Journal of Applied Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Estimation of accelerated hazards models based on case *K* informatively interval-censored failure time data

Rui Ma, Shishun Zhao, Jianguo Sun & Shuying Wang

To cite this article: Rui Ma, Shishun Zhao, Jianguo Sun & Shuying Wang (2023): Estimation of accelerated hazards models based on case *K* informatively interval-censored failure time data, Journal of Applied Statistics, DOI: 10.1080/02664763.2023.2196752

To link to this article: https://doi.org/10.1080/02664763.2023.2196752

Published online: 05 Apr 2023.

JOURNAL OF APPLIED STATISTICS
https://doi.org/10.1080/02664763.2023.2196752

Taylor & Francis
Taylor & Francis Group

Check for updates

Estimation of accelerated hazards models based on case *K* informatively interval-censored failure time data

Rui Ma[a], Shishun Zhao[a], Jianguo Sun[b] and Shuying Wang[c]

[a]Center for Applied Statistical Research and College of Mathematics, Jilin University, Changchun, People's Republic of China; [b]Department of Statistics, University of Missouri, Columbia, MO, USA; [c]School of Mathematics and Statistics, Changchun University of Technology, Changchun, People's Republic of China

ABSTRACT

The accelerated hazards model is one of the most commonly used models for regression analysis of failure time data and this is especially the case when, for example, the hazard functions may have monotonicity property. Correspondingly a large literature has been established for its estimation or inference when right-censored data are observed. Although several methods have also been developed for its inference based on interval-censored data, they apply only to limited situations or rely on some assumptions such as independent censoring. In this paper, we consider the situation where one observes case *K* interval-censored data, the type of failure time data that occur most in, for example, medical research such as clinical trials or periodical follow-up studies. For inference, we propose a sieve borrow-strength method and in particular, it allows for informative censoring. The asymptotic properties of the proposed estimators are established. Simulation studies demonstrate that the proposed inference procedure performs well. The method is applied to a set of real data set arising from an AIDS clinical trial.

ARTICLE HISTORY
Received 6 May 2022
Accepted 23 March 2023

KEYWORDS
Accelerated hazards model; informative observation process; case *K* interval-censored data; sieve maximum likelihood; Bernstein polynomials

1. Introduction

This paper discusses the estimation of or reference on the accelerated hazards (AH) model, one of the most commonly used models for regression analysis of failure time data. Unlike the proportional or additive hazards model, which focuses on the effect of covariates on the hazard function, the AH model concerns the direct effect of covariates on the failure time itself. Although many methods have been proposed in the literature for inference on the AH model, most of them apply only to right-censored data [2,3,5,36]. In the following, we will consider interval-censored data, a more general type of failure time data.

Interval-censored failure time data usually refer to the data in which the failure time of interest is observed only to belong to an interval instead of being known exactly [22]. It is easy to see that such data can naturally occur in many medical research and studies such as clinical trials and periodic follow-up studies among others. Also interval-censored data can arise in different formats, including case I, case II and case *K* interval-censored

CONTACT Shuying Wang wangshuying0601@163.com

data. Case I interval-censored data are often referred as current status data [8,13], meaning that each subject is only observed one time and thus the failure time is either left- or right-censored. In the case of case II interval-censored data, there exist two observation time on each subject and in this case, the failure time of interest can be left-, right- or interval-censored. Besides two previous situations, in an experiments, individuals may experience several observation times and the failure event is only known to occur between two consecutive observation time points. This kind of interval-censored data is often referred as case K interval-censored data [11,12,19,21].

Many researchers have investigated regression analysis of interval-censored failure time data. For example, [8,17,18,35] considered the fitting of the proportional hazards model to interval-censored data and developed some maximum likelihood estimation (MLE) procedures. For the same problem, [4,31] studied the use of the proportional odds model and the accelerated failure time model. Note that sometimes these models may not fit data well and instead one may prefer a different model or a different model such as the AH model may be preferred [34]. One such situation, for example, is when it is known that the hazard or survival functions corresponding to different covariates may cross each other.

As mentioned above, a great deal of literature has been established for inference about the AH model based on right-censored data. For example, [3,36] developed some estimating equation-based methods, and [5] gave a Bayesian approach with a spline prior for the smooth survival density under parametric distribution families. Also [2] compared the AH model to several other popular classes of regression models and provided some test procedures. A couple of methods also exist for fitting the AH model to interval-censored data [14,24]. However, these methods have one major limitation or restriction that they apply only to the situation of independent interval censoring. It is well known that dependent or informative censoring, meaning that the censoring mechanism may be related to the failure time of interest, can occur sometimes. In these cases, the analysis that ignores its presence would result in biased or misleading estimators or results [30,32,33]. In the following, we will consider case K informatively interval-censored data.

An example of informative censoring is given by a medical follow-up study where the patients may pay clinical visits based on their own health conditions instead of pre-specified schedules. To deal with informative censoring, two types of commonly used methods are the copula model approach [22,27,34] and the frailty model approach [28–30,32,33]. The former employs a copula model to describe the correlation between the censoring mechanism and the failure time of interest, while the latter uses latent variables to characterize the relationship between them [28–30]. In the following, we will adopt the frailty model approach and develop a two-step sieve maximum likelihood estimation procedure with the use of Bernstein polynomials.

The remainder of this paper is organized as follows. In Section 2, some notation, models and assumptions that will be used throughout the paper will be introduced along with the data structure. The proposed estimation procedure will be described in Section 3 and Section 4 will establish the asymptotic properties of the resulting estimators. Some results from a simulation study will be presented in Section 5 and suggest that the proposed method works well in practical situations. The proposed methodology is applied to a set of real data arising from an AIDS clinical trial in Section 6, and Section 7 gives some discussion and concluding remarks.

2. Notation, models and assumptions

Consider a failure time study that consists of n independent subjects and let T_i represent the failure time of interest for subject i. Also let x_i denote a p-dimensional vector of covariates and $U_{i0} = 0 < U_{i1} < U_{i2} < \cdots < U_{iK_i}$ represent a sequence of observation times, where K_i denotes the number of observations on the subject. We assume that given x_i and a latent variable b_i, T_i follows the following AH frailty model:

$$\lambda_i(t|x_i, b_i) = \lambda_0(t \exp(\boldsymbol{\beta_1}^\top x_i + \beta_2 b_i)). \tag{1}$$

Here $\lambda_0(\cdot)$ denotes an unknown baseline hazard function and $\boldsymbol{\beta} = (\boldsymbol{\beta_1}^\top, \beta_2)^\top$ the vector of regression parameters. Correspondingly, the cumulative hazard function of T_i has the form

$$\Lambda_i(t|x_i, b_i) = \Lambda_0(t \exp(\boldsymbol{\beta_1}^\top x_i + \beta_2 b_i)) \exp(-\boldsymbol{\beta_1}^\top x_i - \beta_2 b_i), \tag{2}$$

where $\Lambda_0(\cdot) = \int_0^t \lambda_0(s)\, \mathrm{d}s$, the unknown baseline cumulated baseline hazard function.

To describe the observation process, define the point process $\tilde{N}_i(t) = \sum_{j=1}^{K_i} I(U_{ij} \le t)$, which only jumps at each observation time point for individual i. In the following, we will assume that $\tilde{N}_i(t)$ is a non-homogeneous Poisson process satisfying the proportional rate model

$$E(\mathrm{d}\tilde{N}_i|x_i, b_i) = \lambda_{0h}(t) \exp(\boldsymbol{\alpha}^\top x_i + b_i)\, \mathrm{d}t \tag{3}$$

given x_i and b_i. In the above, $\lambda_{0h}(t)$ denotes an unknown continuous baseline rate function and $\boldsymbol{\alpha}$ a vector of regression parameters. Some comments on this will be given below. Also it will be assumed that given x_i and b_i, T_i and $\tilde{N}_i(t)$ are independent.

Let $\delta_{ij} = I(U_{i,j-1} < T_i \le U_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, K_i$, and τ_i denote the follow-up or stopping time for individual i, which is also assumed to be independent of T_i given x_i and b_i. Then the observed data have the form

$$O = \{O_i = (\tau_i, U_{ij}, \delta_{ij}, x_i, j = 1, \ldots, K_i), i = 1, \ldots, n\}.$$

That is, one only observes case K interval-censored data. Under the assumptions above and given the b_i's, the conditional observed likelihood function has the form

$$L(\boldsymbol{\beta}, \Lambda_0|b_i) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{K_i} \left[S_i(U_{i,j-1}) - S_i(U_{i,j})\right]^{\delta_{ij}} S_i(U_{iK_i})^{1-\sum_{j=1}^{K_i} \delta_{ij}} \right\}, \tag{4}$$

where $S_i(t) = \exp[-\Lambda_0(t \exp(\boldsymbol{\beta_1}^\top x_i + \beta_2 b_i)] \exp(-\boldsymbol{\beta_1}^\top x_i - \beta_2 b_i)$. In the next section, we will discuss the estimation of the regression parameter $\boldsymbol{\beta}$ or inference on models (1) and (3).

3. Inference procedure

Now we consider inference about models (1) and (3) and for this, we propose a two-step sieve maximum likelihood estimation procedure by using the strength-borrowing idea discussed in [9,28] among others. In the method, we first consider estimation about model (3) and then the estimation of $\boldsymbol{\beta}$.

To describe the first step, define $N_i(t) = \tilde{N}_i(t \wedge \tau_i)$ for subject i, $i = 1, 2, \ldots, n$. Then we have that

$$N_i(t) = \int_0^{t\wedge\tau_i} \mathrm{d}\tilde{N}_i(u) = \int_0^t I(\tau_i \geq u)\, \mathrm{d}\tilde{N}_i(u),$$

and one can easily show that given x_i and b_i,

$$E\{\mathrm{d}N_i(t)|x_i, b_i\} = E\{I(\tau_i \geq t)|x_i, b_i\} \cdot \exp(x_i^\top \boldsymbol{\alpha} + b_i)\, \mathrm{d}\Lambda_{oh}(t).$$

It follows that $E\{\mathrm{d}N_i(t)\} = E\{\exp(x_i^\top \boldsymbol{\alpha} + b_i)I(\tau_i \geq t)\}\, \mathrm{d}\Lambda_{oh}(t)$ or

$$\mathrm{d}\Lambda_{oh}(t) = \frac{1}{E\{\exp(x_i^\top \boldsymbol{\alpha} + b_i)I(\tau_i \geq t)\}\, \mathrm{d}E\{N_i(t)\}}. \tag{5}$$

In addition, note that $E\{\mathrm{d}N_i(t)|x_i, b_i\} = \Lambda_{0h}(t) \cdot \exp(x_i^\top \boldsymbol{\alpha} + b_i)E\{I(\tau_i \geq t)|x_i, b_i\}$. Then by taking the expectation of $E\{\mathrm{d}N_i(t)|x_i, b_i\}$ and after some calculations, we have that

$$\Lambda_{0h} = \frac{1}{E\{\exp(x_i^\top \boldsymbol{\alpha} + b_i)I(\tau_i \geq t)\}} E\{N_i(t)I(\tau_i \geq t)\} \tag{6}$$

and

$$\log \Lambda_{0h}(\tau_0) - \log \Lambda_{0h}(t) = \int_t^{\tau_i} \frac{1}{N_i(s)I(\tau_i \geq s)}\, \mathrm{d}E\{N_i(s)\}.$$

