Statistics and Its Interface Volume 8 (2015) 391–402

# Efficient estimation for the additive hazards model in the presence of left-truncation and interval censoring

Peijie Wang, Xingwei Tong, Shishun Zhao, and Jianguo Sun*

The additive hazards model is one of the most commonly used regression models in failure time data analysis and many authors have discussed its inference under various situations (Lin and Ying, 1994; Lin et al., 1998; Zeng et al., 2006; Wang et al., 2010). In this paper, we consider it when one faces left-truncated and interval-censored data, which often occur in, for example, epidemiological and medical follow-up studies. For inference, an efficient sieve maximum likelihood estimation procedure is developed and assessed by simulation studies, which indicate that the proposed method works well in practical situations. An illustrative example is also provided.

Keywords and phrases: Additive hazards model, Interval-censoring, Left-truncation, Sieve maximum likelihood estimation.

## 1. INTRODUCTION

The additive hazards model is one of the most commonly used regression models in failure time data analysis, especially when one is interested in excess risk or risk difference (Kalbfleisch and Prentice, 2002; Klein and Moeschberger, 2003). For the inference about the model, many estimation procedures have been proposed under various situations (Lin and Ying, 1994; Lin et al., 1998; Zhou and Sun, 2003; Zeng et al., 2006; Wang et al., 2010). For example, some of the early works were given by Lin and Ying (1994) and Lin et al. (1998), who considered the situations of right-censored data and current status data described below, respectively. More recently, both Zeng et al. (2006) and Wang et al. (2010) discussed the fitting of the model to interval-censored data.

By interval-censored failure time data, we mean that the failure time of interest is observed only to belong to an interval or a window instead of observed exactly or right-censored (Finkelstein, 1986; Sun, 2006; Chen et al., 2012). It is easy to see that such data naturally occur in many fields including medical follow-up studies such as clinical trials. In these situations, it is usually the case that study subjects are given a set of prespecified clinical or observation times for checking the status or occurrence of a certain disease or medical condition. However, it is well-known that it is common that the real observation times will be different from subject to subject and the occurrence time of the disease or medical condition is known only to be between some observation times. For the case where each subject is observed only once, the observed data are usually referred to as current status data (Jewell and van der Laan, 2004ab). Other fields that often produce interval-censored data include demographical studies, economic and financial studies, epidemiological studies, social sciences and tumorigenicity experiments.

Many authors have investigated regression analysis of interval-censored failure time data and some early references include Finkelstein (1986), Huang (1996) and Huang and Wellner (1996). For relatively complete and recent references, the readers are referred to Sun (2006) and Chen et al. (2012). In addition to censoring, a failure time study may also involve the left-truncation, which occurs if a subject has to satisfy certain conditions or experience some initial events to be included in a study. It is apparent that the truncation can make the analysis much more complicated and the analysis would yield biased results if one ignores the truncation as discussed below (Pan and Chappell, 1999).

Several methods have been proposed for regression analysis of left-truncated and interval-censored data. For example, Kim (2003) considered left-truncated and current status data and Pan and Chappell (2002) and Shen (2014) investigated the general situation, all under the proportional hazards model. However, it does not seem to exist an established procedure for fitting the additive hazards model to such data. It is well-known that sometimes the proportional hazards model may not fit the data well or be appropriate and the additive hazards model describes a different aspect (Lin and Ying, 1994; Kulich and Lin, 2000). In particular, one may want to employ the additive hazards model when the interest is on additive or excess risk as often in, for example, social sciences, for which the proportional hazards model is clearly not appropriate. In the following, an efficient estimation procedure is presented for the additive hazards model with left-truncated and interval-censored data.

The remainder of the paper is organized as follows. We will begin in Section 2 with introducing some notation, the model and the assumptions used throughout the paper. A sieve maximum likelihood estimation procedure is

*Corresponding author.

then developed in Section 3 for estimation of unknown parameters. In the method, piecewise linear functions are used to approximate the baseline cumulative hazard function (Huang and Rossini, 1997). The resulting estimators of regression parameters are shown to be consistent and have asymptotic normal distribution and furthermore, we show that they are efficient. Section 4 presents some results obtained from a simulation study conducted to evaluate the proposed estimation procedure and they suggest that the approach works well for practical situations. An illustrative example is provided in Section 5 and Section 6 contains some discussion and concluding remarks.

## 2. NOTATION, MODEL AND ASSUMPTIONS

Consider a failure time study that involves $n$ independent subjects as well as both left-truncation and interval-censoring. For subject $i$, let $T_i$ denote the failure time of interest and suppose that there is a vector of covariates denoted by $Z_i$, $i = 1, ..., n$. Also let $X_i$ denote the left-truncation time and $U_i$ and $V_i$ the interval-censored observation times associated with subject $i$ such that $X_i < U_i < V_i$ and one only knows $X_i \leq T_i \leq U_i$, $U_i < T_i \leq V_i$ or $T_i > V_i$. Define $\delta_{1i} = I(T_i \leq U_i)$, $\delta_{2i} = I(U_i < T_i \leq V_i)$, $\delta_{3i} = I(T_i > V_i)$, $i = 1, ..., n$. Then the observed data have the form

$$\{ Y_i = (\delta_{1i}, \delta_{2i}, X_i, U_i, V_i, Z_i) ; \; i = 1, ..., n \} .$$

In the following, we suppose that the main objective is to make inference about the effects of the $Z_i$'s on the $T_i$'s.

To describe the covariate effects, we will assume that given $Z_i$, the hazard function of $T_i$ has the form

$$\lambda(t|Z) = \lambda_0(t) + \beta^T Z , \tag{1}$$

where $\lambda_0(t)$ denotes an unknown baseline hazard function and $\beta$ is a vector of regression parameters. That is, the $T_i$'s follow the additive hazards model (Lin and Ying, 1994). Define $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$, the baseline cumulative hazard function, and $S_0(t) = \exp\{ -\Lambda_0(t) \}$, the baseline survival function. Then the survival function of $T_i$ has the form $S(t) = S_0(t) \exp\{ -\beta^T Z_i t \}$.

In the following, we will assume that given $Z_i$, $(X_i, U_i, V_i)$ are independent of $T_i$. Then the conditional likelihood function of $\beta$ and $\Lambda_0$ given $T_i \geq X_i$ can be written as

$$L_n(\beta, \Lambda_0) = \prod_{i=1}^{n} \{ P(X_i \leq T_i \leq U_i | T_i \geq X_i) \}^{\delta_{1i}}$$
$$\{ P(U_i < T_i \leq V_i | T_i \geq X_i) \}^{\delta_{2i}} \{ P(T_i > V_i | T_i \geq X_i) \}^{\delta_{3i}} .$$

The resulting log-likelihood function has the form

$$\begin{aligned} l_n(\beta, \Lambda_0) = \sum_{i=1}^{n} & \{ \Lambda_0(X_i) + \beta^T Z_i X_i \} \\ & + \delta_{1i} \log \Big\{ \exp\{-\Lambda_0(X_i) - \beta^T Z_i X_i\} \\ & - \exp\{-\Lambda_0(U_i) - \beta^T Z_i U_i\} \Big\} \\ & + \delta_{2i} \log \Big\{ \exp\{-\Lambda_0(U_i) - \beta^T Z_i U_i\} \\ & - \exp\{-\Lambda_0(V_i) - \beta^T Z_i V_i\} \Big\} \\ & + (1 - \delta_{1i} - \delta_{2i}) \left\{ -\Lambda_0(V_i) - \beta^T Z_i V_i \right\} . \end{aligned} \tag{2}$$

In the next section, we will discuss the maximization of the log-likelihood function above with the focus on the inference about regression parameters $\beta$.

## 3. EFFICIENT SIEVE MAXIMUM LIKELIHOOD ESTIMATION

Now we consider the estimation of $\beta$ and $\Lambda_0$. For this, it is apparent that a nature approach is to maximize the log-likelihood function $l_n$ directly. On the other hand, it is well-known that the maximization is not easy or straightforward due to the dimension of $\Lambda_0(t)$. To deal with this, following Huang and Rossini (1997) and others, we propose to employ the sieve approach that approximates $\Lambda_0(t)$ by using linear functions.

