is the indicator function. Current status data are sometimes referred to as “case 1” interval censored data (Groeneboom and Wellner, 1992).

For current status data, one sample problem had been studied, for example, by Turnbull (1976) as a special case in a general censoring scheme and Groeneboom and Wellner (1992) by using nonparametric likelihood function. When the main interest lays on the relationship between certain regressor and the failure time, many semi-parametric regression methods have been proposed. For example, Huang (1996) studied the proportional hazards model; Rossini and Tsiatis (1994) studied the proportional odds model; Lin, Oakes and Ying (1998) and Martinussen and Scheike (2002) studied semi-parametric additive hazards model. Most of the existing methods are based on maximizing certain likelihood functions. For general interval censored data, inference for the proportional hazards (Huang and Wellner, 1996; Satten, 1996, Cai and Betensky, 2003) and proportional odds model (Shen, 1998; Rabinowitz, Betensky and Tsiatis, 2000) have been investigated.

An alternative to the proportional hazards and proportional odds models is the accelerated failure time (AFT) model. This model relates the covariates linearly to the logarithm of the survival time:

$$\log(T) = \beta_0' \mathbf{Z} + \epsilon, \quad (1)$$

where error term ϵ is independent of the p -dimensional covariate vector $\mathbf{Z}$ and its distribution is left unspecified. The simple interpretation of the AFT model makes it a useful alternative to the popular proportional hazards model. The AFT model has been studied extensively in the literature for analyzing right censored data (Buckley and James, 1979; Tsiatis, 1990; Wei, Lin and Ying, 1990; Jin, Lin, Ying and Wei, 2003). Inference for the AFT model with current

status or general interval censored data is more difficult since the NPMLE is not directly applicable. Rabinowitz, Tsiatis and Aragon (1995) first proposed a class of estimators motivated from a score test for AFT model. Along the lines of likelihood based approaches, Murphy, van der Vaart, and Wellner (1999) and Shen (2000) applied penalized NPMLE and random-sieve likelihood method to the AFT model with current status data, respectively. Recently, Betensky, Rabinowitz, and Tsiatis (2001) studied a numerically efficient simple estimation procedure under the general interval censorship. However, the validity of foregoing procedures may depend on additional conditions on the monitoring time and their implementations are numerically difficult, especially for high-dimensional covariates. Lastly, their asymptotic inference often involves with difficult non-parametric functional estimation, where the choice of smoothing parameter is notoriously difficult.

In this paper, we propose a relatively simple approach to the regression analysis of interval censored data using the AFT model. We first consider fitting the AFT model to current status data. The estimator and its asymptotic properties are derived in section 2. We present in section 3 a novel MCMC-based numerical method for simultaneously obtaining the point and variance estimates. We generalize these procedures for analyzing interval censored data in section 4. In section 5, we apply our method to data from a tumorigenicity study and to a breast cosmesis data set. The simulation studies shown in section 6 suggest that the proposed methods perform well in finite samples. We close in section 7 with some remarks.

2 Estimation of Regression Parameters with Current Status Data

Let $\{(C_i, \Delta_i, \mathbf{Z}_i) : i = 1, \dots, n\}$ be n i.i.d copies of $(C, \Delta, \mathbf{Z})$. In the following, we assume that the monitoring time C is independent of ϵ but may depend on $\mathbf{Z}$. Suppose for now β_0 in (1) is known and our interest lies in testing the validity of the AFT model specification. The AFT model essentially assumes that the distribution of the residual $\epsilon = \log(T) - \beta_0' \mathbf{Z}$ is independent of the covariates $\mathbf{Z}$. This is equivalent to postulating a null proportional hazards

assumption about the residual and the covariates with

$$\lambda_{\epsilon}(t \mid \mathbf{Z}) = \lambda_0(t),$$

where $\lambda_{\epsilon}(\cdot \mid \mathbf{Z})$ is the hazard function of ϵ conditioning on the covariate $\mathbf{Z}$ and $\lambda_0(\cdot)$ is some unknown baseline hazard function. One approach to testing this assumption is to fit a working proportional hazards model

$$\lambda_{\epsilon}(t \mid \mathbf{Z}) = \lambda_0(t)e^{\gamma_0' \mathbf{Z}}, \quad (2)$$

to the residuals data $\{(\tilde{C}_i(\beta_0), \Delta_i, \mathbf{Z}_i) : i = 1, \dots, n\}$, where $\tilde{C}_i(\beta_0) = \log(C_i) - \beta_0' \mathbf{Z}_i$ is the corresponding “monitoring time” for $\epsilon_i = \log(T_i) - \beta_0' \mathbf{Z}_i$ and $\Delta_i = I_{\{T_i \geq C_i\}} = I_{\{\epsilon_i \geq \tilde{C}_i(\beta_0)\}}$, and test the hypothesis $H_0 : \gamma_0 = 0$ based on an estimator of γ_0 . To estimate γ_0 in the Cox model with current status data, various methods have been proposed (Huang, 1996; Shiboski, 1998). One estimator of particular interest is the non-parametric maximum likelihood estimator (NPMLE) proposed by Huang (1996). It has been established that the NPMLE $\hat{\gamma}_n$ is consistent, asymptotical normal and semi-parametric efficient under mild regularity conditions (Huang, 1996). It follows that under $H_0 : \gamma_0 = 0$,

$$n^{1/2} \hat{\gamma}_n \sim N(0, \mathbf{B}_0^{-1}),$$

where

$$\mathbf{B}_0 = E \left[R(\tilde{C}(\beta_0), \mathbf{Z}) \left\{ \mathbf{Z} - \frac{E\{\mathbf{Z}H^2(\tilde{C}(\beta_0)|\mathbf{Z})|\tilde{C}(\beta_0)\}}{E\{H^2(\tilde{C}(\beta_0)|\mathbf{Z})|\tilde{C}(\beta_0)\}} \right\}^{\otimes 2} \right],$$

$\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}'$, for $a \in R^p$, $R(\tilde{C}(\beta_0), \mathbf{Z}) = H^2\{\tilde{C}(\beta_0)|\mathbf{Z}\}[1 - F_0\{\tilde{C}(\beta_0)\}]/F_0\{\tilde{C}(\beta_0)\}$, $H(\cdot|\mathbf{Z})$ is the cumulative hazard function of $\tilde{C}(\beta_0) = \log(C) - \beta_0' \mathbf{Z}$ given $\mathbf{Z}$, and $F_0(\cdot)$ is the cumulative distribution function of ϵ . Thus, under the AFT model (1), we expect that $\hat{\gamma}_n$ obtained based on the residual data is close to zero (in the order of $O_p(n^{-\frac{1}{2}})$).