Let s_l's denote the ordered and distinct values of the observation times U_{ij}'s, $d_{(l)}$ the number of the observation times equal to $s_{(l)}$, and $R_{(l)}$ the total number of observation events with observation times and observation terminating time satisfying $U_{ij} \leq s_{(l)} \leq \tau_i$. The equation above suggests that one can estimate $\lambda_{0h}(t)$ by

$$\hat{\lambda}_{0h}(t) = \prod_{t\leq s\leq\tau_i} \left(1 - \frac{1}{E(N_i(s)I(\tau_i \geq s))}\right) \mathrm{d}E\{N_i(s)\},$$

and estimate $\Lambda_{0h}(t)$ by the nonparametric maximum likelihood estimator

$$\hat{\Lambda}_{0h}(t) = \prod_{s_{(l)}>t} \left(1 - \frac{d_{(l)}}{R_{(l)}}\right).$$

For estimation of the regression parameter $\boldsymbol{\alpha}$, note that

$$E[K_i|x_i, b_i, \tau_i] = \Lambda_{0h}(\tau_i) \exp(\boldsymbol{\alpha}^\top x_i + b_i),$$

which gives $E[K_i\Lambda_{0h}^{-1}(\tau_i) - E(e^{b_i}) \exp(\boldsymbol{\alpha}^\top x_i)] = 0$. It follows that one can define a class of estimating equations

$$\sum_{i=1}^n \omega_i\tilde{x}_i(K_i\hat{\Lambda}_{0h}^{-1}(\tau_i) - E(e^{b_i}) \exp(\boldsymbol{\alpha}^\top x_i)) = 0,$$

where $\tilde{x}_i^\top = (1, x_i^\top)$, and the ω_i's are some weights. Let $\hat{\boldsymbol{\alpha}}$ denote the estimator of $\boldsymbol{\alpha}$ given by the equation above. Then it is natural to estimate b_i by

$$\hat{b}_i = \log\left\{\frac{K_i}{\hat{\Lambda}_{0h}(\tau_i) \exp(\hat{\boldsymbol{\alpha}}^\top x_i)}\right\}. \tag{7}$$

Now we discuss the second step or estimation of model (1). For this, define the parameter space $\Phi = \{\theta = (\boldsymbol{\beta}, \Lambda) \in \mathcal{B} \otimes \mathcal{M}\}$ and the sieve space

$$\Phi_n = \{\theta_n = (\boldsymbol{\beta}_n, \Lambda_n) \in \mathcal{B}_n \otimes \mathcal{M}_n\},$$

where $\mathcal{B} = \{\boldsymbol{\beta} | \boldsymbol{\beta} \in R^{p+1}, ||\boldsymbol{\beta}|| \le M\}$, $\mathcal{B}_n = \{\boldsymbol{\beta}_n | \boldsymbol{\beta}_n \in R^{p+1}, ||\boldsymbol{\beta}_n|| \le M\}$, and M is a positive constant. Also define $\mathcal{M}_n = \bigcup_{\beta \in \mathcal{B}} \mathcal{M}_n^{\beta}$, where

$$\mathcal{M}_n^{\boldsymbol{\beta}} = \Bigg\{ \hat{\Lambda}_n(t) = \sum_{k=0}^{m} \phi_k B_k(t \exp(x_i^\top \boldsymbol{\beta_1} + \beta_2 \hat{b}_i), m, c, u) \exp(-x_i^\top \boldsymbol{\beta_1} - \beta_2 \hat{b}_i) :$$

$$\sum_{0 \le k \le m} \phi_k \le M_n, 0 \le \phi_0 \le \phi_1 \le, \dots, \le \phi_m \Bigg\}.$$

In the above,

$$B_k(t, m, c, u) = \binom{m}{k} \left(\frac{t-c}{u-c}\right)^k \left(1 - \frac{t-c}{u-c}\right)^{(m-k)}, \quad 0 \le c \le u < \infty,$$

the Bernstein basis polynomial of the degree $m = O(n^{\upsilon})$ for some $\upsilon \in (0, 1)$ with $[c, u]$ denoting the range of $t \exp(x_i^\top \beta_1 + \beta_2 \hat{b}_i)$.

Let $\boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_m)^\top$. Then over the sieve space, the log of the conditional likelihood function given in (4) can be written as

$$l\left(\boldsymbol{\beta}, \boldsymbol{\phi} \mid \hat{b}_i' s\right) = \sum_{i=1}^{n} \Big[\delta_{1i} \log\left(1 - \hat{S}_i\left(U_{i1}\right)\right) + \delta_{i2} \log\left(\hat{S}_i\left(U_{i1}\right) - \hat{S}_i\left(U_{i2}\right)\right) + \dots$$
$$+ \delta_{i,K_i-1} \log\left(\hat{S}_i\left(U_{i,K_i-1}\right) - \hat{S}_i\left(U_{iK_i}\right)\right) + \delta_{K_i} \log\left(\hat{S}_i\left(U_{iK_i}\right)\right) \tag{8}$$

after replacing the b_i's by the $\hat{b}_i$'s, where

$$\hat{S}_i(t) = \exp\left(- \sum_{k=0}^{m} \phi_k B_k(t \exp(x_i^\top \beta_1 + \beta_2 \hat{b}_i), m, c, u) / \exp(x_i^\top \beta_1 + \beta_2 \hat{b}_i) \right).$$

It is thus natural to estimate $\boldsymbol{\beta}$ and $\Lambda_0(t)$ or $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ by maximizing the log conditional likelihood function given in (8). For the maximization, we suggest to employ the following algorithm:

- *Step* 0. Select initial estimators $\hat{\boldsymbol{\beta}}^{\mathbf{0}}$ and $\hat{\boldsymbol{\phi}}^{\mathbf{0}}$.
- *Step* 1. In the qth iteration of the algorithm with the fixed value of $\hat{\boldsymbol{\beta}}^{(\boldsymbol{q-1})} = (\hat{\boldsymbol{\beta}}_{\mathbf{1}}^{(q-1)\top}, \hat{\beta}_2^{(q-1)})^\top$, calculate $\hat{b}_i$ from (7). Let $D_q = \{U_{ij} \exp(\hat{\boldsymbol{\beta}}_{\mathbf{1}}^{(q-1)\top} x_i + \hat{\beta}_2^{(q-1)} \hat{b}_i)$, $i = 1, \dots, n; j = 1, \dots, K_i\}$ and $[c^{(q)}, u^{(q)}] = [\min(D_q), \max(D_q)]$. Then update $\boldsymbol{\phi}^{(\boldsymbol{q})}$ by maximizing the log-likelihood $l(\hat{\boldsymbol{\beta}}^{(\boldsymbol{q-1})}, \boldsymbol{\phi} | \hat{b}_i' s)$ with respect to $\boldsymbol{\phi}$.
- *Step* 2. Update $\hat{\boldsymbol{\beta}}^{(\boldsymbol{q})}$ by maximizing the log-likelihood $l(\boldsymbol{\beta}, \hat{\boldsymbol{\phi}}^{(\boldsymbol{q})} | \hat{b}_i' s)$ with respect to $\boldsymbol{\beta}$.

- *Step* 3. Update $\hat{\Lambda}^{(q)}(t)$ with the current estimates $\hat{\boldsymbol{\beta}}^{(q)}$ and $\hat{\boldsymbol{\phi}}^{(q)}$ as

$$\hat{\Lambda}^{(q)}(t) = \sum \phi_k B_k \left(t \exp\left(x^\top \boldsymbol{\beta}_1^{(q)} + \beta_2^{(q)} \hat{b} \right), m, c, u \right) \exp\left(-x^\top \boldsymbol{\beta}_1^{(q)} - \beta_2^{(q)} \hat{b} \right);$$

- *Step* 4. Repeat *Step* 1 – *Step* 3 until convergence.

Note that before the implementation of the algorithm above, one may employ the reparameterization $\phi_k = \sum_{l=0}^{j} \exp(\phi_l^*)$ to remove the restriction on the ϕ_k's. Also in the numerical study below, we have tried different initial estimators and it seems that the algorithm is robust to the initial values. Let $\hat{\boldsymbol{\beta}}_n$ and $\hat{\Lambda}_n(t)$ denote the estimators of $\boldsymbol{\beta}$ and $\Lambda_0(t)$ defined above. In the following section, we will establish their asymptotic properties.

4. Asymptotic properties

To establish the asymptotic properties of $\hat{\boldsymbol{\beta}}_n$ and $\hat{\Lambda}_n(t)$, assume that there exist censoring variables $0 < U_1 < U_2 < \cdots U_{n^*}$, which are conditionally independent of T given a p-dimensional covariate vector X and b, where n^* is the number of observation times. Let $L = \max(U_l : U_l < T, l = 1, \dots n^*)$ and $R = \min(U_l : U_l \geq T, l = 1, \dots n^*)$. If T is left-censored, we have $L = 0$ and $R = U_1$, and if T is right-censored, we have $L = U_{n^*}$ and $R = \infty$. Define $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \log(\Lambda(t)))^\top$ and the distance

$$d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||^2 + || \log \Lambda_1 - \log(\Lambda_2)||^2_{\varpi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)})^{1/2}$$

for any two $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ [24], where

$$\begin{aligned} || \log \Lambda_1(t) - \log \Lambda_2(t)||^2_{\varpi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)} = {} & E[(\log \Lambda_1(L \exp(\boldsymbol{\beta}_1{}^\top X)) - \log \Lambda_2(L \exp(\boldsymbol{\beta}_2{}^\top X)))^2 | b] \\ & + E[(\log \Lambda_1(R \exp(\boldsymbol{\beta}_1{}^\top X)) \\ & - \log \Lambda_2(R \exp(\boldsymbol{\beta}_2{}^\top X)))^2 | b]. \end{aligned}$$

We need the following regularity conditions.

(A1) There exists a positive constant M_X such that $|X| \leq M_X$ almost surely, and $E(XX^\top)$ is nonsingular.
(A2) The parameter space $\mathcal{B}$ is compact, and for any $\beta \in \mathcal{B}$, $||\beta|| < M_{\mathcal{B}}$.
(A3) There exists a finite interval $[a_1, a_2] \subset R^+$ such that the supports of observation process are subsets of $[a_1, a_2]$ and there exists a constant $\xi > 0$ such that $P(U_k - U_l > \xi) = 1$ for any $0 \leq k < l \leq n^*$.
(A4) The first derivative of $\log(\Lambda_0)$ is strictly positive. Furthermore, $\log(\Lambda_0)$ has bounded *rth* derivatives on the interval $[a_1 \exp(-M_{\mathcal{B}} M_X), a_2 \exp(M_{\mathcal{B}} M_X)]$.
(A5) The conditional joint density of the observation process given covariate X and the latent variable b is bounded.
(A6) The failure time T and the observation process are conditionally independent given the covariate x and the latent variable b. In addition, the joint distribution of $U_1, \dots U_{n^*}$ and X does not involve the parameters β_0 and Λ_0.
(A7) For the follow-up time τ and latent variable b, we have that $P(\tau \geq \tau_0, \exp(b) > 0) > 0$. In addition, $P(\tau > \tau_\epsilon) = 1$, where $\tau_\epsilon = \inf\{t : \Lambda_{0h} > \epsilon\}$ for some $\epsilon > 0$ and

$E\{\tilde{N}(\tau)^2\} < \infty$. For the latent variable b, the variable $\exp(b)$ is bounded and there exists a positive small constant $\xi > 0$ such that $\exp(b) > \xi$ almost surely.

Note that most of the conditions above are commonly required under the interval censoring. Now we are ready to establish the asymptotic properties of $\hat{\theta}_n$.