Let $0 = t_0 < t_1 < \ldots < t_{q_n} = \tau$ denote a partition of the observation interval $[0, \tau]$, where $\tau$ denotes the largest follow-up time. Here $q_n$ is usually called the sieve number and set to be an increasing integer along with $n$ at the rate $O(n^\kappa)$ with $0 < \kappa < 1/2$. Define $H_n$ to be the set of all linear functions

$$\Lambda_n(t) = \sum_{l=1}^{q_n} I_l(t) \left\{ h_{l-1} + \frac{h_l - h_{l-1}}{t_l - t_{l-1}} (t - t_{l-1}) \right\}$$

with $\Lambda_n(t) \leq M$ for $0 \leq t \leq \tau$. Here $I_l(t) = I(t_{l-1} < t \leq t_l)$, $M$ is a constant, and $0 = h_0 \leq h_1 \leq h_2 \leq \ldots \leq h_{q_n} \leq M$ are unknown parameters. It is easy to see that $\Lambda_n(t_l) = h_l$, $l = 0, 1, \ldots, q_n$. By following the sieve approach, we can estimate $\beta$ and $\Lambda_0$ by maximizing $l_n(\beta, \Lambda_0)$ over $\Theta_n = B \times H_n$, where $B$ denotes the parameter space for $\beta$. In practice, it is more convenient to reparameterize the $h_l$'s by $h_l = \sum_{k=1}^{l} e^{\gamma_k}$ to remove the range limitation, where $\gamma = (\gamma_1, \ldots, \gamma_{q_n})^T$ are some unknown parameters. With respect to $\gamma$, we can rewrite $\Lambda_n(t)$ as

$$\begin{aligned} \Lambda_n(t) = {} & I_1(t) \, e^{\gamma_1} \, \frac{t - t_0}{t_1 - t_0} \\ & + \sum_{l=2}^{q_n} I_l(t) \left( \sum_{k=1}^{l-1} e^{\gamma_k} + e^{\gamma_l} \frac{t - t_{l-1}}{t_l - t_{l-1}} \right) . \end{aligned} \tag{3}$$

It is easy to see that $\Lambda_n(t)$ is a piecewise linear function and by focusing on the space $\Theta_n$, we have a finite parameter estimation problem compared to the original estimation problem. Also it can be easily shown that as $n \to \infty$, $\Theta_n$ converges to the original parameter space $\Theta$ and thus $\Theta_n$ can be used as a sieve space.

For estimation of $\beta$ and $\Lambda_0$, we propose to use the estimator $\widehat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n)$ defined as

$$l_n(\hat{\beta}_n, \hat{\Lambda}_n) = \max_{\Theta_n} l_n(\beta, \Lambda) = \max_{\beta, \gamma_l' s} l_n(\beta, \gamma_l' s).$$

For the determination of $\widehat{\theta}_n$, it is apparent that a natural method is to solve the following score equations

$$(4) \quad U(\beta) = \sum_{i=1}^n \delta_{1i} \frac{S_{ni}(X_i)(-Z_i X_i) - S_{ni}(U_i)(-Z_i U_i)}{S_{ni}(X_i) - S_{ni}(U_i)} + \delta_{2i} \frac{S_{ni}(U_i)(-Z_i U_i) - S_{ni}(V_i)(-Z_i V_i)}{S_{ni}(U_i) - S_{ni}(V_i)} + (1 - \delta_{1i} - \delta_{2i})(-Z_i V_i) + Z_i X_i = 0$$

and

(5)
$$U(\gamma_l) = \sum_{i=1}^n \delta_{1i} \frac{S_{ni}(X_i)(-\frac{\partial \Lambda_n(X_i)}{\partial \gamma_l}) - S_{ni}(U_i)(-\frac{\partial \Lambda_n(U_i)}{\partial \gamma_l})}{S_{ni}(X_i) - S_{ni}(U_i)} + \delta_{2i} \frac{S_{ni}(U_i)(-\frac{\partial \Lambda_n(U_i)}{\partial \gamma_l}) - S_{ni}(V_i)(-\frac{\partial \Lambda_n(V_i)}{\partial \gamma_l})}{S_{ni}(U_i) - S_{ni}(V_i)} + (1 - \delta_{1i} - \delta_{2i})(-\frac{\partial \Lambda_n(V_i)}{\partial \gamma_l}) + \frac{\partial \Lambda_n(X_i)}{\partial \gamma_l} = 0, \quad l = 1, ..., q_n,$$

where

$$S_{ni}(t) = \exp(-\Lambda_n(t) - \beta^T Z_i t),$$
$$\frac{\partial \Lambda_n(t)}{\partial \gamma_l} = \left\{ I(t > t_l) + \frac{t - t_{l-1}}{t_l - t_{l-1}} I_l(t) \right\} e^{\gamma_l}.$$

In the following, we will establish the consistency of $\widehat{\theta}_n$ and the asymptotic normality and efficiency of $\hat{\beta}_n$. For this, let $G(x, u, v)$ denote the joint distribution function of $(X_i, U_i, V_i)$ and define the distance

$$d\{(\beta_1, \Lambda_1), (\beta_2, \Lambda_2)\} = \| \beta_1 - \beta_2 \|_2 + \| \Lambda_1 - \Lambda_2 \|_2,$$

where

$$\| \Lambda_1 - \Lambda_2 \|_2^2 = \int \{\Lambda_1(x) - \Lambda_2(x)\}^2 + \{\Lambda_1(u) - \Lambda_2(u)\}^2 + \{\Lambda_1(v) - \Lambda_2(v)\}^2 dG(x, u, v).$$

All limits below are taken as $n \to \infty$. First we will give a result about the information matrix for $\beta$.

**Theorem 3.1.** *Suppose that the regularity conditions (A1)–(A7) described in the Appendix hold. Then it can be shown that the information matrix $I(\beta)$ given in the Appendix for $\beta$ is a positive definite matrix with finite entries.*

**Theorem 3.2.** *Suppose that the regularity conditions (A1)–(A7) described in the Appendix hold. Then we have*

$$d(\widehat{\theta}_n, \theta_0) = d\{(\widehat{\beta}_n, \widehat{\Lambda}_n), (\beta_0, \Lambda_0)\} \to 0$$

*in probability. Furthermore, it can be shown that*

$$d(\widehat{\theta}_n, \theta_0) = d\{(\widehat{\beta}_n, \widehat{\Lambda}_n), (\beta_0, \Lambda_0)\} = O_p(\max\{n^{-(1-\kappa)/2}, n^{-r\kappa}\})$$

*in probability with $0 < \kappa < 1/2$, $r$ defined in the condition (A5) and $\beta_0$ being the true value of $\beta$.*

**Theorem 3.3.** *Suppose that $\beta_0$ is an interior point of $B$ and $1/4r < \kappa < 1/2$. Also assume that the regularity conditions (A1)–(A7) described in the Appendix hold. Then we have*

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) = -I^{-1}(\beta_0) \sqrt{n} P_n l^*_{\beta_0}(Y) + o_p(1) \to N(0, I^{-1}(\beta_0))$$

*in distribution, where $P_n$ denotes the empirical measure of $\{ Y_i = (\delta_{1i}, \delta_{2i}, X_i, U_i, V_i, Z_i) ; i = 1, \ldots, n \}$ and $l^*_\beta$ the efficient score function for $\beta$ derived in the Appendix.*

The proofs of the results above are sketched in the Appendix. Note that Theorem 3.2. suggests that the proposed estimator $\widehat{\theta}_n$ is not only consistent, but also can be optimal with setting $\kappa = 1/(1 + 2r)$, which gives the convergence rate of $n^{r/(1+2r)}$, equal to $n^{1/3}$ or $n^{2/5}$ for $r = 1$ or 2, respectively. Theorem 3.3. tells us that one can approximate the distribution of $\widehat{\beta}_n$ by the normal distribution and its asymptotic covariance reaches the lower bound. That is, $\widehat{\beta}_n$ is efficient.

To make inference about $\beta$ based on the result above, one needs to estimate the information matrix $I(\beta)$. For this, one common method is to use the observed information matrix, which would involve the calculation of the second derivatives of the log likelihood. Instead we suggest to employ the profile likelihood method (Murphy and van der Vaart, 1999). Specifically, let $pl_n(\beta) = l_n(\beta, \Lambda_\beta)$ denote the profile log likelihood for $\beta$, where $\Lambda_\beta$ denotes $\Lambda$ that maximizes $l_n(\beta, \Lambda)$ for given $\beta$. Also let $h_1, \ldots, h_d$ denote random variables and $e_i$ a $d$-dimensional vector of zeros except the $i$th element being equal to one, where $d$ denotes the dimension of $\beta$. For $i \neq j = 1, ..., d$, define

$$-\widehat{I}_{ij} = \frac{1}{n h_i h_j} (pl_n(\widehat{\beta} + h_i e_i + h_j e_j) - pl_n(\widehat{\beta} + h_i e_i) - pl_n(\widehat{\beta} + h_j e_j) + pl_n(\widehat{\beta}))$$

and

$$-\widehat{I}_{ii} = \frac{1}{n h_i^2} (pl_n(\widehat{\beta} + h_i e_i) - 2 pl_n(\widehat{\beta}) + pl_n(\widehat{\beta} - h_i e_i))$$

for $i = 1, ..., d$. Then one can show that if $h_i \xrightarrow{P} 0$ and both $h_i/h_j$ and $(\sqrt{n} h_i)^{-1}$ are bounded in probability, $\widehat{I}_{ij}$ converges in probability to the $(i, j)$th component of the information matrix $I(\beta)$. In other words, $I(\beta)$ can be con-