Now, we return to the estimation problem under the AFT model where β_0 is unknown. Since the distribution of $\epsilon(\beta) = \log(T) - \beta' \mathbf{Z}$ is independent of $\mathbf{Z}$ if and only if $\beta = \beta_0$, the aforementioned hypothesis testing procedure motivates us to estimate β_0 by solving the following estimating equations

$$\hat{\gamma}_n(\beta) = o_p(n^{-\frac{1}{2}}), \quad (3)$$

where for any given β , $\hat{\gamma}_n(\beta)$ is the NPMLE of γ_0 in the working model (2) based on the data $\{(\tilde{C}_i(\beta), \Delta_i, \mathbf{Z}_i) : i = 1, \dots, n\}$. In the appendix, we show that under appropriate regularity conditions, $\hat{\gamma}_n(\beta)$ converges to a deterministic function $\gamma_0(\beta)$ in probability uniformly in β . Furthermore, any solution to (3), $\hat{\beta}$, is consistent for β_0 provided that β_0 is the unique root of $\gamma_0(\beta) = 0$. To make inference about β_0 , we also provide a sketched proof in the appendix to show that $n^{1/2}(\hat{\beta} - \beta_0)$ can be approximated by $\mathbf{A}_0^{-1}n^{1/2}\hat{\gamma}_n(\beta_0)$, which converges weakly to $N(0, \mathbf{A}_0^{-1}\mathbf{B}_0^{-1}(\mathbf{A}_0')^{-1})$ as $n \rightarrow \infty$, where $\mathbf{A}_0 = d\gamma_0(\beta)/d\beta|_{\beta=\beta_0}$ is a deterministic nonsingular matrix.

3 Implementation of the Inference Procedure

The proposed estimation procedure for β_0 may be carried out in two steps: 1) computing the NPMLEs of the regression parameter in a set of working proportional hazards models indexed by β ; and 2) finding $\hat{\beta}$ such that $\hat{\gamma}_n(\beta) = o_p(n^{-\frac{1}{2}})$. For the first step, algorithms such as the “iterative convex minorant” (Huang, 1996) and the “iterative pool-adjacent-violator” algorithm (Barlow, Bartholomew, Bremner, and Brunk, 1972) can be used. A grid search method can be used to find a solution $\hat{\beta}$ in the second step when the covariates are one dimensional. However, with higher dimensional covariates, grid search methods become infeasible and solving the equations $\hat{\gamma}_n(\beta) = o_p(n^{-\frac{1}{2}})$ turns out to be rather difficult due to the discontinuity of $\hat{\gamma}_n(\beta)$ in β . In addition to the difficulties in solving the estimating equations, estimating the “slope” of $\hat{\gamma}_n(\beta)$, which is needed to make inference about $\hat{\beta}$, involves with complicated numerical derivatives or nonparametric functional estimations (Huang, 1996).

To avoid these difficulties, we propose a MCMC based approach to the implementation of the estimation and inference procedures. Specifically, we draw samples of β^* from the distribution whose density function is proportional to

$$e^{-[M_{\mathbf{b}}\{\hat{\gamma}_n(\mathbf{b})\} - M_{\mathbf{b}}(0)]} \quad (4)$$

and use the empirical distribution of these samples to approximate the distribution of $\hat{\beta}$,

where $M_{\mathbf{b}}(\boldsymbol{\gamma}) = \max_{\Lambda} \text{loglik}_1(\boldsymbol{\gamma}, \Lambda)$,

$$\text{loglik}_1(\boldsymbol{\gamma}, \Lambda) = \sum_{i=1}^n \left(\delta_i \log[1 - \exp\{-e^{\boldsymbol{\gamma}'\mathbf{z}_i} \Lambda(\log(c_i) - \mathbf{b}'\mathbf{z}_i)\}] - (1 - \delta_i) e^{\boldsymbol{\gamma}'\mathbf{z}_i} \Lambda(\log(c_i) - \mathbf{b}'\mathbf{z}_i) \right),$$

and $\{(c_i, \delta_i, \mathbf{z}_i) : i = 1, \dots, n\}$ are the observed data. Note that the foregoing maximization with respect to $\Lambda(\cdot)$ is over all right continuous increasing step functions with jump points at $\{\log(c_i) - \mathbf{b}'\mathbf{z}_i : i = 1, \dots, n\}$. Drawing samples of $\boldsymbol{\beta}^*$ from the target distribution is straightforward to realize with existing algorithms such as the importance sampling and newly developed MCMC methods. We recommend the following procedure for implementation based on the Metropolis algorithm (Metropolis et al., 1953):

1. Select a starting point $\boldsymbol{\beta}_1^*$ and a variance covariance matrix Σ_0 for the proposal distribution.
2. For $k = 2, \dots, L$

(a) Draw $\boldsymbol{\beta}_{\text{new}}^*$ from the proposal distribution $N(\boldsymbol{\beta}_{k-1}^*, \Sigma_0)$,

(b) $\boldsymbol{\beta}_k^* = \begin{cases} \boldsymbol{\beta}_{\text{new}}^* & \text{with probability } p \\ \boldsymbol{\beta}_{k-1}^* & \text{with probability } 1 - p \end{cases}$, where

$$p = \min \left\{ 1, \frac{\exp\{-[M_{\boldsymbol{\beta}_{\text{new}}^*} \{\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_{\text{new}}^*)\} - M_{\boldsymbol{\beta}_{\text{new}}^*}(0)]\}}{\exp\{-[M_{\boldsymbol{\beta}_{k-1}^*} \{\hat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_{k-1}^*)\} - M_{\boldsymbol{\beta}_{k-1}^*}(0)]\}} \right\}.$$

To eliminate the effect of the starting distribution, we discard $\{\boldsymbol{\beta}_k^* : k = 1, \dots, L_0\}$ and use the remaining samples $\{\boldsymbol{\beta}_k^* : k = L_0, \dots, L\}$ to make inference. The efficiency of the algorithm depends on careful choice of Σ_0 and can be monitored by measures such as average rejection rate and autocorrelation coefficient of the resulted chain. A nice review on this topic is given by Gelman, Carlin, Stern and Rubin (section 11.9, 11.10, 2003). Some practical guidelines for the implementation of this algorithm are described in the example section. In the appendix, we show that the conditional distribution of $n^{1/2}(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}})$ approximates the distribution $N(0, \mathbf{A}_0^{-1} \mathbf{B}_0^{-1} (\mathbf{A}'_0)^{-1})$ and thus the sample mean and variance of $\{\boldsymbol{\beta}_k^* : k = L_0, \dots, L\}$ can be used to approximate $\hat{\boldsymbol{\beta}}$ and its variance, respectively. To be more specific, we establish that the conditional distribution of $\boldsymbol{\eta}^* = n^{1/2}(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}})$ restricted at a bounded

region $\{\boldsymbol{\eta} : \|\boldsymbol{\eta}\| \leq c_n\}$ converges to the distribution of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, where c_n is a user specified constant such that $c_n \rightarrow \infty$ and $n^{-1/2}c_n \rightarrow 0$. Based on our limited experience, we found that the extra restriction on $\boldsymbol{\beta}^*$ is rarely needed in practice as long as the targeting function defines an appropriate unimodal density.

