Lifetime Data Analysis
https://doi.org/10.1007/s10985-020-09496-z

Semiparametric efficient estimation for additive hazards regression with case II interval-censored survival data

Baihua He[1] · Yanyan Liu[1] · Yuanshan Wu[2] · Xingqiu Zhao[3]

Received: 15 November 2018 / Accepted: 2 March 2020

Abstract
Interval-censored data often arise naturally in medical, biological, and demographical studies. As a matter of routine, the Cox proportional hazards regression is employed to fit such censored data. The related work in the framework of additive hazards regression, which is always considered as a promising alternative, remains to be investigated. We propose a sieve maximum likelihood method for estimating regression parameters in the additive hazards regression with case II interval-censored data, which consists of right-, left- and interval-censored observations. We establish the consistency and the asymptotic normality of the proposed estimator and show that it attains the semiparametric efficiency bound. The finite-sample performance of the proposed method is assessed via comprehensive simulation studies, which is further illustrated by a real clinical example for patients with hemophilia.

Keywords Survival analysis · Interval-censored data · Additive hazards · Sieve maximum likelihood estimator · Semiparametric efficiency bound · Empirical process

1 Introduction

Interval-censoring is encountered in studies when the event of interest cannot be observed and is only known to occur within a time interval. For example, the event time to human immunodeficiency virus (HIV) positive status for transfusion-related

✉ Yanyan Liu
liuyy@whu.edu.cn

Xingqiu Zhao
xingqiu.zhao@polyu.edu.hk

[1] School of Mathematics and Statistics, Wuhan University, Wuhan 430072, Hubei, China

[2] School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan 430072, Hubei, China

[3] Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Published online: 10 March 2020

acquired immune deficiency syndrome (AIDS) patients can never be known exactly, while the only observed information is the change of status among some monitoring times.

When only one monitoring time is applied and each subject experiences the event either before or after the monitoring time, this type of data is referred to as current status or case I interval-censored data (Huang 1996; Li and Zhang 1998). When each subject is known to experience the event between a time interval, or before the first point of the time interval, or after the last point of the time interval, the resulting data are referred to as case II interval-censored data. The analysis of case II interval-censored data are more challenging than that of right censored data due to the more complicated data structure. The counting process and martingale theory which are commonly used for the analysis of right-censored data do not directly apply to interval-censored data.

There have been a rich literature for the regression analysis of current status data and case II interval-censored data (Sun 2006). Some authors developed efficient estimation procedures under different models for current status data (Huang 1996; Martinussen and Scheike 2002; Sun and Shen 2009; Wang et al. 2008; Xue et al. 2004). There are several types of methods for analyzing the case II interval-censored data. One is an imputation-based method where the interval-censored event times are imputed and then some well-known semiparametric regression analysis for right-censored data can be directly applied. The imputation-based method for the Cox model (Cox 1972) was studied by some authors (Satten et al. 1998; Song and Ma 2008; Zhang et al. 2009). However, this method may produce a biased estimator for regression parameters and may not be efficient. Other types of methods include likelihood-based and estimating equation-based approaches. For example, the Cox model for case II interval-censored data was considered by Finkelstein (1986), Goggins and Finkelsten (2000), Kim and Xue (2002), Seaman and Bird (2001), Zhao et al. (2005) among others. As an alternative to the Cox model, the semiparametric accelerated failure time model for interval-censored data was studied by Betensky et al. (2001), Gómez et al. (2003), Li and Zhang (1998), Li and Pu (2003), Rabinowitz et al. (1995), among others. van der Vaart (1998) developed an estimating equation approach for regression analysis of linear transformation models with interval-censored failure time data. Most recently, Zhou et al. (2017) developed an efficient sieve maximum likelihood estimation approach for bivariate interval-censored failure time data.

The aforementioned studies are based on case II interval-censored data with the experience of event between the two monitoring times. In medical studies with periodic follow-ups, the occurrence of the event such as a disease onset is known before the first monitoring time, or between the two monitoring times, or after the second monitoring time. The resulting data are referred also as case II interval-censored data (Huang and Wellner 1997). The studies on this kind of case II interval-censored data are limited. Among those available, Zhang et al. (2001) presented a nonparametric test. Zeng et al. (2006) discussed the efficient estimation of the regression parameters for an additive hazards model using the full likelihood approach. However, the implementation of their approach can be quite complicated because of the need for estimation of the baseline cumulative hazard function, which is time consuming especially when the monitoring variables are continuous. Wang et al. (2010) proposed a relatively easy estimating equation-based approach under the additive hazards model.

Their method does not need estimation of any nuisance baseline hazard functions. However they assume that the monitoring times follow Cox-type models, which may lead to biased estimators when modeled incorrectly. Another drawback is that the approach is unable to offer an estimate of the baseline hazard function. Zhang et al. (2010) proposed a spline-based sieve semiparametric likelihood estimation procedure for the Cox model, in which the log baseline cumulative hazard function is estimated by a monotone B-spline (Schumaker 1981). Their method does not need any assumption about the distribution of monitoring times and the resulted sieve likelihood estimator is asymptotically normal and achieves the semiparametric efficiency bound defined in Bickel et al. (1993). In addition, the spline-based sieve estimator of the baseline hazard function converges at the optimal nonparametric rate. They developed an easy-to-implement method to consistently estimate the standard error of the estimated regression parameter. Considering the merits of Zhang et al. (2010), we propose to fit the additive hazard model to case II interval censored data by using the sieve maximum likelihood estimation. Compared with the Cox model (Cox 1972), the additive risk model is particularly useful for estimating the difference in hazards. The merits of our proposed method have several aspects. Firstly, by directly estimating the log baseline hazard function based on B-spline approximations, we relax the constraints of being non-negative of the baseline hazard function and the monotonicity of the baseline cumulative hazard function. With this relaxation, the standard Newton–Raphson algorithm can be employed directly, and thus our proposed inference procedure is easy to implement compared with Zeng et al. (2006). Secondly, we do not impose any assumption on the models of monitoring times, which is more reasonable in practice. Lastly, we rigorously prove that the proposed estimator is asymptotically normal and semiparametrical efficient.

The rest of the paper is organized as follows. In Sect. 2, we introduce the model and describe the spline-based sieve semiparametric maximum likelihood approach. Furthermore, we obtain an estimator for the hazard function under the additive risk model. Section 3 presents the asymptotic results of the estimator. We provide simulation results in Sect. 4. An application to hemophilia data is illustrated in Sect. 5. Section 6 concludes some remarks. All technical proofs are given in the “Appendix”.

2 Sieve maximum likelihood estimation

The observations of case II interval censored data include two positive monitoring times U and V where U and V are two random variables satisfying $U \leq V$ with probability 1. The true event time T falls into one of the following three exclusive categories: T is between U and V (interval-censored); T is larger than V (right-censored); or T is less than U (left-censored). Define two indicator variables as $\Delta_1 = I(T \leq U)$, $\Delta_2 = I(U < T \leq V)$. The observed data for the ith subject can be summarized as: $\mathbf{O}_i = (\Delta_{1i}, \Delta_{2i}, U_i, V_i, \mathbf{Z}_i)$, $i = 1, 2, \ldots, n$.

Given the covariate $\mathbf{Z}$, the hazard function of the event time T at time t is

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) + \boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{Z}, \tag{1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, and $\boldsymbol{\theta}_0$ is a d-vector of unknown regression parameters. We define the baseline cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(s)\mathrm{d}s$.

Let e be the ending time of study, assume that conditional on $\mathbf{Z}$, T is independent of (U, V) and the distribution of (U, V) is noninformative of T. We further assume that the distribution of $(U, V, \mathbf{Z})$ does not involve in $(\boldsymbol{\theta}, \Lambda_0)$. Then the logarithm of the observed likelihood function in terms of parameters $(\boldsymbol{\theta}, \Lambda_0)$ can be expressed as

$$\ell_n(\boldsymbol{\theta}, \Lambda_0) = \sum_{i=1}^{n} \ell(\boldsymbol{\theta}, \Lambda_0; \mathbf{O}_i),$$

where

$$\begin{aligned}\ell(\boldsymbol{\theta}, \Lambda_0; \mathbf{O}_i) = &\, \Delta_{1i} \log\{1 - \exp(-\Lambda_0(U_i) - \boldsymbol{\theta}^{\mathrm{T}}(U_i\mathbf{Z}_i))\} \\ &+ \Delta_{2i} \log\{\exp(-\Lambda_0(U_i) - \boldsymbol{\theta}^{\mathrm{T}}(U_i\mathbf{Z}_i)) - \exp(-\Lambda_0(V_i) - \boldsymbol{\theta}^{\mathrm{T}}(V_i\mathbf{Z}_i))\} \\ &- (1 - \Delta_{1i} - \Delta_{2i})\{\Lambda_0(V_i) + \boldsymbol{\theta}^{\mathrm{T}}(V_i\mathbf{Z}_i)\}.\end{aligned}$$

The maximum likelihood estimator for $(\boldsymbol{\theta}, \Lambda_0)$ can be obtained by maximizing the observed log-likelihood function $\ell_n(\boldsymbol{\theta}, \Lambda_0)$ over the parametric space $\Theta \times \Omega$ where

$$\Theta = \{\boldsymbol{\theta} : \boldsymbol{\theta} \text{ is in a compact set of } R^d, \|\boldsymbol{\theta}\| \leq M\}$$

with M being a positive constant and

$$\Omega = \{\Lambda_0(t) : \Lambda_0(t) \text{ is a step-function with jumps only at the examination times}\}.$$

The calculation is not easy because it involves a large number of parameters, with upper bound to $d + 2n$ if there are no ties among $\{(U_i, V_i), i = 1, \cdots, n\}$. This high dimensional optimization problem is particularly challenging when the sample size is large.