Table 1. Simulation results on estimation of $\beta$ with discrete covariates

| $\beta_0$ | $n$ | Bias | SSE | ESE | CP |
|---|---|---|---|---|---|
| 20% left-truncation | | | | | |
| 0 | 100 | -0.0110 | 0.2830 | 0.2691 | 0.949 |
| | 200 | -0.0053 | 0.1908 | 0.1871 | 0.952 |
| 0.5 | 100 | 0.0119 | 0.3380 | 0.3442 | 0.955 |
| | 200 | 0.0013 | 0.2388 | 0.2368 | 0.948 |
| 1 | 100 | 0.0169 | 0.4204 | 0.4149 | 0.949 |
| | 200 | 0.0135 | 0.2823 | 0.2860 | 0.948 |
| 40% left-truncation | | | | | |
| 0 | 100 | -0.0086 | 0.2852 | 0.2731 | 0.945 |
| | 200 | -0.0050 | 0.1929 | 0.1858 | 0.952 |
| 0.5 | 100 | 0.0076 | 0.3430 | 0.3397 | 0.956 |
| | 200 | 0.0070 | 0.2403 | 0.2384 | 0.951 |
| 1 | 100 | 0.0233 | 0.4200 | 0.4135 | 0.950 |
| | 200 | 0.0074 | 0.2883 | 0.2856 | 0.947 |

Table 2. Simulation results on estimation of $\beta$ with continuous covariates

| $\beta_0$ | $n$ | Bias | SSE | ESE | CP |
|---|---|---|---|---|---|
| 20% left-truncation | | | | | |
| 0 | 100 | -0.0070 | 0.4952 | 0.4554 | 0.933 |
| | 200 | -0.0057 | 0.3283 | 0.3204 | 0.952 |
| 0.5 | 100 | -0.0166 | 0.6014 | 0.5618 | 0.939 |
| | 200 | 0.0075 | 0.3947 | 0.3968 | 0.947 |
| 1 | 100 | 0.0186 | 0.7219 | 0.6589 | 0.928 |
| | 200 | 0.0088 | 0.4558 | 0.4582 | 0.938 |
| 40% left-truncation | | | | | |
| 0 | 100 | -0.0148 | 0.4876 | 0.4468 | 0.935 |
| | 200 | -0.0034 | 0.3202 | 0.3214 | 0.953 |
| 0.5 | 100 | -0.0178 | 0.6000 | 0.5712 | 0.936 |
| | 200 | 0.0060 | 0.4007 | 0.3999 | 0.956 |
| 1 | 100 | 0.0308 | 0.7210 | 0.6616 | 0.934 |
| | 200 | 0.0206 | 0.4829 | 0.4622 | 0.949 |

sistently estimated by $\hat{I} = (\widehat{I}_{ij})$. For the numerical studies below, we use either $h_i = \max(|\widehat{\beta}_i|, 1) \times \text{sign}(\widehat{\beta}_i)/\sqrt{n}$ or $h_i = \text{sign}(\widehat{\beta}_i)/\sqrt{n}$.

## 4. A SIMULATION STUDY

To evaluate the finite sample performance of the estimation procedure proposed in the previous sections, we conducted a simulation study. In the study, for the covariate, we considered three cases. The first two are to assume that there exists one covariate and the $Z_i$'s follow either the Bernoulli distribution with the success probability being 0.5 or the uniform distribution over $(0, 1)$, respectively. The third case is to assume that there exist two covariates with the first one generated from the Bernoulli distribution with the success probability being 0.5 and the second covariate from the uniform distribution over $(0, 1)$. For the generation of the left-truncated and interval-censored times, we first generated $W_{1i}$, $W_{2i}$ and $W_{3i}$ from the exponential distributions with the parameters $a$, $b$ and $c$, respectively, and then took $X_i = W_{1i}$, $U_i = X_i + W_{2i} + 0.1$ and $V_i = U_i + W_{3i} + 0.1$. The failure times of interest $T_i$'s were generated under model (1) with $\lambda_0(t) = 1$ and the empirical approach was employed to obtain the required truncation and censoring percentages. The results given below are based on 1,000 replications.

Tables 1 and 2 present the results on estimation of $\beta$ based on the simulated data with the $Z_i$'s being either discrete or continuous, respectively, the true value $\beta_0 = 0,\ 0.5$ or 1, and $n = 100$ or 200. The results include the estimated bias (Bias) given by the average of the estimators minus the true value of $\beta$, the sample standard error (SSE), the average of the estimated standard errors (ESE) and the 95% empirical coverage probability (CP). Here we took $q_n$ to be the integer part of $n^{1/3}$. In each of the tables, the top part is for the case of 20% left-truncation, while the bottom part corresponds to 40% left-truncation. These results suggest that the proposed estimator seems to be unbiased and the variance estimation method also seems to work well. As expected, the results became better when the sample size increased.

*Table 3. Simulation results on estimation of $\beta$ with two covariates*

| $\beta_0$ | $n$ | Bias | SSE | ESE | CP |
|---|---|---|---|---|---|
| $(\beta_1, \beta_2) = (0.5, 1)$ | | | | | |
| 100 | $\beta_1$ | 0.0251 | 0.4461 | 0.4427 | 0.955 |
| | $\beta_2$ | 0.0355 | 0.8077 | 0.7468 | 0.933 |
| 200 | $\beta_1$ | -0.0096 | 0.3061 | 0.3086 | 0.953 |
| | $\beta_2$ | -0.0089 | 0.5416 | 0.5243 | 0.939 |
| $(\beta_1, \beta_2) = (0.5, 0.5)$ | | | | | |
| 100 | $\beta_1$ | 0.0338 | 0.4017 | 0.3880 | 0.944 |
| | $\beta_2$ | 0.0167 | 0.7348 | 0.6425 | 0.925 |
| 200 | $\beta_1$ | 0.0115 | 0.2805 | 0.2729 | 0.948 |
| | $\beta_2$ | -0.0044 | 0.4755 | 0.4530 | 0.939 |

*Table 4. Estimated biases with and without taking into account the left-truncation*

| Covariate Type | % of left-truncation | $n$ | With left-truncation | Without (ignoring) left-truncation |
|---|---|---|---|---|
| discrete | 20% | 100 | -0.0124 | -0.2896 |
| | | 200 | -0.0052 | -0.2816 |
| | 40% | 100 | 0.0114 | -0.5282 |
| | | 200 | 0.0097 | -0.5262 |
| continuous | 20% | 100 | 0.0156 | -0.2954 |
| | | 200 | 0.0089 | -0.2767 |
| | 40% | 100 | -0.0260 | -0.5953 |
| | | 200 | -0.0112 | -0.5357 |

The results on estimation of $\beta$ with two covariates are given in Table 3 with $\beta_0 = (0.5, 1)'$ or $(0.5, 0.5)'$, 40% left-truncation percentage and the other set-ups being the same as in Tables 1 and 2. Again these results indicate that the proposed estimation procedure seems to work well. In addition, the CP in all three tables plus the quantile plots of the standardized estimator against the standard normal variable, which are not shown here, suggest that the normal approximation to the distribution of the proposed estimator seems to be appropriate too for the situations considered here. It is interesting to note that the results obtained under two different truncation percentages are close to each other. In other words, for the situations considered here, the truncation percentage does not seem to have significant effects.

To further investigate the truncation effect on the estimation of regression parameters, we repeated the studies above but assuming no truncation in fitting the proposed estimation procedure to the simulated data. In other words, we compared the proposed method to the same method but ignoring the truncation. Table 4 displays the estimated biases for the two approaches under various situations with $\beta_0 = 1$ and the other set-ups being the same as with either Table 1 or 2. It is easy to see that the ignoring of the left-truncation yielded biased estimates. In all results given above, we chose the sieve number $q_n$ to be the integer part of $n^{1/3}$, which is 4 or 5 for $n = 100$ or 200, respectively. As suggested by a referee, it would be useful to evaluate the robustness of the proposed estimation procedure to the sieve number. Table 5 presents the results on estimation of $\beta$ by using different sieve numbers with the other set-ups being the same as in Table 1. One can see from the table that they are similar to each other and suggest that the proposed estimation procedure seems to be robust to the sieve number.

Note that in the above, the focus has been on the interval-censored data that can be described by the two variables $U_i$ and $V_i$. In practice, the observed data may have a different format like those arising from periodic follow-up studies such as clinical trials. To assess the validity of the proposed estimation procedure for these situations, we also performed the study in which each subject was supposed to be observed at a sequence of fixed, equally spaced time points $t_1 < \cdots < t_k$. Furthermore, it was assumed that at each time point, a subject was actually observed with the probability $p$, and we defined $U_i$ and $V_i$ to be the largest and smallest $t_j$ at which subject $i$ was observed and that are smaller and larger than the generated failure time $T_i$, respectively. The other set-ups are the same as before and Table 6 gives the obtained estimation results with 40% left-truncation, the $Z_i$'s following the Bernoulli distribution, $k = 10$, $t_j = 1 + 0.1j$, $j = 1, \ldots, 10$ and $p = 0.6$ or 0.8. They indicate that the proposed approach seems still to be valid and perform well. We also considered other set-ups and obtained similar results.

