Statistics and Its Interface Volume 16 (2023) 553–563

A pairwise pseudo-likelihood approach for the additive hazards model with left-truncated and interval-censored data

Peijie Wang*, Yichen Lou, and Jianguo Sun

Left-truncated and interval-censored data occur commonly and some approaches have been proposed in the literature for their analysis. However, most of the existing methods are based on the conditional likelihood given left-truncation times, which can be inefficient since the information in the marginal likelihood of the truncation times is ignored. To address this, in this paper, a pairwise pseudo-likelihood augmented estimation approach is proposed under the additive hazards model that can fully make use of all available information. The derived estimator is shown to be consistent and asymptotically normal, and simulation studies suggest that the proposed method works well and provides a substantial efficiency gain over the conditional approach. In addition, the method is applied to a set of real data arising from an AIDS cohort study.

Keywords and phrases: Additive hazards model, Bootstrap, Interval-censored data, Left truncation, Pairwise pseudo-likelihood augmented estimation.

1. INTRODUCTION

The analysis of left-truncated and interval-censored failure time data has recently attracted much attention due to their general structure and common occurrence in many areas such as demographical, financial and medical studies. By interval censoring [4, 23], we usually mean that the failure time of interest is observed only to belong to a window or an interval instead of being observed exactly or right-censored. A typical example of interval-censored data is given in a health or medical follow-up study, which commonly examines the status of a disease or medical condition of the study subject from time to time. It is apparent that for the situation, the status of a subject may change between follow-up times and also subjects may miss the scheduled follow-ups, yielding interval-censored observations.

In addition to interval censoring, sometimes one may also have to deal with left truncation that usually occurs due to some sampling biases. One such case is that sometimes only the subjects who satisfy certain conditions or experience some initial events can be included in a study. In prevalent cohort studies, for example, they usually only include the individuals who have survived until the enrollment. A more specific example of left-truncated and interval-censored data that motivated this investigation is given by an AIDS cohort study that will be discussed in details below.

Many authors have investigated regression analysis of right-censored or interval-censored failure time data. The most common and popular models are the proportional hazards model and its generalization. Some early and representative references on these include [1, 5, 7] on right-censored data and [11, 13, 30] on interval-censored data. However, it is well-known that sometimes one may be interested in the additive or excess risk instead of the relative risk. For the situation, the additive hazards model has been commonly used and discussed by many authors [18, 19, 29].

For the analysis of left-truncated and right-censored or interval-censored data, some approaches have been proposed in the literature. However, most of the existing approaches are based on the conditional likelihood given the left-truncation times. For example, [24] discussed regression analysis of left-truncated and right-censored data under the proportional hazards model, while [3] considered the same problem under the semiparametric transformation model. [15] and [21] both considered regression analysis of left-truncated and interval censored data under the proportional hazards model. [26] investigated the same problem with data arising from the additive hazards model.

Note that for all conditional methods, they only make use of partial information on truncation times. To address this and improve the efficiency, some efforts have been made to recover some missing information in the conditional methods. Among others, for example, [14] and [27] developed the pairwise pseudo-likelihood augmented estimation approaches for right-censored data under the additive and proportional hazards model, respectively. [25] considered interval censoring instead of right censoring under the proportional hazards model. [12] developed an EM algorithm through introducing two layers of data augmentation for length-biased interval-censored data, a special case of left-truncated and interval-censored data. However, there does not seem to exist similar work for left-truncated and interval-censored data under the additive hazards model, the focus of this paper.

*Corresponding author.

In this paper, we propose to augment the conditional likelihood with a pairwise likelihood constructed from the marginal likelihood of the truncation times. First we will begin in Section 2 with introducing some notation and describing the assumed models as well as the structure of the observed data. The resulting likelihood function is then presented. In Section 3, the pairwise pseudo-likelihood augmented estimation procedure is proposed for inference and the asymptotic properties of the estimator are established. Section 4 presents some results obtained from an extensive simulation study, which indicate that the proposed method seems to work well for practical situations and provides a substantial efficiency gain over the conditional approach. In Section 5, the proposed approach is applied to a set of real data arising from the AIDS cohort study, and Section 6 contains some discussions and concluding remarks.

2. NOTATION AND MODELS

Consider a failure time study involving a disease and an event of interest and for a patient from the target population, let T^* be the underlying survival time, measuring the time from the disease onset to the event. Suppose that there exists an independent truncation time, denoted as A^*, measuring the time from the disease onset to the study enrollment. In a prevalent cohort, this means that we observe the pairs (A^*, T^*) such that the events happen after the enrollment or we only have realizations from $(A, T) \equiv (A^*, T^*)|A^* \leq T^*$. Note that the above sampling scheme induces a positive correlation between A and T in the observed biased sample. Let Z be a p-dimension vector of covariates for the subject involved in the study. In the following, it will be assumed that for each study subject, there exist two observation times U and V with $U < V$ and the observed data have the form $(A, U, V, \delta_1 = I(T \leq U), \delta_2 = I(U < T \leq V), \delta_3 = I(T > V), Z)$, where $\delta_1 + \delta_2 + \delta_3 = 1$. That is, only interval-censored data are available.

Let f and S denote the density and survival functions of T^* and g the density function of A^*. Then the joint density function $h(t, a)$ of (T, A) evaluated at (t, a) can be expressed as

$$h(t,a) = \frac{f(t)g(a)}{\int_0^\infty S(u)g(u)du}, \qquad 0 \leq a \leq t.$$

To describe the covariate effects, we assume that given Z, the hazard function of T^* 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 β is a vector of regression parameters. That is, the survival time of interest T^* follows the additive hazards model ([18, 19]). Let $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$, the baseline cumulative hazard function. Then the survival function of T^* is given by $S(t|Z) = \exp\{-\Lambda_0(t) - \beta^T Zt\}$.

Suppose that the study consists of n independent subjects and give the observed data

$$O = \{(A_i, U_i, V_i, \delta_{1i}, \delta_{2i}, \delta_{3i}, Z_i), i = 1, \ldots, n\}.$$

Denote $(U, V) = (A, A) + (U_0, V_0)$ and assume that (U_0, V_0) is independent of (A, T) given Z. Then the fully likelihood of the observed data is proportional to

$$\begin{aligned}
L_n &= \prod_{i=1}^n \Big\{ \{S(A_i|Z_i) - S(U_i|Z_i)\}^{\delta_{1i}} \{S(U_i|Z_i) - S(V_i|Z_i)\}^{\delta_{2i}} \\
&\qquad S(V_i|Z_i)^{\delta_{3i}} g(A_i) \Big\} \Big/ \Big\{ \int_0^\infty S(u|Z_i)g(u)du \Big\} \\
&= \prod_{i=1}^n \Big\{ \{S(A_i|Z_i) - S(U_i|Z_i)\}^{\delta_{1i}} \{S(U_i|Z_i) - S(V_i|Z_i)\}^{\delta_{2i}} \\
&\qquad S(V_i|Z_i)^{\delta_{3i}} \Big\} S(A_i|Z_i) \times \prod_{i=1}^n \frac{S(A_i|Z_i)g(A_i)}{\int_0^\infty S(u|Z_i)g(u)du} \\
&\equiv L_n^C \times L_n^M.
\end{aligned}$$

In the above, L_n^C represents the conditional likelihood function of $(U, V, \delta_1, \delta_2, \delta_3)$ given (A, Z) and L_n^M the marginal likelihood function of A given Z. Note that here it is assumed that g, the density function of A, does not depend on the covariates Z.

3. PAIRWISE PSEUDO-LIKELIHOOD AUGMENTED ESTIMATION PROCEDURE

As mentioned above, in the presence of left truncation, most of the existing methods are based on the conditional likelihood function L_n^C. A main drawback of this is that the resulting estimators can be less efficient since they completely ignore the information contained in L_n^M. To address this and improve the efficiency, we propose to supplement L_n^C with the information in L_n^M. Specifically, following [14] and [25], we first apply the pairwise pseudo-likelihood method by [17] to L_n^M in order to eliminate the nuisance function g, and then estimate β and $\Lambda_0(\cdot)$ based on a composite likelihood consisting of L_n^C and L_n^P, the pairwise pseudo-likelihood derived below.