To overcome the computational difficulty in fully nonparametric estimation problems, Geman and Hwang (1982) proposed to approximate the unknown function in the log-likelihood by a linear span of some known basis functions and obtained a sieve log-likelihood. Then the original optimization problem is transferred to maximizing the sieve log-likelihood with respect to the unknown coefficients in the linear span. This, in turn, reduces the dimensionality of the optimization problem significantly as the number of basis functions required to reasonably approximate the unknown function grows slowly as sample size increases.

In the following, we develop a spline-based sieve semiparametric maximum likelihood estimation approach in the context of the additive hazards model with case II interval-censored data. First of all, to overcome the nonnegative constraint on $\lambda_0(\cdot)$, we define $g(t) = \log \lambda_0(t)$. Then the log likelihood is written as

$$\ell_n(\boldsymbol{\theta}, g) = \sum_{i=1}^{n} \ell(\boldsymbol{\theta}, g; \mathbf{O}_i),$$

where

$$
\begin{aligned}
&\ell(\boldsymbol{\theta}, g; \mathbf{O}_i) \\
&= \Delta_{1i} \log \left\{1 - \exp\left(-\int_0^{U_i} \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(U_i\mathbf{Z}_i)\right)\right\} \\
&\quad + \Delta_{2i} \log \left\{ \exp\left(-\int_0^{U_i} \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(U_i\mathbf{Z}_i)\right)\right. \\
&\quad \left. - \exp\left(-\int_0^{V_i} \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(V_i\mathbf{Z}_i)\right)\right\} \\
&\quad - \left(1 - \Delta_{1i} - \Delta_{2i}\right)\left\{ \int_0^{V_i} \exp\{g(s)\}\mathrm{d}s + \boldsymbol{\theta}^{\mathrm{T}}(V_i\mathbf{Z}_i)\right\}.
\end{aligned}
$$

Suppose that U and V take values in $[a, b]$ where a and b are finite numbers. Let $a \equiv t_0 < t_1 < \ldots < t_{K_n} < t_{K_n+1} \equiv b$ be a partition of $[a, b]$ with $K_n \approx O(n^{\nu})$ and $\max_{0\le j\le K_n} |t_{j+1} - t_j| = O(n^{-\nu})$ for $\nu \in (0, 0.5)$. Denote the set of partition points by $T_{K_n} = \{t_1, \ldots, t_{K_n}\}$, and $S_n(T_{K_n}, K_n, m)$ be the space of polynomial splines of order m defined in Schumaker (1981) (page 108, Definition 4.1). According to Schumaker (1981) (page 117, Corollary 4.10), there exists a local basis $\{B_j : 1 \le j \le q_n\}$ with $q_n = K_n + m$ such that for $g \in S_n(T_{K_n}, K_n, m)$, we can write $g(t) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{B}(t) = \sum_{j=1}^{q_n} \beta_j B_j(t)$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{q_n})^{\mathrm{T}}$ and $\mathbf{B}(t) = (B_1(t), \ldots, B_{q_n}(t))^{\mathrm{T}}$. Under some suitable smoothness assumptions, g_0, the true function of g, can be well approximated by $g(t) = \sum_{j=1}^{q_n} \beta_j B_j(t)$ in $S_n(T_{K_n}, K_n, m)$. The resulting sieve log-likelihood can be expressed as

$$
\ell_n(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \ell(\boldsymbol{\theta}, \boldsymbol{\beta}; \mathbf{O}_i)
$$

with

$$
\begin{aligned}
\ell(\boldsymbol{\theta}, \boldsymbol{\beta}; \mathbf{O}_i) &= \Delta_{1i} \log \left\{1 - \exp\left(-\int_0^{U_i} \exp\left\{\sum_{j=1}^{q_n} \beta_j B_j(s)\right\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(U_i\mathbf{Z}_i)\right)\right\} \\
&\quad + \Delta_{2i} \log \left\{ \exp\left(-\int_0^{U_i} \exp\left\{\sum_{j=1}^{q_n} \beta_j B_j(s)\right\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(U_i\mathbf{Z}_i)\right)\right. \\
&\quad \left. - \exp\left(-\int_0^{V_i} \exp\left\{\sum_{j=1}^{q_n} \beta_j B_j(s)\right\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(V_i\mathbf{Z}_i)\right)\right\} \\
&\quad - \left(1 - \Delta_{1i} - \Delta_{2i}\right)\left\{ \int_0^{V_i} \exp\left\{\sum_{j=1}^{q_n} \beta_j B_j(s)\right\}\mathrm{d}s + \boldsymbol{\theta}^{\mathrm{T}}(V_i\mathbf{Z}_i)\right\}.
\end{aligned}
$$

The sieve maximum likelihood estimator $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\beta}}_n)$ can be derived by maximizing $\ell_n(\boldsymbol{\theta}, g)$ over a sieve space $\Theta \times \mathbb{G}_n$, where $\mathbb{G}_n = \{g(t) : g(t) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{B}(t) \in S_n, \|\boldsymbol{\beta}\| \leq B\}$ and B is a positive constant. The sieve estimator for $g_0(t)$ is $\widehat{g}(t) = \sum_{j=1}^{q_n} \widehat{\beta}_j B_j(t)$.

3 Asymptotic properties

To establish the asymptotic properties of sieve estimator $(\widehat{\boldsymbol{\theta}}_n, \widehat{g}_n)$, we first introduce some notation and assumptions. Let $\|\mathbf{a}\|$ denote the Euclidean norm for a vector and define the following metric in a functional space $\mathbb{G}$ which is a class of functions with bounded pth derivative in $[a, b]$ for $p \geq 1$ and p is an integer,

$$\|g_1 - g_2\|_{\mathbb{G}}^2 = E\left(\int_0^U \{g_1(s) - g_2(s)\}^2 \mathrm{d}s\right) + E\left(\int_0^V \{g_1(s) - g_2(s)\}^2 \mathrm{d}s\right)$$

with any elements $g_1, g_2 \in \mathbb{G}$, and (U, V) be random monitoring times defined in Sect. 2. Define the distance between any two elements in $\Theta \times \mathbb{G}$, $\tau_1 = (\boldsymbol{\theta}_1, g_1)$ and $\tau_2 = (\boldsymbol{\theta}_2, g_2)$ as:

$$d(\tau_1, \tau_2) = \|\tau_1 - \tau_2\|_\tau = \{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|^2 + \|g_1 - g_2\|_{\mathbb{G}}^2\}^{1/2}.$$

Let

$$\begin{aligned}\ell(\boldsymbol{\theta}, g; \mathbf{O}) = {} & \Delta_1 \log\left\{1 - \exp\left(-\int_0^U \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(U\mathbf{Z})\right)\right\} \\ & + \Delta_2 \log\left\{\exp\left(-\int_0^U \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(U\mathbf{Z})\right)\right. \\ & \left. - \exp\left(-\int_0^V \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(V\mathbf{Z})\right)\right\} \\ & -(1 - \Delta_1 - \Delta_2)\left\{\int_0^V \exp\{g(s)\}\mathrm{d}s + \boldsymbol{\theta}^{\mathrm{T}}(V\mathbf{Z})\right\}.\end{aligned}$$

Consider a parametric smooth sub-model with parameter $(\boldsymbol{\theta}, g_{(s)})$, where $g_{(0)} = g$ and

$$h = \frac{\partial g_{(s)}}{\partial s}|_{s=0}. \tag{2}$$

Let $\mathcal{H}$ be the class of functions h defined by (2). The score operator for g is

$$\dot{\ell}_2(\boldsymbol{\theta}, g; \mathbf{O})[h] = \frac{\partial}{\partial s}\ell(\boldsymbol{\theta}, g_{(s)}; \mathbf{O})|_{s=0}.$$

For a d-dimensional $\boldsymbol{\theta}$, $\dot{\ell}_1(\boldsymbol{\theta}, g; \mathbf{O})$ is the vector of partial derivatives of $\ell(\boldsymbol{\theta}, g; \mathbf{O})$ with respect to the components of $\boldsymbol{\theta}$. For each component of $\dot{\ell}_1$, a score operator for g is defined as an equation above. So the score operator for g corresponding to $\dot{\ell}_1$ is

$$\dot{\ell}_2(\boldsymbol{\theta}, g; \mathbf{O})[\mathbf{h}] = \left\{\dot{\ell}_2(\boldsymbol{\theta}, g; \mathbf{O})[h_1], \ldots, \dot{\ell}_2(\boldsymbol{\theta}, g; \mathbf{O})[h_d]\right\}^{\mathrm{T}},$$

where $\mathbf{h} = (h_1, \ldots, h_d)^{\mathrm{T}}$ with $h_k \in \mathcal{H}, 1 \leq k \leq d$.

According to Bickel et al. (1993) (Theorem 1), the efficient score vector for $\boldsymbol{\theta}$ is $\dot{\ell}_1(\boldsymbol{\theta}, g; \mathbf{O}) - \dot{\ell}_2(\boldsymbol{\theta}, g; \mathbf{O})[\mathbf{h}_0]$, where $\mathbf{h}_0$ is an element of $\mathcal{H}^d$ that minimizes $E\|\dot{\ell}_1(\boldsymbol{\theta}, g; \mathbf{O}) - \dot{\ell}_2(\boldsymbol{\theta}, g; \mathbf{O})[\mathbf{h}]\|^2$ and $\mathbf{h}_0$ is called the least favorable direction. Denote the efficient score by $\ell^*(\boldsymbol{\theta}, g; \mathbf{O}) = \dot{\ell}_1(\boldsymbol{\theta}, g; \mathbf{O}) - \dot{\ell}_2(\boldsymbol{\theta}, g; \mathbf{O})[\mathbf{h}_0]$. Then the information for θ is

$$I(\boldsymbol{\theta}) = E\|\ell^*(\tau; \mathbf{O})\|^2 = E\|\dot{\ell}_1(\boldsymbol{\theta}, g; \mathbf{O}) - \dot{\ell}_2(\boldsymbol{\theta}, g; \mathbf{O})[\mathbf{h}_0]\|^2. \quad (3)$$

By similar arguments as Huang and Wellner (1997), we can prove that the least favorable direction $\mathbf{h}_0$ is the unique solution to the integral equation

$$m(t) - \int Q(t, x) m(x) \mathrm{d}x = -\frac{r(t)}{\varphi(t)} \quad (4)$$

where $m(t) = \int_0^t \exp\{g(s)\} h(s) \mathrm{d}s$, the notations $Q(t, x)$, $r(t)$ and $\varphi(t)$ are deferred to the "Appendix".