Theorem 4.1: *Suppose that the regularity conditions* (A1)–(A4) *and* (A7) *given above hold. Then the estimator $\hat{\theta}_n$ converges to θ_0, the true value of θ, in probability as $n \to \infty$.*

Theorem 4.2: *Suppose that the regularity conditions* (A1)–(A4) *and* (A7) *given above hold. Then we have that $d(\theta_n, \theta_0) = O_p(n^{-\frac{1}{2}(v-1)} + n^{-\frac{rv}{2}})$ as $n \to \infty$.*

Theorem 4.3: *Suppose that the regularity conditions* (A1)–(A7) *given above hold. Then as $n \to \infty$, we have that*

$$n^{-1/2}(\hat{\beta}_n - \beta_0) = n^{-1/2} I^{-1}(\beta_0) \sum_{i=1}^{n} l^*(\beta_0, \xi_0(\cdot, \beta_0)|O) + o_p(1) \xrightarrow{d} N(0, I^{-1}(\beta_0)),$$

where $I^{-1}(\beta_0) = P\{l^*(\beta_0, \xi_0(\cdot, \beta_0)|O)^{\otimes 2}\}$, $l^*(\beta_0, \xi_0(\cdot, \beta_0)|O)$ is the efficient score function of β defined in the Appendix, P is the probability measure, and $a^{\otimes 2} = aa^\top$ for any arbitrary vector a.

The proof of the results above will be sketched in the Appendix. For inference about $\boldsymbol{\beta}$, it is apparent that we need to estimate the covariance matrix of $\hat{\boldsymbol{\beta}}_n$. For this, a commonly used approach would be to derive a consistent estimator of $I^{-1}(\beta_0)$ or apply the profile likelihood approach but both would have no explicit expressions. Corresponding to this, we suggest to apply the simple bootstrap approach [10] to estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$ by

$$\hat{\Sigma} = \frac{1}{B-1} \sum_{b^*=1}^{B} \left\{ \hat{\boldsymbol{\beta}} - \frac{1}{B} \sum_{b^*=1}^{B} \hat{\boldsymbol{\beta}}^{*(b^*)} \right\}^{\otimes 2}.$$

In the above, B denotes the number of bootstrap samples, $\hat{\boldsymbol{\beta}}^{*(b^*)}$ the estimator of β based on the b^*th bootstrap sample of size n drawn randomly from the observed data O with replacement, and $a^{\otimes 2} = aa^\top$.

5. A simulation study

In this section, we conduct a simulation study to examine the performance of the proposed method. In the study, we focus on the estimation of the regression parameter $\boldsymbol{\beta}$. For each individual i, we generate covariates x_i under two scenarios. More specifically, in scenario 1, we only consider one single covariate generated from either $B(1, 0.5)$, the Bernoulli distribution with the success probability 0.5, or the standard normal distribution $N(0, 1)$. In scenario 2, we consider two covariates with one following $B(1, 0.5)$ and the other following $B(1, 0.5)$ or $N(0, 1)$. To generate the true failure times T_i's, we also need to

Table 1. Simulation results on estimation of β with one single covariate.

	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
	$n = 300$				$n = 500$			
Para.	$x \sim B(1, 0.5)$							
$\beta_1 = -0.5$	0.0095	0.3622	0.3307	0.948	−0.0079	0.2856	0.2704	0.939
$\beta_2 = -0.5$	−0.0127	0.2951	0.2914	0.954	0.0043	0.1964	0.2121	0.959
$\beta_1 = -0.5$	0.0198	0.3151	0.3295	0.952	0.0094	0.2555	0.2659	0.942
$\beta_2 = 0$	−0.0200	0.1968	0.2019	0.960	−0.0073	0.1480	0.1519	0.950
$\beta_1 = 0$	−0.0454	0.2928	0.3005	0.967	−0.0399	0.2167	0.2274	0.963
$\beta_2 = -0.5$	0.0080	0.2861	0.2923	0.930	0.0051	0.1942	0.2225	0.939
$\beta_1 = 0$	0.0166	0.3072	0.3067	0.955	−0.0053	0.2305	0.2416	0.953
$\beta_2 = 0.5$	−0.0054	0.1474	0.1733	0.971	0.0045	0.1010	0.1201	0.939
$\beta_1 = 0.5$	−0.0526	0.2488	0.2722	0.959	−0.0477	0.1802	0.2031	0.967
$\beta_2 = 0$	0.0243	0.2030	0.2131	0.933	0.0070	0.1665	0.1610	0.928
$\beta_1 = 0.5$	−0.0405	0.3097	0.3084	0.949	−0.0317	0.2309	0.2337	0.946
$\beta_2 = 0.5$	0.0174	0.1823	0.1948	0.954	0.0131	0.1230	0.1363	0.967
	$x \sim N(0, 1)$							
$\beta_1 = -0.5$	0.0395	0.1144	0.1378	0.971	0.0263	0.0747	0.0927	0.970
$\beta_2 = -0.5$	0.0307	0.2759	0.2875	0.942	0.0223	0.1970	0.2140	0.942
$\beta_1 = -0.5$	0.0439	0.1050	0.1320	0.966	0.0289	0.0681	0.0885	0.969
$\beta_2 = 0$	0.0069	0.2303	0.2566	0.948	0.0023	0.1783	0.1824	0.929
$\beta_1 = 0$	−0.0267	0.1286	0.1297	0.959	−0.0208	0.0963	0.0951	0.933
$\beta_2 = -0.5$	0.0292	0.2786	0.2990	0.934	−0.0193	0.1931	0.2250	0.955
$\beta_1 = 0.5$	−0.0372	0.1107	0.1306	0.958	−0.0291	0.0701	0.0872	0.967
$\beta_2 = 0$	0.0288	0.2223	0.2315	0.938	0.0277	0.1773	0.1781	0.941
$\beta_1 = 0$	0.0136	0.1599	0.1510	0.930	0.0137	0.1223	0.1143	0.947
$\beta_2 = 0.5$	−0.0098	0.1522	0.1836	0.979	0.0023	0.0940	0.1218	0.924
$\beta_1 = 0.5$	−0.0301	0.1383	0.1452	0.966	−0.0291	0.0919	0.1013	0.957
$\beta_2 = 0.5$	0.0266	0.1921	0.2041	0.950	0.0251	0.1470	0.1497	0.929

generate the latent variables b_i's from the normal distribution with mean 0 and the standard error $\sigma = 1$. Then given the x_i's and b_i's, the T_i's are generated under model (2) with $\Lambda_0(t) = (2/3)t^{3/2}$.

For the generation of the observation times. we first generate the number of observation times K_i for subject i based on the Poisson distribution with mean

$$\Lambda_{ih}(\tau_i|x_i, b_i) = \frac{\tau_i \exp(x_i^\top \alpha + b_i)}{4}.$$

where the τ_i's are follow-up times from the uniform distribution over [1,2]. Given K_i, we set $U_{i1}, \ldots, U_{iK_i}$ to be the order statistics of a random sample of size K_i from the uniform distribution over $(0, \tau_i)$. We choose $n = 300$ or 500 as sample size, $B = 100$ as the bootstrap sample size and $m = [n^{1/4}]$ for the degree of freedoms of for the Bernstein polynomials. All results below are based on 1000 replications.

Table 1 presents the results on the estimation of $\boldsymbol{\beta}$ given by the proposed estimation approach for the situation of one single covariate with the true values of β_1 and β_2 being −0.5, 0 or 0.5 and the true value of α being 0.15. In the table, 'Bias' represents the empirical bias given by the average of the estimates minus the true value, 'SSE' the sample standard deviation of the estimates, 'ESE' the average of the estimated standard errors, and 'CP' the 95% empirical coverage probability. The results suggest that the proposed estimator seems to be unbiased and the suggested variance estimation procedure also appears to be reasonable. In addition, the empirical coverage probabilities are close to 0.95, indicating that

Table 2. Simulation results on estimation of $\boldsymbol{\beta}$ with two covariates, x_{11} and x_{12}. $x_{11} \sim B(1, 0.5)$, $x_{12} \sim B(1, 0.5)$ or $N(0, 1)$.

	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
Para.	$x_{12} \sim B(1, 0.5)$				$x_{12} \sim N(0, 1)$			
$\beta_{11} = -0.5$	0.0295	0.3440	0.3149	0.943	0.0406	0.3324	0.3387	0.953
$\beta_{12} = -0.5$	−0.0247	0.3755	0.3208	0.936	0.0399	0.1277	0.1468	0.963
$\beta_2 = -0.5$	0.0310	0.3575	0.3674	0.941	0.0486	0.2509	0.3338	0.942
$\beta_{11} = -0.5$	0.0387	0.3073	0.3020	0.943	0.0548	0.3123	0.3315	0.946
$\beta_{12} = -0.5$	0.0163	0.3220	0.3081	0.962	0.0502	0.1197	0.1407	0.962
$\beta_2 = 0$	0.0139	0.1915	0.1855	0.956	0.0201	0.2398	0.2639	0.942
$\beta_{11} = -0.5$	0.0263	0.3457	0.3327	0.945	0.0644	0.3143	0.3387	0.950
$\beta_{12} = -0.5$	0.0379	0.3549	0.3259	0.929	0.0543	0.1221	0.1437	0.968
$\beta_2 = 0.5$	0.0066	0.1572	0.1690	0.972	0.0187	0.2074	0.2148	0.948
$\beta_{11} = 0$	−0.0359	0.2753	0.2941	0.943	−0.0573	0.2922	0.3031	0.964
$\beta_{12} = 0$	−0.0173	0.3055	0.2971	0.936	−0.0276	0.1362	0.1387	0.940
$\beta_2 = -0.5$	0.0685	0.3042	0.3025	0.941	0.0217	0.2566	0.3739	0.963
$\beta_{11} = 0$	−0.0034	0.3102	0.3051	0.950	−0.0151	0.2937	0.3058	0.957
$\beta_{12} = 0$	−0.0007	0.2902	0.3055	0.957	0.0129	0.1541	0.1411	0.938
$\beta_2 = 0.5$	−0.0161	0.1589	0.1763	0.963	−0.0206	0.1602	0.1794	0.966
$\beta_{11} = 0.5$	−0.0480	0.3035	0.3194	0.956	0.0244	0.3122	0.3287	0.950
$\beta_{12} = 0.5$	−0.0447	0.3079	0.3151	0.951	0.0367	0.1413	0.1453	0.956
$\beta_2 = -0.5$	0.0327	0.2448	0.3101	0.949	−0.0217	0.2266	0.2739	0.953
$\beta_{11} = 0.5$	−0.0503	0.2700	0.2933	0.959	−0.0370	0.2911	0.3014	0.950
$\beta_{12} = 0.5$	−0.0541	0.2880	0.2957	0.952	−0.0467	0.1313	0.1458	0.958
$\beta_2 = 0$	0.0329	0.2156	0.2252	0.937	0.0208	0.2340	0.2508	0.943
$\beta_{11} = 0.5$	−0.0528	0.3196	0.3176	0.949	−0.0451	0.3113	0.3169	0.949
$\beta_{12} = 0.5$	−0.0225	0.3168	0.3173	0.947	−0.0431	0.1386	0.1533	0.946
$\beta_2 = 0.5$	0.0038	0.2044	0.2097	0.949	−0.0133	0.2074	0.2182	0.950

Table 3. Simulation results on estimation of $\boldsymbol{\beta}$ with different m.

		Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
	Para.	$x \sim B(1, 0.5)$				$x \sim N(0, 1)$			
$m = 3$	β_1=0.5	−0.0362	0.2828	0.3089	0.956	−0.0321	0.2979	0.3125	0.950
	β_2=0.5	0.1030	0.1728	0.1966	0.969	0.0101	0.1680	0.1889	0.948
$m = 4$	β_1=0.5	−0.0313	0.2943	0.3116	0.966	−0.0331	0.2989	0.3134	0.951
	β_2=0.5	0.0104	0.1852	0.1986	0.957	0.0101	0.1770	0.1939	0.965
$m = 5$	β_1=0.5	−0.0222	0.3084	0.3171	0.950	−0.0314	0.2309	0.2337	0.946
	β_2=0.5	0.0157	0.1959	0.1958	0.951	0.0130	0.1750	0.1833	0.951

the normal approximation to the distribution of the proposed estimator of the regression parameters appears to be appropriate. As expected, the results become better in general when the sample size increased.

The results on estimation of $\boldsymbol{\beta}$ given by the proposed estimation approach for the two covariate situation with $n = 300$ are given in Table 2. One can see from the table that they are similar to those given in Table 1 and again suggest that the proposed estimation procedures seems to work well for the situations considered. Note that in the above, we have used the fixed degree of freedoms m for the Bernstein polynomials and in practice, one may be interested in seeing how this affects the proposed estimator. To investigate this, we repeated the study giving Table 1 with the true values of β_1 and β_2 being 0.5 and $n = 300$ except setting $m = 3, 4$ or 5 and present the obtained results in Table 3. It is apparent that the results seem to be similar to each other and suggest that the proposed estimation procedure seems to be stable with respect to m.

Table 4. Simulation results on estimation of β with the mixed Poisson observation process.

	Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
Para.	$x \sim B(1, 0.5)$				$x \sim N(0, 1)$			
$\beta_1 = -0.5$	0.0051	0.3493	0.3418	0.938	0.0336	0.1208	0.1372	0.963
$\beta_2 = -0.5$	0.0023	0.2907	0.2860	0.933	0.0492	0.2498	0.2922	0.938
$\beta_1 = -0.5$	0.0439	0.1050	0.1320	0.966	0.0480	0.1058	0.1333	0.964
$\beta_2 = 0$	0.0006	0.1992	0.1945	0.955	0.0326	0.2255	0.2500	0.941
$\beta_1 = 0$	−0.0400	0.2803	0.2904	0.963	−0.0141	0.1298	0.1300	0.952
$\beta_2 = -0.5$	0.0038	0.2721	0.2820	0.952	0.0186	0.2757	0.3014	0.948
$\beta_1 = 0$	0.0164	0.3026	0.3193	0.961	0.0263	0.1571	0.1542	0.937
$\beta_2 = 0.5$	−0.0211	0.1467	0.1538	0.964	−0.0116	0.1539	0.1823	0.971
$\beta_1 = 0.5$	−0.0345	0.2636	0.2901	0.966	−0.0418	0.1226	0.1329	0.940
$\beta_2 = 0$	0.0216	0.2516	0.2887	0.943	0.0109	0.2349	0.2402	0.935
$\beta_1 = 0.5$	−0.0197	0.3009	0.3109	0.957	−0.0354	0.1274	0.1437	0.957
$\beta_2 = 0.5$	−0.0166	0.1882	0.1962	0.959	0.0071	0.1991	0.2038	0.948

Table 5. Simulation results on estimation of β with the covariate generated from $Ga(1, 1)$.

Para.	Bias	SSE	ESE	CP
$\beta_1 = -0.5$	0.0223	0.2419	0.2560	0.960
$\beta_2 = -0.5$	0.0062	0.1895	0.2700	0.940
$\beta_1 = -0.5$	0.0112	0.3536	0.3165	0.960
$\beta_2 = 0$	−0.0244	0.1935	0.2062	0.970
$\beta_1 = 0$	−0.0351	0.1478	0.1669	0.970
$\beta_2 = -0.5$	−0.0140	0.3750	0.2940	0.925
$\beta_1 = 0$	−0.0003	0.1742	0.1876	0.925
$\beta_2 = 0.5$	−0.0171	0.1507	0.1722	0.965
$\beta_1 = 0.5$	−0.0201	0.0388	0.0445	0.965
$\beta_2 = 0$	−0.0131	0.2342	0.2489	0.940
$\beta_1 = 0.5$	−0.0169	0.0462	0.0446	0.965
$\beta_2 = 0.5$	−0.0196	0.3146	0.2887	0.930

Note that in the proposed method, it has been assumed that the observation process is the non-homogeneous Poisson process and it is apparent that this may not be true sometimes. To check its possible effect on the proposed estimator, as suggested by a reviewer, we repeated the study giving Table 1 except using the mixed-Poisson process to generate observation times. Table 4 gives the results obtained with $n = 300$ and one can see that they are similar to those given in Table 1, indicating that the proposed method seems to be robust to the assumption on the observation process. Also by following the suggestion from a reviewer, we again repeated the study yielding Table 1 except generating the covariate from Gamma distribution $Ga(1, 1)$, which is non-symmetric. The obtained results with $n = 300$ are presented in Table 5 and it is easy to see that they gave the same conclusions as before or again indicate that the proposed method works well. In other words, it seems that the skewness has no effects on the proposed method.

6. An application

Now we apply the proposed method to the data arising from an AIDS clinical trial, ACTG 359, which was designed to compare six different antiretroviral treatment regimens for AIDS patients [7,22]. In the study, the patient's blood samples were supposed to be collected and tested at the pre-specified schedule. However, many patients did not follow the pre-specified test times and thus their actual visits differ from each other. More specifically, they may miss their scheduled visits or drop out early. Among others, one variable that was measured at each patient's clinical visit is the number of RNA copies, which is usually used to measure the viral load level of the patient. On the number of RNA copies, one event or time of interest is the first time at which the number of RNA copies drops below the threshold of 500 viral copies/ml, an important indicator for the patient's disease status. However, due to the nature of the study described above, the exact occurrence time of the event was not observed and instead only interval-censored data are available.

For the analysis here, by following [7] and others, we will focus on the 271 patients who had at least one clinical visit during the 12 months follow-up period. All patients are divided into two groups based on their initial numbers of RNA copies below or above 20,000 viral copies/ml. We set $x_i = 1$ if patient i's initial number of RNA was above 20,000 viral copies/ml and otherwise $x_i = 0$. Also for individual i, we define T_i to be the first time when his or her number of RNA copies dropped below 500 viral copies/ ml.

The application of the proposed method with $m = 3$ and $B = 100$ gave $\hat{\beta}_1 = 3.1162$ and $\hat{\beta}_2 = -1.1531$ with the standard errors of 1.1673 and 0.4350, respectively. These correspond to the p-values being 0.0038 and 0.0040, respectively, for testing $\beta_1 = 0$ and $\beta_2 = 0$. These results indicate that the patients in the two groups were significantly different in terms of the rate at which the number of RNA copies drops below the threshold. More specifically, it seems that the patients with higher initial number of RNA copies tended to have a significantly lower rate to drop below the threshold than the others. The results above also suggest that the dropping rate was significantly related to the clinical visit process. For comparison, we also applied the method given in [24], who considered the same situation here but assumed non-informative interval censoring and obtained the estimated covariate effect $\hat{\beta}^* = 2.7555$ with the standard error of 0.9376. Although both analyses gave similar results, it is apparent the latter seems to underestimate the covariate effect.

7. Discussion and concluding remarks

This paper discussed the estimation of or inference on the AH model when one observes case K informatively interval-censored failure time data. For the problem, the frailty-based joint models were proposed to describe the covariate effects and the relationship between the failure time of interest and the informative observation process. Furthermore, a two-step sieve maximum likelihood estimation procedure was developed for estimation and the asymptotic properties of the resulting estimator were established. A simulation study was performed to assess the finite sample performance of the proposed methodology and suggested that it works well in practical situations. An application to an AIDS clinical trial was provided.

As mentioned above, several methods have been proposed for the same or similar problem but all of these existing methods apply only to independent interval censoring. In particular, [24] gave a sieve maximum likelihood estimation approach for the AH model based on case II interval-censored data in addition to assuming non-informative censoring or observation processes. In contrast, the estimation approach proposed above is much more flexible or applies to much more general situations. In the proposed method, for generality, no distribution was assumed for the latent variables b_i's. An alternative to this is to assume a parametric distribution and it is apparent that this may yield more efficient estimation procedure if the assumed distribution is correct. On the other hand, it may be difficult to check such assumption in practice.

There exist several directions on related topics for future research. One is that instead of purely interval-censored data, sometimes one may face partly interval-censored data in which some study subjects may provide exact observations on the failure time of interest. It is apparent that it would be useful to generalize the proposed method to this latter situation. A similar direction is to generalize the proposed estimation procedure to allow for the existence of a cured subgroup. A third direction is that instead of the AH model, sometimes one may prefer a different model or a more general model such as the generalized AH model [14], which includes the proportional hazards model and the AH model as well as other models as special cases. It is easy to see that for this, one would need to develop a new or different estimation procedure.

Acknowledgments

The authors wish to thank the Editor, the Associate Editor and two reviewers for their many comments and suggestions that greatly improved the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was partly supported by the National Natural Science Foundation of China (No. 12071176), China Postdoctoral Science Foundation (No. 2021M700536), Outstanding Youth Fund Project of Jilin Natural Science Foundation (No. 20230101371JC), and Science and Technology Developing Plan of Jilin Province (Nos. 20200201258JC, 20190902012TC, 20190304129YY).

References

[1] Y.Q. Chen, *A Class of Semiparametric Scale-Change Hazards Regression Models and Its Adequacy for Censored Survival Data*, Bepress, 2000.

[2] Y.Q. Chen, *Accelerated hazards regression model and its adequacy for censored survival data*, Biometrics 57 (2001), pp. 853–860.

[3] Y.Q. Chen and M.C. Wang, *Analysis of accelerated hazards models*, J. Am. Stat. Assoc. 95 (2000), pp. 608–618.

[4] M. Chen, X. Tong, and J. Sun, *The proportional odds model for multivariate interval-censored failure time data*, Stat. Med. 26 (2007), pp. 5147–5161.

[5] Y. Chen, T. Hanson, and J. Zhang, *Accelerated hazards model based on parametric families generalized with Bernstein polynomials.*, Biometrics 70 (2014), pp. 192–201.

[6] Y. Ding and B. Nan, *A sieve m-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data*, Ann. Stat. 39 (2011), pp. 2795–3443.

[7] G. Gómez, A. Espinal, and S. Lagakos, *Inference for a linear regression model with an interval-censored covariate*, Stat. Med. 22 (2003), pp. 409–425.

[8] J. Huang, *Efficient estimation for the proportional hazards model with interval censoring*, Ann. Stat. 24 (1996), pp. 540–568.

[9] C.Y. Huang and M.C. Wang, *Joint modeling and estimation for recurrent event processes and failure time data*, J. Am. Stat. Assoc. 99 (2004), pp. 1153–1165.