Suppose that a sample $\{(A_i, Z_i), (A_j, Z_j) : i < j\}$ is available. Following the argument in [17], the pseudo-likelihood of the pair (i, j), conditional on (Z_i, Z_j) and the order statistic of (A_i, A_j), is given by

$$\begin{aligned}
&\Big\{ \frac{S(A_i|Z_i)g(A_i)}{\int_0^\infty S(u|Z_i)g(u)du} \times \frac{S(A_j|Z_j)g(A_j)}{\int_0^\infty S(u|Z_j)g(u)du} \Big\} \Big/ \\
&\quad \Big\{ \frac{S(A_i|Z_i)g(A_i)}{\int_0^\infty S(u|Z_i)g(u)du} \times \frac{S(A_j|Z_j)g(A_j)}{\int_0^\infty S(u|Z_j)g(u)du} + \\
&\quad \frac{S(A_i|Z_j)g(A_i)}{\int_0^\infty S(u|Z_j)g(u)du} \times \frac{S(A_j|Z_i)g(A_j)}{\int_0^\infty S(u|Z_i)g(u)du} \Big\} = \frac{1}{1 + R_{ij}(\beta)},
\end{aligned}$$

where $R_{ij}(\beta)$ denotes the generalized odds ratio and has the form

$$R_{ij}(\beta) = \frac{S(A_i|Z_j)S(A_j|Z_i)}{S(A_i|Z_i)S(A_j|Z_j)} = \exp\{\beta^T(Z_i - Z_j)(A_i - A_j)\}$$

under the additive hazards model. The pairwise pseudo-likelihood L_n^P of all pairs is then given by

$$L_n^P = \prod_{i<j} \frac{1}{1 + R_{ij}(\beta)}.$$

It is worth noting that L_n^P is a function of β only, not depending on Λ_0 nor g by canceling them out, whereas L_n^M is a function of (β, Λ_0, g).

Furthermore, [17] showed that the pairwise likelihood can retain the majority of the information in the likelihood from which it is derived, and that the efficiency loss may not be substantial, depending on the model as well as the values of the parameters by the simulation studies. Therefore, to estimate (β, Λ_0), we propose using L_n^P as a reasonable good surrogate for L_n^M in the full likelihood approach. The analogous idea has been exploited in the analysis of left-truncated data by [14, 25, 27].

To account for the different magnitudes of L_n^C and L_n^P, we propose to maximize the following composite log-likelihood function:

$$\begin{aligned} l_n(\beta, \Lambda_0) = \frac{1}{n}\sum_{i=1}^{n} &\{\delta_{1i}\log\{S(A_i|Z_i) - S(U_i|Z_i)\} \\ &+ \delta_{2i}\log\{S(U_i|Z_i) - S(V_i|Z_i)\} + \delta_{3i}\log\{S(V_i|Z_i)\} \\ &- \log\{S(A_i|Z_i)\}\} - \frac{2}{n(n-1)}\sum_{i<j}\log\{1 + R_{ij}(\beta)\}. \end{aligned}$$

It is easy to see that this is not an easy task since the composite log-likelihood function involves both finite-dimensional and infinite-dimensional parameters. To deal with this, following [31], we employ the sieve maximum likelihood estimation method and use the Bernstein polynomials-based function to approximate $\Lambda_0(\cdot)$.

Specifically, define the Bernstein polynomials-based function $\Lambda^*(t) = \sum_{k=0}^{m} \phi_k B_k(t, m, t_l, t_u)$, where

$$B_k(t, m, t_l, t_u) = \binom{m}{k}\left(\frac{t - t_l}{t_l - t_u}\right)^k \left(1 - \frac{t - t_l}{t_l - t_u}\right)^{m-k}$$

with $\sum_{0\le k\le m} |\phi_k| \le M_n$, $0 \le \phi_0 \le \phi_1 \le \cdots \le \phi_m$, $m = o(n^v)$ for some $v \in (0, 1)$ and $0 \le t_l < t_u < \infty$. In practice, $[t_l, t_u]$ is usually taken as the range of the observed times. Then it is nature to define the estimator $\widehat{\theta}_n = (\widehat{\beta}_n^T, \widehat{\Lambda}_n)^T$ by maximizing $l_n(\beta, \Lambda^*)$. Specifically, differentiating $l_n(\beta, \Lambda^*)$ with respect to (β, ϕ) yields the following score functions:

$$\begin{aligned} \frac{\partial l_n(\beta, \Lambda^*)}{\partial \beta} = \frac{1}{n}\sum_{i=1}^{n} &\{\delta_{1i}\frac{\frac{\partial S(A_i|Z_i)}{\partial \beta} - \frac{\partial S(U_i|Z_i)}{\partial \beta}}{S(A_i|Z_i) - S(U_i|Z_i)} + \\ &\delta_{2i}\frac{\frac{\partial S(U_i|Z_i)}{\partial \beta} - \frac{\partial S(V_i|Z_i)}{\partial \beta}}{S(U_i|Z_i) - S(V_i|Z_i)} + \delta_{3i}\frac{\frac{\partial S(V_i|Z_i)}{\partial \beta}}{S(V_i|Z_i)} - \frac{\frac{\partial S(A_i|Z_i)}{\partial \beta}}{S(A_i|Z_i)}\} \\ &- \frac{2}{n(n-1)}\sum_{i<j}\frac{\frac{\partial R_{ij}(\beta)}{\partial \beta}}{1 + R_{ij}(\beta)}, \end{aligned}$$

$$\begin{aligned} \frac{\partial l_n(\beta, \Lambda^*)}{\partial \phi_k} = \frac{1}{n}\sum_{i=1}^{n} &\{\delta_{1i}\frac{\frac{\partial S(A_i|Z_i)}{\partial \phi_k} - \frac{\partial S(U_i|Z_i)}{\partial \phi_k}}{S(A_i|Z_i) - S(U_i|Z_i)} + \\ &\delta_{2i}\frac{\frac{\partial S(U_i|Z_i)}{\partial \phi_k} - \frac{\partial S(V_i|Z_i)}{\partial \phi_k}}{S(U_i|Z_i) - S(V_i|Z_i)} + \delta_{3i}\frac{\frac{\partial S(V_i|Z_i)}{\partial \phi_k}}{S(V_i|Z_i)} - \frac{\frac{\partial S(A_i|Z_i)}{\partial \phi_k}}{S(A_i|Z_i)}\} \\ &- \frac{2}{n(n-1)}\sum_{i<j}\frac{\frac{\partial R_{ij}(\beta)}{\partial \phi_k}}{1 + R_{ij}(\beta)}, \end{aligned}$$

where

$$\begin{aligned} \frac{\partial S(t|Z)}{\partial \beta} &= S(t|Z)(-Zt), \\ \frac{\partial S(t|Z)}{\partial \phi_k} &= S(t|Z)(-B_k(t, m, t_l, t_u)), \\ \frac{\partial R_{ij}(\beta)}{\partial \beta} &= R_{ij}(\beta)(Z_i - Z_j)(A_i - A_j), \\ \frac{\partial R_{ij}(\beta)}{\partial \phi_k} &= 0. \end{aligned}$$

To establish asymptotic properties of the proposed estimators above, define the distance between $\theta_1 = (\beta_1^T, \Lambda_1)^T$ and $\theta_2 = (\beta_2^T, \Lambda_2)^T$ as $d(\theta_1, \theta_2) = (\|\beta_1 - \beta_2\|^2 + \|\Lambda_1 - \Lambda_2\|_2^2)^{1/2}$, where $\|\cdot\|$ is the Euclidean norm, and $\|\Lambda_1 - \Lambda_2\|_2^2 = E[(\Lambda_1(U) - \Lambda_2(U))^2 + (\Lambda_1(V) - \Lambda_2(V))^2 + (\Lambda_1(A) - \Lambda_2(A))^2]$. Let $\theta_0 = (\beta_0^T, \Lambda_0)^T$ denote the true value of θ. The following theorems give the asymptotic consistency of $\widehat{\theta}_n$ and the asymptotic normality of $\widehat{\beta}_n$.

Theorem 3.1. *Suppose that the regularity conditions 1-4 given in the Appendix hold. Then $\widehat{\theta}_n$ is a strong consistent estimator of θ_0.*

Theorem 3.2. *Suppose that the regularity conditions 1-4 given in the Appendix hold. Then $d(\widehat{\theta}_n, \theta_0) = O_p(n^{-\min\{rv/2,(1-v)/2\}})$, where $0 < v < 1$ such that the Bernstein polynomial degree $m = o(n^v)$ and r is defined in condition 1.*

Theorem 3.3. *Suppose that the regularity conditions 1-6 given in the Appendix hold. If $r > 2$, $v > 1/2r$, then $\sqrt{n}(\widehat{\beta}_n - \beta_0) \to N(0, \Sigma)$, where $\Sigma = I_*^{-1}(\beta_0)$, denoting the semiparametric efficient bound for β with respect to the composite log-likelihood function, with $I_*(\beta_0)$ given in the last paragraph of the Appendix.*

The proofs of the theorems above are sketched in the Appendix. For the application of the method proposed above, one needs to choose m, the degree of freedoms for Bernstein polynomials. Many authors have discussed this ([31]) and suggested to set $m = [n^{1/4}]$, the largest integer smaller than $n^{1/4}$, which will be used below. Also to make inference about the parameter β, it is obvious that we also need to estimate the covariance matrix of $\widehat{\beta}_n$. Since the covariance matrix of $\widehat{\beta}_n$ is complex and difficult to estimate immediately, we propose to apply the simple bootstrap procedure ([10]) described below. It is well-known that the bootstrap procedure can usually provide a direct and easy tool to estimation of covariances and confidence intervals among others when there is no explicit formula available for them.