Some regular conditions are needed for the asymptotic results.

A1 (a) There exists a positive number c such that $P(V - U \geq c) = 1$; (b) the union of the supports of U and V is contained in an interval $[a, b]$, where $0 < a < b < \infty$, and $0 < \Lambda_0(a) < \Lambda_0(b) < \infty$.

A2 (a) $E(\mathbf{Z}\mathbf{Z}^{\mathrm{T}})$ is nonsingular; (b) $\mathbf{Z}$ is uniformly bounded, that is, there exists $z_0 > 0$ such that $P(\|\mathbf{Z}\| \leq z_0) = 1$.

A3 The parametric space Θ is a compact subset of $\mathbb{R}^d$.

A4 $g_0 = \log \lambda_0$ belongs to $\mathbb{G}$, a class of functions with bounded pth derivative in $[a, b]$ for $p \geq 1$ and p is an integer.

A5 The conditional density $f(u, v|\mathbf{z})$ of (U, V) given $\mathbf{z}$, has bounded partial derivatives with respect to (u, v). The bounds of these partial derivatives do not depend on $(u, v, \mathbf{z})$.

A6 For some $\eta \in (0, 1)$ and for all $\mathbf{a} \in \mathbb{R}^d$, $\mathbf{a}^{\mathrm{T}}\mathrm{var}(\mathbf{Z}|U)\mathbf{a} \geq \eta \mathbf{a}^{\mathrm{T}} E(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|U)\mathbf{a}$ and $\mathbf{a}^{\mathrm{T}}\mathrm{var}(\mathbf{Z}|V)\mathbf{a} \geq \eta \mathbf{a}^{\mathrm{T}} E(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|V)\mathbf{a}$ a.s..

A7 (Smoothness of the model). Denote $Pf = \int f(\mathbf{O}) \mathrm{d}P(\mathbf{O})$, for some $\alpha > 1$ satisfying $\alpha p \nu > \frac{1}{2}$ and $(\boldsymbol{\theta}, g)$ in the neighborhood $\{(\boldsymbol{\theta}, g) : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq \eta_n, \|g - g_0\|_{\mathbb{G}} \leq C n^{-p\nu}\}$,

(a1) $|P\dot{\ell}_1(\boldsymbol{\theta}, g; \mathbf{O}) - P\dot{\ell}_1(\boldsymbol{\theta}_0, g_0; \mathbf{O}) - P\ddot{\ell}_{11}(\boldsymbol{\theta}_0, g_0; \mathbf{O})[\boldsymbol{\theta} - \boldsymbol{\theta}_0] - P\ddot{\ell}_{12}(\boldsymbol{\theta}_0, g_0; \mathbf{O})$

$[g - g_0]| = o(|\boldsymbol{\theta} - \boldsymbol{\theta}_0|) + O(\|g - g_0\|^{\alpha})$, where $\ddot{\ell}_{11}(\boldsymbol{\theta}_0, g_0; \mathbf{O}) = -\dot{\ell}_1(\boldsymbol{\theta}_0, g_0; \mathbf{O})$

$\dot{\ell}_1^{\mathrm{T}}(\boldsymbol{\theta}_0, g_0; \mathbf{O})$, $\ddot{\ell}_{12}(\boldsymbol{\theta}_0, g_0; \mathbf{O})[h] = -\dot{\ell}_1(\boldsymbol{\theta}_0, g_0; \mathbf{O})\dot{\ell}_2(\boldsymbol{\theta}_0, g_0; \mathbf{O})[h]$;

(a2) $|P\dot{\ell}_2(\boldsymbol{\theta}, g; \mathbf{O})[\mathbf{h}_0] - P\dot{\ell}_2(\boldsymbol{\theta}_0, g_0; \mathbf{O})[\mathbf{h}_0] - (P\ddot{\ell}_{21}(\boldsymbol{\theta}_0, g_0; \mathbf{O})[\mathbf{h}_0])(\boldsymbol{\theta} - \boldsymbol{\theta}_0) -$

$P\ddot{\ell}_{22}(\boldsymbol{\theta}_0, g_0; \mathbf{O})[\mathbf{h}_0, (g - g_0)]| = o(|\boldsymbol{\theta} - \boldsymbol{\theta}_0|) + O(\|g - g_0\|^{\alpha})$, where $\ddot{\ell}_{21}(\boldsymbol{\theta}_0, g_0; \mathbf{O})$

$[h] = -\dot{\ell}_1(\boldsymbol{\theta}_0, g_0; \mathbf{O})\dot{\ell}_2(\boldsymbol{\theta}_0, g_0; \mathbf{O})[h], \ddot{\ell}_{22}(\boldsymbol{\theta}_0, g_0; \mathbf{O})[h_1, h_2] = -\dot{\ell}_2(\boldsymbol{\theta}_0, g_0, \mathbf{O})[h_1]\dot{\ell}_2(\boldsymbol{\theta}_0, g_0, \mathbf{O})[h_2]$.

A8 (Positive information). $P(\ddot{\ell}_{12}(\boldsymbol{\theta}_0, g_0; \mathbf{O})[h] - P\ddot{\ell}_{22}(\boldsymbol{\theta}_0, g_0; \mathbf{O})[\mathbf{h}_0, h]) = 0$, for all $h \in \mathcal{H}$.

Conditions A1–A5 are the regular conditions for the consistency. Condition A6 can be justified. Taking U as an example, define $\lambda_1^* = \max\{\text{eigenvalue}(E(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|U))\}$ and $\lambda_d^* = \min\{\text{eigenvalue}(\text{var}(\mathbf{Z}|U))\}$. Under condition A2 that $E(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|U)$ is a positive definite matrix, if $\text{var}(\mathbf{Z}|U)$ is a positive definite matrix, then it can be proved that for any $\mathbf{a} \in R^d$, $\mathbf{a}^{\mathrm{T}}\text{var}(\mathbf{Z}|U)\mathbf{a} \geq \mathbf{a}^{\mathrm{T}}\lambda_d^*\mathbf{a} \geq \frac{\lambda_d^*}{\lambda_1^*}\mathbf{a}^{\mathrm{T}}\lambda_1^*\mathbf{a} \geq \frac{\lambda_d^*}{\lambda_1^*}\mathbf{a}^{\mathrm{T}}E(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|U)\mathbf{a}$. Therefore, Condition A6 holds by taking $\eta \leq \lambda_d^*/\lambda_1^*$. These conditions have been assumed in many papers on analysis of Case II interval-censored, e.g. Zhang et al. (2010). We can justify for the rest similarly. As can be seen in Condition A7, the faster convergence rate $p\nu$, the less smoothness of the model α required. Condition A8 is closely related to the information matrix for a semiparametric model. It can be proved that

$$\begin{aligned}&P(\ddot{\ell}_{12}(\boldsymbol{\theta}_0, g_0; \mathbf{O})[h] - P\ddot{\ell}_{22}(\boldsymbol{\theta}_0, g_0; \mathbf{O})[\mathbf{h}_0, h])\\&\quad = -P\{(\dot{\ell}_1(\boldsymbol{\theta}_0, g_0; \mathbf{O}) - \dot{\ell}_2(\boldsymbol{\theta}_0, g_0; \mathbf{O})[\mathbf{h}_0])\dot{\ell}_2(\boldsymbol{\theta}_0, g_0; \mathbf{O})[h]\}.\end{aligned}$$

So for $k = 1, \ldots, d$, if $\dot{\ell}_2(\boldsymbol{\theta}_0, g_0; \mathbf{O})[h_{0k}]$ is the projection of the kth element of $\dot{\ell}_1(\boldsymbol{\theta}_0, g_0; \mathbf{O})$ in the closure of the space generated by $\{\dot{\ell}_2(\boldsymbol{\theta}_0, g_0; \mathbf{O})[h], h \in \mathcal{H}\}$, then we have condition A8. Huang (1996) presents more detailed explanations of conditions A7 and A8. Furthermore,

$$I(\boldsymbol{\theta}_0) = P(-\ddot{\ell}_{11}(\boldsymbol{\theta}_0, g_0; \mathbf{O}) + P\ddot{\ell}_{21}(\boldsymbol{\theta}_0, g_0; \mathbf{O})[\mathbf{h}_0]) = E(\ell^*(\boldsymbol{\theta}_0, g_0; \mathbf{O}))^{\otimes 2}.$$

Under the above conditions, we can obtain the following theorems.

Theorem 1 (Rate of convergence) *Let $K_n = O(n^{\nu})$, where ν satisfies the restriction $\{2(1+p)\}^{-1} < \nu < \{2p\}^{-1}$. Under conditions A1–A6, we have*

$$d(\widehat{\tau}_n, \tau_0) = O_p\{n^{-\min(p\nu, (1-\nu)/2)}\}.$$

The proof of Theorem 1 can be obtained by verifying the conditions of Theorem 1 in Shen and Wong (1994). This Theorem implies that, if $\nu = (2p+1)^{-1}$, $d(\widehat{\tau}_n, \tau_0) = O_p(n^{-p/(1+2p)})$, which is the optimal convergence rate in the nonparametric setting. Although the overall convergence rate is lower than $n^{-\frac{1}{2}}$, the proposed estimator for the regression $\boldsymbol{\theta}_0$ is still asymptotically normal at the rate of $n^{-\frac{1}{2}}$ and attains the semiparametrical efficiency.

Theorem 2 (Asymptotic normality and efficiency) *Suppose conditions A1–A8 hold and ν satisfies the restriction $\{2(1+p)\}^{-1} < \nu < \{2p\}^{-1}$. Then*

$$n^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \longrightarrow N\{0, I^{-1}(\boldsymbol{\theta}_0)\}$$

in distribution, where $I(\boldsymbol{\theta}_0)$ is defined in (3).