[10] C.Y. Huang and M.C. Wang, *Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring*, Biometrics 66 (2010), pp. 39–49.
[11] L. Li and Z. Pu, *Regression models with arbitrarily interval-censored observations*, Commun. Stat. Theory Methods 28 (1999), pp. 1547–1563.
[12] L. Li and Z. Pu, *Rank estimation of log-linear regression with interval-censored data*, Lifetime Data Anal. 9 (2003), pp. 57–70.
[13] D.Y. Lin, D. Oakes, and Z. Ying, *Additive hazards regression with current status data*, Biometrika 85 (1998), pp. 289–298.
[14] X. Liu and L. Xiang, *Generalized accelerated hazards mixture cure models with interval-censored data*, Comput. Stat. Data Anal. 161 (2021), pp. 107248–107268.
[15] R. Oller, G. Gómez, and M.L. Calle, *Interval censoring: model characterizations for the validity of the simplified likelihood*, Canad. J. Stat. 32 (2004), pp. 315–326.
[16] R. Oller, G. Gómez, and M.L. Calle, *Interval censoring: identifiability and the constant-sum property*, Biometrika 94 (2007), pp. 61–70.
[17] W. Pan, *Extending the iterative convex minorant algorithm to the Cox model for interval-censored data*, J. Comput. Graph. Stat. 8 (1999), pp. 109–120.
[18] P.S. Rosenberg, *Hazard function estimation using B-Splines*, Biometrics 51 (1995), pp. 874–887.
[19] A. Schick. and Q. Yu, *Consistency of the GMLE with mixed case interval-censored data*, Scand. J. Stat. 27 (2000), pp. 45–55.
[20] X. Shen and W.H. Wong, *Convergence rate of sieve estimates*, Ann. Stat. 22 (1994), pp. 580–615.
[21] S. Song, *Estimation with univariate mixed case interval censored data*, Stat. Sin. 14 (2004), pp. 269–282.
[22] J. Sun, *The Statistical Analysis of Interval-censored Failure Time Data*, Springer, 2006.
[23] T. Sun and Y. Ding, *Copula-based semiparametric regression method for bivariate data under general interval censoring*, Biostatistics 22 (2021), pp. 315–330.
[24] Z. Szabo, X. Liu, and L. Xiang, *Semiparametric sieve maximum likelihood estimation for accelerated hazards model with interval-censored data*, J. Stat. Plan. Inference 205 (2020), pp. 175–192.
[25] A.W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge, 1998.
[26] A.W. van der Vaart and J.A. Wellner, *Weak Convergence and Empirical Processes*, Springer-Verlag, New York, 1996.
[27] L. Wang, J. Sun, and X. Tong, *Efficient estimation for the proportional hazards model with bivariate current status data*, Lifetime Data Anal. 14 (2008), pp. 134–153.
[28] P. Wang, H. Zhao, and J. Sun, *Regression analysis of case K interval-censored failure time data in the presence of informative censoring*, Biometrics 72 (2016), pp. 1103–1112.
[29] S. Wang, C. Wang, P. Wang, and J. Sun, *Semiparametric analysis of the additive hazards model with informatively interval-censored failure time data*, Comput. Stat. Data Anal. 125 (2018), pp. 1–9.
[30] S. Wang, D. Xu, C. Wang, and J. Sun, *Semiparametric analysis of case k interval-censored failure time data in the presence of a cured subgroup and informative censoring*, J. Stat. Comput. Simul. 91 (2021), pp. 1–16.
[31] M. Zhang and M. Davidian, *'Smooth' semiparametric regression analysis for arbitrarily censored time-to-event data*, Biometrics 64 (2008), pp. 567–576.
[32] Z. Zhang, J. Sun, and L. Sun, *Statistical analysis of current status data with informative observation times*, Stat. Med. 24 (2005), pp. 1399–1407.
[33] Z. Zhang, L. Sun, J. Sun, and D.M. Finkelstein, *Regression analysis of failure time data with informative interval censoring*, Stat. Med. 26 (2007), pp. 2533–2546.
[34] B. Zhang, X. Tong, and J. Sun, *Efficient estimation for the proportional odds model with bivariate current status data*, Far. East J. Theor. Stat. 27 (2009), pp. 113–132.
[35] Y. Zhang, L. Hua, and J. Huang, *A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data*, Scand. J. Stat. 37 (2010), pp. 338–354.
[36] J. Zhang, Y. Peng, and O. Zhao, *A new semiparametric estimation method for accelerated hazard model*, Biometrics 67 (2011), pp. 1352–1360.

[37] Q. Zhou, T. Hu, and J. Sun, *A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data*, J. Amer. Stat. Assoc. 112 (2017), pp. 664–672.
[38] G.G Lorentz. Bernstein Polynomials (2nd ed.), New York: Chelsea Publishing, 1986.

Appendices

In this Appendix, we will sketch the proofs of Theorems 4.1–4.3 described above mainly by using the empirical process theory developed in van der [27–29]. First of all we need proof a lemma.

Lemma A.1: *Model* (2) *is identifiable if given b, Λ_0 is not a linear function but it is continuous and X is not a degenerate variable.*

Proof of Lemma A.1.: Assume there are two parameter vectors $(\boldsymbol{\beta}^\top, \Lambda_0)$ and $(\tilde{\boldsymbol{\beta}}^\top, \tilde{\Lambda}_0)$ in Φ. And for every one sample data set O_i and latent variable b_i, we have $(\boldsymbol{\beta}^\top, \Lambda_0) \neq (\tilde{\boldsymbol{\beta}}^\top, \tilde{\Lambda}_0)$ but $l(\boldsymbol{\beta}, \Lambda_0|O_i, b_i) = l(\tilde{\boldsymbol{\beta}}, \tilde{\Lambda}_0|O_i, b_i)$.

Note that given X_i and b_i, L_i and R_i are conditionally independent of T_i the censoring is non-informative, which fit the condition of [15] and imply that model (2) satisfies the constant-sum property. Following Theorem 1 of [16], it means $l(\boldsymbol{\beta}, \Lambda_0|O_i, b_i) = l(\tilde{\boldsymbol{\beta}}, \tilde{\Lambda}_0|O_i, b_i)$ is equivalent with $S(\boldsymbol{\beta}, \Lambda_0|X_i, b_i) = S(\tilde{\boldsymbol{\beta}}, \tilde{\Lambda}_0|X_i, b_i)$. By taking the negative logarithm of the survival function and assuming Λ_0 and $\tilde{\Lambda}_0$ are differentiable with respect to t, we get

$$\lambda_0(t\exp(\boldsymbol{\beta}_1^\top X_i + \beta_2 b_i)) = \tilde{\lambda}_0(t\exp(\tilde{\boldsymbol{\beta}}_1^\top X_i + \tilde{\beta_2} b_i)).$$

Take $t' = t\exp(\boldsymbol{\beta}_1^\top X_i + \beta_2 b_i)$ into the equation above, we get $\lambda_0(t') = \tilde{\lambda}_0(t'\exp((\boldsymbol{\beta}_1 - \tilde{\beta}_1)^\top X_i + (\beta_2 - \tilde{\beta}_2)b_i))$. Note that for fixed t' the left side is constant which does not depend on the change of X and b_i'. To take more discussion about that, let $X_i' \neq X_i$, $b_i' \neq b$ and substitute $t'' = t'\exp((\boldsymbol{\beta}_1 + \tilde{\boldsymbol{\beta}}_1)^\top X_i' - (\beta_2 - \tilde{\beta}_2)b_i')$ into the equation, we get

$$\lambda_0(t'') = \lambda_0(t''\exp((\boldsymbol{\beta}_1 - \tilde{\boldsymbol{\beta}}_1)^\top (X_i' - X_i) + (\beta_2 - \tilde{\beta}_2)(b_i' - b_i))).$$

For simplify, let $\varrho = (\boldsymbol{\beta}_1 - \tilde{\boldsymbol{\beta}}_1)^\top (X_i' - X_i) + (\beta_2 - \tilde{\beta}_2)(b_i' - b_i)$. We have the following equation by applying the same scaling method previous k times

$$\lambda_0(t'') = \tilde{\lambda}_0(t''\exp(k\varrho)).$$

It is clear that $\varrho \neq 0$. Without loss of generality, we assume that $\varrho \leq 0$. Then as $n \to 0$, $t''\exp(k\varrho) \to \infty$. By the assumption that $\tilde{\lambda}_0(t'')$ is continuous, $\lambda_0(0) = \tilde{\lambda}_0(t'')$ for any $t'' \in [0, \infty)$ and $\lambda_0(t)$ is a constant function with respect to t. It violate the assumption $\Lambda_0(\cdot)$ is not linear function. So model (2) is identifiable. ■

Proof of Theorem 4.1.: The proof of consistency of $\boldsymbol{\theta}_n$ can be built by verifying three conditions of Theorem 5.7 of [25]. To this aim, let $l(\theta|O, b)$ be the log-likelihood based on the observed data and latent variable b in Section 2. $Pl(\boldsymbol{\theta}|O, b)$ denotes the expectation of the likelihood function of one element sample with respect to the true model and $P_n l(|O, \hat{b})$ the expectation with respect to the empirical distribution based on the n-element sample and the estimated b. Throughout the proof, we denote an arbitrary positive constant by C, which does not depend on n, θ_0. Denote $M(\theta) = pl(\boldsymbol{\theta}|O, b)$ and $M_n(\boldsymbol{\theta}) = p_n l(\boldsymbol{\theta}|O, \hat{b})$.

The first condition in Theorem 5.7 to verify is that

$$\sup_{\boldsymbol{\theta}\in\Theta} |M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| \to 0.$$

Note that $|M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| \leq |p_n l(\boldsymbol{\theta}|O, \hat{b}) - pl(\boldsymbol{\theta}|O, \hat{b})| + |pl(\boldsymbol{\theta}|O, \hat{b}) - pl(\boldsymbol{\theta}|O, b)|$. For the second part of right side $\hat{b} \xrightarrow{p} b$ and by delta method, we have $|pl(\boldsymbol{\theta}|O, \hat{b}) - pl(\boldsymbol{\theta}|O, b)| \xrightarrow{p} 0$.

For the first part of the right side, we find that when b is replaced by $\hat{b}$ it has the similar condition in [24]. Under condition (A2) and for any $\epsilon > 0$, $\boldsymbol{\beta} \in \boldsymbol{B}$, there exists $\boldsymbol{\beta_s}$ with such that $|\boldsymbol{\beta_s} - \boldsymbol{\beta}| < \epsilon$, where $1 \leq s \leq C(1/\epsilon)^{p+1}$ and C is a positive number. Let $X^* = (X^\top, b)^\top$. Since condition (A1), $|\boldsymbol{\beta}^\top \boldsymbol{X}^* - \boldsymbol{\beta}_s^\top \boldsymbol{X}^*| < \epsilon M_{\boldsymbol{X}^*}$. In addition, for some small ϵ, $1 \leq s \leq C(1/\epsilon^p)$ for positive constant C. Since $|X^*|$ is bounded by M_{X^*}, $|\boldsymbol{\beta}^\top \boldsymbol{X^*} - \boldsymbol{\beta}_s^\top \boldsymbol{X^*}| < \epsilon M_{\boldsymbol{X^*}}$. In addition, for some $1 < \kappa < 1 + C\epsilon$, there are $\exp(\boldsymbol{\beta}_s^\top \boldsymbol{X^*})\kappa^{-1} \leq \exp(\boldsymbol{\beta}^\top \boldsymbol{X^*}) \leq \exp(\boldsymbol{\beta}_s^\top \boldsymbol{X^*})\kappa$. According Shen and Wong [20], using $1/(2r+2) < \nu < 1/(2r)$, for each $f \in \mathcal{M}_n$, there exists a brackets $[\log f_1, \log f_2]$ such that $\log f_1 \leq \log f \leq \log f_2$, which imply that $P_n|\log f_1(U) - \log f_1(U)| \leq C\epsilon$ and $P_n|\log f_1(V) - \log f_1(V)| \leq C\epsilon$.