## 5. AN ILLUSTRATIVE EXAMPLE

In this section, we illustrate the inference procedure proposed in the previous sections by applying it to the AIDS cohort study of hemophiliacs discussed in DeGruttola and

*Table 5. Simulation results with different sieve numbers*

| $n$ | $q_n$ | Bias | SSE | ESE | CP |
|---|---|---|---|---|---|
| 20% left-truncation | | | | | |
| 100 | 3 | 0.0166 | 0.3414 | 0.3392 | 0.954 |
| | 4 | -0.0182 | 0.3406 | 0.3392 | 0.947 |
| | 5 | -0.0199 | 0.3197 | 0.3338 | 0.942 |
| 200 | 4 | 0.0127 | 0.2407 | 0.2349 | 0.953 |
| | 5 | 0.0113 | 0.2364 | 0.2349 | 0.944 |
| | 6 | -0.0029 | 0.2321 | 0.2309 | 0.926 |
| 40% left-truncation | | | | | |
| 100 | 3 | 0.0156 | 0.3402 | 0.3381 | 0.956 |
| | 4 | 0.0198 | 0.3462 | 0.3396 | 0.948 |
| | 5 | 0.0157 | 0.3513 | 0.3427 | 0.954 |
| 200 | 4 | 0.0143 | 0.2312 | 0.2350 | 0.957 |
| | 5 | 0.0103 | 0.2429 | 0.2353 | 0.947 |
| | 6 | 0.0045 | 0.2457 | 0.2358 | 0.946 |

*Table 6. Simulation results based on the simulated data from periodic follow-up studies*

| $\beta_0$ | $n$ | Bias | SSE | ESE | CP |
|---|---|---|---|---|---|
| $p = 0.8$ | | | | | |
| 1 | 100 | 0.0501 | 0.3843 | 0.3724 | 0.957 |
| | 200 | 0.0157 | 0.2569 | 0.2545 | 0.950 |
| 0.5 | 100 | 0.0145 | 0.3158 | 0.3000 | 0.950 |
| | 200 | 0.0115 | 0.2050 | 0.2053 | 0.953 |
| 0 | 100 | -0.0086 | 0.2539 | 0.2422 | 0.950 |
| | 200 | -0.0025 | 0.1641 | 0.1655 | 0.953 |
| $p = 0.6$ | | | | | |
| 1 | 100 | 0.0407 | 0.4010 | 0.3763 | 0.954 |
| | 200 | 0.0158 | 0.2524 | 0.2560 | 0.955 |
| 0.5 | 100 | 0.0097 | 0.3067 | 0.2972 | 0.959 |
| | 200 | 0.0084 | 0.2111 | 0.2061 | 0.950 |
| 0 | 100 | -0.0096 | 0.2561 | 0.2416 | 0.948 |
| | 200 | 0.0050 | 0.1642 | 0.1664 | 0.956 |

Lagakos (1989) and Kim et al. (1993) among others. The original study consists of 257 patients with Type A or B hemophilia and these patients were at risk for HIV-1 infection due to the contaminated blood factor that they received for their treatment. Also the patients are classified into two groups, lightly and heavily treated groups, according to the amount of blood they received and one of the original study goals is to assess the effect of the amount of blood or the group effect. For both HIV-1 infection and AIDS diagnosis times, only interval-censored data are available. For the analysis below, following others, we will focus on the 188 patients who were found to be infected by HIV-1 at the time of the analysis and among them, 41 were diagnosed to have AIDS. Our interest is on the group effect on the AIDS diagnosis time.

For the analysis, define $T_i$ to be the AIDS diagnosis time for patient $i$ and $Z_i = 0$ if the $i$th patient belongs to the lightly treated group and $Z_i = 1$ otherwise. By following Kim (2003), we will use the midpoint of the observed interval for HIV-1 infection as the left-truncation time for the AIDS diagnosis time. Some comments on this are given below. To apply the proposed estimation procedure, we tried several sieve numbers. With $q_n = 5$, we obtained $\widehat{\beta}_n = 0.0172$ with the estimated standard error of 0.0080, which gives the $p$-value of 0.0324 for testing no group effect. If using $q_n = 9$, the method yielded $\widehat{\beta}_n = 0.0197$ with the estimated standard error being 0.0086, corresponding to the $p$-value of 0.0224 for the same test. The results with other $q_n$ are similar and they all suggest that the patients in the heavily treated group had significantly shorter AIDS diagnosis time or higher risk of developing AIDS than in the lightly treated group. For comparison, we also analyzed the data by using the proposed method but ignoring the truncation as discussed in the previous section. The resulting estimates with $q_n = 5$ and $q_n = 9$ for the group effect are $-0.0088$ or $-0.0521$, respectively. The negative sign of the estimated treatment effect clearly suggests that the ignoring of left-truncation can lead to some misleading or completely different results or conclusions as suggested by the simulation study.

## 6. CONCLUDING REMARKS

This paper discussed regression analysis of left-truncated and interval-censored failure time data arising from the additive hazards model. For estimation, an efficient sieve maximum likelihood estimation procedure was developed with the use of piecewise linear functions to approximate the unknown baseline cumulative hazard function. In addition, we established the asymptotic properties of the proposed estimators including the efficiency of the estimated regression parameters. The simulation study indicates that the proposed method seems to work well for practical situations.

Note that in the proposed estimation approach, we have employed piecewise linear functions to approximate the baseline cumulative hazard function and one advantage of such approximation is its simplicity. Alternatively, one may use any other smooth functions and develop similar estimation procedures. Among others, for example, one may apply $I$-spline functions or the seminonparametric approximation discussed in Zhang and Davidian (2008). However, the implementation may be more complicated. Also it is worth noting that for the implementation of the proposed estimation procedure, as usual, one faces a constrained estimation problem as only values of $\beta$ that makes $\lambda(t|Z)$ in model (1) nonnegative are valid.

As pointed out above, the focus of this paper has been on the left-truncated and interval-censored data that can be characterized by variables $X_i$, $U_i$ and $V_i$. In particular, it has been assumed that the truncation time is exactly known. Sometimes this may not be the case (Turnbull, 1976) and instead the truncation time may be known only to belong to an interval. In this latter situation, it is easy to see that a simple and natural approach is to employ an imputation procedure to impute the truncation time and then apply the proposed method as in Section 5. However, this clearly may not yield a valid analysis and an appropriate approach should be to apply some new methods that can deal with both interval-truncated and interval-censored data, whose development is beyond the scope of this paper. With respect to the format of censoring intervals, in reality, one could face the interval-censored data given or described by other formats such as those arising from clinical trials or periodic follow-up studies in general. The simulation study given in Section 4 indicates that the proposed estimation procedure seems to be valid for these types of interval-censored data too. On the other hand, of course, it would be useful to provide some theoretical justification to it.

There exist several other directions for more research too. For example, it would be helpful to develop a data-driven method for the selection of the sieve number in the proposed approach. In addition to left-truncation, sometimes one may also face right-truncation or a truncation interval (Turnbull, 1976) as mentioned above and thus it would be useful to generalize the presented method to these situations. Another direction for future research is the investigation of the model adequacy and the development of some test procedures for it. In the case of right-censored failure time data, some procedures have been developed based on various types of residuals. For interval-censored data, however, there seems to exist little research except Ghosh (2003), who considered the current status data situation only. Of course, a related question is the robustness of the proposed estimation procedure to model (1). For this, we conducted a small simulation study and the results indicate that as expected, the bias of the estimated regression parameter increased when the underlying model moved away from model (1).

## ACKNOWLEDGEMENTS

The authors wish thank the Associate Editor and three reviewers for their many insightful and valuable comments and suggestions, which greatly improved the paper. This work was partly supported by both a NSF grant and a NIH grant to the fourth author.

## APPENDIX A. PROOFS OF THE THEOREMS

Before presenting the proof, we will first define some notation and describe the regularity conditions needed. Let the parameter space be $\Theta = B \times \Phi$, where $B$ is a bounded subset of $R^d$ and $\Phi$ is defined by

$$\Phi = \{\Lambda : [\tau_0, \tau_1] \to [m_0, M_0] \text{ and } \Lambda \text{ is nondecreasing}\}.$$

Condition (A1). Given $Z$, $(X, U, V)$ are independent of $T$.

Condition (A2). The joint distribution of $(X, U, V, Z)$ does not depend on $(\beta, \Lambda_0)$.