Let B be a prespecified positive integer and for each $b = 1, ..., B$, draw a simple random sample $O^{(b)} = \left\{ O_i^{(b)}; i = 1, \ldots, n \right\}$ of size n with replacement from the observed data $O = \left\{ O_i; i = 1, \ldots, n \right\}$. Let $\widehat{\beta}^{(b)}$ denote the proposed estimator of β based on the resampled data set $O^{(b)}$ defined above. Then a nature estimator of the covariance matrix of $\widehat{\beta}_n$ is given by

$$\widehat{\Sigma} = \frac{1}{B-1} \sum_{b=1}^{B} \left\{ \widehat{\beta}^{(b)} - \frac{1}{B} \sum_{b=1}^{B} \widehat{\beta}^{(b)} \right\}^{\otimes 2},$$

where $a^{\otimes 2} = aa^T$. The simulation study below indicates that this procedure works well.

4. SIMULATION STUDY

In this section, we present some results obtained from a simulation study performed to investigate the finite-sample properties of the proposed pairwise pseudo-likelihood estimation procedure and compare it to the conditional approach. In the study, we considered the situation with two covariates where Z_1 was assumed to follow a Bernoulli distribution with success probability 0.5 and Z_2 to follow the uniform distribution over $(0, 1)$. The failure time T^* was generated from model (2.1) with $\Lambda_0(t) = t$ and $Z = (Z_1, Z_2)^T$. The underlying truncation time A^* was independently generated from the exponential distribution or the uniform distribution, where the parameter was determined by the percentage of the truncation proportion (TP).

To form a prevalent cohort of sample size (SS) n, the realizations of A^*, T^*, Z were generated until n subjects satisfied the sampling constraint $A^* \leq T^*$. To generate the observation times U and V, we assumed that 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. Then we defined U and V to be the largest and smallest t_j at which subject was observed and that are smaller and larger than the generated failure time T, respectively. We set $k = 10$, $t_j = A + 0.1j$, $j = 1, \ldots, 10$ and $p = 0.8$. We generated 500 data sets with $n = 200$ *or* 400, $B = 20$ and $m = [n^{1/4}]$, the largest integer smaller than $n^{1/4}$.

Tables 1 and 2 give the estimation results obtained for the regression parameters $\beta = (0.5, 0.5)^T$ with A^* from the exponential distribution and the uniform distribution, respectively. They include the estimated bias (Bias) given by the average of the estimates minus the true value, the sample standard error (SSE), the average of the estimated standard errors (ESE), and the 95% empirical coverage probability (CP). The relative efficiency (RE) given by the sample variance of the conditional approach divided by that of the proposed method is also shown in the tables. One can see from the tables that the estimates seem to be unbiased and gave better performance when SS increased. Also for the variance estimation, the bootstrap method seems to perform well, and the results on CP suggest that the normal approximation to the distribution of the estimators appears to be appropriate. In addition, as expected, the proposed pairwise pseudo-likelihood estimator provided considerable efficiency gains compared to the corresponding conditional approach estimator under all scenarios considered, especially under the high truncation proportion scenarios. Note that in the above, we took $B = 20$. We also considered different values for B such as $B = 50$ or 100 and obtained similar results. In other words, the bootstrap variance estimation does not seem to be sensitive to B.

To see the performance on the proposed method further, Figures 1 and 2 display the estimated baseline cumulative hazards functions for the situations corresponding to Tables 1 and 2 with SS $n = 400$, respectively. They include the true baseline cumulative hazards function (black and solid curves), the averages of the obtained estimates (red and dashed curves) and the 95% empirical confidence bands (blue and dash-dot curves). One can see from the figures that the proposed method seems to work well.

For the problem discussed above, one may be also interested in the performance of either the proposed method or the conditional approach that simply ignores the truncation. To investigate this, we repeated the studies and compared the three methods. Table 3 displays the estimated biases for the three approaches under various situations with $n = 400$ and the set-ups being same as with either Table 1 or Table 2. It is easy to see that the ignoring of the left-truncation can yield biased estimates.

5. AN APPLICATION

Now we apply the inference procedure proposed in the previous sections to the AIDS cohort study of hemophiliacs discussed in [9] and [16] 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. For both HIV-1 infection and AIDS diagnosis times, only interval-censored data are available. For the analysis

Table 1. Simulation results on estimation of β with A^* following the exponential distribution

TP	SS	Parameter	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP	RE
			Proposed approach				Conditional approach				
40 %	n=200	β_1	0.0075	0.2312	0.2293	0.930	0.0137	0.2530	0.2554	0.934	1.197
		β_2	0.0176	0.3639	0.4055	0.946	0.0378	0.4117	0.4565	0.932	1.280
	n=400	β_1	−0.0024	0.1561	0.1583	0.940	0.0050	0.1795	0.1752	0.938	1.322
		β_2	0.0065	0.2555	0.2679	0.940	0.0156	0.2810	0.2985	0.940	1.210
60 %	n=200	β_1	0.0112	0.2113	0.2088	0.936	0.0082	0.2566	0.2501	0.940	1.475
		β_2	0.0362	0.3669	0.3606	0.940	0.0569	0.4389	0.4385	0.934	1.431
	n=400	β_1	0.0151	0.1429	0.1460	0.934	0.0085	0.1774	0.1770	0.942	1.541
		β_2	−0.0064	0.2282	0.2416	0.948	0.0029	0.2832	0.2938	0.940	1.540
80 %	n=200	β_1	0.0181	0.1843	0.1938	0.956	0.0201	0.2459	0.2583	0.956	1.780
		β_2	0.0567	0.3097	0.3206	0.944	0.0805	0.4201	0.4480	0.944	1.840
	n=400	β_1	0.0088	0.1286	0.1344	0.948	0.0149	0.1708	0.1819	0.948	1.764
		β_2	0.0179	0.2136	0.2168	0.932	0.0238	0.2842	0.2945	0.946	1.770

Table 2. Simulation results on estimation of β with A^* following the uniform distribution

TP	SS	Parameter	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP	RE
			Proposed approach				Conditional approach				
40 %	n=200	β_1	0.0151	0.2355	0.2385	0.950	0.0197	0.2499	0.2565	0.944	1.126
		β_2	0.0203	0.3966	0.4278	0.942	0.0228	0.4173	0.4568	0.938	1.107
	n=400	β_1	−0.0057	0.1635	0.1662	0.946	−0.0022	0.1707	0.1769	0.940	1.090
		β_2	0.0067	0.2729	0.2804	0.946	−0.0050	0.2899	0.2991	0.950	1.128
60 %	n=200	β_1	0.0287	0.2137	0.2136	0.932	0.0463	0.2559	0.2593	0.952	1.434
		β_2	0.0222	0.3490	0.3667	0.942	0.0171	0.4161	0.4484	0.938	1.421
	n=400	β_1	−0.0072	0.1499	0.1483	0.948	−0.0105	0.1720	0.1778	0.952	1.317
		β_2	0.0026	0.2332	0.2419	0.946	−0.0016	0.2682	0.2923	0.950	1.323
80 %	n=200	β_1	0.0075	0.1788	0.1779	0.946	0.0227	0.2459	0.2614	0.944	1.891
		β_2	0.0195	0.2865	0.2999	0.952	0.0473	0.4173	0.4519	0.962	2.122
	n=400	β_1	0.0141	0.1284	0.1244	0.944	0.0040	0.1763	0.1781	0.952	1.885
		β_2	0.0095	0.1969	0.2025	0.944	0.0156	0.2881	0.3002	0.958	2.141

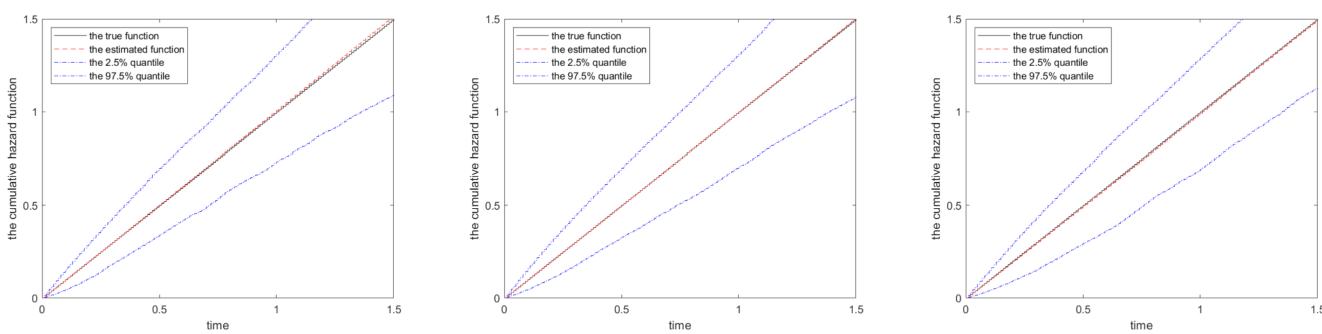

Figure 1. Estimates of the baseline cumulative hazards function Λ with A^* following the exponential distribution: the left for TP $= 40\%$, the middle for TP $= 60\%$ and the right for TP $= 80\%$.

below, 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. The patients are classified into two groups, lightly and heavily treated groups, according to the amount of blood they received for hemophilia. Also the age indicators that indicate if the age of a subject was below 20 at his or her HIV-1 infection were recorded. Our interest is to assess the treatment and age effects on the AIDS diagnosis time.