Based on the general semiparametric information theory which is described in Bickel et al. (1993), $I(\boldsymbol{\theta}_0)$ is the information matrix evaluated at $\boldsymbol{\theta}_0$. As have been disscused in Zhang et al. (2010), it is not straightforward to estimate the information matrix $I(\boldsymbol{\theta}_0)$, they proposed a least-squares approach to estimate $I(\boldsymbol{\theta}_0)$ based on (3). Specifically, with a random sample $\mathbf{O}_1, \ldots, \mathbf{O}_n$ and the consistent estimator $(\widehat{\boldsymbol{\theta}}_n, \widehat{g}_n)$, we can obtain $I(\boldsymbol{\theta}_0)$ by the minimum value of

$$\rho_n(\mathbf{h}) = \frac{1}{n}\sum_{i=1}^{n} \left\| \dot{\ell}_1(\widehat{\boldsymbol{\theta}}_n, \widehat{g}_n; \mathbf{O}_i) - \dot{\ell}_2(\widehat{\boldsymbol{\theta}}_n, \widehat{g}_n; \mathbf{O}_i)[\mathbf{h}] \right\|^2$$

over $\mathcal{H}$. The detailed descriptions are given in Huang et al. (2008). Theorem 2 can be shown by checking the conditions presented in Appendix B of Zhang et al. (2010).

4 Simulation study

We conducted simulation studies to evaluate the behavior of the proposed sieve estimator with finite sample size in this section. For comparison, we also considered Wang et al. method (2010) under their settings, referred as WST (2010) in the following. The middle-point imputation approach is also considered for comparison, referred as MD. The failure times T were generated from

$$\lambda(t|Z) = \lambda_0(t) + \theta_0 Z(t).$$

The monitoring variables U and V were modeled with the Cox-type hazard functions

$$\lambda^U(t|Z) = \lambda_1(t)e^{\gamma_0 Z(t)}, \qquad \text{and}$$
$$\lambda^V(t|Z) = I(t > U)\lambda_2(t)e^{\gamma_0 Z(t)}$$

respectively. Here we took $Z \sim$ Bernoulli(0.5), the baseline hazard functions $\lambda_0(t) = 2$ or $t^2/2$, $\lambda_1(t) = 4$ and $\lambda_2(t) = 2$. Set the true parameters $\theta_0 = 0.5, 0,$ or -0.5 and $\gamma_0 = 0.5, 0,$ or -0.5.

The proposed sieve estimates were computed by using the cubic B-splines, where the number of knots is chosen to be 2 and the knots are placed at the 25th and 75th quantiles of the distinct observation times of the set $\{(U_i, V_i), i = 1, \ldots, n\}$. We used the R software function "nloptr" to calculate the maximum likelihood estimates of $\boldsymbol{\beta}$ and θ. All simulation results are based on 1000 repetitions with the sample size of $n = 100$ and $n = 200$. Tables 1 and 2 represent the simulation results for the proposed estimator as well as Wang et al. (2010). The MD method performs worse than the others, therefore, we only present the results for "MD" in Table 1 for comparison. We present the bias (Bias), the sample standard error (SSE), and the estimated standard error (ESE) obtained by the least squares method based on the B-splines given in Huang et al. (2008). Column "CP" in the tables stands for the coverage proportion of the 95% confidence intervals. As seen from Table 1, the proposed method and Wang et al. (2010) are comparable. From both Tables 1 and 2, the standard deviation of the estimates are slightly higher than the sample standard errors, which leads the empirical coverage probabilities to exceed the nominal level in most cases. These can be lessened a bit as the sample size increases. These indicate that the proposed estimate is asymptotically unbiased and the proposed variance approximation is reasonable. The method proposed by Wang et al. (2010) yielded nearly unbiased estimates and the values of CP are far away from 95% in Table 2. In this case, the proposed method outperforms the WST method.

Considering the special type of generating U and V, we employ a more general way. The two examination times (U, V) are generated as follows,

$$U \sim \text{Uniform}(c_1, c_2), \;\; V \sim \text{Uniform}(U + 0.01, c_3).$$

We vary c_1, c_2 and c_3 to generate two types of censoring rates. The proportions of left-, interval- and right-censored are (0.25, 0.50, 0.25) and (0.25, 0.25, 0.50). Covariate $Z \sim \text{Uniform}(0, 1)$. We set $\lambda_0(t) = t^2/2$ or $e^t/10$. The true regression parameter $\theta_0 = 0.5$, 1 or 1.5 and consider $n = 100$ and 200. The simulation results are reported in Tables 3 and 4.

It can be seen from Tables 3 and 4 that the proposed estimates are essentially unbiased in all settings regardless of the proportions of left-, right-, interval-censoring rate, expect for the case when $n = 100$ and $\lambda_0(t) = e^t/10$ with right censoring rate 0.5, while the bias is approaching to 0 as the sample size increase. Although the overestimation of the standard deviation results in the coverage probabilities large than 95%, this phenomenon can be eliminated as sample size increases. The WTS estimates are nearly unbiased, but the CP values are much smaller than 95%. In addition, we plot in Fig. 1 the averages of estimates of the true cumulative baseline hazard function (red solid line) by proposed method (black dash line) and MD (blue dash line) method respectively. It can be seen that the proposed method performs better than MD method. The estimation bias decreases as the sample size increases from $n = 100$ to $n = 200$ and then to $n = 400$.

5 Real data analysis

We applied the proposed method to a HIV data set. The HIV data arise from a 16-center prospective studies to investigate the risk of HIV-1 infection among people with hemophilia. These patients were at risk of HIV-1 infection because they received for their treatments blood products such as factor VIII and factor IX concentrate made from the plasma of thousands of donors. In the study, for patients' HIV-1 infection times, only interval-censored data are available, and the patients were placed into different groups according to the average annual dose of the blood products they received. Patients were categorized as high ($>$ 50,000 U), medium (20,001–50,000 U), low (1–20,000 U), and none (no factor use) at the time of abstraction. The numbers of patients in the four groups are 74, 102, 132, and 236, respectively. For the data, the time unit is year and observation 0 means January 1, 1978, the start of the epidemic and the time

Table 1 Simulation results and comparison of the proposed sieve MLE with the estimating equation-based estimator presented in WST (2010) and middle-point approach when $\lambda_0(t) \equiv 2$, and U and V follow the Cox models

γ	θ	Method	$n = 100$				$n = 200$			
			Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
− 0.5	− 0.5	Proposed	0.061	0.387	0.410	0.975	0.025	0.261	0.308	0.945
		WST	− 0.009	0.581	0.536	0.946	− 0.003	0.384	0.375	0.944
		MD	− 0.410	0.245	0.501	1.000	− 0.370	0.173	0.421	1.000
	0	Proposed	0.048	0.482	0.505	0.955	0.019	0.356	0.339	0.946
		WST	− 0.035	0.661	0.592	0.924	− 0.020	0.430	0.412	0.940
		MD	− 0.425	0.253	0.507	0.994	− 0.440	0.177	0.425	0.993
	0.5	Proposed	− 0.005	0.447	0.625	0.969	0.003	0.349	0.420	0.963
		WST	− 0.004	0.736	0.658	0.938	− 0.002	0.449	0.456	0.954
		MD	− 0.853	0.257	0.511	0.737	− 0.868	0.178	0.429	0.437
0	− 0.5	Proposed	0.065	0.426	0.421	0.976	0.019	0.281	0.309	0.948
		WST	− 0.067	0.568	0.529	0.934	− 0.002	0.357	0.369	0.956
		MD	0.348	0.294	0.529	0.991	0.330	0.199	0.443	1.000
	0	Proposed	0.055	0.494	0.532	0.963	0.025	0.359	0.357	0.948
		WST	− 0.049	0.625	0.586	0.934	0.019	0.386	0.410	0.960
		MD	0.017	0.305	0.541	1.000	0.004	0.211	0.453	1.000
	0.5	Proposed	− 0.006	0.460	0.637	0.969	0.005	0.356	0.433	0.968
		WST	− 0.002	0.705	0.651	0.938	0.003	0.453	0.455	0.936
		MD	− 0.355	0.312	0.550	0.989	− 0.370	0.221	0.460	0.996
0.5	− 0.5	Proposed	0.075	0.446	0.475	0.985	0.031	0.333	0.319	0.967
		WST	− 0.085	0.668	0.589	0.916	0.023	0.420	0.412	0.950
		MD	0.749	0.346	0.578	0.896	0.738	0.229	0.485	0.828

Table 1 continued

γ	θ	Method	$n = 100$				$n = 200$			
			Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
	0	Proposed	0.039	0.510	0.583	0.967	0.018	0.375	0.394	0.952
		WST	− 0.006	0.696	0.654	0.930	0.003	0.454	0.457	0.952
		MD	0.524	0.369	0.602	0.970	0.510	0.234	0.504	0.989
	0.5	Proposed	− 0.004	0.468	0.689	0.978	0.002	0.360	0.466	0.965
		WST	0.055	0.769	0.727	0.926	0.041	0.503	0.506	0.958
		MD	0.254	0.386	0.620	0.993	0.235	0.251	0.519	1.000

SSE sample standard error, *ESE* estimated standard error, and *CP* coverage proportion of the 95% confidence intervals

Table 2 Simulation results and comparison of the proposed sieve MLE with the estimating equation-based estimator presented in Wang et al. (2010) when $\lambda_0(t) = t^2/2$, and U and V follow the Cox models