Condition (A3). There exist $0 < \tau_0 < \tau_1$ and $0 < m_0 < M_0 < \infty$, such that $P(\tau_0 \leq X < U < V \leq \tau_1) = 1$ and $m_0 < \Lambda_0(\tau_0) < \Lambda_0(\tau_1) < M_0$.

Condition (A4). There exists a positive number $\eta$ such that $P(U - X \geq \eta) = 1$ and $P(V - U \geq \eta) = 1$.

Condition (A5). For $r = 1$ or 2, the $r$th derivative of $\Lambda_0(t)$ is bounded and continuous over $[\tau_0, \tau_1]$.

Condition (A6). (a) $Z$ is bounded, that is, there exists $z_0 > 0$, such that $P(|Z| \leq z_0) = 1$. (b) The distribution of $Z$ is not concentrated on any proper affine subspace of $R^d$, (i.e. of dimension $d - 1$ or smaller).

Condition (A7). The conditional density $g(x, u, v|z)$ of $(X, U, V)$ given $Z$ has uniformly bounded partial derivatives with respect to $x$, $u$ and $v$. The bounds of these partial derivatives do not depend on $z$.

Note that the regularity conditions described above are commonly used in the interval-censored data literature (Huang, 1996; Huang and Wellner, 1996). In particular, conditions (A1) and (A2) are to ensure the independent censorship and conditions (A3) and (A6)(a) usually hold in typical medical studies. The condition (A4) is commonly required to ensure the existence of interval censoring and (A6)(b) is simply for the identification of the model.

## A.1 Proof of Theorem 3.1

To calculate the information for $\beta$. Let $l(\beta,\Lambda)$ be the log-likelihood function for a sample size $n=1$. Define functions $A_i$, $i=1,2,3,4$, by

$$A_1(x,u,v,z)=\frac{\exp(-\Lambda(x)-\beta^T zx)}{\exp(-\Lambda(x)-\beta^T zx)-\exp(-\Lambda(u)-\beta^T zu)},$$
$$A_2(x,u,v,z)=\frac{\exp(-\Lambda(u)-\beta^T zu)}{\exp(-\Lambda(x)-\beta^T zx)-\exp(-\Lambda(u)-\beta^T zu)},$$
$$A_3(x,u,v,z)=\frac{\exp(-\Lambda(u)-\beta^T zu)}{\exp(-\Lambda(u)-\beta^T zu)-\exp(-\Lambda(v)-\beta^T zv)},$$

and

$$A_4(x,u,v,z)=\frac{\exp(-\Lambda(v)-\beta^T zv)}{\exp(-\Lambda(u)-\beta^T zu)-\exp(-\Lambda(v)-\beta^T zv)}.$$

By conditions (A3) and (A4), $A_1$, $A_2$, $A_3$ and $A_4$ are positive functions of $(x,u,v,z)$. Let $y=(\delta_1,\delta_2,x,u,v,z)$. Then the score function for $\beta$ is

$$\begin{aligned}\dot{l}_\beta(y)=\frac{\partial}{\partial\beta}l(y;\beta,\Lambda)&=z\{\delta_1(-xA_1+uA_2)\\&+\delta_2(-uA_3+vA_4)-(1-\delta_1-\delta_2)v+x\}.\end{aligned}$$

The score operator for $\Lambda$ is

$$\begin{aligned}\dot{l}_\Lambda\phi(y)&=\frac{\partial}{\partial s}l(y;\beta,\Lambda+s\phi)\mid_{s=0}\\&=\delta_1(-\phi(x)A_1+\phi(u)A_2)+\delta_2(-\phi(u)A_3+\phi(v)A_4)\\&\quad-(1-\delta_1-\delta_2)\phi(v)+\phi(x).\end{aligned}$$

Let $F$ is the distribution corresponding to $\Lambda$ and $P$ is the joint probability measure of $(\delta_1,\delta_2,X,U,V,Z)$, then the score operator $\dot{l}_\Lambda$ maps $L_2^0(F)$ to $L_2^0(P)$, where $L_2^0(F)\equiv\{a:\int a dF=0 \quad\text{and}\quad \int a^2dF<\infty\}$, and $L_2^0(P)$ is defined similarly as $L_2^0(F)$. Let $\dot{l}_\Lambda^T$: $L_2^0(P)\to L_2^0(F)$ be the adjoint operator of $\dot{l}_\Lambda$, i.e., for any $a\in L_2^0(F)$ and $b\in L_2^0(P)$,

$$\langle b,\dot{l}_\Lambda a\rangle_P\ =\ \langle\dot{l}_\Lambda^T b,a\rangle_F\,,$$

where $\langle\cdot,\cdot\rangle_P$ and $\langle\cdot,\cdot\rangle_F$ are the inner products in $L_2^0(P)$ and $L_2^0(F)$, respectively. We need to find $\phi^*$ such that $\dot{l}_\beta-\dot{l}_\Lambda\phi^*$ is orthogonal to $\dot{l}_\Lambda\phi$ in $L_2^0(P)$. This amounts to solving the following normal equation:

$$\text{(A1)}\qquad \dot{l}_\Lambda^T\dot{l}_\Lambda\phi^*\ =\ \dot{l}_\Lambda^T\dot{l}_\beta.$$

First note that we have

$$\begin{aligned}\dot{l}_\Lambda^T\dot{l}_\Lambda\phi(t)\ &=\ E[\dot{l}_\Lambda\phi(Y)|T\ =\ t]\\&=\ E_ZE[\dot{l}_\Lambda\phi(Y)|T\ =\ t,Z\ =\ z]\end{aligned}$$

by Groeneboom and Wellner (1992), pages 8–9, or Bickel et al. (1993), pages 271–272. Also by conditions (A1) and (A4),

$$\begin{aligned}E[\dot{l}_\Lambda\phi(Y)|T\ =\ t,Z\ =\ z]\ &=\ \int_{x=\tau_0}^{t}\int_{u=t}^{\tau_1}\int_{v=u+\eta}^{\tau_1}[-\phi(x)A_1\\&+\phi(u)A_2]g(x,u,v|z)1_{[u-x\geq\eta]}dvdudx+\int_{x=\tau_0}^{t}\int_{u=x+\eta}^{t}\int_{v=t}^{\tau_1}\\&[-\phi(u)A_3+\phi(v)A_4]g(x,u,v|z)1_{[v-u\geq\eta]}dvdudx\\&-\int_{x=\tau_0}^{t}\int_{u=x+\eta}^{t}\int_{v=u+\eta}^{t}\phi(v)g(x,u,v|z)dvdudx\\&+\int_{x=\tau_0}^{\tau_1}\int_{u=x+\eta}^{\tau_1}\int_{v=u+\eta}^{\tau_1}\phi(x)g(x,u,v|z)dvdudx,\end{aligned}$$

where $g(x,u,v|z)$ is the conditional density of $(X,U,V)$ given $Z$. Let

$$\begin{aligned}B_1(x,u,v)\ &=\ E_Z[A_1(x,u,v,Z)g(x,u,v|Z)],\\B_2(x,u,v)\ &=\ E_Z[A_2(x,u,v,Z)g(x,u,v|Z)],\\B_3(x,u,v)\ &=\ E_Z[A_3(x,u,v,Z)g(x,u,v|Z)],\\B_4(x,u,v)\ &=\ E_Z[A_4(x,u,v,Z)g(x,u,v|Z)],\end{aligned}$$

and $B_5(x,u,v)\ =\ E_Z[g(x,u,v|Z)]$. By the definition of $A$'s, $B$'s are all positive functions, and

$$\begin{aligned}&\text{(A2)}\quad B_1(x,u,v)\ =\ B_2(x,u,v)+B_5(x,u,v),\\&\text{(A3)}\quad B_3(x,u,v)\ =\ B_4(x,u,v)+B_5(x,u,v),\\&\text{(A4)}\quad B_1(x,u,v)+B_4(x,u,v)\ =\ B_2(x,u,v)+B_3(x,u,v).\end{aligned}$$

We calculate

$$\begin{aligned}L(t)\ \equiv\ \dot{l}_\Lambda^T\dot{l}_\Lambda\phi(t)\ &=\ \int_{x=\tau_0}^{t}\int_{u=t}^{\tau_1}\int_{v=u+\eta}^{\tau_1}[-\phi(x)B_1(x,u,v)\\&+\phi(u)B_2(x,u,v)]1_{[u-x\geq\eta]}dvdudx+\int_{x=\tau_0}^{t}\int_{u=x+\eta}^{t}\int_{v=t}^{\tau_1}\\&[-\phi(u)B_3(x,u,v)+\phi(v)B_4(x,u,v)]1_{[v-u\geq\eta]}dvdudx\\&-\int_{x=\tau_0}^{t}\int_{u=x+\eta}^{t}\int_{v=u+\eta}^{t}\phi(v)B_5(x,u,v)dvdudx\\&+\int_{x=\tau_0}^{\tau_1}\int_{u=x+\eta}^{\tau_1}\int_{v=u+\eta}^{\tau_1}\phi(x)B_5(x,u,v)dvdudx.\end{aligned}$$