For the analysis, define T_i to be the AIDS diagnosis time for patient i. Denote $Z_i = (Z_{1i}, Z_{2i})^T$, where $Z_{1i} = 0$ if the ith patient belongs to the lightly treated group and $Z_{1i} = 1$ otherwise, $Z_{2i} = 0$ if the ith patient had age below 20 at HIV-1 infection and $Z_{2i} = 1$ otherwise. By following [15] and [26], we will use the midpoint of the observed interval for HIV-1 infection as the left-truncation time for the AIDS

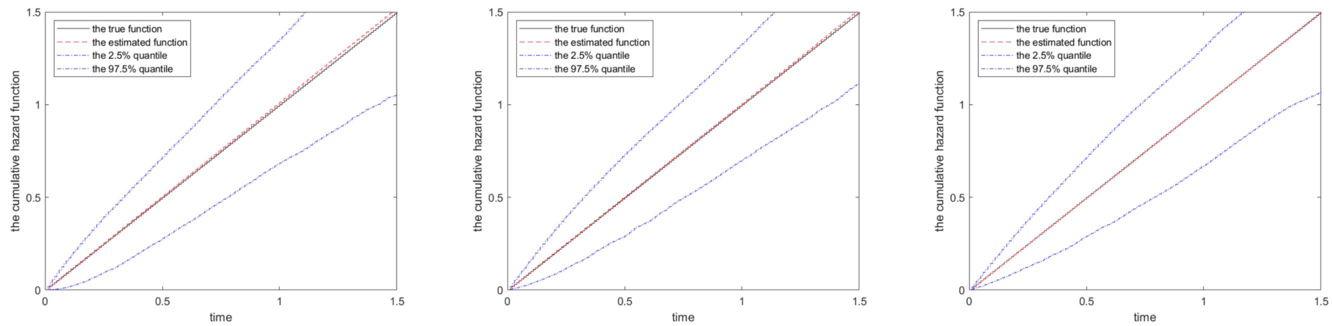

Figure 2. Estimates of the baseline cumulative hazards function Λ with A^* following the uniform distribution: the left for TP $= 40\%$, the middle for TP $= 60\%$ and the right for TP $= 80\%$.

Table 3. Estimated biases of regression parameters with $n = 400$

TP	Parameter	proposed approach	conditional approach	ignoring truncation	proposed approach	conditional approach	ignoring truncation
			A^* exponential			A^* uniform	
40 %	β_1	0.0035	0.0067	−0.1781	0.0155	0.0164	−0.2619
	β_2	0.0077	0.0157	−0.1808	−0.0040	−0.0059	−0.3333
60 %	β_1	0.0169	0.0203	−0.2077	0.0159	0.0226	−0.2859
	β_2	0.0143	0.0132	−0.2564	−0.0019	−0.0049	−0.3954
80 %	β_1	0.0081	0.0061	−0.2227	0.0131	0.0205	−0.2468
	β_2	0.0161	0.0247	−0.2644	−0.0034	0.0119	−0.3534

Table 4. The estimated covariate effects for the AIDS cohort study

covaraites	proposed approach	conditional approach	ignoring truncation
Group	0.0162	0.0168	0.0003
ESE	0.0060	0.0053	0.0073
p-value	0.0069	0.0015	0.9672
Age	0.0023	0.0034	−0.0048
ESE	0.0064	0.0062	0.0130
p-value	0.7193	0.5834	0.7120

diagnosis time. We estimated the covariate effect using the proposed approach with $m = 3$ and $B = 100$. For comparison, we also analyzed the data by using the conditional approach and the approach ignoring truncation as discussed in the previous section.

Table 4 presents the analysis results, including the estimated covariate effects, the estimated standard error, and the p-value for testing the covariate effect being zero. One can see from the table that the proposed approach and the conditional approach suggest that there seems no age effect on AIDS diagnosis time. Furthermore, they indicate 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. Although the conclusions are similar, it seems that the conditional approach overestimated the group and age effects. In addition, the approach ignoring truncation seems to give some misleading or completely different results or conclusions as suggested by the simulation study. We also tried other values for m and B and obtained similar results.

6. DISCUSSION

This paper discussed regression analysis of left-truncated and interval-censored data under the additive hazards model and a pairwise pseudo-likelihood estimation approach was proposed. In the proposed method, by augmenting a pairwise likelihood from the marginal likelihood of the truncation times, one can recover some missing information in the conditional methods. Furthermore, the asymptotic properties of the proposed estimator were established, and the simulation study indicated that the proposed procedure performed well and gave the efficiency gain over the conditional approach. Also the approach was applied to a set of real data from the AIDS cohort study.

As mentioned above, the focus of this paper has been on the left-truncated and interval-censored data where the truncation time is known exactly. Sometimes this may be not true since the truncation time may also be interval-censored or one may face the right-truncation case. It would be useful to generalize the proposed approach to these situations. For the assumptions used in our paper, one is that the underlying truncation time and the observation times are independent of the interested survival time given the covariates. Sometimes this may be inappropriate since the patient in a health study or clinical trials may tend to see doctors and be examined according to his feelings. For the situation, it would be necessary to consider the approach under the dependent truncation and censoring assumptions. Also note that for simplicity, it has been assumed that covariates have no effect on the left truncation variable, which may be not true sometimes. It will be straightforward to generalize the proposed method to this latter situation.

There exist some other directions for future research. One is the development of model checking procedures for the use of the additive hazards model in the situation discussed above. In general, the additive hazards model is used when one is interested in the additive or excess risk. However, it may not be easy for the development based left-truncated and interval-censored data since there seem to exist little research on the problem even based on interval-censored data. Another direction is that in the implementation of the proposed procedure, the Bernstein polynomials are used to approximate the baseline cumulative hazards function. One needs to choose the degree of the polynomials and one nature way is to try different values and compare the results. An alternative way is to develop some likelihood principles such as AIC, BIC or data-driven methods for its selection.

APPENDIX A. PROOFS OF THE ASYMPTOTIC RESULTS

In this appendix, we will sketch the proofs of the theorems given above. First we need to give some more notations and conditions. Define the parameter space $\Theta = \{\theta = (\beta^T, \Lambda)^T \in \mathcal{B} \otimes \mathcal{M}\}$, where $\mathcal{B}$ is a subset of R^p and $\mathcal{M}$ is the collection of all bounded and continous noncreasing, nonnegative functions over $[0, \tau]$. Define the sieve space as $\Theta_n = \{\theta = (\beta^T, \Lambda_n)^T \in \mathcal{B} \otimes \mathcal{M}_n\}$, where $\mathcal{M}_n$ is the collection of the Bernstein polynomials-based function defined in Section 3. Let $l_i^c(\theta)$ and $l_{i,j}^p(\theta)$ denote the log $\mathcal{L}_n^C$ and $\mathcal{L}_n^P$, respectively, corresponding to subject i and pair sample (i,j). For two independent observations O and O', denote $W = (O, O')$, then rewrite $l_n(\theta) = [2/\{n(n-1)\}] \sum_{1 \le i<j \le n} r(\theta, W_{ij})$ with $W_{ij} = (O_i, O_j)$ and $r(\theta, W_{ij}) = r_{ij}(\theta) = \{l_i^c(\theta) + l_j^c(\theta)\}/2 + l_{i,j}^p(\theta)$. The letter C represents a constant, and it does not necessarily represent the same value each time. Let $P^2 = P \otimes P$ denote the product probability measure. Then, we establish the asymptotic properties of $\hat{\theta}_n = (\hat{\beta}_n^T, \hat{\Lambda}_n)^T$ under the following regularity conditions.