Now we can construct the brackets $[l_s^1, l_s^2], 1 \leq s \leq C(1/\epsilon)^p, 1 \leq i \leq \epsilon^{-Cm}$ for $\mathcal{L}_1 = \{l(\boldsymbol{\theta} \mid \mathcal{O}), \boldsymbol{\theta} \in \Theta_n\}$, where

$$\begin{aligned}
l_s^1(\mathcal{O}) = {} & \delta_1^* \log\left[1 - \exp\left\{-\exp\left(-\boldsymbol{\beta}_s^\top \boldsymbol{X^*} + \log f_1\left(U \exp\left(\boldsymbol{\beta}_s^\top \boldsymbol{X^*}\right)\kappa^{-1}\right) - C\epsilon\right)\right\}\right] \\
& + \delta_2^* \log\left[\exp\left\{-\exp\left(-\boldsymbol{\beta}_s^\top \boldsymbol{X^*} + \phi_i^2\left(U \exp\left(\boldsymbol{\beta}_s^\top \boldsymbol{X^*}\right)\varrho\right) + C\epsilon\right)\right\}\right. \\
& \left. - \exp\left\{-\exp\left(-\boldsymbol{\beta}_s^\top \boldsymbol{X^*} + \log f_1\left(U \exp\left(\boldsymbol{\beta}_s^\top \boldsymbol{X^*}\right)\kappa^{-1}\right) - C\epsilon\right)\right\}\right] \\
& - \delta_3^* \exp\left(-\boldsymbol{\beta}_s^\top \boldsymbol{X^*} + \log f_2\left(U \exp\left(\boldsymbol{\beta}_s^\top \boldsymbol{X^*}\right)\kappa\right) + C\epsilon\right), \\
l_s^2(\mathcal{O}) = {} & \delta_1^* \log\left[1 - \exp\left\{-\exp\left(-\boldsymbol{\beta}_s^\top \boldsymbol{X^*} + \log f_2\left(U \exp\left(\boldsymbol{\beta}_s^\top \boldsymbol{X^*}\right)\kappa^{-1}\right) - C\epsilon\right)\right\}\right] \\
& + \delta_2^* \log\left[\exp\left\{-\exp\left(-\boldsymbol{\beta}_s^\top \boldsymbol{X^*} + \phi_i^2\left(U \exp\left(\boldsymbol{\beta}_s^\top \boldsymbol{X^*}\right)\varrho\right) + C\epsilon\right)\right\}\right. \\
& \left. - \exp\left\{-\exp\left(-\boldsymbol{\beta}_s^\top \boldsymbol{X^*} + \log f_1\left(U \exp\left(\boldsymbol{\beta}_s^\top \boldsymbol{X^*}\right)\kappa^{-1}\right) - C\epsilon\right)\right\}\right] \\
& - \delta_3^* \exp\left(-\boldsymbol{\beta}_s^\top \boldsymbol{X^*} + \log f_1\left(U \exp\left(\boldsymbol{\beta}_s^\top \boldsymbol{X^*}\right)\kappa\right) + C\epsilon\right),
\end{aligned}$$

where $\delta_1^* = I\{L = 0\}$, $\delta_2^* = I\{0 < L < R < \infty\}$ and $\delta_3^* = I\{R = \infty\}$.

For $l(\boldsymbol{\theta}|\mathcal{O}, b) \in \mathcal{L}_1$, these brackets imply that there exist s such that $l_s^1(\mathcal{O}) \leq l(\boldsymbol{\theta} \mid \mathcal{O}, b) \leq l_s^2(\mathcal{O})$, By Taylor expansion and mean value theorem, we can prove that $P_n|l_s^1(\mathcal{O}) - l_s^2(\mathcal{O})| < C\epsilon$. Take $N(\epsilon, \mathcal{L}_1, L_1(P_n))$ and $N(\epsilon, \mathcal{L}_1, L_1(P_n))$ as ϵ-bracketing number and ϵ-covering number such that ϵ-bracketing number of is bounded by $C(1/\epsilon)^{p+1+Cm}$. By Theorem 2.4.3 of [26] and the fact that $N(\epsilon, \mathcal{L}_1, L_1(P_n)) \leq N_{[]}(2\epsilon, \mathcal{L}_1, L_1(P_n))$, it follows Glivenko–Cantelli class and such that $\sup_{\boldsymbol{\theta}\in\Theta} |M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| \to 0$. And it provides that $\sup_{\boldsymbol{\theta}\in\Theta} |p_n l(\boldsymbol{\theta}|O, \hat{b}) - pl(\boldsymbol{\theta}|O, \hat{b})| \to 0$. Hence we have $\sup_{\boldsymbol{\theta}\in\Theta} |M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| \to 0$.

The second and third conditions we need to verify are that

$$\mathrm{Sup}_{\boldsymbol{\theta},\, \mathrm{d}(\boldsymbol{\theta},\boldsymbol{\theta}_0)>\epsilon}\, M(\boldsymbol{\theta}) < M(\boldsymbol{\theta}_0)$$

and

$$M_n(\boldsymbol{\theta}_n) \geq M_n(\boldsymbol{\theta}_0) - o_p(1).$$

When b is estimated as $\hat{b}$, it has the same condition with [24] and we can verify them through the conclusion of [24] directly. For this, $M_n(\boldsymbol{\theta}_n) = pl_n(\boldsymbol{\theta}|O, \hat{b})$ and proof is completed. ■

Proof of Theorem 4.2.: The proof of convergence rate is based on Theorem 3.4.1 of [26]. Let Mn be the random function and M_n be the deterministic function in the theorem for all n. In the following, we verify the conditions of that theorem along the same line as that of Theorem 2 in [37].

Note that by Lorentz [38], there exists function $\phi_{0,n} \in M_n$ such that $|\phi - \phi_{0,n}| = Cqn_r = O(n^{-r\nu/2})$. Let $\phi_{0,n} = (\beta_0, \Lambda_n))$. We have $d(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{0,n}) = O(n^{-r\nu/2})$. By similar calculations to the number of brackets in the proof of Theorem 4.2, one can show that $M(\boldsymbol{\theta}_0) - M(\boldsymbol{\theta}_{0,n}) \leq sd^2(\boldsymbol{\theta} - \boldsymbol{\theta}_{0,n}) = O(n^{-r\nu})$.

Let $\eta > 0$ be a small constant and define $L_\eta = \{l(\boldsymbol{\theta}|O,b) - l(\boldsymbol{\theta}_{0,n}|O,b) : \boldsymbol{\theta} \in \mathcal{B}\bigotimes\mathcal{M}, \eta/2 < d(\boldsymbol{\theta},\boldsymbol{\theta}_{0,n}) < \eta\}$.Using the previous statement, and the fact that $M(\boldsymbol{\theta}_0) - M(\boldsymbol{\theta}) > Cd^2(\boldsymbol{\theta}_0,\boldsymbol{\theta})$ in a neighbourhood of $\boldsymbol{\theta}_0$, for all $l(\boldsymbol{\theta}|O,b) - l(\boldsymbol{\theta}_{0,n}|O,b) \in L_n$ and for large enough n,

$$pl(\boldsymbol{\theta}|O,b) - pl(\boldsymbol{\theta}|O,b) = M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_0) + M(\boldsymbol{\theta}_0) - M(\boldsymbol{\theta}_{0,n}) \leq -C\eta^2 + Cn^{-rv} \leq -c\eta^2.$$

Under Conditions (C1)–(C3), L_η is uniformly bounded. Moreover, with some algebraic manipulations similar to the one in case of brackets and for small enough η, $(pl(\boldsymbol{\theta}|O,b) - pl(\boldsymbol{\theta}|O,b))^2 \leq c\eta^2$ for some positive constant c. By applying Lemma 3.4.2 of [26], we get

$$E[||\sqrt{n}(M_n - M)||_{L_n}] \leq CJ_{[]}(\eta, L_n, L_2(P))\left[1 + \frac{J_{[]}(\eta, L_n, L_2(P))}{\eta^2 n^{1/2}}\right],$$

where $||\cdot||_{L_n}$ is the uniform bound of the operator over the space L_η and $J_{[]}(\eta, L_\eta, L_2(P)) = \int_0^\eta [1 + \log N_{[]}(\epsilon, L_n, L_2(P))]^{1/2}\, d\epsilon$. Using the similar calculation, we get

$$\log N_{[]}(\epsilon, L_n, L_2(P))]^{1/2} \leq C(m+p+1)\log(\eta/\epsilon)$$

for $0 < \epsilon < \eta$. So $J_{[]}(\eta, L_\eta, L_2(P)) \leq C\eta(m+p+1)^{1/2}$. Let $\psi_n(\eta) = C\eta(m+p+1)^{1/2} + c\eta(m+p+1)^{-1/2}$, where $\psi_n(\eta)/\eta$ is a decreasing function of η. Choose $r_n = (m+p+1)^{-1/2}n^{1/2}$ then $r_n^2\psi_n(1/r_n) = Cn^{1/2}$.

Note that θ_n is MLE estimator. So $M_n(\boldsymbol{\theta}_n) \geq M_n(\boldsymbol{\theta}_{0,n})$ is satisfied. Also $d(\boldsymbol{\theta}_n - \boldsymbol{\theta}_{0,n}) < d(\boldsymbol{\theta},\boldsymbol{\theta}_0) + d(\boldsymbol{\theta}_0,\boldsymbol{\theta}_{0,n})$ and both of the two terms on the right hand side tend to 0 when $n \to \infty$.

By Theorem 3.4.1 of [26], we get that $r_n d(\boldsymbol{\theta}_n, \boldsymbol{\theta}_{0,n} = O_p(1))$. Together with $d(\boldsymbol{\theta}_0,\boldsymbol{\theta}_{0,n}) = O(n^{-rv/2})$, we get $d(\boldsymbol{\theta}_n,\boldsymbol{\theta}_0) = O_p(r_n^{-1} + n^{-rv/2}) = O_p(n^{-\frac{1}{2}(v-1)} + n^{-\frac{rv}{2}})$. ■

Proof of Theorem 4.3.: The proof of Theory 4 is mainly based on Theorem 2.1 and Theorem 4.2 in [6]. We define a few new notation first. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \beta_2)^\top, X^* = (X^\top, b)^\top$,

$$H^p = \{\xi(\cdot,\boldsymbol{\beta}) = \phi(g(t, X^*, \boldsymbol{\beta})), \phi \in \phi^p, t \in [a,b]\},$$

where $g(t, X^*, \beta)) = t\exp((\beta - \beta_0)^\top X^*)$, ξ is a composite function of ϕ composed with g, and $\xi(t, X^*, \boldsymbol{\beta}) = \phi(t)$. Also define

$$H = \{h : h(\cdot,\boldsymbol{\beta}) = \frac{\partial \xi_\eta(\cdot,\boldsymbol{\beta})}{\partial\eta}|_{\eta=0} = \omega(g(\cdot,\boldsymbol{\beta}))\xi_\eta \in H^p\}.$$

To obtain the asymptotic normality of parameter $\hat{\boldsymbol{\theta}}_n$, it suffices to verify the following assumptions to ensure that Theorem 2.1 of [6] holds.

(B1) $d(\hat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0) = O_p(n^{-\rho})$ for some $\rho > 0$.
(B2) $Pl'_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \xi_0(\cdot,\boldsymbol{\beta}_0|O,b)) = 0$ and $Pl'_{\xi}(\boldsymbol{\beta}_0, \xi_0(\cdot,\boldsymbol{\beta}_0|O,b)) = 0$ for all $h \in H$.