Let

$$\begin{aligned}C_1(x,u,v)\ &=\ E_Z[ZA_1(x,u,v,Z)g(x,u,v|Z)],\\C_2(x,u,v)\ &=\ E_Z[ZA_2(x,u,v,Z)g(x,u,v|Z)],\\C_3(x,u,v)\ &=\ E_Z[ZA_3(x,u,v,Z)g(x,u,v|Z)],\\C_4(x,u,v)\ &=\ E_Z[ZA_4(x,u,v,Z)g(x,u,v|Z)],\end{aligned}$$

and $C_5(x,u,v)\ =\ E_Z[Zg(x,u,v|Z)]$. Then further calculation yields

$$\begin{aligned}R(t)\ \equiv\ \dot{l}_\Lambda^T\dot{l}_\beta(t)\ &=\ \int_{x=\tau_0}^{t}\int_{u=t}^{\tau_1}\int_{v=u+\eta}^{\tau_1}[-xC_1(x,u,v)\\&+uC_2(x,u,v)]1_{[u-x\geq\eta]}dvdudx+\int_{x=\tau_0}^{t}\int_{u=x+\eta}^{t}\int_{v=t}^{\tau_1}\\&[-UC_3(x,u,v)+VC_4(x,u,v)]1_{[v-u\geq\eta]}dvdudx\end{aligned}$$

$$- \int_{x=\tau_0}^{t} \int_{u=x+\eta}^{t} \int_{v=u+\eta}^{t} VC_5(x,u,v)dvdudx$$
$$+ \int_{x=\tau_0}^{\tau_1} \int_{u=x+\eta}^{\tau_1} \int_{v=u+\eta}^{\tau_1} XC_5(x,u,v)dvdudx.$$

Let

$$\begin{aligned} b(t) \;=\; & \int_{u=t+\eta}^{\tau_1} \int_{v=u+\eta}^{\tau_1} B_1(t,u,v)dvdu \\ & + \int_{x=\tau_0}^{t-\eta} \int_{v=t+\eta}^{\tau_1} B_2(x,t,v)dvdx \\ & + \int_{x=\tau_0}^{t-\eta} \int_{v=t+\eta}^{\tau_1} B_3(x,t,v)dvdx \\ & + \int_{x=\tau_0}^{t-2\eta} \int_{u=x+\eta}^{t-\eta} B_4(x,u,t)dudx \\ & + \int_{x=\tau_0}^{t-2\eta} \int_{u=x+\eta}^{t-\eta} B_5(x,u,t)dudx. \end{aligned}$$

After some straightforward calculations, the derivative of $L(t)$ is

$$\begin{aligned} L'(t) = & -b(t)\phi(t) + \int_{u=t+\eta}^{\tau_1} \int_{v=u+\eta}^{\tau_1} \phi(u)B_2(t,u,v)dvdu \\ & + \int_{x=\tau_0}^{t-\eta} \int_{v=t+\eta}^{\tau_1} \phi(x)B_1(x,t,v)dvdx \\ & + \int_{x=\tau_0}^{t-\eta} \int_{v=t+\eta}^{\tau_1} \phi(v)B_4(x,t,v)dvdx \\ & + \int_{x=\tau_0}^{t-2\eta} \int_{u=x+\eta}^{t-\eta} \phi(u)B_3(x,u,t)dudx. \end{aligned}$$

Similarly, let

$$\begin{aligned} c(t) \;=\; & \int_{u=t+\eta}^{\tau_1} \int_{v=u+\eta}^{\tau_1} C_1(t,u,v)dvdu \\ & + \int_{x=\tau_0}^{t-\eta} \int_{v=t+\eta}^{\tau_1} C_2(x,t,v)dvdx \\ & + \int_{x=\tau_0}^{t-\eta} \int_{v=t+\eta}^{\tau_1} C_3(x,t,v)dvdx \\ & + \int_{x=\tau_0}^{t-2\eta} \int_{u=x+\eta}^{t-\eta} C_4(x,u,t)dudx \\ & + \int_{x=\tau_0}^{t-2\eta} \int_{u=x+\eta}^{t-\eta} C_5(x,u,t)dudx. \end{aligned}$$

The derivative of $R(t)$

$$\begin{aligned} r(t) \equiv R'(t) \;=\; & -c(t)t + \int_{u=t+\eta}^{\tau_1} \int_{v=u+\eta}^{\tau_1} uC_2(t,u,v)dvdu \\ & + \int_{x=\tau_0}^{t-\eta} \int_{v=t+\eta}^{\tau_1} xC_1(x,t,v)dvdx \\ & + \int_{x=\tau_0}^{t-\eta} \int_{v=t+\eta}^{\tau_1} vC_4(x,t,v)dvdx \\ & + \int_{x=\tau_0}^{t-2\eta} \int_{u=x+\eta}^{t-\eta} uC_3(x,u,t)dudx. \end{aligned}$$

By conditions (A3)–(A7), $r$ has a bounded derivative $r'$ on $[\tau_0, \tau_1]$. So equation (A1) reduces to

(A5)
$$\begin{aligned} & -b(t)\phi(t) + \int_{u=t+\eta}^{\tau_1} \int_{v=u+\eta}^{\tau_1} \phi(u)B_2(t,u,v)dvdu \\ & + \int_{x=\tau_0}^{t-\eta} \int_{v=t+\eta}^{\tau_1} \phi(x)B_1(x,t,v)dvdx \\ & + \int_{x=\tau_0}^{t-\eta} \int_{v=t+\eta}^{\tau_1} \phi(v)B_4(x,t,v)dvdx \\ & + \int_{x=\tau_0}^{t-2\eta} \int_{u=x+\eta}^{t-\eta} \phi(u)B_3(x,u,t)dudx \;=\; r(t). \end{aligned}$$

By conditions (A3) and (A4), we have $\inf_{\tau_0 \le t \le \tau_1} b(t) > 0$. Let $d(t) = -r(t)/b(t)$ and

$$\begin{aligned} K(t,w) = ( & \int_{w+\eta}^{\tau_1} B_2(t,w,v)1_[t+\eta \le w \le \tau_1]dv \\ & + \int_{t+\eta}^{\tau_1} B_1(w,t,v)1_[\tau_0 \le w \le t-\eta]dv \\ & + \int_{\tau_0}^{t-\eta} B_4(x,t,w)1_[t+\eta \le w \le \tau_1]dx \\ & + \int_{\tau_0}^{t-2\eta} B_3(x,w,t)1_[x+\eta \le w \le t-\eta]dx)/b(t). \end{aligned}$$

Then we obtain a Fredholm integral equation of the second kind,

$$\phi^*(t) - \int K(t,w)\phi^*(w)dw \;=\; d(t).$$

Furthermore, note the facts that $K(t,w)$ is a bounded kernel ($K$ being a $L_2$ kernel suffices) and if $\phi(t) - \int K(t,w)\phi(w)dw = 0$, then we have $\phi(t) \equiv 0$ on $[\tau_0, \tau_1]$. By these facts and the classical results on Fredholm integral equations (Kanwal (1971), Sections 4.2 and 4.3; or Kress (1989), Chapter 4), there exists a resolvent $\Gamma(t,w)$ (completely determined by $K$) such that

(A6)
$$\phi^*(t) \;=\; d(t) + \int \Gamma(t,w)d(w)dw.$$

From equations (A5) and (A6), we can derive properties of $\phi^*$. By (A6), $\phi^*$ is bounded on $[\tau_0, \tau_1]$. By (A5), this implies $\phi^*$ is continuous. This in turn implies $\phi^*$ is differentiable. Since $b$ is bounded away from zero, the partial derivative of each component of $K$ with respect to t is bounded on its integral interval and the derivative of $r$ is bounded. It follows that the derivative of $\phi^*$ is bounded. This completes the proof.

## A.2 Proof of Theorem 3.2

Before proving Theorem 3.2., define a semi distance $\rho$ on $\Theta$ by

$$\rho^2(\theta,\theta_0) \;=\; E\{(\beta-\beta_0)^T\dot{l}_\beta(Y)+\dot{l}_\Lambda(Y)[\Lambda-\Lambda_0]\}^2,$$

where $\dot{l}_\beta$ is the score function for $\beta$, $\dot{l}_\Lambda$ is the score function for $\Lambda$, and both are evaluated at the true parameter value $(\beta_0,\Lambda_0)$. Let $\phi^*$ be the least favorable direction calculated above. So $\rho^2(\theta,\theta_0) = (\beta-\beta_0)^T I(\beta_0)(\beta-\beta_0) + E\{(\beta-\beta_0)^T\dot{l}_\Lambda(Y)[\phi^*]+\dot{l}_\Lambda(Y)[\Lambda-\Lambda_0]\}^2$, where $I(\beta_0)$ is the information matrix for $\beta_0$. Because $I(\beta_0)$ is positive definite, $\dot{l}_\beta$ and $\dot{l}_\Lambda$ are bounded away from 0 and $\infty$, this implies that $d(\theta,\theta_0)=O(\rho(\theta,\theta_0))$.