$\lambda_0(t)$	γ	θ	Method	$n = 100$				$n = 200$			
				Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
$t^2/2$	− 0.5	0	Proposed	− 0.039	0.244	0.361	0.987	− 0.020	0.117	0.189	0.959
			WST	− 0.001	0.341	0.092	0.398	0.014	0.221	0.067	0.456
		0.5	Proposed	0.041	0.183	0.263	0.982	− 0.013	0.131	0.170	0.965
			WST	0.012	0.425	0.117	0.394	0.029	0.284	0.086	0.456
	0	0	Proposed	0.053	0.283	0.424	0.971	− 0.034	0.131	0.194	0.952
			WST	− 0.001	0.386	0.102	0.422	− 0.004	0.232	0.075	0.504
		0.5	Proposed	− 0.056	0.212	0.339	0.987	− 0.020	0.141	0.227	0.968
			WST	− 0.008	0.478	0.127	0.386	0.008	0.299	0.094	0.456
	0.5	0	Proposed	− 0.066	0.360	0.431	0.952	0.054	0.350	0.372	0.949
			WST	− 0.031	0.463	0.115	0.374	− 0.016	0.280	0.086	0.492
		0.5	Proposed	− 0.049	0.232	0.350	0.989	− 0.013	0.145	0.249	0.978
			WST	− 0.028	0.557	0.139	0.374	− 0.008	0.331	0.103	0.468

SSE sample standard error, *ESE* estimated standard error, and *CP* coverage proportion of the 95% confidence intervals

Table 3 Simulation results and comparison of the proposed sieve MLE with the estimating equation-based estimator presented in Wang et al. (2010) when $\lambda_0(t) = t^2/2$ or $e^t/10$, and the proportions of left-, interval- and right-censored are (0.25, 0.50, 0.25)

$\lambda_0(t)$	θ	Method	$n = 100$				$n = 200$			
			Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
$t^2/2$	0.5	Proposed	− 0.031	0.214	0.276	0.968	− 0.015	0.160	0.154	0.951
		WST	− 0.043	0.394	0.123	0.464	0.019	0.295	0.089	0.464
	1	Proposed	0.015	0.278	0.340	0.988	0.000	0.210	0.241	0.966
		WST	− 0.047	0.493	0.146	0.440	0.050	0.374	0.107	0.436
	1.5	Proposed	0.079	0.355	0.454	0.988	− 0.010	0.274	0.295	0.948
		WST	− 0.016	0.599	0.170	0.420	0.068	0.451	0.123	0.420
$e^t/10$	0.5	Proposed	− 0.040	0.216	0.347	0.972	− 0.011	0.167	0.188	0.975
		WST	− 0.023	0.340	0.104	0.444	0.011	0.239	0.075	0.450
	1	Proposed	− 0.067	0.281	0.342	0.968	− 0.008	0.216	0.217	0.933
		WST	0.050	0.430	0.126	0.408	0.029	0.335	0.092	0.422
	1.5	Proposed	− 0.065	0.323	0.422	0.986	0.009	0.258	0.297	0.977
		WST	− 0.018	0.541	0.148	0.398	0.060	0.404	0.107	0.434

SSE sample standard error, *ESE* estimated standard error, and *CP* coverage proportion of the 95% confidence intervals

Table 4 Simulation results and comparison of the proposed sieve MLE with the estimating equation-based estimator presented in Wang et al. (2010) when $\lambda_0(t) = t^2/2$ or $e^t/10$, and the proportions of left-, interval- and right-censored are (0.25, 0.25, 0.50)

$\lambda_0(t)$	θ	Method	$n = 100$				$n = 200$			
			Bias	SSE	ESE	CP	Bias	SSE	ESE	CP
$t^2/2$	0.5	Proposed	− 0.026	0.194	0.276	0.975	− 0.015	0.136	0.207	0.961
		WST	0.017	0.299	0.088	0.466	0.021	0.211	0.064	0.442
	1	Proposed	0.042	0.253	0.340	0.973	− 0.025	0.184	0.271	0.960
		WST	0.042	0.426	0.125	0.426	0.035	0.300	0.091	0.434
	1.5	Proposed	− 0.063	0.303	0.504	0.976	− 0.035	0.228	0.319	0.955
		WST	0.057	0.575	0.171	0.438	0.051	0.408	0.124	0.458
$e^t/10$	0.5	Proposed	− 0.061	0.236	0.389	0.976	− 0.015	0.172	0.232	0.950
		WST	0.019	0.313	0.092	0.448	0.021	0.224	0.067	0.434
	1	Proposed	− 0.107	0.335	0.530	0.987	− 0.047	0.243	0.254	0.962
		WST	0.040	0.473	0.141	0.450	0.039	0.339	0.102	0.460
	1.5	Proposed	− 0.158	0.390	0.516	0.988	− 0.070	0.239	0.298	0.966
		WST	0.060	0.638	0.192	0.452	0.063	0.464	0.139	0.444

SSE sample standard error, *ESE* estimated standard error, and *CP* coverage proportion of the 95% confidence intervals

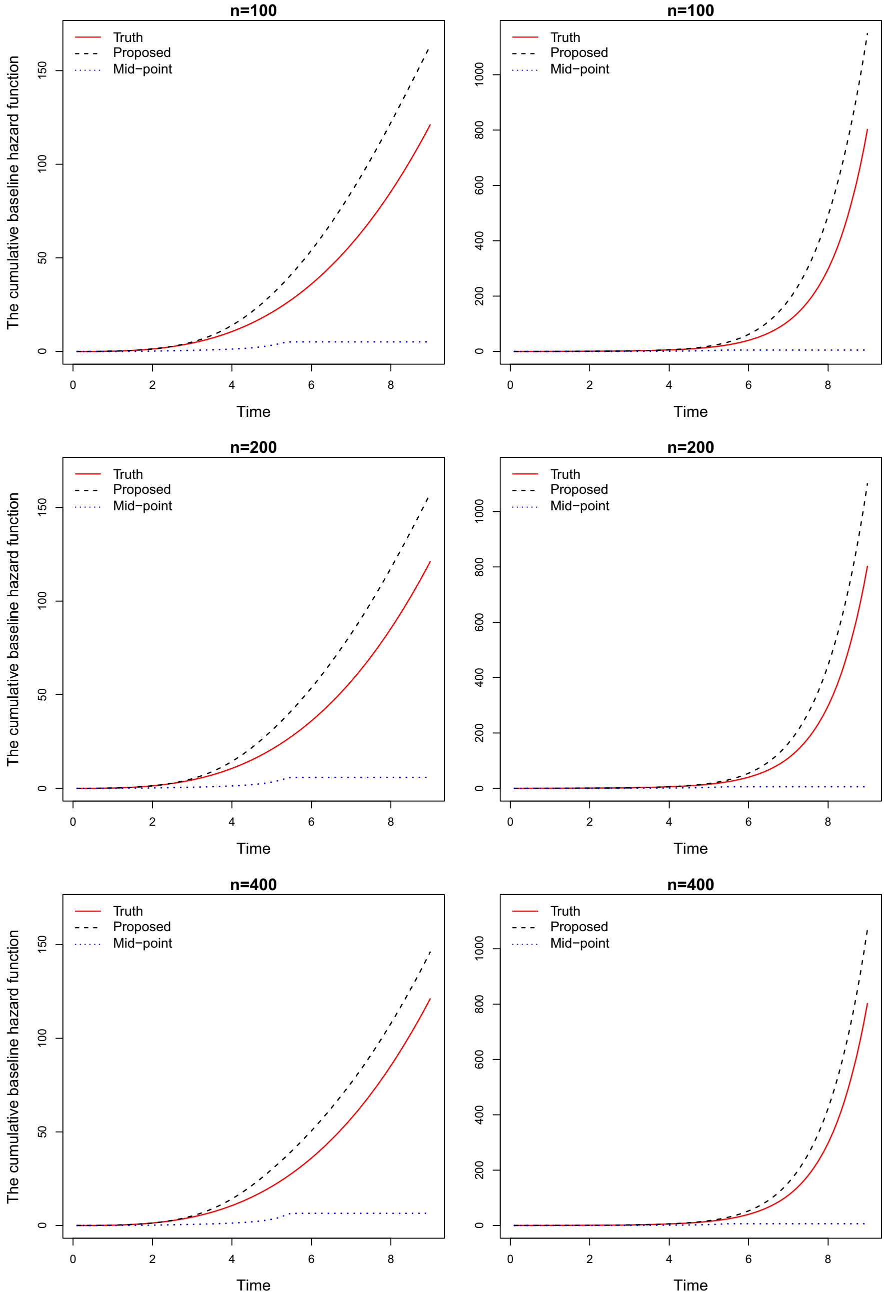

Fig. 1 Estimates of the cumulative of baseline hazard function, the left side is $\lambda_0(t) = t^2/2$ and the right side is $\lambda_0(t) = e^t/10$

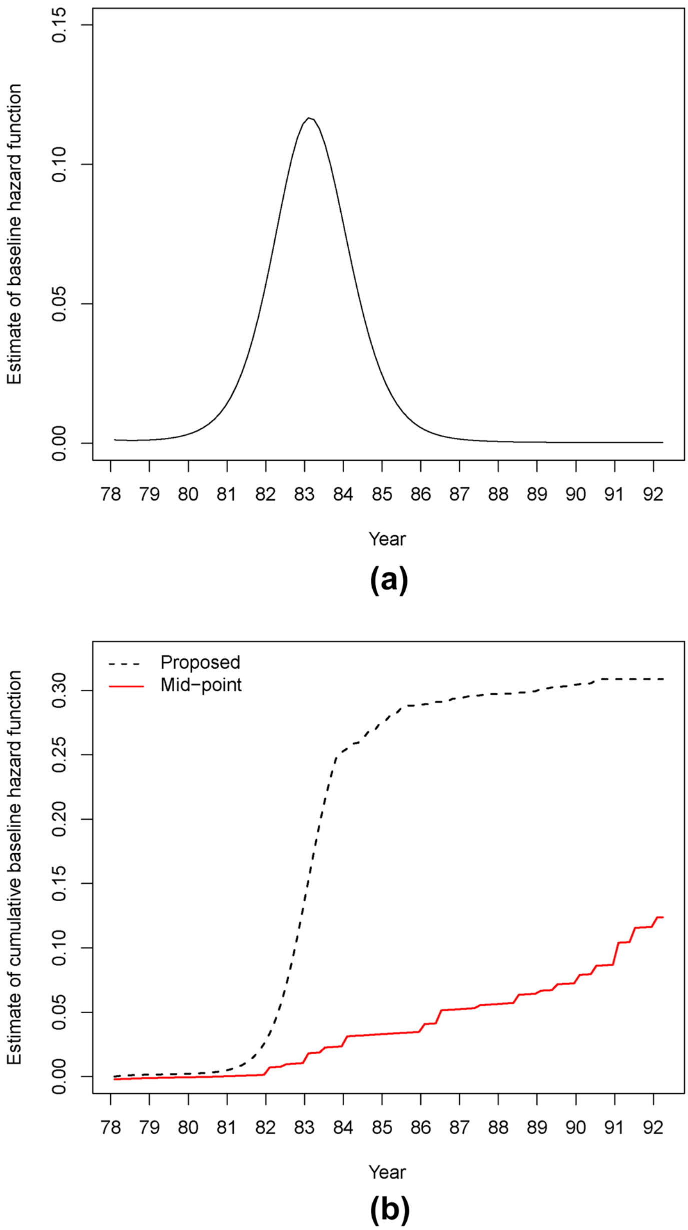

Fig. 2 **a** The estimate of baseline hazard function for hemophilia data using the proposed cubic B-spline sieve MLE. **b** The two estimates of cumulative baseline hazard function for hemophilia data using the proposed cubic B-spline sieve MLE (black dashed curve) and MD method (red solid curve)