4 General Interval Censoring

One advantage of the proposed methods is its easy extension to incorporate general interval censoring. We will illustrate this by extending the methods described in the previous sections to allow for the most common type of interval censoring, sometimes referred to as *case 2* interval censoring (Huang and Wellner, 1997). When case 2 interval censoring occurs, T is only known to have or have not occurred by two monitoring times U and V . Typically data for analysis are organized as $\{(U_i, V_i, \Delta_{1i}, \Delta_{2i}, \mathbf{Z}_i) : i = 1, \dots, n\}$, where $\Delta_{1i} = I_{\{T_i < U_i\}}$ and $\Delta_{2i} = I_{\{U_i \leq T_i < V_i\}}$.

For case 2 interval censored data, the NPMLE of γ_0 in the Cox proportional hazards model has also been studied (Huang and Wellner, 1996). For our purpose, we again use the Cox model as the working model and for any given $\boldsymbol{\beta}$, we obtain $\hat{\gamma}_n(\boldsymbol{\beta})$, the NPMLE of γ_0 , based on the transformed data $\{(\log(U_i) - \boldsymbol{\beta}'\mathbf{Z}_i, \log(V_i) - \boldsymbol{\beta}'\mathbf{Z}_i, \Delta_{1i}, \Delta_{2i}, \mathbf{Z}_i) : i = 1, \dots, n\}$. If $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and thus the Cox model (2) holds with $\gamma_0 = 0$, it can be shown in the view of Theorem 4.2 of Huang and Wellner (1997) that $n^{1/2}\hat{\gamma}_n(\boldsymbol{\beta}_0)$ converges in distribution to mean zero Gaussian distribution. Therefore, in the presence of interval censoring, we also can estimate $\boldsymbol{\beta}_0$ in the AFT model by solving the estimating equations $\hat{\gamma}_n(\boldsymbol{\beta}) = o_p(n^{-1/2})$.

The inference of $\hat{\boldsymbol{\beta}}$ can be carried out in the same fashion as for the current status data. The only modification needed is that based on the observed data $\{(u_i, v_i, \delta_{1i}, \delta_{2i}, \mathbf{z}_i), i = 1, \dots, n\}$, the density function for the target distribution becomes proportional to

$$e^{-[K_{\mathbf{b}}\{\hat{\gamma}_n(\mathbf{b})\} - K_{\mathbf{b}}(0)]},$$

where $K_{\mathbf{b}}(\boldsymbol{\gamma}) = \max_{\Lambda} \loglik_2(\boldsymbol{\gamma}, \Lambda)$, and $\loglik_2(\boldsymbol{\gamma}, \Lambda) =$

$$\sum_{i=1}^n \left[\delta_{1i} \log \left\{ 1 - \exp\{-e^{\boldsymbol{\gamma}'\mathbf{z}_i} \Lambda(\log(u_i) - \mathbf{b}'\mathbf{z}_i)\} \right\} + \delta_{2i} \log \left\{ \exp\{-e^{\boldsymbol{\gamma}'\mathbf{z}_i} \Lambda(\log(u_i) - \mathbf{b}'\mathbf{z}_i)\} \right. \right. \\ \left. \left. - \exp\{-e^{\boldsymbol{\gamma}'\mathbf{z}_i} \Lambda(\log(v_i) - \mathbf{b}'\mathbf{z}_i)\} \right\} - (1 - \delta_{1i} - \delta_{2i}) e^{\boldsymbol{\gamma}'\mathbf{z}_i} \Lambda(\log(v_i) - \mathbf{b}'\mathbf{z}_i) \right].$$

5 Real Data Example

We use two real examples to illustrate our estimation procedure. The first data set is from a tumorigenicity study described fully in Dinse and Lagakos (1983). In this study, 319 Fisher rats were given 125 doses of polybrominated biphenyl mixture (PBB) over a six-month period. The main objective of the study is to examine the relationship between the dose levels of PBB measured in milligrams per kilogram of body weight and the occurrence of bile duct hyperplasia. The time to the occurrence of bile duct hyperplasia cannot be observed directly because tumors can only be found once the animal has died. Instead, the failure time is only observed to have or have not occurred by time of nature death or sacrifice. In addition to the dose level, there are three other covariates recorded for each rat: sex, baseline weight and number of the tier in which the rat was housed.

We propose to assess the dose effect of PBB on the tumor occurrence by fitting an AFT model with all 4 covariates. Inference about the regression coefficients were be made based on samples of $\boldsymbol{\beta}^*$ drawn from the density function given in (4). Due to the complicated nature of the density function, we employed a flexible Metropolis algorithm as described in section 3 for drawing samples from the target distribution. We specify the proposal distribution required by the algorithm as multivariate normal with appropriate covariance matrix since the target distribution is asymptotically normal (section 11.10, Gelman, Carlin, Stern, and Rubin, 2003). To ascertain the covariance matrix, we fit a linear regression working model based on $\{(\log(C_i), \Delta_i, \mathbf{Z}_i), i = 1, \dots, n\}$ with the error distribution being log(Weibull) and obtain the estimated variance-covariance matrix of the regression coefficients, denoted by $\tilde{\Sigma}$. We hope that $\tilde{\Sigma}$ is in the similar magnitude of the variance-covariance matrix of the targeting distribution, even though the parametric Weibull regression model is likely to be misspecified.

At the $(k + 1)$ th step of the Metropolis algorithm, $N(\boldsymbol{\beta}_k^*, c_0 \tilde{\Sigma}_0)$ is used as the initial proposal distribution, where $c_0 = 1$. Following the suggestion of Liu (2001), c_0 was then adaptively updated to yield desired rejection rate of the chain. Within each draw, we use iterative convex minorant algorithm to maximize the objective function. We found that if $c_0 = 6.25$, then the rejection rate is around 65%, which is regarded as satisfactory for generating a multivariate normal distribution. Therefore we fix $c_0 = 6.25$ and let $\Sigma_0 = c_0 \tilde{\Sigma}$ to generate a chain with 15000 draws. It seems that the produced chain converges to its equilibrium distribution (the target distribution) after initial 3000 iterations. Figure 1 plotted the trace for each of the four covariates. From the figure, the chain mixed reasonably well. Figure 2 plotted the histogram for marginal distribution of each of the four covariates. While all the marginal distributions are approximately normal, the slight skewness presented in some of the four marginal distributions suggests that the mode of the distribution may be closer to the solution of the estimating equation than the sample mean. Therefore, we used the mode of the kernel-based density function estimators with last 12000 samples as the point estimator $\hat{\boldsymbol{\beta}}$. Furthermore, the variance covariance matrix of the resulted estimator was estimated by $11999^{-1} \sum_{i=3001}^{15000} (\boldsymbol{\beta}_i^* - \hat{\boldsymbol{\beta}})^{\otimes 2}$. $L = 15000$ was selected because after the burn in period, 12000 metropolis samples correspond to $12000 \times 0.3/4 \approx 900$ independent draws using the rule of thumb given by Gelman, Carlin, Stern, and Rubin (section 11.1, 2003) and independent draws more than 750 were regarded as sufficient for estimating the mean and standard error of a four dimensional normal distribution as discussed in Tian, Liu, Zhao and Wei (2004). Complete results are shown in Table 3. The estimated dose effect is -0.505 with standard error 0.235 indicating that rats receiving higher doses of PBB develop bile duct hyperplasia earlier than those receiving lower doses controlling for other covariates.