Condition 1. *The true value of β, denoted by β_0, lies in the interior of a compact set $\mathcal{B}$ in R^p. The true cumulative baseline hazard function $\Lambda_0(t)$ is rth continuously differentiable for $r \ge 2$ and strictly increasing on $[0, \tau]$ with $\Lambda_0(0) = 0$.*

Condition 2. *The covariate vector Z has bounded support in R^p and $cov(Z)$ is nonsingular.*

Condition 3. *There exists a positive η such that $P(U - A > \eta) = 1$ and $P(V - U > \eta) = 1$. Furthermore, the union support of A, U and V is contained in an interval $[a, b]$, where $0 < a < b < \infty$ with $0 < \Lambda_0(a) < \Lambda_0(b) < \infty$.*

Condition 4. *For every θ in a neighborhood of θ_0, $P^2\{r(\theta, W) - r(\theta_0, W)\} \le -Cd^2(\theta, \theta_0)$.*

Condition 5. *The matrix $E(S_\beta S_\beta^T)$ is finite and positive definite, where S_β is defined below.*

Condition 6. $0 < P^2\{r'(\theta_0, W)[\iota]\} < \infty$ *for all $\iota \ne 0, \iota \in V$, where V denotes a linear span of $\Theta - \theta_0$; For $\theta \in \{\theta \in \Theta, d(\theta, \theta_0) = O(\delta_n)\}$, $P^2\{r''(\theta, W)[\theta - \theta_0, \theta - \theta_0] - r''(\theta_0, W)[\theta - \theta_0, \theta - \theta_0]\} = O(\delta_n^3)$ and $\delta_n^3 = o(n^{-1})$.*

Proof of Theorem 1. Define $\mathcal{L}_n = \{r(\theta, W) : \theta \in \Theta_n\}$. Under Conditions $1-4$, we have that for any θ_1 and $\theta_2 \in \Theta_n$,

$$|r(\theta_1, W) - r(\theta_2, W)| \le \frac{1}{2}\{|l^c(\theta_1, O) - l^c(\theta_2, O)| + |l^c(\theta_1, O') - l^c(\theta_2, O')|\} + |l^p(\theta_1, W) - l^c(\theta_2, W)|.$$

Define $\varphi_t(\omega) = \Lambda_\omega(t) + t\beta_\omega^T Z$ and $\varphi_{t,t'}(\omega) = (t - t')(\beta_\omega^T Z - \beta_\omega^T Z')$, with $\Lambda_\omega = \omega\Lambda_1 + (1-\omega)\Lambda_2$ and $\beta_\omega = \omega\beta_1 + (1-\omega)\beta_2$. Then, by the mean value theorem, we have that there exist $0 \le \xi, \zeta \le 1$, such that

$$|l^c(\theta_1, O) - l^c(\theta_2, O)| \le |\delta_1\{\varphi'_A(\xi) - \varphi'_U(\xi)\} + \delta_2\{\varphi'_U(\xi) - \varphi'_V(\xi)\} + \delta_3(\varphi'_V(\xi) - \varphi'_A(\xi))|$$

and

$$|l^p(\theta_1, W) - l^p(\theta_2, W)| \le \varphi_{A,A'}(\zeta)$$

where $\varphi'_t(\xi) = (\Lambda_1 - \Lambda_2)(t) + t(\beta_1 - \beta_2)^T Z$ and $\varphi'_{t,t'}(\zeta) = (t - t')[(\beta_1 - \beta_2)^T(Z - Z')]$. Under Conditions $1-3$, we have that $|r(\theta_1, W) - r(\theta_2, W)| \le C||\beta_1 - \beta_2|| + C||\Lambda_1 - \Lambda_2||_\infty$ where $||\Lambda_1 - \Lambda_2||_\infty = \sup_t |\Lambda_1(t) - \Lambda_2(t)| \le \max_{0 \le k \le m} |\phi_{1k} - \phi_{2k}| = ||\phi_1 - \phi_2||$, $\phi_i = (\phi_{i0}, \ldots, \phi_{im})^T$ denote the Bernstein coefficients corresponding to $\Lambda_i, i = 1, 2$.

Combine the results, let $\theta^{(1)}, \ldots, \theta^{(N)}$ be the minimum collection of points in Θ_n, for any $\theta \in \Theta_n$, there exists k such that

$$\frac{2}{n(n-1)} \sum_{1 \le i<j \le n} |r(\theta, W_{ij}) - r(\theta^{(k)}, W_{ij})| \le C||\beta - \beta^{(k)}|| + C||\phi - \phi^{(k)}|| < \epsilon.$$

It can be checked that $\mathcal{B}$, the compact subset of R^p, is covered by $(C/\epsilon)^p$ balls with radius $\epsilon/2C$ and $\{\phi \in R^{m+1}, \sum_{k=0}^m |\phi_k| \le M_n\}$ is covered by $(CM_n/\epsilon)^{m+1}$ balls with radius $\epsilon/2C$. Then the covering number is

$$N(\epsilon, \mathcal{L}_n, L^1(U_n)) \le C(M_n/\epsilon)^{m+1}(1/\epsilon)^p.$$

So

$$\frac{\log N(\epsilon, \mathcal{L}_n, L^1(U_n))}{n} \xrightarrow{P} 0.$$

By Corollary 3.2 in [2], we obtain

$$\sup_{\theta\in\Theta_n} |U_n r(\theta,W) - P^2 r(\theta,W)| \to 0 \quad a.s..$$

Let $g(\theta,W) = -r(\theta,W)$. Define $\mathcal{K}_\epsilon = \{\theta : d(\theta,\theta_0) \geq \epsilon, \theta \in \Theta_n\}$ for $\epsilon > 0$ and

$$\begin{aligned}\kappa_{1n} &= \sup_{\theta\in\Theta_n} |U_n g(\theta,W) - P^2 g(\theta,W)| \\ \kappa_{2n} &= U_n g(\theta_0,W) - P^2 g(\theta_0,W).\end{aligned}$$

We can show that

$$\begin{aligned}\inf_{\mathcal{K}_\epsilon} P^2 g(\theta,W) &= \inf_{\mathcal{K}_\epsilon}\{P^2 g(\theta,W) - U_n g(\theta,W) + U_n g(\theta,W)\} \\ &\leq \kappa_{1n} + \inf_{\mathcal{K}_\epsilon} U_n g(\theta,W).\end{aligned}$$

If $\hat\theta_n \in \mathcal{K}_\epsilon$, since θ_0 is the extreme point of the likelihood, we have

$$\begin{aligned}\inf_{\mathcal{K}_\epsilon} U_n g(\theta,W) = U_n g(\hat\theta_n,W) &\leq U_n g(\theta_0,W) \\ &= \kappa_{2n} + P^2 g(\theta_0,W).\end{aligned}$$

Define $\delta_n^\epsilon = \inf_{\mathcal{K}_\epsilon}\{P^2 g(\theta,W) - P^2 g(\theta_0,W)\}$. With Condition 4, $P^2\{r(\theta,W) - r(\theta_0,W)\} \leq -Cd^2(\theta,\theta_0)$, we can conclude that

$$\inf_{\mathcal{K}_\epsilon} P^2 g(\theta,W) \leq \kappa_{1n} + \kappa_{2n} + P^2 g(\theta_0,W) = \kappa_n + P^2 g(\theta_0,W)$$

where $\kappa_n = \kappa_{1n} + \kappa_{2n}$, which satisfy $\kappa_n \geq \delta_n^\epsilon$. Combine the results, we have $\{\hat\theta_n \in \mathcal{K}_\epsilon\} \subseteq \{\kappa_n \geq \delta_n^\epsilon\}$, which gives that $d(\hat\theta_n,\theta_0)$ converges almost surely towards 0.

□

Proof of Theorem 2. First, by Theorem 1.6.2 of [20], there must be a Bernstein polynomial $\Lambda_{0,n}$ which has

$$||\Lambda_{0,n} - \Lambda_0||_\infty = O(m^{-r/2}) = O(n^{-rv/2}).$$

Define $\theta_{0,n} = (\beta_0, \Lambda_{0,n})$, where $\phi_0 = (\phi_{00},\ldots,\phi_{0m})^T$ denote the coefficients corresponding to $\Lambda_{0,n}$ and $\phi_{0k} = \Lambda_0\{t_l + (k/m)(t_u - t_l)\}$. Then, we have $d(\theta_{0,n},\theta_0) = O(n^{-rv/2})$.