at which all patients are considered to be nonnegative. Of these, 63 patients were left-censored, 204 patients were interval-censored, and the remaining 277 patients were right-censored. Among patients with left-censored event times, the average length of time from entry to the left censoring time was 6.47 years. Among patients with interval-censored event times, the average length from entry to the left monitoring time was 3.44 years and the average length from entry to the right monitoring time was 5.57 years; and among patients with right-censored event times, the average length from entry to the right censoring time was 12.49 years.

We treat different level of doses as dummy variables, introducing high, medium, low variables. For example, the variable high is 1, if the patients falls in the high dose and 0 otherwise, other variables analogy to high. We fit the semiparametric additive hazard model

$$\lambda(t) = \lambda_0(t) + \boldsymbol{\theta}_0^T Z,$$

to analyze the difference of the hazard for the time until the appearance of HIV-1 infection between two groups. Here $\boldsymbol{\theta}_0 = (\theta_{01}, \theta_{02}, \theta_{03})$ represents the effect of high, medium and low level of dose respectively. Applying the proposed method, the cubic B-splines sieve semiparametric maximum likelihood estimate of $\boldsymbol{\theta}_0$ is (0.240, 0.148, 0.026) with the estimated standard error of (0.048, 0.027, 0.008) respectively. Further the p values for testing $\boldsymbol{\theta}_0 = \mathbf{0}$ are all much smaller than 0.001. It can be found that patients in the high group have 0.240 higher risks, mediums have 0.148 higher risks and lows have 0.026 higher risks of being infected by HIV-1 than those receiving no factor VIII concentrate. The corresponding estimated coefficients by middle-point imputation approach are (0.015, 0.049, 0.068).

Figure 2a shows the estimate of baseline hazard function $\lambda_0(t)$ using the proposed sieve MLE method. Patients first suffer from drastically increasing risk and then the hazard gradually decreased with time, which agrees with the results in Kroner et al. (1994). In Fig. 2b, we plot the cumulative baseline hazard function using the proposed cubic B-spline sieve MLE (black dashed curve) and the MD method (red solid curve). The two curves indicate that the MD method presents a lower risk estimate than our proposed method.

6 Conclusions

This article considered a spline-based sieve semiparametric maximum likelihood approach for an additive hazards model with case II interval censored data. Employing the B-splines to approximate the log baseline hazard function directly, we reduced the dimensionality of the estimation problem and removed the constraint of being non-negative of the baseline hazard function and the monotonicity of the baseline cumulative hazard function. Hence, the R function "nloptr" can implement the optimization problem for the semiparametric likelihood inference procedure and eased the burden of computation. We showed the consistency of the proposed estimator and derived the rate of convergence. Furthermore, the sieve estimator for the regression $\boldsymbol{\theta}_0$ was asymptotically normal and attained the semiparametric efficiency bound. However, the asymptotic distribution of $\lambda_0(t)$ is still under investigation.

Acknowledgements The authors would like to thank the Editor, the Associate Editor and the two reviewers for their constructive and insightful comments and suggestions that greatly improved the paper. This research is partly supported by the National Natural Science Foundation of China (Nos. 11571263, 11671311, 11771366) and the Research Grant Council of Hong Kong (15301218, 15303319).

Appendix: Proofs of Theorems

First we derive the integral equation for the least favorable direction. Denote

$$
\begin{aligned}
\xi_1(U, V, g) =& \frac{\exp\left(-\int_0^U \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(U\mathbf{Z})\right)}{1-\exp\left(-\int_0^U \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(U\mathbf{Z})\right)}\\
\xi_2(U, V, g) =& \frac{\exp\left(-\int_0^U \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(U\mathbf{Z})\right)}{\exp\left(-\int_0^U \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(U\mathbf{Z})\right) - \exp\left(-\int_0^V \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(V\mathbf{Z})\right)}\\
\xi_3(U, V, g) =& \frac{\exp\left(-\int_0^V \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(V\mathbf{Z})\right)}{\exp\left(-\int_0^U \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(U\mathbf{Z})\right) - \exp\left(-\int_0^V \exp\{g(s)\}\mathrm{d}s - \boldsymbol{\theta}^{\mathrm{T}}(V\mathbf{Z})\right)}\\
\varphi_1(U, V) =& E_{\mathbf{z}}\{\xi_1(U, V, g) f(U, V|\mathbf{Z})\},\ \varphi_2(U, V) = E_{\mathbf{z}}\{\xi_2(U, V, g) f(U, V|\mathbf{Z})\}\\
\varphi_3(U, V) =& E_{\mathbf{z}}\{\xi_3(U, V, g) f(U, V|\mathbf{Z})\},\ \varphi_4(U, V) = E_{\mathbf{z}}\{f(U, V|\mathbf{Z})\}\\
\psi_1(U, V) =& E_{\mathbf{z}}\{\mathbf{Z}\xi_1(U, V, g) f(U, V|\mathbf{Z})\},\ \psi_2(U, V) = E_{\mathbf{z}}\{\mathbf{Z}\xi_2(U, V, g) f(U, V|\mathbf{Z})\}\\
\psi_2(U, V) =& E_{\mathbf{z}}\{\mathbf{Z}\xi_3(U, V, g) f(U, V|\mathbf{Z})\},\ \psi_4(U, V) = E_{\mathbf{z}}\{\mathbf{Z} f(U, V|\mathbf{Z})\},
\end{aligned}
$$

where $E_{\mathbf{z}}$ means taking expectation with respect $\mathbf{Z}$. Follow the similar steps of Huang et al. (2008), define function,

$$
\begin{aligned}
\varphi(t) =& \int_{t+c}^{b} \varphi_1(t, x)\mathrm{d}x + \int_{t+c}^{b} \varphi_2(t, x)\mathrm{d}x + \int_{a}^{t-c} \varphi_3(x, t)\mathrm{d}x + \int_{a}^{t-c} \varphi_4(x, t)\mathrm{d}x\\
\psi(t) =& \int_{t+c}^{b} \psi_1(t, x)\mathrm{d}x + \int_{t+c}^{b} \psi_1(t, x)\mathrm{d}x + \int_{a}^{t-c} \psi_3(x, t)\mathrm{d}x + \int_{a}^{t-c} \psi_4(x, t)\mathrm{d}x\\
r(t) =& -\psi(t)t + \int_{a}^{t-c} x\psi_2(x, t)\mathrm{d}x + \int_{t+c}^{b} x\psi_3(t, x)\mathrm{d}x\\
Q(t, x) =& \{\varphi_2(x, t) I_{a\le x\le t-c} + \varphi_3(t, x) I_{t+c\le x\le b}\}/\varphi(t).
\end{aligned}
$$

Then we can attain (4).

Next we present the proof for Theorem 1 and 2. Throughout the following proofs, for notation simplicity, we denote $P_n f = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{O}_i)$, $M(\tau) = P\ell(\tau; \mathbf{O}) = P\ell(\boldsymbol{\theta}, g; \mathbf{O})$ and $M_n(\tau) = P_n\ell(\tau; \mathbf{O}) = P_n\ell(\boldsymbol{\theta}, g; \mathbf{O})$, let C represent a generic constant that may vary from place to place.

Proof of Theorem 1 To show the consistency and derive the convergence rate, we just need to verify the following conditions C1–C3 in Theorem 1 of Shen and Wong (1994), which are presented as follows:

C1 $\inf_{\{d(\tau,\tau_0)\geq\epsilon,\tau\in\Theta\times\mathbb{G}_n\}} M(\tau_0)-M(\tau)\geq C\inf_{\{d(\tau,\tau_0)\geq\epsilon,\tau\in\Theta\times\mathbb{G}_n\}} d^2(\tau,\tau_0)$ where $\tau_0=(\boldsymbol{\theta}_0,g_0)$, and C1 holds with $\alpha=1$.

C2 $\sup_{\{d(\tau,\tau_0)\leq\epsilon,\tau\in\Theta\times\mathbb{G}_n\}} \text{var}(\ell(\tau_0;\mathbf{O})-\ell(\tau;\mathbf{O}))\leq\sup_{\{d(\tau,\tau_0)\leq\epsilon,\tau\in\Theta\times\mathbb{G}_n\}} d^2(\tau,\tau_0)$, and C2 holds with $\beta=1$.

C3 Let $\mathcal{F}_n=\{\ell(\tau;\cdot):\tau\in\Theta\times\mathbb{G}_n\}$, $H(\epsilon,\mathcal{F}_n)\leq Cn^{2\gamma_0}\log(1/\epsilon)$, where $H(\epsilon,\mathcal{F}_n)$ is the L_∞-metric entropy of the space $\mathcal{F}_n$ and C3 holds with $2\gamma_0=\nu$ and $\gamma=0^+$.