To examine if the $\hat{\boldsymbol{\beta}}$ obtained from the resulted chain is a solution to the estimating equation, we calculated the likelihood ratio test statistic, $\mathcal{R}(\hat{\boldsymbol{\beta}})$, of the proportional hazards working model based on the transformed data $\{(\tilde{C}_i(\hat{\boldsymbol{\beta}}), \Delta_i, \mathbf{Z}_i) : i = 1, \dots, n\}$. In an ideal situation, when $\hat{\boldsymbol{\beta}}$ solves the “exact” estimating equation, i.e. $\hat{\gamma}_n(\hat{\boldsymbol{\beta}}) = 0$, the test statistics should be zero. In practice, due to the discontinuity nature of the estimating equation, we would only expect the statistics to be close to zero. To measure the closeness to zero,

we compared the observed $\mathcal{R}(\hat{\beta})$ with the asymptotic distribution of test statistic $\mathcal{R}(\beta_0)$. It follows from Murphy and Van der Vaart (2000) that $\mathcal{R}(\beta_0)$ is approximately χ_p^2 . In this example, the observed test statistics is 0.03 corresponding to the 10^{-4} quantile of χ_4^2 . It implies that $\hat{\gamma}_n(\hat{\beta})$ is sufficiently close to zero and $\hat{\beta}$ is a root of the equation $\hat{\gamma}_n(\beta) = o_p(n^{-1/2})$.

For the second example, we consider the breast cosmesis data set described fully in Finkelstein and Wolfe (1985) and Finkelstein (1986). This data set consists of 94 observations from a retrospective study looking at the time to cosmetic deterioration measured by the appearance of breast retraction on early breast cancer patients. The objective of the study is to compare the event times among patients receiving radiotherapy alone ($Z = 1$) to the event times among those receiving primary radiotherapy in combination with chemotherapy ($Z = 0$). The event times were subject to interval censoring due to the irregular observation times. We aim to assess the therapy effect on the event time by fitting an AFT model with a single covariate Z . Using similar procedure as described above, we generated a Markov chain with length 5000. Using the last 4000 samples in the chain, we obtained the estimate of β as 0.52 with standard error 0.16 indicating that the group receiving radiotherapy in combination with chemotherapy experienced a significantly early appearance of breast retraction. Since model only contains a scalar covariate, we also solved the estimating equation by a grid search to examine if $\hat{\beta}$ solves the estimating equation. The grid search method gives essentially the same answer with $\hat{\beta} = 0.53$. Figure 3 plotted the histogram of $\beta_i^*, i = 1001, \dots, 5000$ versus the estimating function $\hat{\gamma}_n(\beta)$. From figure 3, the mode of the distribution of β^* overlaps with the zero crossing of the estimating function very well.

6 Simulation study

We conducted extensive simulation studies to examine the validity of the large sample approximations for making inference in finite sample sizes. For the case with current status data, in one of the studies, we generated 3 covariates Z_1 , Z_2 and Z_3 , independently from a $\text{Uniform}(-0.5, 0.5)$, and the failure time T from the AFT model

$$\log(T) = -Z_1 + Z_3 + \epsilon, \quad (5)$$

with ϵ following $N(0, 1/4)$. The monitoring time C was generated from a log-normal distribution $\exp\{N(0, 1/4)\}$. Sample sizes of 150 and 300 were considered. This configuration resulted in, on average, about 50% of the failure times being left censored, i.e., $P(T \geq C) \approx 0.5$. For each simulated data set, we implemented the Metropolis algorithm to produce a Markov chain with length 5000. The proposal distribution was also chosen adaptively to yield a desired rejection level (between 50% and 80%). Statistical inferences about the true regression parameter $\beta_0 = (-1, 0, 1)'$ were made based on the last 4000 MCMC samples of β^* . In particular, we estimate β_0 by the sample mean of the last 4000 realizations of β^* and the standard error of this estimator is estimated by the empirical standard error of these realizations. In Table 1, we present empirical biases, sampling standard errors, averages of the standard error estimates and empirical coverage probabilities of the 95% confidence intervals for $\hat{\beta}$. The results suggest that the parameter estimators have negligible biases. The standard error estimates are close to the sampling standard errors. In addition, the confidence intervals have appropriate coverage probabilities.

To examine the performance of the estimation procedures for interval censored data, we generated the first monitoring time U from the log-normal distribution $\exp\{N(-0.35, 1/4)\}$ and let the second monitoring time $V = e^{0.7}U$. The failure time T was also generated using the model (5). This choice of (U, V) resulted in equal proportions of T falling into $(0, U)$, (U, V) and (V, ∞) . For each simulated data set, the inference procedures were carried out in a fashion similar to that for the current status data. As reported in Table 2, results from this study suggest that the proposed methods for analyzing interval censored data also perform well in finite samples.

7 Discussion

In this paper, we proposed a new method for analyzing current status data and more generally, interval censored data, using the accelerated failure time model. The new estimation approach is derived from inverting a statistic for testing a null proportional hazards model. This technique has previously been considered. For example, Tsiatis's rank-based estimating

equation (Tsiatis, 1990) for AFT model with right censored data can be viewed as inverting the score test for null proportional hazards working model. Depending on the distribution of the error term, more efficient estimators could be constructed similarly by inverting tests for other semi-parametric working models such as the proportional odds or additive hazards model. The choice of the working model depends on the computational cost and the efficiency associated with the resulting estimator.

For the ease of implementation, we also provided a novel MCMC-based numerical procedure. The Monte Carlo method provides a simple solution for obtaining point estimates and variance estimates for the parameters of interest. Due to the possible abnormality of the distribution of β^* and potential outliers in the Markov chain, robust versions of the sample mean and variance are sometimes preferred for finite sample case. Our simulation results suggest that the proposed procedure works well at moderate sample sizes and has the advantage of allowing for multiple covariates.

The proposed method is related to the simulated annealing approach of Lin and Geyer (1992). Lin and Geyer (1992) proposed to find the solution to a set of complicated estimating equations by generating a Markov chain based on the Metropolis algorithm with appropriately shrinkaging proposal distribution. Our proposal aims to generate a Markov chain from a specific target distribution and use the chain to simultaneously locate the root to the estimating equations and its variance estimator. The use of Monte Carlo methods as an alternative to directly solving estimating equations has also been considered by Tian, Liu, Zhao and Wei (2004) in the scenario of importance sampling.

To implement the MCMC-based resampling method in practice, we need to determine the appropriate length of the Markov chain. The optimal choice of the length depends on the dimension of the unknown parameter, the desired accuracy and the efficiency of the Metropolis algorithm. Some guidelines for choosing an appropriate length can be found in Tian, Liu, Zhao and Wei (2004).