(B3) There exists direction $h^* = (h_1^*, \ldots, h_d^*)^\top$,where $h_j^* \in H$ for $j = 1, \ldots, d$ such that

$$Pl''_{\boldsymbol{\beta}\boldsymbol{\xi}}(\boldsymbol{\beta}_0, \xi_0(\cdot, \boldsymbol{\beta}_0|O, b))[h] - Pl''_{\xi\xi}(\boldsymbol{\beta}_0, \xi_0(\cdot, \boldsymbol{\beta}_0|O, b))[h^*, h] = 0.$$

(B4) $P_n l'_\beta(\hat{\beta}_n, \xi_0(\cdot, \beta_0)) = o_p(n^{-1/2})$ and $P_n l'_\xi(\hat{\beta}_n, \xi_0(\cdot, \beta_0)) = o_p(n^{-1/2})$.

(B5) For some $c > 0$, $G_n = n^{1/2}(P_n - P)$

$$\sup_{d(\boldsymbol{\theta},\boldsymbol{\theta}_0) \le cn^{-\rho}, \boldsymbol{\theta} \in \Theta_n} |G_n l'_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \xi \cdot, \boldsymbol{\beta}|O) - G_n l'_\xi(\boldsymbol{\beta}_0, \xi \cdot, \boldsymbol{\beta}|O)| = o_p(1)$$

and

$$\sup_{d(\boldsymbol{\theta},\boldsymbol{\theta}_0) \le cn^{-\rho}, \boldsymbol{\theta} \in \Theta_n} |G_n l^{\xi}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \xi \cdot, \boldsymbol{\beta}|O) - G_n l'_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \xi \cdot, \boldsymbol{\beta}|O)| = o_p(1).$$

(B6) For some $\varsigma > 1$ satisfying $\varsigma\rho > 1/2$, and for $\boldsymbol{\theta}$ in neighbourhood of $\boldsymbol{\theta}_0 : \{\boldsymbol{\theta}, d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \le cn^{-\rho}, \boldsymbol{\theta} \in \Theta_n\}$,

$$\begin{aligned} &Pl'_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \xi(\cdot, \boldsymbol{\beta})|O, b) - Pl'_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \xi(\cdot, \boldsymbol{\beta})|O, b) \\ &\quad - Pl''_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}, \xi(\cdot, \beta)|O, b)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ &\quad - Pl''_{\boldsymbol{\beta}\xi}(\boldsymbol{\beta}, \xi(\cdot, \boldsymbol{\beta})|O, b)(\xi(\cdot, \boldsymbol{\beta}) - \xi(\cdot, \boldsymbol{\beta})) \\ &= O(d^{\varsigma}(\boldsymbol{\theta}, \boldsymbol{\theta}_0)) \end{aligned}$$

and

$$\begin{aligned} &Pl'_\xi(\boldsymbol{\beta}, \xi(\cdot, \boldsymbol{\beta})|O, b)[h^*(\cdot, \boldsymbol{\beta})] - Pl'_\xi(\boldsymbol{\beta}, \xi(\cdot, \boldsymbol{\beta})|O, b)[h^*(\cdot, \boldsymbol{\beta}_0)] \\ &\quad - Pl''_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}, \xi(\cdot, \boldsymbol{\beta})|O, b)[h^*(\cdot, \boldsymbol{\beta}_0)](\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ &\quad - Pl''_{\boldsymbol{\beta}\xi}(\boldsymbol{\beta}, \xi(\cdot, \boldsymbol{\beta})|O, b)[h^*(\cdot, \boldsymbol{\beta}_0)](\xi(\cdot, \boldsymbol{\beta}) - \xi(\cdot, \boldsymbol{\beta})) \\ &= O(d^{\varsigma}(\boldsymbol{\theta}, \boldsymbol{\theta}_0)). \end{aligned}$$

First, we have assumption (B1) hold if we let $\rho = n^{-\frac{1}{2}(\nu-1)} + n^{-\frac{r\nu}{2}}$. And (B2) also hold because of the fact of the score function is mean zero.

Next we need to verify (B3). More specific it is to find $h^* = (h_1^*, \ldots, h_d^*)^\top$ with $h^*(t, x, \boldsymbol{\beta}) = w^*(t)$ such that

$$Pl''_{\boldsymbol{\beta}\xi}(\boldsymbol{\beta}_0, \xi_0(\cdot, \boldsymbol{\beta}_0|O, b))[h] - Pl''_{\xi\xi}(\boldsymbol{\beta}_0, \xi_0(\cdot, \boldsymbol{\beta}_0|O, b))[h^*, h] = 0.$$

For simplicity, let $X^* = (X^\top, b)^\top, \boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \beta_2), m(t) = \phi t \exp(\beta^\top X^*) - \beta^\top X^*, j(t) = \exp\{-m(t)\}$ $m(t)$, $\phi'(t) = \phi'\{t \exp(\boldsymbol{\beta}^\top X^*) t \exp(\boldsymbol{\beta}^\top X^*)\}X^*$,$\omega(t) = \omega(t \exp(\boldsymbol{\beta}^\top X^*))$ and $\omega'(t) = \omega'(t \exp(\boldsymbol{\beta}^\top X^*)) \exp(\boldsymbol{\beta}^\top X^*)X^*$. After calculations, we get

$$\begin{aligned} &Pl''_{\beta\xi}(\beta_0, \xi_0(\cdot, \beta_0|O, b))[h] - Pl''_{\xi\xi}(\boldsymbol{\beta}_0, \xi_0(\cdot, \boldsymbol{\beta}_0|O, b))[h^*, h] \\ &= E\left\{\delta_1 \frac{1 - m(L) - \exp\{-m(L)\}j(L)\omega(L)[\phi'(L) - X^* - \omega^*(L)]}{[1 - \exp\{-m(L)\}]^2}\right. \\ &\quad + \delta_2 \left\{ \frac{[\exp(-m(L)) - m(L)\exp(-m(R)) - \exp(-m(R))][\phi'(R) - X^* - \omega^*(R)]h(R)\omega(R)}{[exp\{m(L)\} - exp\{-m(R)\}]^2}\right. \\ &\quad - \left\{ \frac{[\exp(-m(L)) + m(L)\exp(-m(R)) - \exp(-m(R))][\phi'(V) - X^* - \omega^*(R)]h(L)\omega(U)}{[exp\{m(L)\} - exp\{-m(R)\}]^2}\right. \\ &\quad \left. + \frac{h(L)h(R)\omega(L)[\phi'(V) - X^* - \omega^*(R)] + h(R)h(R)\omega(R)[\phi'(L) - X - \omega^*(L)]}{[exp\{m(L)\} - exp\{-m(R)\}]^2}\right\} \\ &\quad \left. - \delta_3 m(R)\omega(R)[\phi'(R) - X^* - \omega^*(R)]\right\} \end{aligned}$$

$$+ E\{\delta_1 \frac{\omega'(L)j(L)}{1-\exp(-m(L))} + \delta_2 \frac{\omega' h(R) - \omega'(L)h(L)}{\exp(-m(L)) - \exp(-m(R))} - \delta_3 m(R)\omega'(R)\Big\}$$
$$= \Delta_1 + \Delta_2,$$

where $\Delta_2 = E\{\delta_1 \frac{\omega'(L)j(L)}{1-\exp(-m(L))} + \delta_2 \frac{\omega' h(R)-\omega'(L)h(L)}{\exp(-m(L))-\exp(-m(R))} - \delta_3 m(R)\omega'(R)\}$. Under (A6), T is conditionally independent of U and V given X and b, we get $\Delta_2 = E\{\delta_1 \frac{\omega'(L)j(L)}{1-\exp(-m(L))} + \delta_2 \frac{\omega' h(R)-\omega'(L)h(L)}{\exp(-m(L))-\exp(-m(R))} - \delta_3 m(R)\omega'(R)\} = 0$.

Next we need to show that $\Delta_1 = 0$ too. Note that

$$h^*(t) = \omega^*(t) = \phi^*(t) = \phi'\{t\exp(\boldsymbol{\beta}^\top X^*)\}t\exp(\beta X^*)X^* - X^*$$

is a solution of the equation.

By the fact of zero-mean for a score function, it is easy to verify the following equalities:

$$Pl''_{\boldsymbol{\beta}\xi}(\boldsymbol{\beta}_0, \xi(\cdot,\boldsymbol{\beta}_0)|O)[h] = -P\{l'_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \xi(\cdot,\boldsymbol{\beta}_0)|O)l'^{\top}_{\xi}(\boldsymbol{\beta}_0, \xi(\cdot,\boldsymbol{\beta}_0)|O)[h]\}$$
$$Pl''_{\xi\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \xi(\cdot,\boldsymbol{\beta}_0)|O)[h] = -P\{l'_{\boldsymbol{\beta}}(\beta_0, \xi(\cdot,\boldsymbol{\beta}_0)|O)[h]l'^{\top}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \xi(\cdot,\boldsymbol{\beta}_0)|O)\}$$
$$Pl''_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \xi(\cdot,\boldsymbol{\beta}_0)|O) = -P\{l'_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \xi(\cdot,\boldsymbol{\beta}_0)|O)[h]l'^{\top}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \xi(\cdot,\boldsymbol{\beta}_0)|O)\}.$$

Together with the fact that

$$Pl''_{\boldsymbol{\beta}\xi}(\boldsymbol{\beta}_0, \xi(\cdot,\boldsymbol{\beta}_0)|O)[h^*] - Pl''_{\xi\xi}(\boldsymbol{\beta}_0, \xi(\cdot,\boldsymbol{\beta}_0)|O)[h^*, h^*] = 0,$$

we get

$$P\{-l''_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}_0|O) + l''_{\xi\boldsymbol{\beta}}(\boldsymbol{\theta}_0|O)[h^*] + l''_{\boldsymbol{\beta}\xi}(\boldsymbol{\theta}_0|O)[h^*] - l''_{\xi\xi}(\boldsymbol{\theta}_0|O)[h^*, h^*]\}$$
$$= P\{l'_{\boldsymbol{\beta}}(\boldsymbol{\theta}_0|O) - l'_{\xi}(\boldsymbol{\theta}_0|O)[h^*]\}^{\bigotimes 2}$$
$$= Pl^*_{\boldsymbol{\beta}_0}(O)^{\bigotimes 2},$$

which is a non-singular matrix.

To verify (B4). We need to verify $P_n l'_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \xi_0(\cdot,\boldsymbol{\beta}_0)) = o_p(n^{-1/2})$ and $P_n l'_{\xi}(\hat{\boldsymbol{\beta}}_n, \xi_0(\cdot,\boldsymbol{\beta}_0)) = o_p(n^{-1/2})$.

$P_n l'_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n, \xi_0(\cdot,\boldsymbol{\beta}_0)) = o_p(n^{-1/2})$ can be get by the property of Sieve MLE.

For verify $P_n l'_{\xi}(\hat{\boldsymbol{\beta}}_n, \xi_0(\cdot,\beta_0)) = o_p(n^{-1/2})$. Note that $P_n l_{\xi}(\hat{\boldsymbol{\beta}}_n, \hat{\xi}(\cdot,\hat{\boldsymbol{\beta}}_n)|O)[h^*_n] = 0$, which is the directional derivative for $l(\hat{\boldsymbol{\beta}}_n; X, b)$, it is sufficient to prove $P_n l_{\xi}(\hat{\boldsymbol{\beta}}_n, \hat{\xi}(\cdot,\hat{\boldsymbol{\beta}}_n)|O)[h^*_j] = 0$, where j is the *jth* component of h^*.