We first prove the consistency of $\widehat{\theta}_n$, and then the convergence rate. Let $p(Y;\widehat{\theta}_n) = p(Y;\widehat{\beta}_n,\widehat{\Lambda}_n)$, $p(Y;\theta_{0n}) = p(Y;\beta_0,\Lambda_{0n})$ and $p(Y;\theta_0)=p(Y;\beta_0,\Lambda_0)$, where $p$ is defined as followed,

$$\begin{aligned}&p(y;\beta,\Lambda)\\ &=\big\{\exp\{-\Lambda(x)-\beta^T zx\}-\exp\{-\Lambda(u)-\beta^T zu\}\big\}^{\delta_1}\\ &\quad\big\{\exp\{-\Lambda(u)-\beta^T zu\}-\exp\{-\Lambda(v)-\beta^T zv\}\big\}^{\delta_2}\\ &\quad\big\{\exp\{-\Lambda(v)-\beta^T zv\}\big\}^{1-\delta_1-\delta_2}/\exp\{-\Lambda(x)-\beta^T zx\}.\end{aligned}$$

Also let $\theta_{0n}=(\beta_0,\Lambda_{0n})$ be the projection of the true parameter $\theta_0\in\Theta$ into $\Theta_n$. Since $(\widehat{\beta}_n,\widehat{\Lambda}_n)$ maximizes the likelihood function over $\Theta_n$, and $(\beta_0,\Lambda_{0n})\in\Theta_n$, thus we have

$$\sum_{i=1}^n \log p(Y_i;\widehat{\theta}_n)\ge\sum_{i=1}^n\log p(Y_i;\theta_{0n})\,,\;\sum_{i=1}^n\log\frac{p(Y_i;\widehat{\theta}_n)}{p(Y_i;\theta_{0n})}\ge 0\,.$$

By the concavity of the log function, for any $0<\alpha<1$,

(A7)
$$\frac{1}{n}\sum_{i=1}^n\log\{1-\alpha+\alpha\frac{p(Y_i;\widehat{\theta}_n)}{p(Y_i;\theta_{0n})}\}\ge\frac{1}{n}\sum_{i=1}^n\alpha\log\frac{p(Y_i;\widehat{\theta}_n)}{p(Y_i;\theta_{0n})}\ge 0.$$

The left-hand side can be written as

$$\text{(A8)}\qquad \int\log\{1-\alpha+\alpha\frac{p(y;\widehat{\theta}_n)}{p(y;\theta_{0n})}\}d(P_n-P)(y) + \int\log\{1-\alpha+\alpha\frac{p(y;\widehat{\theta}_n)}{p(y;\theta_{0n})}\}dP(y),$$

where $P_n$ is the empirical measure of $(\delta_{1i},\delta_{2i},X_i,U_i,V_i,Z_i)$, $i=1,\ldots,n$; $P$ is the joint probability measure of $(\delta_1,\delta_2,X,U,V,Z)$.

Note that the class of functions $\{\log(1-\alpha+\alpha p(y;\theta)/p(y;\theta_{0n}))\;:\;\theta=(\beta,\Lambda)\in\Theta_n\}$ is uniformly bounded and uniformly Lipschitz of order 1, which is thus Glivenko−Cantelli class. This leads

$$\text{(A9)}\qquad \int\log\{1-\alpha+\alpha\frac{p(y;\widehat{\theta}_n)}{p(y;\theta_{0n})}\}d(P_n-P)(y)\to 0$$

in probability. By Jensen's Inequality,

$$\begin{aligned}\text{(A10)}\qquad &\int\log\{1-\alpha+\alpha\frac{p(y;\widehat{\theta}_n)}{p(y;\theta_{0n})}\}dP(y)\\ &\le\log\{\int 1-\alpha+\alpha\frac{p(y;\widehat{\theta}_n)}{p(y;\theta_{0n})}dP(y)\}\\ &=\log\{1-\alpha+\alpha\int\frac{p(y;\widehat{\theta}_n)}{p(y;\theta_{0n})}dP(y)\}=o_p(1).\end{aligned}$$

Combing (A7)–(A10), we obtain that

$$\int\log\{1-\alpha+\alpha\frac{p(y;\widehat{\theta}_n)}{p(y;\theta_{0n})}\}dP(y)=o_p(1).$$

It follows that $p(y;\widehat{\theta}_n) = p(y;\theta_{0n})+o_p(1)$. This implies $d(\widehat{\theta}_n,\theta_{0n})\to 0$ in probability. Since $\Lambda_{0n}\in H_n$, condition (A5) and (A7) immediately yields that $d(\theta_0,\theta_{0n})\to 0$ in probability. By the triangle inequality, we can obtain that $d(\widehat{\theta}_n,\theta_0)\to 0$ in probability.

Now we give the proof of the convergence rate. Let $\varepsilon$ be a small and fixed positive number. Define a function class $\Psi(\varepsilon)=\{l(y;\theta)\,:\,d(\theta,\theta_{0n})\le\varepsilon,\theta\in\Theta_n\}$. It is obvious that the log-likelihood function satisfies the Lipschitz conditions with respect to its conditions. Then, for all $0<\xi<\varepsilon$, $\log N_{[]}(\xi,\Psi,d)\le A_1(\varepsilon/\xi+q_n\log(\varepsilon/\xi))$, where $\log N_{[]}(\xi,\Psi,d)$ is the bracketing number and $A_1$ is a constant; see van der Vaart and Wellner (1996). Let $J_{[]}(\varepsilon,\Psi,d) = \int_0^\varepsilon\{1+\log N_{[]}(\xi,\Psi,d)\}^{1/2}d\xi$ be the integral entropy. It follows that $J_{[]}(\varepsilon,\Psi,d)\le A_2 q_n^{1/2}\varepsilon$, where $A_2$ is a constant. Then together with Lemma 3.4.2 in van der Vaart and Wellner (1996), for sufficiently large $n$,

$$E^*[\sup_{\varepsilon/2\le d(\theta,\theta_{0n})\le\varepsilon,\theta\in\Theta_n}|\sqrt{n}(P_n-P)(l(y;\theta)-l(y;\theta_{0n}))|]\le O(1)\varepsilon q_n^{1/2},$$

where $E^*$ is the outer expectation.

By triangle inequality, one can derive that

$$E|l(Y;\theta)-l(Y;\theta_{0n})|\le O(1)(d^2(\theta,\theta_{0n})+2d^2(\theta_{0n},\theta_0)),$$

and thus

$$\sup_{\varepsilon/2\le d(\theta,\theta_{0n})\le\varepsilon,\theta\in\Theta_n}(El(Y;\theta)-El(Y;\theta_{0n}))\le O(1)\varepsilon^2,$$

for sufficiently large n. Let $\phi(\varepsilon)=\varepsilon q_n^{1/2}$ and $r_n=(n/q_n)^{1/2}$. Furthermore, $\theta_n$ is consistent, so we can verify that the conditions of Lemma 3.4.1 in van der Vaart and Wellner (1996) are all satisfied. It then follows that $d(\widehat{\theta}_n,\theta_{0n}) = O_p\{(n/q_n)^{-1/2}\}=O_p(n^{-(1-\kappa)/2})$.

Since $\Lambda_{0n}\in H_n$, condition (A5) yields that

$$\int_0^\tau|\Lambda_0(t)-\Lambda_{0n}(t)|^2dt=O(n^{-2r\kappa})\,,$$

$$E\{l_n(y;\theta_0) - l_n(y;\theta_{0n})\} = O(n^{-2r\kappa}),$$

which, together with (A7), immediately yields that $d(\theta_0, \theta_{0n}) = O(n^{-r\kappa})$. By triangle inequality, we can obtain that $d(\widehat{\theta}_n, \theta_0) = O_p(\max\{n^{-(1-\kappa)/2}, n^{-r\kappa}\})$.

## A.3 Proof of Theorem 3.3

For the proof, we will employ Theorem 6.1 of Huang(1996). Because $\widehat{\beta}_n$ maximizes $l_n(\beta, \widehat{\Lambda}_n)$, we have

$$S_{1n}(\widehat{\beta}_n, \widehat{\Lambda}_n) \equiv \frac{1}{n}\sum_{i=1}^n \dot{l}_\beta(Y_i; \widehat{\beta}_n, \widehat{\Lambda}_n) = 0$$

with probability one for sufficiently large $n$. The score function for the nonparametric component evaluated at $(\widehat{\beta}_n, \widehat{\Lambda}_n)$ along the least favorable direction $\phi^*$ will generally not be 0, because maximization is carried out with monotonicity constraints. However, it suffices to show that this score function (A11)

$$S_{2n}(\widehat{\beta}_n, \widehat{\Lambda}_n)[\phi^*] \equiv \frac{1}{n}\sum_{i=1}^n \dot{l}_\Lambda(Y_i; \widehat{\beta}_n, \widehat{\Lambda}_n)[\phi^*] = o_p(n^{-1/2}),$$

where $\dot{l}_\Lambda(y;\beta,\Lambda)[\phi^*] = \dot{l}_\Lambda\phi^*(y)$ and $\dot{l}_\Lambda\phi^*(y)$ is defined in the information calculation.