8 Appendix

The appendix is organized as the following: At first, we stated necessary notations and regularity conditions used in the proof, and then we provide a sketch of justification of consistency, asymptotical normality and resampling methods for the proposed estimator. The mathematically rigorous proofs need further work and currently under investigation.

The following notations are used in the proof. Let $\tilde{C}(\beta) = \log(C) - \beta'Z$, $\Theta = \{\gamma, \Lambda, \beta\}$ denote the whole set of unknown parameters, $\mathcal{D}$ denote the observation (C, Δ, Z) , $S_Z(C; \Theta) = \exp\left\{-e^{\gamma'Z}\Lambda(\tilde{C}(\beta))\right\}$,

$$r_1(\Theta; \mathcal{D}) = \Delta e^{\gamma'Z} \frac{S_Z(C; \Theta)}{\{1 - S_Z(C; \Theta)\}^2}, \quad r_2(\Theta; \mathcal{D}) = \frac{\Delta}{1 - S_Z(C; \Theta)} - 1,$$

$$m(\Theta; \mathcal{D}) = \Delta \log\{1 - S_Z(C; \Theta)\} + (1 - \Delta) \log\{S_Z(C; \Theta)\}, \quad m_0(\Theta) = E\{m(\Theta; \mathcal{D})\},$$

$$\dot{m}_\gamma(\Theta; \mathcal{D}) = \left. \frac{\partial m(\Theta; \mathcal{D})}{\partial \gamma} \right|_{\eta=0} = Ze^{\gamma'Z} \Lambda\{\tilde{C}(\beta)\} r_2(\Theta; \mathcal{D}),$$

$$\dot{m}_\Lambda(\Theta; \mathcal{D})[h] = \left. \frac{\partial m(\gamma, \Lambda_\eta, \beta; \mathcal{D})}{\partial \eta} \right|_{\eta=0} = e^{\gamma'Z} h_1\{\tilde{C}(\beta)\} r_2(\Theta; \mathcal{D}),$$

$$\dot{S}_1(\Theta) = E\{\dot{m}_\gamma(\Theta; \mathcal{D})\}, \quad \dot{S}_2(\Theta)[h] = E\{\dot{m}_\Lambda(\Theta; \mathcal{D})[h]\},$$

$$\dot{S}_{11}(\Theta) = E\left[Z^{\otimes 2} e^{\gamma'Z} \Lambda\{\tilde{C}(\beta)\} \left\{r_2(\Theta; \mathcal{D}) - \Lambda\{\tilde{C}(\beta)\} r_1(\Theta; \mathcal{D})\right\}\right],$$

$$\dot{S}_{12}(\Theta)[h_1] = \dot{S}_{21}(\Theta)[h_1] = E\left[Ze^{\gamma'Z} h_1\{\tilde{C}(\beta)\} \left\{r_2(\Theta; \mathcal{D}) - \Lambda\{\tilde{C}(\beta)\} r_1(\Theta; \mathcal{D})\right\}\right],$$

and

$$\dot{S}_{22}(\Theta)[h_1, h_2] = E\left[-e^{\gamma'Z} h_1\{\tilde{C}(\beta)\} h_2\{\tilde{C}(\beta)\} r_1(\Theta; \mathcal{D})\right],$$

where $h \in \mathcal{H}_\Lambda = \{h : h = \partial \Lambda_\eta / \partial \eta|_{\eta=0}\}$ and Λ_η is a parametric path (Murphy and Van der Vaart, 2000) through Λ , i.e. $\Lambda_\eta|_{\eta=0} = \Lambda$, where Note that $\dot{S}_{11}(\Theta)$, $\dot{S}_{12}(\Theta)[h] = \dot{S}_{21}(\Theta)[h]$, and $\dot{S}_{22}(\Theta)[h_1, h_2]$ are the “partial derivatives” of $\dot{S}_1(\Theta)$ and $\dot{S}_2(\Theta)[h]$ with respect to γ and Λ .

We assume the following regularity conditions hold true throughout the section:

A.1 The (unobservable) failure time is independent of the examination times given the covariates.

A.2 Z is uniformly bounded and not concentrate in any R^{p-1} hyperplane.

- A.3 The cumulative distribution function of $\epsilon = \log(T) - \beta'_0 \mathbf{Z}$, $F_0(\cdot)$ has a strictly positive and continuously differentiable density function on the support.
- A.4 The support of transformed monitoring time $\tilde{C}(\beta_0) = \log(C) - \beta'_0 \mathbf{Z}$ is an finite interval $[\tau_1, \tau_2]$ with $0 < F_0(\tau_1) < F_0(\tau_2) < 1$. The density function of $C|\mathbf{Z} = \mathbf{z}$ is uniformly bounded.
- A.5 For any $\beta \in \Omega_\beta$, $m_0(\Theta)$ has unique maximizer $(\gamma_0(\beta), \Lambda_0(\cdot; \beta))$, where $\gamma_0(\beta)$ is continuously differentiable with respect to β and $\Lambda_0(u; \beta)$ is continuously differentiable with respect to both u and β . Furthermore, for any $\epsilon > 0$, there is a $\delta > 0$, such that $\sup_{\|\gamma - \gamma_0(\beta)\| \geq \delta} m_0(\Theta) < m_0\{\Theta_0(\beta)\} - \epsilon$, for any $\beta \in \Omega_\beta$, where $\Theta_0(\beta) = (\gamma_0(\beta), \Lambda_0(\cdot; \beta), \beta)$ and Ω_β , the parameter space of β , is a compact set in R^p containing β_0 as an interior point.
- A.6 β_0 is the unique solution of $\gamma_0(\beta) = 0, \beta \in \Omega_\beta$.
- A.7 $d\gamma(\beta)/d\beta|_{\beta=\beta_0} = \mathbf{A}_0$ is nonsingular.

A.1-A.4 are analogue to regularity conditions used for proportional hazards model with current status data. A.5 assumes that for any $\beta \in \Omega_\beta$, there exist an isolated maximizer of the objective function with respect to γ . A.6 and A.7 guarantee that the estimating equation has unique consistent root and nonzero information (slope) at the true parameter, respectively.