Since there exist $h^*_{j,n} \in H_n$ such that $||h^*_j - h^*_{j,n}|| = O(n^{-\nu})$. Replace $h^*_j = h^*_j - h^*_{j,n}$ and rewrite $P_n l'_{\xi}(\hat{\boldsymbol{\beta}}_n, \hat{\xi}(\cdot,\hat{\boldsymbol{\beta}}_n)|O)[h^*_j - h^*_{j,n}] = I_{1,n} + I_{2,n}$, where

$$I_{1,n} = (P_n - P)l'_{\xi}(\hat{\boldsymbol{\beta}}_n, \xi(\cdot,\hat{\boldsymbol{\beta}}_n)|O)[h^*_j - h^*_{j,n}] = 0$$

and

$$I_{2,n} = P\{l'_{\xi}(\hat{\boldsymbol{\beta}}_n, \hat{xi}(\cdot,\hat{\boldsymbol{\beta}}_n)|O)[h^*_j - h^*_{j,n}] - l'_{\xi}(\beta_0, \hat{xi}(\cdot,\beta_0)|O)[h^*_j - h^*_{j,n}]\}.$$

Let $L_3 = \{l'_{\xi}(\boldsymbol{\theta}|O)[h^*_j - h_j] : \boldsymbol{\theta} \in \Theta_n, h_j \in H_n, ||h^*_j - h_{j,n}|| \le n^{-\nu}\}$ and the ϵ-bracketing number associated with $L_2(P)$-norm is bounded ny $C(1/\epsilon)^d(1/\epsilon)^{Cm}(1/\epsilon)^{Cm}$ such that L_3 is P-Donsker. Since $l'_{\xi}(\hat{\boldsymbol{\beta}}_n, \hat{\xi}(\cdot,\hat{\boldsymbol{\beta}}_n)|O)[h^*_j - h^*_{j,n}] \in L_3$ and according Corollary 2.3.12 of van der Vaart and Wellner [25] we have $I_{1,n} = O_p(n^{-1/2})$.

For $I_{2,n}$, under condition (A1)–(A4) and some calculation with Cauchy–Schwarz inequality, we get

$$\begin{aligned} I_{2,n} &= P\left\{ l'_{\xi}\left(\hat{\boldsymbol{\beta}}_n, \hat{\xi}\left(\cdot, \hat{\boldsymbol{\beta}}_n\right) \mid \mathcal{O}\right)\left[h_j^* - h_{j,n}^*\right] - l'_{\xi}\left(\hat{\boldsymbol{\beta}}_0, \hat{\xi}\left(\cdot, \hat{\boldsymbol{\beta}}_0\right) \mid \mathcal{O}\right)\left[h_j^* - h_{j,n}^*\right]\right\} \\ &\le C\,\mathrm{d}\left(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{\mathbf{0}}\right)\left\|h_j^* - h_{j,n}^*\right\|_{\infty} = o_p\left(n^{-1/2}\right). \end{aligned}$$

Together with $I_{1,n} = O_p(n^{-1/2})$, (B4) holds.

Next we need to check (B5). Define $L_4 = \{l'_{\xi}(\boldsymbol{\beta}, \xi(\cdot, \boldsymbol{\beta}) - l'_{\xi}(\boldsymbol{\beta}, \xi(\cdot, \boldsymbol{\beta}_0) : \boldsymbol{\theta} \in \Theta_n, d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \le \eta\}$. Note a $\epsilon-$ bracketing number associated with $L_2(P)-$ norm is bounded by $(\eta/\epsilon)^{Cq_n+p}$ such that $L_4(\eta)$ is also P-Donsker. Choose $\eta = O(n^{-\frac{1}{2}(\nu-1)} + n^{-\frac{r\nu}{2}})$.$l'_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \xi(\cdot, \boldsymbol{\beta}) \mid \mathcal{O}) - l'_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \xi(\cdot, \boldsymbol{\beta}_0) \mid \mathcal{O}) \in L_4(\eta)$. By the convergence rate of $\hat{\theta}_n$ we have

$$P\left\{l'_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \xi(\cdot, \boldsymbol{\beta}) \mid \mathcal{O}) - l'_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_0, \xi\left(\cdot, \boldsymbol{\beta}_0\right) \mid \mathcal{O}\right)\right\}^2 \to 0 \quad \text{as } n \to \infty.$$

In a similar manner, we can proof $\mathcal{L}_5(\eta) = \{l'_{\xi}(\boldsymbol{\beta}, \xi(\cdot, \boldsymbol{\beta}) \mid \mathcal{O}) - l'_{\xi}(\boldsymbol{\beta}_0, \xi(\cdot, \boldsymbol{\beta}_0) \mid \mathcal{O}) : \boldsymbol{\theta} \in \Theta_n, d(\boldsymbol{\theta}, \boldsymbol{\theta}_{\mathbf{0}}) \le \eta\}$ is P-Donsker and for any $\tilde{f} \in \mathcal{L}_5$ $P\tilde{f}^2 \xrightarrow{n\to 0} 0$, such that (B5) hold.

To verify (B6), we will use Taylor expansion to proof the first equation, and the second is similar. Use Multi-Taylor expansion of $l'_{\beta}(\theta \mid \mathcal{O})$ at θ_0

$$\begin{aligned} l'_{\boldsymbol{\beta}}(\boldsymbol{\theta} \mid \mathcal{O}) &= l'_{\beta}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right) + l''_{\beta\beta}(\tilde{\boldsymbol{\theta}} \mid \mathcal{O})\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right) + l''_{\boldsymbol{\beta}\xi}(\tilde{\boldsymbol{\theta}} \mid \mathcal{O})\left[\xi(\cdot, \boldsymbol{\beta}) - \xi_0\left(\cdot, \boldsymbol{\beta}_0\right)\right] \\ &= l'_{\beta}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right) + l''_{\beta\boldsymbol{\beta}}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right)\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right) + l''_{\beta\xi}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right)\left[\xi(\cdot, \boldsymbol{\beta}) - \xi_0\left(\cdot, \boldsymbol{\beta}_0\right)\right] \\ &\quad + \left\{l''_{\beta\boldsymbol{\beta}}(\tilde{\boldsymbol{\theta}} \mid \mathcal{O})\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right) - l''_{\beta\beta}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right)\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right)\right\} \\ &\quad + \left\{l''_{\boldsymbol{\beta}\xi}(\tilde{\boldsymbol{\theta}} \mid \mathcal{O})\left[\xi(\cdot, \boldsymbol{\beta}) - \xi_0\left(\cdot, \boldsymbol{\beta}_0\right)\right] - l''_{\boldsymbol{\beta}\xi}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right)\left[\xi(\cdot, \boldsymbol{\beta}) - \xi_0\left(\cdot, \boldsymbol{\beta}_0\right)\right]\right\}. \end{aligned}$$

Thus

$$\begin{aligned} &P\left\{l'_{\boldsymbol{\beta}}(\boldsymbol{\theta} \mid \mathcal{O}) - l'_{\boldsymbol{\beta}}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right) - l''_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right)\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right) - l''_{\boldsymbol{\beta}\xi}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right)\left[\xi(\cdot, \boldsymbol{\beta}) - \xi_0\left(\cdot, \boldsymbol{\beta}_0\right)\right]\right\} \\ &\quad = P\left\{\left[l''_{\boldsymbol{\beta}\boldsymbol{\beta}}(\tilde{\boldsymbol{\theta}} \mid \mathcal{O}) - l''_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right)\right]\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right)\right\} \\ &\qquad + P\left\{l''_{\boldsymbol{\beta}\xi}(\tilde{\boldsymbol{\theta}} \mid \mathcal{O})\left[\xi(\cdot, \boldsymbol{\beta}) - \xi_0(\cdot, \boldsymbol{\beta})\right] - l''_{\boldsymbol{\beta}\xi}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid O\right)\left[\xi(\cdot, \boldsymbol{\beta}) - \xi_0(\cdot, \boldsymbol{\beta})\right]\right\}. \end{aligned}$$

let $\mathcal{L}_6(\eta) = \{l''_{\beta\beta}(\tilde{\boldsymbol{\theta}} \mid \mathcal{O}) - l''_{\beta\beta}(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}) : \boldsymbol{\theta} \in \Theta_n, d(\boldsymbol{\theta}, \boldsymbol{\theta}_{\mathbf{0}}) \le \eta\}$ and choose $\eta = n^{-\min(\frac{1}{2}(\nu-1), \frac{1}{2}r\nu)}$. it can be proved that L_6 is P-Donsker. Moreover,

$$\begin{aligned} &P\left\{\left[l''_{\beta\beta}(\tilde{\boldsymbol{\theta}} \mid \mathcal{O}) - l''_{\beta\beta}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right)\right]\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right)\right\} \\ &\quad = n^{-\min(\frac{1}{2}r\nu, (1-\nu)/2)+\epsilon} P\left\{\left[l''_{\boldsymbol{\beta}\boldsymbol{\beta}}(\tilde{\boldsymbol{\theta}} \mid \mathcal{O}) - l''_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right)\right]\frac{\boldsymbol{\beta} - \boldsymbol{\beta}_0}{n^{-\min(\frac{1}{2}r\nu, (1-\nu)/2)+\epsilon}}\right\}. \end{aligned}$$

Thus

$$\begin{aligned} &P\left\{\left[l''_{\beta\boldsymbol{\beta}}(\tilde{\boldsymbol{\theta}} \mid \mathcal{O}) - l''_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\boldsymbol{\theta}_{\mathbf{0}} \mid \mathcal{O}\right)\right]\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right)\right\} \\ &\quad = O_p\left(n^{-\min(\frac{1}{2}r\nu, (1-\nu)/2)+\epsilon} n^{-1/2}\right) = O_p\left(n^{-2\min(\frac{1}{2}r\nu, (1-\nu)/2)}\right). \end{aligned}$$

By a similar approach, we get

$$P\left\{l''_{\boldsymbol{\beta}\xi}(\tilde{\boldsymbol{\theta}} \mid \mathcal{O})\left[\xi(\cdot, \boldsymbol{\beta}) - \xi_0\left(\cdot, \boldsymbol{\beta}_0\right)\right] - l''_{\boldsymbol{\beta}\xi}\left(\boldsymbol{\theta}_0 \mid O\right)\left[\xi(\cdot, \boldsymbol{\beta}) - \xi_0\left(\cdot, \boldsymbol{\beta}_0\right)\right]\right\} = O_p\left(n^{-2\min(r\nu, (1-\nu)/2)}\right).$$

Therefore

$$P\left\{l'_{\beta}(\boldsymbol{\theta} \mid \mathcal{O}) - l'_{\beta}\left(\boldsymbol{\theta}_0 \mid \mathcal{O}\right) - l''_{\beta\beta}\left(\boldsymbol{\theta}_0 \mid \mathcal{O}\right)\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right) - l''_{\beta\xi}\left(\boldsymbol{\theta}_0 \mid \mathcal{O}\right)\left[\xi(\cdot, \boldsymbol{\beta}) - \xi_0\left(\cdot, \boldsymbol{\beta}_0\right)\right]\right\}$$
$$= O\left(n^{-\iota\eta}\right),$$

where $\eta = O(n^{-\frac{1}{2}(\nu-1)} + n^{-\frac{r\nu}{2}})$, $\iota = 2 > 1$. Therefore the first equation in Assumption (B6) is hold.

By using Theorem 2.1 in [6], we get

$$n^{1/2}\left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) = n^{-1/2} I^{-1}\left(\beta_0\right) \sum_{i=1}^{n} I^*\left(\boldsymbol{\beta}_{\mathbf{0}}, \xi_0\left(\cdot, \boldsymbol{\beta}_0\right) \mid \mathcal{O}\right) + o_p(1) \xrightarrow{d} N\left(0, I^{-1}\left(\boldsymbol{\beta}_0\right)\right).$$

■