Define the following three classes of function:

$$\begin{aligned}\mathcal{F}_\eta &= \{r(\theta,W) - r(\theta_{0,n},W) : \theta \in \Theta_n, \eta/2 \leq d(\theta,\theta_{0,n}) \leq \eta\}, \\ \mathcal{G}_\eta^{(1)} &= \{l^c(\theta,O) - l^c(\theta_{0,n},O) : \theta \in \Theta_n, \eta/2 \leq d(\theta,\theta_{0,n}) \leq \eta\}, \\ \mathcal{G}_\eta^{(2)} &= \{l^p(\theta,W) - l^p(\theta_{0,n},W) : \theta \in \Theta_n, \eta/2 \leq d(\theta,\theta_{0,n}) \leq \eta\}.\end{aligned}$$

Using Condition 4, it is easy to show that for large n,

$$\begin{aligned}\sup_{\substack{\theta\in\Theta_n \\ \eta/2<d(\theta,\theta_{0,n})\leq\eta}} & P^2\{r(\theta,W) - r(\theta_{0,n},W)\} \leq \\ \sup_{\substack{\theta\in\Theta_n \\ \eta/2<d(\theta,\theta_{0,n})\leq\eta}} & P^2\{r(\theta,W) - r(\theta_0,W)\} + \\ & P^2\{r(\theta_0,W) - r(\theta_{0,n},W)\} \leq -C\eta^2.\end{aligned}$$

By the definition of the above classes, we have

$$\begin{aligned}E^*||n^{1/2}(U_n - P^2)||_{\mathcal{F}_\eta} \leq{}& E^*||n^{1/2}(P_n - P)||_{\mathcal{G}_\eta^{(1)}} \\ &+ E^*||n^{1/2}(U_n - P^2)||_{\mathcal{G}_\eta^{(2)}}.\end{aligned}$$

Define $\varphi(\beta,\Lambda) = \beta^T(Z - Z')(A - A')$, then $l^p(\theta,W) = -\log[1 + \exp\{\varphi(\beta,\Lambda)\}]$ is a Lipschitz transformation of φ with finite bound. Further define

$$\mathcal{H}_\eta = \{\varphi(\beta,\Lambda) - \varphi(\beta_0,\Lambda_{0,n}) : \theta \in \Theta_n, \eta/2 \leq d(\theta,\theta_{0,n}) \leq \eta\},$$

which leads to $E^*||n^{1/2}(U_n - P^2)||_{\mathcal{G}_\eta^{(2)}} \leq CE^*||n^{1/2}(U_n - P^2)||_{\mathcal{H}_\eta}$.

Next, denote $\hat U_n\varphi(\beta,\Lambda) = \sum_{i=1}^n E\{U_n\varphi(\beta,\Lambda) - P^2\varphi(\beta,\Lambda) \mid O_i\}$. It can be verified that $P^2\varphi(\beta,\Lambda) = 2Cov\{\beta^T Z, A\}$. Moreover, since the pair O_i and O_j are i.i.d.,

$$\begin{aligned}E\{\varphi_{ij}(\beta,\Lambda) \mid O_i\} ={}& E\{(\beta^T Z_i - \beta^T Z_j)(A_i - A_j) \mid A_i, Z_i\} \\ ={}& \beta^T Z_i A_i - A_i E(\beta^T Z_i) - \beta^T Z_i E(A_i) \\ &+ E(\beta^T Z_i A_i).\end{aligned}$$

Thus, we have

$$\begin{aligned}\hat U_n\varphi(\beta,\Lambda) ={}& \sum_{i=1}^n E\left\{\binom{n}{2}^{-1}\sum_{j<k}\varphi_{jk}(\beta,\Lambda) - P^2\varphi(\beta,\Lambda) \mid O_i\right\} \\ ={}& \frac{2}{n}\sum_{i=1}^n\left\{\beta^T Z_i A_i - A_i E(\beta^T Z_i) - \beta^T Z_i E(A_i)\right. \\ &\left. + E(\beta^T Z_i A_i)\right\} - 4Cov(\beta^T Z, A).\end{aligned}$$

By i.i.d. property of the observations and the definition of covariance, direct calculation gives

$$\tilde U_n \equiv U_n\varphi(\beta,\Lambda) - P^2\varphi(\beta,\Lambda) - \hat U_n\varphi(\beta,\Lambda) = \binom{n}{2}^{-1}\sum_{i<j}\tilde U_n^{(i,j)}$$

where

$$\begin{aligned}\tilde U_n^{(i,j)} ={}& \beta^T Z_i A_i - \beta^T Z_j A_i - \beta^T Z_i A_j + \beta^T Z_j A_j \\ & - 2Cov(\beta^T Z, A) - \left\{\beta^T Z_i A_i - A_i E(\beta^T Z_i)\right. \\ & \left. - \beta^T Z_i E(A_i) + E(\beta^T Z_i A_i) - 2Cov(\beta^T Z, A)\right\} \\ & - \left\{\beta^T Z_j A_j - A_j E(\beta^T Z_j) - \beta^T Z_j E(A_j)\right. \\ & \left. + E(\beta^T Z_j A_j) - 2Cov(\beta^T Z, A)\right\} \\ ={}& -(\beta^T Z_i - E(\beta^T Z_i))(A_j - E(A_j)) \\ & - (\beta^T Z_j - E(\beta^T Z_j))(A_i - E(A_i)).\end{aligned}$$

Therefore, we have

$$\begin{aligned}\tilde U_n ={}& -\binom{n}{2}^{-1}\sum_{i=1}^n\sum_{j=1}^n(\beta^T Z_i - E(\beta^T Z_i))(A_j - E(A_j)) \\ \asymp{}& -\frac{2}{n}\sum_{i=1}^n(\beta^T Z_i - E(\beta^T Z_i))\frac{1}{n}\sum_{j=1}^n(A_j - E(A_j)),\end{aligned}$$

where $\asymp$ denotes asympotototically equivalent. Similar arguments as in the proof of Theorem 1 in [25], we have

$$(U_n - P^2)\varphi(\beta,\Lambda) = \frac{2}{n}\sum_{i=1}^{n}\varphi^{(1)}(\theta,O_i) - \frac{2}{n^2}\{\sum_{i=1}^{n}\varphi^{(2)}(\theta,O_i)\}\{\sum_{j=1}^{n}\varphi^{(3)}(\theta,O_j)\},$$

where $\varphi^{(1)}(\theta,O_i) = \beta^T Z_i[A_i - E\{A_i\}] - A_i E\{\beta^T Z_i\} + E\{\beta^T Z_i A_i\} - 2Cov\{\beta^T Z_i, A_i\}$, $\varphi^{(2)}(\theta,O_i) = \beta^T Z_i - E\{\beta^T Z_i\}$ and $\varphi^{(3)}(\theta,O_j) = A_j - E\{A_j\}$.

Then, define the following three classes:

$$\mathcal{H}_\eta^{(1)} = \{\varphi^{(1)}(\theta,O) - \varphi^{(1)}(\theta_{0,n},O) : \theta\in\Theta_n, \eta/2 \le d(\theta,\theta_{0,n}) \le \eta\},$$
$$\mathcal{H}_\eta^{(2)} = \{\varphi^{(2)}(\theta,O) - \varphi^{(2)}(\theta_{0,n},O) : \theta\in\Theta_n, \eta/2 \le d(\theta,\theta_{0,n}) \le \eta\},$$
$$\mathcal{H}_\eta^{(3)} = \{\varphi^{(3)}(\theta,O) - \varphi^{(3)}(\theta_{0,n},O) : \theta\in\Theta_n, \eta/2 \le d(\theta,\theta_{0,n}) \le \eta\}.$$

So, we have $E^*||n^{1/2}(U_n - P^2)||_{\mathcal{H}_\eta} \le CE^*||n^{1/2}(P_n - P)||_{\mathcal{H}_\eta^{(1)}} + CE^*||n^{1/2}(P_n-P)||_{\mathcal{H}_\eta^{(2)}} + CE^*||n^{1/2}(P_n-P)||_{\mathcal{H}_\eta^{(3)}}$. We can obtain that $N_{[]}(\epsilon,\mathcal{G}_\eta^{(1)},L_2(P))$, $N_{[]}(\epsilon,\mathcal{H}_\eta^{(1)},L_2(P))$, $N_{[]}(\epsilon,\mathcal{H}_\eta^{(2)},L_2(P))$ and $N_{[]}(\epsilon,\mathcal{H}_\eta^{(3)},L_2(P)) \le C(m+1+p)\log(\eta/\epsilon)$ with the same idea of [22]. By Lemma 3.4.2 of [28], we can conclude

$$E^*||n^{1/2}(U_n - P^2)||_{\mathcal{F}_\eta} \le C\{(m+1)^{1/2}\eta + (m+1)/n^{1/2}\}.$$

Now, take $\gamma_n = n^{(1-v)/2}$, with Theorem 3.2.5 of [28], we have that $\gamma_n d(\hat\theta_n,\theta_{0,n}) = O_p(1)$. So $d(\theta_n,\theta_0) = O_p(n^{-(1-v)/2} + n^{-rv/2})$. □