8.1 Consistency

Under the listed regularity conditions, following the similar argument in Huang (1996), for any fixed β , let $(\hat{\gamma}_n(\beta), \hat{\Lambda}_n(\cdot, \beta))$ denote the maximizer of $n^{-1} \sum_{i=1}^n m(\Theta; \mathcal{D}_i)$ with respect to γ and Λ , and $(\hat{\gamma}_n(\beta), \hat{\Lambda}_n(\cdot, \beta))$ converges to $(\gamma_0(\beta), \Lambda_0(\cdot, \beta))$ in probability, in the sense that

$$\|\hat{\gamma}_n(\beta) - \gamma_0(\beta)\| + \left\{ \int_{R^1} [\hat{\Lambda}_n(u, \beta) - \Lambda_0(u, \beta)]^2 dG_\beta(u) \right\}^{1/2} = o_p(1), \quad (6)$$

where $G_\beta(\cdot)$ is the cumulative distribution function of $\log(C) - \beta'_0 \mathbf{Z}$. The $o_p(1)$ in (6) can be further refined as $O_p(n^{-1/3})$ by theorem 3.2 and 3.3 in Huang (1996). We will show that this convergence is uniform in $\beta \in \Omega_\beta$. With condition A.5, we only need to show that

$n^{-1} \sum_{i=1}^n m(\Theta; \mathcal{D}_i)$ converges to $m_0(\Theta)$ uniformly in $\Theta \in \Omega_{\Theta} = \Omega_{\gamma} \times \Omega_{\Lambda} \times \Omega_{\beta}$, where Ω_{γ} is a compact set in R^p containing $\{\gamma_0(\beta), \beta \in \Omega_{\beta}\}$, $\Omega_{\Lambda} = \{\Lambda(\cdot) : \text{increasing and } 0 < 1/M < \Lambda(u) < M < \infty, \forall u \in \bigcup_{\beta \in \Omega_{\beta}} \text{support of } \log(C) - \beta' \mathbf{Z}\}$, and M is a positive constant.

In the following, we will show that the class of functions $\mathcal{M} = \{m(\Theta; c, \delta, \mathbf{z}) : \Theta \in \Omega_{\Theta}\}$ is Glivenko-Cantelli. Firstly, by theorem 2.7.5 and similar argument in the proof for theorem 2.4.1 of van der Vaart and Wellner (1996), we can show that $\{\Lambda(c - \beta' \mathbf{z}) : (\beta, \Lambda) \in \Omega_{\beta} \times \Omega_{\Lambda}\}$ is Glivenko-Cantelli. Then, using Lemma 2.6.5 and 2.6.18 of van der Vaart and Wellner (1996), the class of functions

$$\{\log(1 - S_{\mathbf{z}}(c; \Theta)) : \Theta \in \Omega_{\Theta}\} = \phi \circ \{\gamma' \mathbf{z} + \log(\Lambda(c - \beta' \mathbf{z})) : \Theta \in \Omega_{\Theta}\},$$

where $\phi(x) = \log[1 - \exp\{-\exp(x)\}]$, is Glivenko-Cantelli. Similar arguments can be used to show that $\{\log(S_{\mathbf{z}}(c; \Theta)) : \Theta \in \Omega_{\Theta}\}$ is Glivenko-Cantelli. Therefore, the class $\mathcal{M}$ is Glivenko-Cantelli with a constant envelope and the uniform convergence of $\hat{\gamma}_n(\beta)$ follows. This, coupled with condition A.6 and A.7, implies that $\hat{\beta} \rightarrow \beta_0$ in probability.

8.2 Asymptotic Normality

To show the asymptotically normality of $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$, it suffices to show the following local linearity condition:

$$\sup_{\|\beta - \beta_0\| \leq \epsilon_n} \frac{n^{1/2} \|\hat{\gamma}_n(\beta) - \hat{\gamma}_n(\beta_0) - \mathbf{A}_0(\beta - \beta_0)\|}{1 + n^{1/2} \|\beta - \beta_0\|} = o_P(1), \quad (7)$$

for any $\epsilon_n \rightarrow 0$. To this end, let

$$\mathbf{h}^*(u; \beta) = - \frac{E \left[\mathbf{Z} e^{\gamma_0' \mathbf{Z}} \{r_2(\Theta_0(\beta); \mathcal{D}) - \Lambda_0(u; \beta) r_1(\Theta_0(\beta); \mathcal{D})\} \middle| \tilde{C}(\beta) = u \right]}{E \left\{ e^{\gamma_0' \mathbf{Z}} r_1(\Theta_0(\beta); \mathcal{D}) \middle| \tilde{C}(\beta) = u \right\}}.$$

Then $\mathbf{h}^*(u; \beta)$ satisfies

$$\dot{S}_{12}\{\Theta_0(\beta)\}[h] - \dot{S}_{22}\{\Theta_0(\beta)\}[h, \mathbf{h}^*(\cdot; \beta)] = 0,$$

for any $h \in \mathcal{H}_{\Lambda_0(\cdot; \beta)}$. Using similar arguments as given in the proof of theorem 6.1 in Huang (1996), one can show that

$$n^{1/2} \{\hat{\gamma}(\beta) - \gamma_0(\beta)\} = \Sigma_1^{-1}(\beta) n^{-1/2} \sum_{i=1}^n S^*(\beta; \mathcal{D}_i) + o_p(1),$$

uniformly in β , where

$$\Sigma_1(\beta) = \dot{S}_{11}\{\Theta_0(\beta)\} - \dot{S}_{21}\{\Theta_0(\beta)\}[\mathbf{h}^*(\cdot; \beta)],$$

and

$$S^*(\beta; \mathcal{D}) = \dot{m}_\gamma\{\Theta_0(\beta); \mathcal{D}\} - \dot{m}_\Lambda\{\Theta_0(\beta); \mathcal{D}\}[\mathbf{h}^*(\cdot; \beta)],$$

assuming $\Sigma_1(\beta)$ is nonsingular for $\beta \in \Omega_\beta$.

Note that $S^*(\beta; c, \delta, \mathbf{z})$ is continuously differentiable with respect to $\beta \in \Omega_\beta$ for any $(c, \delta, \mathbf{z})$. By theorem 7.7.4 of Van der Vaart and Wellner (1996), the class of functions $n^{1/2}\{\hat{\gamma}(\beta) - \gamma_0(\beta)\}$ indexed by $\beta \in \{\beta : \|\beta - \beta_0\| \leq \delta\}$ for some $\delta > 0$, is Donsker and hence the local linearity condition (7) is satisfied.

8.3 Resampling Method

Similar to (1.6) in Murphy and Van der Vaart (2000), for $\mathbf{b} \in \mathcal{N}_{\epsilon_n} = \{\mathbf{b} : \|\mathbf{b} - \beta_0\| \leq \epsilon_n\}$, where ϵ_n is any positive sequence of the order of $o(1)$,

$$M_{\mathbf{b}}\{\hat{\gamma}_n(\mathbf{b})\} - M_{\mathbf{b}}(0) = \frac{n}{2}\hat{\gamma}_n(\mathbf{b})'\mathbf{I}(\mathbf{b})\hat{\gamma}_n(\mathbf{b}) + o_p(1 + n\|\hat{\gamma}_n(\mathbf{b})\|^2),$$

where $\mathbf{I}(\mathbf{b}) = \text{ES}^*(\beta; \mathcal{D})^{\otimes 2}$. Furthermore, from the stochastic equicontinuity property of the process $n^{1/2}\{\hat{\gamma}_n(\beta) - \gamma_0(\beta)\}$, we have

$$n^{1/2}\|\{\hat{\gamma}_n(\mathbf{b}) - \gamma_0(\mathbf{b})\} - \{\hat{\gamma}_n(\hat{\beta}) - \gamma_0(\hat{\beta})\}\| = o_p(1).$$