The proof of (A11) goes as follows. Construct a step function $\phi_n^*$ so it has jump points the same as the points where $\widehat{\Lambda}_n$ changes slopes and also approximates $\phi^*$ with precision

$$\int_{\tau_0}^{\tau_1} |\phi_n^* - \phi^*|(t)dt \le O(\|\widehat{\Lambda}_n - \Lambda_0\|_2).$$

Following from the similar arguments in the proof of Theorem 5.3 in Huang and Rossini(1997), we have $S_{2n}(\widehat{\beta}_n, \widehat{\Lambda}_n)[\phi_n^*] \equiv o_p(n^{-1/2})$. To show (A11), it suffices to show that $S_{2n}(\widehat{\beta}_n, \widehat{\Lambda}_n)[\phi^*] - S_{2n}(\widehat{\beta}_n, \widehat{\Lambda}_n)[\phi_n^*] \equiv o_p(n^{-1/2})$. Note that $P\{\dot{l}_\Lambda(y;\beta_0,\Lambda_0)[\phi^* - \phi_n^*]\} = 0$, we obtain that

$$\begin{aligned}
&S_{2n}(\widehat{\beta}_n, \widehat{\Lambda}_n)[\phi^*] - S_{2n}(\widehat{\beta}_n, \widehat{\Lambda}_n)[\phi_n^*] \\
&\quad = P_n\dot{l}_\Lambda(y; \widehat{\beta}_n, \widehat{\Lambda}_n)[\phi^* - \phi_n^*] \\
&\quad = (P_n - P)\dot{l}_\Lambda(y; \widehat{\beta}_n, \widehat{\Lambda}_n)[\phi^* - \phi_n^*] \\
&\qquad + P\{\dot{l}_\Lambda(y; \widehat{\beta}_n, \widehat{\Lambda}_n)[\phi^* - \phi_n^*] - \dot{l}_\Lambda(y;\beta_0,\Lambda_0)[\phi^* - \phi_n^*]\}.
\end{aligned}$$

The first term in the last line is $o_p(n^{-1/2})$ by uniform asymptotic equicontinuity of empirical processes indexed by a Donsker class of functions. By Theorem 3.2. and the Cauchy-Schwartz inequality, the second term is

$$O_p(\max\{n^{-(1-\kappa)/2}, n^{-r\kappa}\})^2 = O_p(\max\{n^{-(1-\kappa)}, n^{-2r\kappa}\}),$$

where $r = 1, 2$. So if we choose $1/4r < \kappa < 1/2$, then the second term is $o_p(n^{-1/2})$.

By Theorem 3.1., the Fisher information matrix for $\beta$ is positive definite. By Theorem 3.2., the rate of convergence is proved. And there are two more facts. Define $S_1$ and $S_2$ be the limits of $S_{1n}$ and $S_{2n}$. One is the uniform asymptotic equicontinuity of $S_{1n}(\beta,\Lambda) - S_1(\beta,\Lambda)$ and $S_{2n}(\beta,\Lambda)[\phi^*] - S_2(\beta,\Lambda)[\phi^*]$ in a small neighborhood of $(\beta_0, \Lambda_0)$, and this follows from uniform asymptotic equicontinuity of empirical processes indexed by a Donsker class of functions. The other is the smoothness of $S_1(\beta,\Lambda)$ and $S_2(\beta,\Lambda)[\phi^*]$ in a small neighborhood of $(\beta_0,\Lambda_0)$, which follows from the Taylor expansion. Thus all conditions in Theorem 6.1 of Huang(1996) have been confirmed and the proof is complete.

*Received 27 June 2014*

## REFERENCES


Bickel, P. J., Klaassen, C. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore. MR1245941

Chen, D., Sun, J. and Peace, K. E. (2012). *Interval-Censored Time-to-Event Data: Methods and Applications.* Chapman & Hall/CRC. MR3053014

DeGruttola, V. G. and Lagakos S. W. (1989). Analysis of doubly censored survival data, with application to AIDS. *Biometrics* **45**, 1–11. MR0999438

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42**, 845–854. MR0872963

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* DMV Seminar Band 19, Birkhäuser, Basel. MR1180321

Ghosh, D. (2003). Goodness-of-fit methods for additive-risk models in tumorigenicity experiments. *Biometrics* **59**, 721–726. MR2004278

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics* **24**, 540–568. MR1394975

Huang, J. and Rossini, A. J. (1997). Sieve estimation for the proportional odds failure time regression model with interval censoring. *Journal of the American Statistical Association* **92**, 960–967. MR1482126

Huang, J. and Wellner, J. A. (1996). Regression models with interval censoring. *Probability Theory and Mathematical Statistics (St. Petersburg, 1993),* 269–296. MR1661714

Jewell, N. P. and van der Laan, M. J. (2004a). Current status data: review, recent developments and open problems. *Advances in Survival Analysis*, Elsevier, Amsterdam, 625–642. MR2065792

Jewell, N. P. and van der Laan, M. J. (2004b). Case-control current status data. *Biometrika* **91**, 529–541. MR2090620

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data.* Wiley, New York, 2nd Edition. MR1924807

Kanwal, R. P. (1971). *Linear Integral Equations.* Academic Press, New York.

Kim, J. S. (2003). Efficient estimation for the proportional hazards model with left-truncated and case I interval-censored data. *Statistica Sinica* **13**, 519–537. MR1977741

Kim, M. Y., DeGruttola, V. G. and Lagakos, S. W. (1993). Analyzing doubly censored data with covariates, with application to AIDS. *Biometrics* **49**, 13–22.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis*, Springer-Verlag, New York.

Kress, R. (1989). *Linear Integral Equations.* Springer-Verlag, Berlin. MR1007594

Kulich, M. and Lin, D. Y. (2000). Additive hazards regression with covariate measurement error. *Journal of the American Statistical Association* **95**, 238–248. MR1803152

Lin, D. Y., Oakes, D. and Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika* **85**, 289–298. MR1649115

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71. MR1279656

Murphy, S. A. and van der Vaart, A. W. (1999). Observed information in semi-parametric models. *Bernoulli* **5(3)**, 381–412. MR1693616

Pan, W. and Chappell, R. (1999). A note on inconsistency of NPMLE of the distribution function from left truncated and case I interval censored data. *Lifetime Data Analysis* **5**, 281–291. MR1749608

Pan, W. and Chappell, R. (2002). Estimation in the Cox proportional hazards model with left truncated and interval censored data. *Biometrics* **58**, 64–70. MR1891044

Shen, P. S. (2014). Proportional hazards regression with interval-censored and left-truncated data. *Journal of Statistical Computation and Simulation* **84**, 264–272. MR3169325

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data.* Springer-Verlag, New York. MR2287318

Turnbull, B. W. (1976). The empirical distribution with arbitrarily grouped censored and truncated data. *Journal of the Royal Statistical Society, Series B* **38**, 290–295. MR0652727

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer-Verlag: New York. MR1385671

Wang, L., Sun, J. and Tong X. (2010). Regression analysis of case II interval-censored failure time data with the additive hazards model. *Statistica Sinica* **20**, 1709–1723. MR2777342

Zeng, D., Cai, J. and Shen, Y. (2006). Semiparametric additive risks model for interval-censored data. *Statistica Sinica* **16**, 287–302. MR2256093

Zhang, M. and Davidian, M. (2008). Smooth semiparametric regression analysis for arbitrarily censored time-to-event data. *Biometrics* **64**, 567–576. MR2432427

Zhou, X. and Sun, L. (2003). Additive hazards regression with missing censoring information. *Statistica Sinica* **13**, 1237–1257. MR2026071

Peijie Wang
Institute of Mathematics
Jilin University
Changchun 130012
China
E-mail address: wangpj12@mails.jlu.edu.cn

Xingwei Tong
School of Mathematical Sciences
Beijing Normal University
Beijing 100875
China
E-mail address: xweitong@bnu.edu.cn

Shishun Zhao
Institute of Mathematics
Jilin University
Changchun 130012
China
E-mail address: zhaoss@jlu.edu.cn

Jianguo Sun
Department of Statistics
University of Missouri
Columbia, Missouri 65211
USA
E-mail address: sunj@missouri.edu

Institute of Mathematics
Jilin University
Changchun 130012
China