Condition C1 with $\alpha=1$ can be verified by similar contexts as in Zhang et al. (2010). Condition C2 can be easily obtained through a Taylor expansion combined with conditions A1–A5. By inequality $\log(x)\leq x-1$, we have the following results, for $\tau\in\Theta\times\mathbb{G}_n$,

$$
\begin{aligned}
&E\{\ell(\tau_0)-\ell(\tau)\}^2\\
&=E\Big(\Delta_1\log\frac{1-\exp\{-\phi_0(\mathbf{Z},U)\}}{1-\exp\{-\phi(\mathbf{Z},U)\}}+\Delta_2\log\frac{\exp\{-\phi_0(\mathbf{Z},U)\}-\exp\{-\phi_0(\mathbf{Z},V)\}}{\exp\{-\phi(\mathbf{Z},U)\}-\exp\{-\phi(\mathbf{Z},V)\}}\\
&\quad-(1-\Delta_1-\Delta_2)\{\phi_0(\mathbf{Z},V)-\phi(\mathbf{Z},V)\}\Big)^2\\
&\leq CE\Big(\Delta_1\frac{\exp\{-\phi(\mathbf{Z},U)\}-\exp\{-\phi_0(\mathbf{Z},U)\}}{1-\exp\{-\phi(\mathbf{Z},U)\}}\\
&\quad+\Delta_2\frac{\exp\{-\phi_0(\mathbf{Z},U)\}-\exp\{-\phi_0(\mathbf{Z},V)\}-(\exp\{-\phi(\mathbf{Z},U)\}-\exp\{-\phi(\mathbf{Z},V)\})}{\exp\{-\phi(\mathbf{Z},U)\}-\exp\{-\phi(\mathbf{Z},V)\}}\\
&\quad(1-\Delta_1-\Delta_2)\{\phi_0(\mathbf{Z},V)-\phi(\mathbf{Z},V)\}\Big)^2\\
&\leq CE\Bigg[\Big(\Delta_1\frac{\exp\{-\phi(\mathbf{Z},U)\}-\exp\{-\phi_0(\mathbf{Z},U)\}}{1-\exp\{-\phi(\mathbf{Z},U)\}}\Big)^2\\
&\quad+\Big(\Delta_2\frac{\exp\{-\phi_0(\mathbf{Z},U)\}-\exp\{-\phi_0(\mathbf{Z},V)\}-(\exp\{-\phi(\mathbf{Z},U)\}-\exp\{-\phi(\mathbf{Z},V)\})}{\exp\{-\phi(\mathbf{Z},U)\}-\exp\{-\phi(\mathbf{Z},V)\}}\Big)^2\\
&\quad+[(1-\Delta_1-\Delta_2)\{\phi_0(\mathbf{Z},V)-\phi(\mathbf{Z},V)\}]^2\Bigg]\\
&\leq CE\Big[(\exp\{-\phi(\mathbf{Z},U)\}-\exp\{-\phi_0(\mathbf{Z},U)\})^2\\
&\quad+(\exp\{-\phi_0(\mathbf{Z},U)\}-\exp\{-\phi_0(\mathbf{Z},V)\}-(\exp\{-\phi(\mathbf{Z},U)\}-\exp\{-\phi(\mathbf{Z},V)\}))^2\\
&\quad+(\{\phi_0(\mathbf{Z},V)-\phi(\mathbf{Z},V)\})^2\Big]\\
&\leq CE\Big[(\exp\{-\phi(\mathbf{Z},U)\}-\exp\{-\phi_0(\mathbf{Z},U)\})^2\\
&\quad+(\exp\{-\phi_0(\mathbf{Z},V)\}-\exp\{-\phi(\mathbf{Z},V)\})^2+(\{\phi_0(\mathbf{Z},V)-\phi(\mathbf{Z},V)\})^2\Big]\\
&\leq CE\Bigg[\|\boldsymbol{\theta}_0-\boldsymbol{\theta}\|^2+\left\{\int_0^U[\exp\{g_0(s)\}-\exp\{g(s)\}]\mathrm{d}s\right\}^2\\
&\quad+\left\{\int_0^V[\exp\{g_0(s)\}-\exp\{g(s)\}]\mathrm{d}s\right\}^2\Bigg]\\
&\leq CE\Big[\|\boldsymbol{\theta}_0-\boldsymbol{\theta}\|^2+\int_0^U[\exp\{g_0(s)\}-\exp\{g(s)\}]^2\mathrm{d}s
\end{aligned}
$$

$$
\begin{aligned}
&+\int_0^V [\exp\{g_0(s)\} - \exp\{g(s)\}]^2 \mathrm{d}s\Big] \\
&\leq CE\Big[\|\boldsymbol{\theta}_0 - \boldsymbol{\theta}\|^2 + \int_0^U [\exp\{g^\star(s)\}]^2\{g_0(s) - g(s)\}^2 \mathrm{d}s \\
&+\int_0^V [\exp\{g^\star(s)\}]^2\{g_0(s) - g(s)\}^2 \mathrm{d}s\Big] \\
&\leq Cd^2(\tau_0, \tau),
\end{aligned}
$$

where the second and the fourth inequality follow from the inequality $(a+b)^2 \leq C(a^2+b^2)$, the sixth inequality is obtained by Cauchy–Schwartz inequality and $g^\star(s)$ is a value between $g_0(s)$ and $g(s)$. With condition C1 which we have already shown, we can verify condition C2 with $\beta = 1$.

Next we verify the condition C3. Let $L_1 = \{\ell(\tau; \mathbf{O}) : \tau \in \Theta \times \mathbb{G}_n\}$. We can easily construct a set of brackets $\{[\ell^L_{s,i}(\mathbf{O}), \ell^U_{s,i}(\mathbf{O})] : s = 1, 2, \ldots, [C(1/\epsilon)^d]; i = 1, 2, \ldots, [C(1/\epsilon)^{Cq_n}]\}$ for any $\ell(\tau; \mathbf{O}) \in L_1$, Specifically,

$$
\begin{aligned}
\ell^L_{si}(\mathbf{O}) = \Delta_1 \log\Big\{1 - \exp\Big(-\int_0^U \exp\{g_i(t)^L\}\mathrm{d}t - ((U\mathbf{Z}^{\mathrm{T}})\boldsymbol{\theta}_s - UC\epsilon)\Big)\Big\} \\
+\Delta_2 \log\Big\{\exp\Big(-\int_0^U \exp\{g_i(t)^U\}\mathrm{d}t - ((U\mathbf{Z}^{\mathrm{T}})\boldsymbol{\theta}_s + UC\epsilon)\Big) \\
-\exp\Big(-\int_0^V \exp\{g_i(t)^L\}\mathrm{d}t - ((V\mathbf{Z}^{\mathrm{T}})\boldsymbol{\theta}_s - VC\epsilon)\Big)\Big\} \\
-\Big(1 - \Delta_1 - \Delta_2\Big)\Big\{\int_0^V \exp\{g_i(t)^U\}\mathrm{d}t + ((V\mathbf{Z}^{\mathrm{T}})\boldsymbol{\theta}_s + VC\epsilon)\Big\}
\end{aligned}
$$

and

$$
\begin{aligned}
\ell^U_{si}(\mathbf{O}) = \Delta_1 \log\Big\{1 - \exp\Big(-\int_0^U \exp\{g_i(t)^U\}\mathrm{d}t - ((U\mathbf{Z}^{\mathrm{T}})\boldsymbol{\theta}_s + UC\epsilon)\Big)\Big\} \\
+\Delta_2 \log\Big\{\exp\Big(-\int_0^U \exp\{g_i(t)^L\}\mathrm{d}t - ((U\mathbf{Z}^{\mathrm{T}})\boldsymbol{\theta}_s - UC\epsilon)\Big) \\
-\exp\Big(-\int_0^V \exp\{g_i(t)^U\}\mathrm{d}t - ((V\mathbf{Z}^{\mathrm{T}})\boldsymbol{\theta}_s + VC\epsilon)\Big)\Big\} \\
-\Big(1 - \Delta_1 - \Delta_2\Big)\Big\{\int_0^V \exp\{g_i(t)^L\}\mathrm{d}t + ((V\mathbf{Z}^{\mathrm{T}})\boldsymbol{\theta}_s - VC\epsilon)\Big\}.
\end{aligned}
$$

where $\{[g^L_i, g^U_i] : i = 1, \ldots, [(1/\epsilon)]^{Cq_n}\}$ is the brackets set for any $g \in S_n$. Then, using a Taylor expansion along with conditions A1–A3, we can conclude that the ϵ-bracketing number for L_1 with $L_1(P)$-norm is bounded by $C(1/\epsilon)^{Cq_n+d}$ and $H(\epsilon, L_1) \leq Cn^{-\nu} \log(1/\epsilon)$. Hence, condition C3 in Theorem 1 of Shen and Wong (1994) holds with $2\gamma_0 = \nu$ and $r = 0^+$.

With condition $A4$, for $g_0 \in \mathbb{G}$, employing Corollary 6.21 in Schumaker (1981), there exists a function $g_{0n} \in S_n$ of order $m \geq p+2$ such that $\|g_{0n} - g_0\|_\infty = O(n^{-p\nu})$, where $\|\cdot\|_\infty$ is the sup-norm, which also means $\|g_{0n} - g_0\|_{\mathbb{G}} = O(n^{-p\nu})$. Now denote $\tau_{0,n} = (\boldsymbol{\theta}_0, g_{0,n})$. Then we have

$$\begin{aligned} M_n(\widehat{\tau}_n) - M_n(\tau_0) &= M_n(\widehat{\tau}_n) - M_n(\tau_{0,n}) + M_n(\tau_{0,n}) - M_n(\tau_0) \\ &\geq P_n \ell(\tau_{0,n}; \mathbf{O}) - P_n \ell(\tau_0; \mathbf{O}) \\ &= (P_n - P)\{\ell(\tau_{0,n}; \mathbf{O}) - \ell(\tau_0; \mathbf{O})\} + M(\widehat{\tau}_{0,n}) - M(\tau_0). \end{aligned}$$

Similar as (Zhang et al. 2010), we can conclude that

$$M_n(\widehat{\tau}_n) - M_n(\tau_0) \geq o_p(n^{-1/2}) - o(1) = -o_p(1),$$

and then $\widehat{\tau}_n$ satisfies inequality (1.1) in Shen and Wong (1994).