This, coupled with the fact that $\sup_{\mathcal{N}_{\epsilon_n}} \|\mathbf{I}(\mathbf{b}) - \mathbf{B}_0\| = \sup_{\mathcal{N}_{\epsilon_n}} \|\mathbf{I}(\mathbf{b}) - \mathbf{I}(\beta_0)\| = o_p(1)$, implies that for $\mathbf{b} \in \mathcal{N}_{\epsilon_n}$,

$$\begin{aligned} M_{\mathbf{b}}\{\hat{\gamma}_n(\mathbf{b})\} - M_{\mathbf{b}}(0) &= \frac{n}{2} \left\{ \gamma_0(\hat{\beta}) - \gamma_0(\mathbf{b}) \right\}' \mathbf{I}(\mathbf{b}) \left\{ \gamma_0(\hat{\beta}) - \gamma_0(\mathbf{b}) \right\} + o_p(1 + n\|\mathbf{b} - \hat{\beta}\|^2) \\ &= \frac{n}{2} (\mathbf{b} - \hat{\beta})' \mathbf{A}_0' \mathbf{B}_0 \mathbf{A}_0 (\mathbf{b} - \hat{\beta}) + o_p(1 + n\|\mathbf{b} - \hat{\beta}\|^2). \end{aligned} \quad (8)$$

Let random variable $\boldsymbol{\eta}_n^* = n^{1/2}(\beta^* - \hat{\beta})I_{\{\|\beta^* - \hat{\beta}\| \leq n^{-1/2}c_n\}}$, where $c_n \rightarrow \infty$ and $n^{-1/2}c_n \rightarrow 0$. Then the density function of $\boldsymbol{\eta}_n^*$ conditioning on the observed data is

$$f_{\boldsymbol{\eta}_n^*}(\boldsymbol{\eta}) = C_{\boldsymbol{\eta}} I_{\{\|\boldsymbol{\eta}\| \leq c_n\}} \exp \left\{ - [M_{\hat{\beta} + \frac{\boldsymbol{\eta}}{\sqrt{n}}} \{\hat{\gamma}_n(\hat{\beta} + \frac{\boldsymbol{\eta}}{\sqrt{n}})\} - M_{\hat{\beta} + \frac{\boldsymbol{\eta}}{\sqrt{n}}}(0)] \right\},$$

where $C_{\boldsymbol{\eta}}$ is the appropriate normalizing constant. It follows from (8) that for any $\epsilon > 0$, $\|\boldsymbol{\eta}\| \leq c_n$, and large n ,

$$P \left\{ \left| M_{\hat{\boldsymbol{\beta}} + \frac{\boldsymbol{\eta}}{\sqrt{n}}} \{ \hat{\boldsymbol{\gamma}}_n (\hat{\boldsymbol{\beta}} + \frac{\boldsymbol{\eta}}{\sqrt{n}}) \} - M_{\hat{\boldsymbol{\beta}} + \frac{\boldsymbol{\eta}}{\sqrt{n}}} (0) - \frac{1}{2} \boldsymbol{\eta}' \mathbf{A}'_0 \mathbf{B}_0 \mathbf{A}_0 \boldsymbol{\eta} \right| \leq \epsilon (1 + \|\boldsymbol{\eta}\|) \right\} > 1 - \epsilon,$$

where the probability is with respect to $\{\mathcal{D}_i : i = 1, \dots, n\}$. This, coupled with some elementary algebra, implies that $(1 - K_0\epsilon)f_2(\boldsymbol{\eta}; \epsilon) \leq f_{\boldsymbol{\eta}_n^*}(\boldsymbol{\eta}) \leq (1 + K_0\epsilon)f_1(\boldsymbol{\eta}; \epsilon)$ for some positive constant K_0 , where $f_1(\boldsymbol{\eta}; \epsilon)$ and $f_2(\boldsymbol{\eta}; \epsilon)$ are the respective density functions of $N(0, (\mathbf{A}'_0 \mathbf{B}_0 \mathbf{A}_0 + \epsilon \mathbf{I}_0)^{-1})$ and $N(0, (\mathbf{A}'_0 \mathbf{B}_0 \mathbf{A}_0 - \epsilon \mathbf{I}_0)^{-1})$, and $\mathbf{I}_0$ is the $p \times p$ identity matrix. As a consequence, for any bounded Lipschitz continuous function $\psi(\cdot)$ on R^p ,

$$\begin{aligned} E\{\psi(\boldsymbol{\eta}_n^*)\} &= \int_{\|\boldsymbol{\eta}\| \leq c_n} \psi(\boldsymbol{\eta}) f_{\boldsymbol{\eta}_n^*}(\boldsymbol{\eta}) d\boldsymbol{\eta} \\ &\leq \int_{\psi(\boldsymbol{\eta}) > 0, \|\boldsymbol{\eta}\| \leq c_n} \psi(\boldsymbol{\eta}) f_1(\boldsymbol{\eta}; \epsilon) (1 + K_0\epsilon) d\boldsymbol{\eta} + \int_{\psi(\boldsymbol{\eta}) < 0, \|\boldsymbol{\eta}\| \leq c_n} \psi(\boldsymbol{\eta}) f_2(\boldsymbol{\eta}; \epsilon) (1 - K_0\epsilon) d\boldsymbol{\eta} \\ &\leq \int_{\|\boldsymbol{\eta}\| \leq c_n} \psi(\boldsymbol{\eta}) f_0(\boldsymbol{\eta}) d\boldsymbol{\eta} + K_1 \epsilon^{\alpha_1}, \end{aligned}$$

where K_1 and α_1 are positive constants and $f_0(\cdot)$ is the density function of $N(0, \mathbf{A}_0^{-1} \mathbf{B}_0^{-1} (\mathbf{A}'_0)^{-1})$. Similarly, $E\{\psi(\boldsymbol{\eta}_n^*)\} \geq \int_{\|\boldsymbol{\eta}\| \leq c_n} \psi(\boldsymbol{\eta}) f_0(\boldsymbol{\eta}) d\boldsymbol{\eta} - K_2 \epsilon^{\alpha_2}$, for some constant $K_2, \alpha_2 > 0$. Since $c_n \rightarrow \infty$, as $n \rightarrow \infty$, with probability approximating to one, the conditional distribution $\boldsymbol{\eta}_n^* = n^{1/2}(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) I_{\{n^{1/2}\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\| \leq c_n\}}$ can be used to approximate the distribution of $N(0, \mathbf{A}_0^{-1} \mathbf{B}_0^{-1} (\mathbf{A}'_0)^{-1})$, which is the limiting distribution of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$.