Proof of Theorem 3. Let Υ denote a linear span of $\Theta - \theta_0$. Define the first order directional derivative of $r(\theta,W)$ at the direction $\iota\in\Upsilon$ and the second order directional derivative as

$$r'(\theta,W)[\iota] = \frac{dr(\theta+s\iota,W)}{ds}\Big|_{s=0},$$
$$r''(\theta,W)[\iota,\tilde\iota] = \frac{dr(\theta+s\iota+\tilde s\tilde\iota,W)}{d\tilde s ds}\Big|_{s=0}\Big|_{\tilde s=0}$$

for any $\theta\in\{\theta\in\Theta : d(\theta,\theta_0) = O(\delta_n)\}$, where $\delta_n = n^{-\min\{(1-v)/2, rv/2\}}$. Let $r'(\theta_0,W)[\iota]$ and $r''(\theta_0,W)[\iota,\tilde\iota]$ denote $r'(\theta,W)[\iota]$ and $r''(\theta,W)[\iota,\tilde\iota]$ evaluated at θ_0, respectively. Define Fisher inner product on the space Υ as $<\iota,\tilde\iota> = P^2\{r'(\theta_0,W)[\iota]r'(\theta_0,W)[\tilde\iota]\}$ and Fisher norm $||\iota|| =<\iota,\iota>$. Let $\bar\Upsilon$ be the closed linear span of Υ under the Fisher norm. Then $(\bar\Upsilon, ||\cdot||)$ is a Hilbert space. Define $h(\theta) = b^T\beta$ for any $\theta\in\Theta$ with $||b|| \le 1$. For any $\iota = (\iota_\beta,\iota_\Lambda)\in\Upsilon$, define

$$h'(\theta_0)[\iota] = \frac{dh(\theta_0+s\iota)}{ds}\Big|_{s=0} = b^T\iota_\beta.$$

Following the idea from section 3.2 of [6], by the definition of $r(\theta,W)$, then

$$r'(\theta_0,W)[\iota] = \lim_{s\to 0}\frac{r(\theta_0+s\iota,W) - r(\theta_0,W)}{s} = \frac{\partial r(\theta_0,W)}{\partial\beta}[\iota_\beta] + \frac{\partial r(\theta_0,W)}{\partial\Lambda}[\iota_\Lambda]$$

where

$$\frac{\partial r(\theta_0,W)}{\partial\beta} = \frac{1}{2}\frac{\partial l^c(\theta_0,O)}{\partial\beta} + \frac{1}{2}\frac{\partial l^c(\theta_0,O')}{\partial\beta} + \frac{\partial l^p(\theta_0,W)}{\partial\beta}$$

and

$$\frac{\partial r(\theta_0,W)}{\partial\Lambda}[\iota_\Lambda] = \frac{1}{2}\frac{\partial l^c(\theta_0,O)}{\partial\Lambda}[\iota_\Lambda] + \frac{1}{2}\frac{\partial l^c(\theta_0,O')}{\partial\Lambda}[\iota_\Lambda] + \frac{\partial l^p(\theta_0,W)}{\partial\Lambda}[\iota_\Lambda].$$

So the conditions for Riesz representation theory hold with

$$\sup_{\iota\in\bar\Upsilon:||\iota||>0}\frac{|h'(\theta_0)[\iota]|^2}{||\iota||^2} = \sup_{\iota\in\bar\Upsilon:||\iota||>0}\frac{|b^T\iota_\beta|^2}{||\iota||^2} = \sup_{\iota\in\bar\Upsilon:||\iota||>0}\frac{|b^T\iota_\beta|^2}{P^2\{r'(\theta_0,W)[\iota]\}^2} < \infty.$$

Now define $\rho(\theta-\theta_0) = r(\theta,W) - r(\theta_0,W) - r'(\theta_0,W)[\theta-\theta_0]$. Let ϵ_n be any positive sequence satisfying $\epsilon_n = o(n^{-1/2})$. For any $\iota\in\Upsilon$, by the rth derivatives of Λ is bounded and Theorem 1.6.2 of [20], there exists $\Pi_n\iota^*\in\Theta_n - \theta_0$ such that $||\Pi_n\iota^* - \iota^*|| = O(n^{-rv/2})$ and $\delta_n||\Pi_n\iota^* - \iota^*|| = o(n^{-1/2})$ when $r > 2$ and $v > 1/2r$. By the definition of $\hat\theta_n$ and some calculations, we have

$$0 \le U_n\{r(\hat\theta_n,W) - r(\hat\theta_n \pm \epsilon_n\Pi_n\iota^*,W)\} = \mp\epsilon_n U_n r'(\theta_0,W)[\iota^*] \mp I_1 + I_2 + I_3,$$

where $I_1 = \epsilon_n U_n r'(\theta_0,W)[\Pi_n\iota^* - \iota^*]$, $I_2 = (U_n - P^2)\{\rho(\hat\theta_n - \theta_0,W) - \rho(\hat\theta_n \pm \epsilon_n\Pi_n\iota^* - \theta_0,W)\}$, ans $I_3 = P^2\{\rho(\hat\theta_n - \theta_0,W) - \rho(\hat\theta_n \pm \epsilon_n\Pi_n\iota^* - \theta_0,W)\}$.

To bound I_1, we define $h_1(O) = E\{r'(\theta_0,W)[\Pi_n\iota^* - \iota^*] \mid O\}$, $\sigma_1^2 = E\{h_1(O)\}^2$ and $\sigma_2^2 = E\{r'(\theta_0,W)[\Pi_n\iota^* - \iota^*]\}^2$, then the variance of $U_n r'(\theta_0,W)[\Pi_n\iota^* - \iota^*]$ is $\{4(n-2)\sigma_1^2 + 2\sigma_2^2\}/\{n(n-1)\}$. Further that $||\Pi_n\iota^* - \iota^*|| = o(1)$, then $\sigma_1^2 = o(1)$ and $\sigma_2^2 = o(1)$. By Chebyshev's inequality,

$$P(|U_n r'(\theta_0,W)[\Pi_n\iota^* - \iota^*]| \ge \epsilon_n) \le \frac{4(n-2)\sigma_1^2 + 2\sigma_2^2}{n(n-1)\epsilon_n^2}.$$

Then we have $U_n r'(\theta_0,W)[\Pi_n\iota^* - \iota^*] = o_p(n^{-1/2})$ and $I_1 = \epsilon_n \times o_p(n^{-1/2})$.

In order to show $I_2 = \epsilon_n \times o_p(n^{-1/2})$, we rewrite I_2 as follow:

$$
\begin{aligned}
I_2 &= \mp \epsilon_n (U_n - P^2)\{r'(\tilde{\theta}, W)[\Pi_n \iota^*] - r'(\theta_0, W)[\Pi_n \iota^*]\} \\
&= \mp \epsilon_n (P_n - P)\{(l^c)'(\tilde{\theta}, O)[\Pi_n \iota^*] - (l^c)'(\theta_0, O)[\Pi_n \iota^*]\} \\
&\quad \mp \epsilon_n (U_n - P^2)\{(l^p)'(\tilde{\theta}, W)[\Pi_n \iota^*] - (l^p)'(\theta_0, W)[\Pi_n \iota^*]\}
\end{aligned}
$$

where $\tilde{\theta}$ lies between $\hat{\theta}_n$ and $\hat{\theta}_n \pm \epsilon_n \Pi_n \iota^*$. By theorem 2.8.4 of [28], we know that $\mathcal{F}^c = \{(l^c)'(\theta, O)[\Pi_n \iota^*]\}$ is Donsker. And define $\mathcal{F}_{\delta_n^c} = \{(l^c)'(\tilde{\theta}, O)[\Pi_n \iota^*] - (l^c)'(\theta_0, O)[\Pi_n \iota^*], ||\tilde{\theta} - \theta_0|| = o(\delta_n)\}$, then by Corollary 2.3.12 of [28], we have

$$
(P_n - P)\{(l^c)'(\tilde{\theta}, O)[\Pi_n \iota^*] - (l^c)'(\theta_0, O)[\Pi_n \iota^*]\} = o_p(n^{-1/2}).
$$

By the Lemma 3 of [27], the preservation of theorem of Lipschitz functions and Theorem 5.3.1 of [8], we conclude that $\mathcal{F}^p = \{(l^p)'(\theta, W)[\Pi_n \iota^*]\}$ satisfies the CLT that, for any $f \in \mathcal{F}^p$, $n^{1/2}(U_n - P^2)(f)$ converges weakly to a mean zero Gaussian process G_f. Then with Theorem 4.1(b) of [2], we know that

$$
\lim_{\delta_n \to 0} \limsup_{n \to \infty} E||n^{1/2}(U_n - P^2)(f)||_{\mathcal{F}^p_{\delta_n}} = 0
$$

where $\mathcal{F}^p_{\delta_n} = \{(l^p)'(\tilde{\theta}, W)[\Pi_n \iota^*] - (l^p)'(\theta_0, W)[\Pi_n \iota^*], ||\tilde{\theta} - \theta_0|| = o(\delta_n)\}$. It is easy to see that

$$
(U_n - P^2)\{(l^p)'(\tilde{\theta}, W)[\Pi_n \iota^*] - (l^p)'(\theta_0, W)[\Pi_n \iota^*]\} = o_p(n^{-1/2}).
$$

Thus we get $I_2 = \epsilon_n \times o_p(n^{-1/2})$.