Next, we derive the convergence rate. We have obtained that condition C3 in Theorem 1 of Shen and Wong (1994) holds with constants $2\gamma_0 = \nu$ and $r = 0^+$ in their notation. Furthermore, the constant τ in Theorem 1 of Shen and Wong (1994) is $(1-\nu)/2 - (\log\log n)/(2\log n)$. On the other hand, we can pick a $\bar{\nu}$ slightly greater than ν such that $(1-\bar{\nu})/2 \leq (1-\nu)/2 - (\log\log n)/(2\log n)$ for large n. We still denote $\bar{\nu}$ by ν and then $\tau = (1-\nu)/2$. The Kullback-Leibler distance between $\tau_0 = (\boldsymbol{\theta}_0, g_0)$ and $\tau_{0,n} = (\boldsymbol{\theta}_0, g_{0n})$ is given by

$$\begin{aligned} &K(\tau_0, \tau_{0,n}) \\ &\quad = P(l(\tau_0; X) - l(\tau_{0n}; X)) \\ &\quad = E\Big(\Big[1 - \exp\{-\phi_{0n}(\mathbf{Z}, U)\}\Big] m\Big[\frac{1 - \exp\{-\phi_0(\mathbf{Z}, U)\}}{1 - \exp\{-\phi_{0n}(\mathbf{Z}, U)\}}\Big] \\ &\qquad + \Big[\exp\{-\phi_{0n}(\mathbf{Z}, U)\} - \exp\{-\phi_{0n}(\mathbf{Z}, V)\}\Big] \\ &\qquad \times m\Big[\frac{\exp\{-\phi_0(\mathbf{Z}, U)\} - \exp\{-\phi_0(\mathbf{Z}, V)\}}{\exp\{-\phi_{0n}(\mathbf{Z}, U)\} - \exp\{-\phi_{0n}(\mathbf{Z}, V)\}}\Big] \\ &\qquad + \exp\Big\{-\phi_{0n}(\mathbf{Z}, V)\Big\} m\Big[\frac{\exp\{-\phi_0(\mathbf{Z}, V)\}}{\exp\{-\phi_{0n}(\mathbf{Z}, V)\}}\Big]\Big) \\ &\quad \leq CE([\exp\{-\phi_0(\mathbf{Z}, U)\} - \exp\{-\phi_{0n}(\mathbf{Z}, U)\}]^2 \\ &\qquad + [\exp\{-\phi_0(\mathbf{Z}, V)\} - \exp\{-\phi_{0n}(\mathbf{Z}, V)\}]^2 \\ &\qquad + [\phi_0(\mathbf{Z}, V) - \phi_{0n}(\mathbf{Z}, V)]^2) \\ &\quad \leq C\|g_0 - g_{0n}\|_2^2 \leq C\|g_0 - g_{0n}\|_\infty^2 = O(n^{-2p\nu}), \end{aligned}$$

where $m(x) = x\log x - x + 1 \leq x(x-1) - x + 1 \leq (x-1)^2$. Then, we can obtain $K^{\frac{1}{2}}(\tau_0, \tau_{0n}) = O(n^{-p\nu})$. Following Theorem 1 of Shen and Wong (1994), we have $d(\widehat{\tau}_n, \tau_0) = O_p\{n^{-\min(p\nu,(1-\nu)/2)}\}$, which completes the proof of Theorem 1. □

Proof of Theorem 2 By Zhang et al. (2010), it is sufficient to derive the asymptotic normality for $\widehat{\boldsymbol{\theta}}_n$ by verifying the following conditions.

B1 $P_n\dot{\ell}_1(\widehat{\tau}_n; \mathbf{O}) = o_p(n^{-1/2})$ and $P_n\dot{\ell}_2(\widehat{\tau}_n; \mathbf{O})[h_0] = o_p(n^{-1/2})$.
B2 $(P_n - P)\{\ell^*(\widehat{\tau}_n; \mathbf{O}) - \ell^*(\tau_0; \mathbf{O})\} = o_p(n^{-1/2})$.

B3 $P\{\ell^*(\widehat{\tau}_n; \mathbf{O}) - \ell^*(\tau_0; \mathbf{O})\} = -I(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + O_p(\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|) + o_p(n^{-1/2})$.

Conditions B1 and B2 can be verified by similar arguments as Zhang et al. (2010). As for condition B3, using (a1) minus (a2) of conditions A7 and A8, we have

$$P\{\ell^*(\widehat{\boldsymbol{\theta}}_n, \widehat{g}_n; \mathbf{O}) - \ell^*(\boldsymbol{\theta}_0, g_0; \mathbf{O})\}$$
$$= -I(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|) + O(\|\widehat{g}_n - g_0\|^\alpha).$$

By Theorem 1 and the fact $\alpha p \nu > \frac{1}{2}$, we have, $O(\|\widehat{g}_n - g_0\|^\alpha) = o_p(n^{-1/2})$. So $B3$ holds. Then Theorem 2 can be established follow the general procedure which has stated in Zhang et al. (2010). □

References

Betensky RA, Rabinowitz D, Tsiatis AA (2001) Computationally simple accelerated failure time regression for interval censored data. Biometrika 88:703–711

Bickel PJ, Klaassen CAJ, Ritov Y, Wellner JA (1993) Efficient and adaptive estimation for semiparametric models. Johns Hopkins University Press, Baltimore

Cox DR (1972) Regression models and life tables (with discussion). J R Stat Soc B 34:187–220

Finkelstein DM (1986) A proportional hazards model for interval-censored failure time data. Biometrics 42:845–854

Geman S, Hwang C-R (1982) Nonparametric maximum likelihood estimation by the method of sieves. Ann Stat 10:401–414

Goggins WB, Finkelsten DM (2000) A proportional hazards model for multivariate interval-censored failure time data. Biometrics 56:940–943

Gómez G, Espinal A, Lagakos SW (2003) Inference for a linear regression model with an interval-censored covariate. Stat Med 22:409–425

Huang J (1996) Efficient estimation for the proportional hazards model with interval censoring. Ann Stat 24:540–568

Huang J, Wellner JA (1997) Interval censored survival data: a review of recent process. In: Lin DY, Fleming TR (eds) Proceedings of the first Seattle symposium in biostatistics: survival analysis. Springer, New York, pp 123–169

Huang J, Zhang Y, Hua L (2008) A least-squares approach to consistent information estimation in semiparametric models. Technical Report, Department of Biostatistics, University of Iowa

Kim M, Xue X (2002) The analysis of multivariate interval-censored survival data. Stat Med 21:3715–3726

Kroner BL, Rosenberg PS, Aledort LM, Alvord WG, Goedert JJ (1994) HIV-1 infection incidence among persons with hemophilia in the United States and Western Europe, 1978–1990. J Acquir Immune Defic 7:279–286

Li G, Zhang C-H (1998) Linear regression with interval censored data. Ann Stat 26:1306–1327

Li L, Pu Z (2003) Rank estimation of log-linear regression with interval-censored data. Lifetime Data Anal 9:57–70

Martinussen T, Scheike TH (2002) Efficient estimation in additive hazards regression with current status data. Biometrika 89:649–658

Rabinowitz D, Tsiatis A, Aragon J (1995) Regression with interval-censored data. Biometrika 82:501–513

Satten GA, Datta S, Williamson JM (1998) Inference based on imputed failure times for the proportional hazards model with interval-censored data. J Am Stat Assoc 93:318–327

Schumaker L (1981) Spline function: basic theory. Wiley, New York

Seaman SR, Bird SM (2001) Proportional hazards model for interval-censored failure times and time-dependent covariates: application to hazard of HIV infection of injecting drug users in prison. Stat Med 20:1855–1870

Shen X, Wong WH (1994) Convergence rate of sieve estimates. Ann Stat 22:580–615

Song X, Ma S (2008) Multiple augmentation for interval-censored data with measurement error. Stat Med 27:3178–3190

Sun J (2006) The statistical analysis of interval-censored failure time data. Springer, New York
Sun J, Shen J (2009) Efficient estimation for the proportional hazards model with competing risks and current status data. Can J Stat 37:592–606
van der Vaart AW (1998) Asymptotic statistics. Cambridge University Press, Cambridge
Wang L, Sun J, Tong X (2008) Efficient estimation for the proportional hazards with bivariate current status data. Lifetime Data Anal 14:134–153
Wang L, Sun J, Tong X (2010) Regression analysis of case II interval-censored failure time data with the additive hazards model. Stat Sin 20:1709–1723
Xue H, Lam KF, Li G (2004) Sieve maximum likelihood estimator for semiparametric regresion models with current status data. J Am Stat Assoc 99:346–356
Zeng D, Cai J, Shen Y (2006) Semiparametric additive risks model for interval-censored data. Stat Sin 16:287–302
Zhang W, Zhang Y, Chaloner K, Stapleton JT (2009) Imputation methods for doubly censored HIV data. J Stat Comput Sim 79:1245–1257
Zhang Y, Hua L, Huang J (2010) A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. Scand J Stat 37:338–354
Zhang Y, Liu W, Zhan Y (2001) A nonparametric two-sample test of the failure function with interval censoring case 2. Biometrika 88:677–686
Zhao X, Lim H, Sun J (2005) Estimating equation approach for regression analysis of failure time data in the presence of interval-censoring. J Stat Plan Inference 129:145–157
Zhou Q, Hu T, Sun J (2017) A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. J Am Stat Assoc 112:664–672