REFERENCES

- BARLOW, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972), Statistical Inference under Order Restriction, Wiley, London.
- BETENSKY, R.A., RABINOWITZ, D., and TSIATIS, A.A. (2001), Computationally simple accelerated failure time regression for interval censored data, *Biometrika*, **88**, 703-711.
- BUCKLEY, J. and JONATHAN, I. (1979), Linear regression with censored data, *Biometrika*, **66**, 429-436.
- CAI, T., and BETENSKY, R. A. (2003), Hazard regression for interval-censored data with penalized spline, *Biometrics*, **59**, 570-579.
- DINSE, G.E. and LAGAKOS, S. W. (1983), Regression analysis of tumour prevalence data, *Applied Statistics*, **32**, 236-248.
- FINKELSTEIN, D. M. and WOLFE R. A. (1985). A semiparametric model for regression analysis of interval censored failure time data, *Biometrics*, **41**, 933-945.
- FINKELSTEIN, D. M. (1986). A proportional hazards model for interval-censored failure time data, *Biometrics*, **42**, 845-854.
- GELMAN, A. and CARLIN, J., STERN, H., and RUBIN, D. (2003). Bayesian Data Analysis, Chapman & Hall, New York.
- GROENEBOOM, P. and WELLNER, J. A. A. (1992) Information Bounds and Nonparametric Maximum Likelihood Estimation, DMV Seminar Band 19, Birkhauser, Basel.
- HUANG, J. (1996) Efficient estimation for the Cox model with interval censoring, *Annals of Statistics*, **24**, 540-568.
- HUANG, J. and WELLNER, J. A. (1997) Interval censored survival data: a review of recent progress, *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, dds.

Lin, D. and Fleming, T. Springer-Verlag, New York.

JIN, Z., LIN, D.Y., WEI, L.J. and YING Z. (2003), Rank-based inference for the accelerated failure time model, *Biometrika*, **90**, 341-353.

LIU, J. (2001), Monte Carlo Strategies in Scientific Computing, Springer, New York.

LIN, D. Y. and GEYER, C.J. (1992), Computational methods for semiparametric linear regression with censored data, *J. Comput. Graph. Statist.*, **1**, 77-90.

LIN, D. Y., OAKES, D. and YING, Z. (1998), Additive hazards regression with current status data, *Biometrika*, **85**, 289-298.

MARTINUSSEN, T. and SCHEIKE, T.H. (2002), Efficient estimation in additive hazards regression with current status data, *Biometrika*, **89**, 649-658.

METROPOLIS, N., ROSENBLUTH,, A.W., TELLER, A.H., and TELLER, E. (1953), Equation of state calculations by fast computing machines, *Journal of Chemical Physics*, **21**, 1087-1092.

MURPHY, S.A., VAN DER VAART, A.W., and WELLNER, J.A. (1999), Current status regression, *Mathematical Methods of Statistics*, **8**, 407-425.

MURPHY, S.A. and VAN DER VAART, A.W. (2000), On profile likelihood, *Journal of American Statistical Association*, **95**, 449-465.

RABINOWITZ, D., BETENSKY, R.A., and TSIATIS, A.A. (2000), Using conditional logistic regression to fit proportional odds models to interval censored data, *Biometrics*, **56**, 511-518.

RABINOWITZ, D., TSIATIS, A.A., and ARAGON, J. (1995), Regression with interval-censored data, *Biometrika*, **82**, 501-513.

ROSSINI, A. and TSIATIS, A.A. (1996), A semiparametric proportional odds regression model for the analysis of current status data, *J. Am. Statist. Assoc.*, **91**, 713-721.

SATTEN, G. A. (1996), Rank-based inference in the proportional hazards model for interval

censored data, *Biometrika*, **83**, 355-370.

SHEN, X. (1998), Proportional odds regression and sieve maximum likelihood estimation, *Biometrika*, **85**, 165-177.

SHEN, X. (2000), Linear regression with current status data, *Journal of American Statistical Association*, **95**, 842-852.

SHIBOSKI, SC. (1998), Generalized additive models for current status data, *Lifetime Data Analysis*, **4**, 29-50.

TIAN, L., LIU, J., ZHAO, Y. and WEI, L.J. (2004), Statistical inferences based on non-smooth estimating functions, to be appeared in *Biometrika*.

TSIATIS, A.A. (1990), Estimating regression parameters using linear rank tests for censored data, *Annals of Statistics*, **18**, 354-372.

TURNBULL, B.W. (1976), The empirical distribution function with arbitrarily grouped censored and truncated data, *J. Roy. Statist. Soc. B*, **38**, 290-295.

VAN DER VAART, A.W. and WELLNER, J.A.(1996), Weak Convergence and Empirical Processes, Springer-Verlag, New York.

WEI, L.J., YING, Z. and LIN, D. Y. (1990), Linear regression analysis of censored survival data based on rank tests, *Biometrika*, **77**, 845-851.

Table 1: Fitted regression coefficients in the AFT model for assessing the covariate effects on the occurrence of bile duct hyperplasia in the tumorigenicity study.

Covariate	Point Estimator	Estimated SD	95% CI interval
Dose level of PBB	-0.505	0.235	(-0.967, -0.043)
Baseline weight	-0.063	0.031	(-0.124, -0.002)
Number of Tier	-0.196	0.222	(-0.631, 0.238)
Sex ¹	0.789	0.673	(-0.530, 2.109)

¹: 1 male, 0 female.

Table 2: Empirical bias (Bias), average of the estimated standard errors (Ase), sampling standard errors (Sse), empirical coverage probabilities (ECP) of 95% and 90% confidence intervals for $\hat{\beta}$ from the simulation study for the AFT model for current status data. Results are based on 500 simulated data sets.

n	Parameter	Bias	Ase	Sse	ECP (95%)	ECP (90%)
150	β_1	0.011	0.269	0.265	95.2%	89.8%
	β_2	0.001	0.234	0.224	94.6%	90.6%
	β_3	-0.019	0.266	0.243	96.2%	91.0%
300	β_1	0.030	0.177	0.177	93.6%	89.2%
	β_2	-0.007	0.158	0.163	93.6%	89.8%
	β_3	-0.009	0.177	0.183	93.8%	89.0%

Table 3: Empirical bias (Bias), average of the estimated standard errors (Ase), sampling standard errors (Sse), empirical coverage probabilities (ECP) of 95% and 90% confidence intervals for $\hat{\beta}$ from the simulation study for the AFT model for interval censored data. Results are based on 500 simulated datasets.

n	Parameter	Bias	Ase	Sse	ECP (95%)	ECP (90%)
150	β_1	-0.011	0.219	0.211	94.6%	91.2%
	β_2	0.005	0.202	0.199	95.8%	90.2%
	β_3	0.022	0.216	0.213	95.6%	90.6%
300	β_1	-0.001	0.145	0.143	96.0%	90.4%
	β_2	0.009	0.137	0.136	96.4%	91.4%
	β_3	0.012	0.145	0.149	94.0%	89.2%

Figure 1: Trace Plot of the Generated Markov Chain for Each of the Four Covariates

Figure 2: Histogram of the MCMC Output for Each of the Four Covariates

Figure 3: Histograms of MCMC output versus the estimating function $\hat{\gamma}_n(\beta)$ in breast cosme-
sis example