For I_3, by mean value theorem and Taylor expansion and Condition 6, we have

$$
\begin{aligned}
& P^2\{\rho(\theta - \theta_0, W)\} = P^2\{r(\theta, W) - r(\theta_0, W) \\
&\quad - r'(\theta_0, W)[\theta - \theta_0]\} = \frac{1}{2} P^2\{r''(\tilde{\theta}, W)[\theta - \theta_0, \theta - \theta_0] \\
&\quad - r''(\theta_0, W)[\theta - \theta_0, \theta - \theta_0]\} \\
&\quad + \frac{1}{2} P^2\{r''(\theta_0, W)[\theta - \theta_0, \theta - \theta_0]\} \\
&= \frac{1}{2} P^2\{r''(\theta_0, W)[\theta - \theta_0, \theta - \theta_0]\} + \epsilon_n \times o_p(n^{-1/2})
\end{aligned}
$$

where $\tilde{\theta}$ lies between θ_0 and θ. Therefore we have

$$
\begin{aligned}
I_3 &= -\frac{1}{2}\{||\hat{\theta}_n - \theta_0||^2 - ||\hat{\theta}_n \pm \epsilon_n \Pi_n \iota^* - \theta_0||^2\} \\
&\quad + \epsilon_n \times o_p(n^{-1/2}) \\
&= \pm \epsilon_n < \hat{\theta}_n - \theta_0, \Pi_n \iota^* > + \frac{1}{2}||\epsilon_n \Pi_n \iota^*||^2 + \epsilon_n \times o_p(n^{-1/2}) \\
&= \pm \epsilon_n < \hat{\theta}_n - \theta_0, \iota^* > \pm \epsilon_n < \hat{\theta}_n, \Pi_n \iota^* - \iota^* > \\
&\quad + \frac{1}{2}||\epsilon_n \Pi_n \iota^*||^2 + \epsilon_n \times o_p(n^{-1/2}).
\end{aligned}
$$

Further that $\delta_n ||\Pi_n \iota^* - \iota^*|| = o(n^{-1/2})$ and $||\Pi_n \iota^*|| \to ||\iota^*||$, then $I_3 = \pm \epsilon_n < \hat{\theta}_n - \theta_0, \iota^* > + \epsilon_n \times o_p(n^{-1/2})$.

Therefore, we obtain that $n^{1/2} < \hat{\theta}_n - \theta_0, \iota^* >= n^{1/2}(U_n - P^2)\{r'(\theta_0, W)[\iota^*]\} + o_p(1) \to N(0, ||\iota^*||^2)$. Let Λ_k^* be the solution to $\inf_{\Lambda^*} E\{\frac{\partial r(\theta_0, W)}{\partial \beta} e_k - \frac{\partial r(\theta_0, W)}{\partial \Lambda}[\Lambda^*]\}^2$. Define the k-th element of S_{β_0} as $\frac{\partial r(\theta_0, W)}{\partial \beta} e_k - \frac{\partial r(\theta_0, W)}{\partial \Lambda}[\Lambda_k^*]$ and $I_*(\beta_0) = P^2(S_{\beta_0} S_{\beta_0}^T)$. Therefore, $\iota^* = (\iota_\beta^*, \iota_\Lambda^*)$ with $\iota_\beta^* = I_*(\beta_0)^{-1} b$ and $\iota_\Lambda^* = -\Lambda^* \iota_\beta^*$, where $\Lambda^* = (\Lambda_1^*, \cdots, \Lambda_p^*)$. So

$$
\begin{aligned}
||\iota^*||^2 &= \sup_{\iota \in \tilde{\Upsilon}: ||\iota|| > 0} \frac{|h'(\theta_0)[\iota]|^2}{||\iota||^2} \\
&= b^T P^2(S_{\beta_0} S_{\beta_0}^T)^{-1} b = b^T I_*^{-1}(\beta_0) b,
\end{aligned}
$$

where $P^2(S_{\beta_0} S_{\beta_0}^T)$ is nonsingular by Condition 5. Combine with Riesz representation theorem, we have that $n^{1/2}(\hat{\beta}_n - \beta_0) \to N(0, \Sigma)$ where $\Sigma = I_*^{-1}(\beta_0)$. □

ACKNOWLEDGEMENTS

The authors wish to thank the Editor-in-Chief, Dr. Yuedong Wang, and two reviewers for their many helpful and insightful comments and suggestions that greatly improved the paper. This work was partially supported by the National Nature Science Foundation of China (Grant No. 11801212, Grant No. 12071176).

Received 21 March 2022

REFERENCES

[1] Andersen, P. K. & Gill, R. D.(1982). Cox regression model for counting processes: a large sample study. *The Annals of Statistics* **10**, 1100–1120. MR0673646

[2] Arcones, M. A. & Giné, E.(1993). Limit theorems for U-processes. *The Annals of Probability* **21**, 1494–1542. MR1235426

[3] Chen, C-M. & Shen, P-S.(2018). Conditional maximum likelihood estimation in semiparametric transformation model with LTRC data. *Lifetime Data Analysis* **24**, 250–272. MR3773859

[4] Chen, D.-G., Sun, J. & Peace, K. E.(2012). *Interval-censored time-to-event data: Methods and applications*. Chapman & Hall/CRC. MR3053014

[5] Chen, K., Jin, Z. & Ying, Z.(2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89**, 659–668. MR1929170

[6] Chen, X., Fan, Y. & Tsyrennikov, V.(2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association* **101**, 1228–1240. MR2328309

[7] Cox, D. R.(1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B* **34**, 187–202. MR0341758

[8] De la Pena, V. & Giné, E.(2012). *Decoupling: from dependence to independence*. Springer Science & Business Media. MR1666908

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

[10] Efron, B. & Tibshirani, R. J.(1993). *An introduction to the bootstrap*. London: Chapman & Hall. MR1270903

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

[12] Gao, F. & Chan K. C. G.(2019). Semiparametric regression analysis of length-biased interval-censored data. *Biometrics* **75**, 121–132. MR3953713

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

[14] Huang, C-Y. & Qin, J.(2013). Semiparametric estimation for the additive hazards model with left-truncated and right-censored data. *Biometrika* **100**, 877–888. MR3142338

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

[16] Kim, M. Y., DeGruttola, V. G. & Lagakos, S. W.(1993). Analyzing doubly censored data with covariates, with application to AIDS. *Biometrics* **49**, 13–22. MR0999438

[17] Liang K-Y. & Qin, J.(2000). Regression analysis under nonstandard situations: a pairwise pseudolikelihood approach. *Journal of the Royal Statistical Society, Series B* **62**, 773–786. MR1796291

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

[19] Lin, D. Y. & Ying, Z.(1994). Semiparametrics analysis of the additive risk model. *Biometrika* **81**, 61–71. MR1279656

[20] Lorentz, G. G.(1986). *Bernstein polynomials*. New York: Chelsea Publishing Co. MR0864976

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

[22] Shen, X. & Wong, W. H.(1994). Convergence rate of sieve estimates. *The Annals of Statistics* **22**, 580–615. MR1292531

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

[24] Wang, M-C., Brookmeyer, R. & Jewell, N. P.(1993). Statistical models for prevalent cohort data. *Biometrics* **49**, 1–11. MR1221402

[25] Wang, P., Li, D. & Sun, J.(2021). A pairwise pseudo-likelihood approach for left-truncated and interval-censored data under the Cox model. *Biometrics* **77**, 1303–1314. MR4357839

[26] Wang, P., Tong, X., Zhao, S., & Sun, J.(2015). Efficient estimation for the additive hazards model in the presence of left-truncation and interval censoring. *Statistics and Its Interface* **8**, 391–402. MR3341336

[27] Wu, F., Kim, S., Qin, J., Saran, R. & Li, Y.(2018). A pairwise likelihood augmented Cox estimator for left-truncated data. *Biometrics* **74**, 100–108. MR3777930

[28] Van Der Vaart, A. W. & Wellner, J. A.(1996). *Weak convergence and empirical processes*. New York: Springer. MR1385671

[29] Zeng, D., Cai, J. & Shen, Y.(2006). Semiparametric additive risk model for interval-censored data. *Statistica Sinica* **16**, 287–302. MR2256093

[30] Zhang, Z., Sun, L., Zhao, X. & Sun, J.(2005). Regression analysis of interval censored failure time data with linear transformation models. *The Canadian Journal of Statistics* **33**, 61–70. MR2154998

[31] Zhou, Q., Hu, T. & Sun, J.(2017). A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. *Journal of the American Statistical Association* **112**, 664–672. MR3671760

Peijie Wang
School of Mathematics
Jilin University, Changchun
China
E-mail address: wangpeijie@jlu.edu.cn
Yichen Lou
School of Mathematics
Jilin University, Changchun
China
E-mail address: louyc19@mails.jlu.edu.cn

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